Markov Chain Models, Time Series Analysis and Extreme Value Theory

D. S. POSKITT and SHIN-HO CHUNG, ***Australian National University***

Reprinted from

ADVANCES IN APPLIED PROBABILITY

Vol. 28, No. 2, pp. 405-425

June 1996

Adv. Appl. Prob. **28**, 405–425 (1996)
Printed in N. Ireland

MARKOV CHAIN MODELS, TIME SERIES ANALYSIS AND EXTREME VALUE THEORY

D. S. POSKITT* AND
SHIN-HO CHUNG,** *Australian National University*

Dedicated to the late Professor E. J. Hannan

Abstract

Markov chain processes are becoming increasingly popular as a means of modelling various phenomena in different disciplines. For example, a new approach to the investigation of the electrical activity of molecular structures known as ion channels is to analyse raw digitized current recordings using Markov chain models. An outstanding question which arises with the application of such models is how to determine the number of states required for the Markov chain to characterize the observed process. In this paper we derive a realization theorem showing that observations on a finite state Markov chain embedded in continuous noise can be synthesized as values obtained from an autoregressive moving-average data generating mechanism. We then use this realization result to motivate the construction of a procedure for identifying the state dimension of the hidden Markov chain. The identification technique is based on a new approach to the estimation of the order of an autoregressive moving-average process. Conditions for the method to produce strongly consistent estimates of the state dimension are given. The asymptotic distribution of the statistic underlying the identification process is also presented and shown to yield critical values commensurate with the requirements for strong consistency.

MARKOV CHAIN; AUTOREGRESSIVE MOVING-AVERAGE; REALIZATION; IDENTIFICATION; CONSISTENCY; EXTREME VALUE

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 62M10
SECONDARY 62M05; 60J10

1. Introduction

Early work on hidden Markov chains can be traced back to Blackwell and Koopmans (1957) and Heller (1965), who were concerned with the probabilistic properties of such processes, and Baum and Petrie (1966), who addressed the consistency and asymptotic normality of the maximum likelihood estimator and provided an early precursor to the EM algorithm associated with Dempster *et al.* (1977). Since that time research into hidden Markov chains has been in abeyance in

Received 22 August 1994; revision received 4 April 1995.

*Postal address: Department of Statistics, Australian National University, Canberra ACT 0200, Australia.

**Postal address: Department of Chemistry, Australian National University, Canberra ACT 0200, Australia.

This work was supported in part by a grant from the National Health & Medical Research Council of Australia.

the statistical literature until very recently (see Qian and Titterington (1991), Leroux (1992)), although in engineering the application of hidden Markov chain models has received considerable attention, particularly in communication applications (Levinson *et al.* 1983). Our interest in the topic is motivated by the need to construct probabilistic models of data obtained from observations on cell membrane channels. Such models are required to help us understand and characterize the microscopic current fluctuations that occur when individual channels in the cell membrane, known as ion channels, open and close. Because all electrical activities in the nervous system, including communication between cells and the influence of hormones and drugs on cell function, are regulated by the opening and closing of ion channels, understanding the gating mechanism at a molecular level is of fundamental importance in neurobiology. The gating mechanism cannot be directly observed, even under an electron-microscope, but measurement of ionic currents flowing across the membrane has been made possible by the development of the giga-seal patch-clamp technique. Thus, detailed information on small ion channel currents, which has hitherto been inaccessible, can now be gleaned from recordings made on real-world processes. Experimental records obtained with this new tool are, however, contaminated by unavoidable noise. The aim of biologists is to recover the true characteristics of the original signal sequence using statistical signal processing techniques of the type described in this paper; see Chung *et al.* (1991) for further details. Such processing is a necessary first step in the construction of a realistic model that describes the molecular mechanisms underlying the opening and closing of ion channels.

The basic model that we wish to consider is defined as follows. Suppose that $\{y_t\}$ is a stationary stochastic process such that

$$y_t = x_t + \zeta_t \tag{1}$$

where:

(A1) $\{x_t\}$ is a regular, homogeneous, discrete-time Markov chain;
(A2) $\{\zeta_t\}$ is a zero-mean, white noise process with variance σ_ζ^2.

Assumption (A2) is imposed for simplicity and can be relaxed without affecting the results presented in the paper. It is necessary, however, to suppose that:

(A3) the noise $\{\zeta_t\}$ is independent of the signal $\{x_t\}$.

The signal component is specified by giving the initial conditions and the transition probability matrix $\boldsymbol{P} = [p_{ij}]$, $i, j = 1, \cdots, n$ where n is the state dimension of the Markov chain and $p_{ij} = \Pr(x_{t+1} = s_i \mid x_t = s_j)$ with $\boldsymbol{s} = (s_1, \cdots, s_n) \in \mathbb{R}^n$ being the state space. Since in what follows we will only be interested in steady state properties we have supposed that $\{x_t\}$ is regular. This guarantees that there exists a stationary distribution $\boldsymbol{p} = (p_{1\cdot}, \cdots, p_{n\cdot})'$, $\sum_{i=1}^n p_{i\cdot} = 1$, such that $\boldsymbol{p} = \boldsymbol{P}\boldsymbol{p}$, the dominant eigenvalue of $\boldsymbol{P}$ being one, with all other eigenvalues being inside the unit circle. Thus we can regard $\{x_t\}$ as being characterized by its transition probability matrix $\boldsymbol{P}$ and state space vector $\boldsymbol{s}$, see Feller (1966) or Karlin (1966).

It is clear that the process $\{y_t\}$ will inherit its stochastic properties from the unobservable signal and noise components and the statistical question that we wish to address is whether the precise nature of this inheritance can be exploited so as to construct a method of identifying the state dimension n from data on $\{y_t\}$. In what follows we will show that this can in fact be done by first establishing a stochastic realization theorem which states that observations on $\{y_t\}$ can be synthesized as data generated from an autoregressive moving-average (ARMA) process in which the order of the process is an elementary function of n. This is the subject matter of the following section. In Section 3 we then use this characterization to motivate the construction of a statistical algorithm for identifying n. Section 4 brings together the subject matter of Sections 2 and 3 to show that the technique being advocated will provide strongly consistent estimates of the state dimension, and Section 5 presents some asymptotic distribution theory that forms a basis for the practical use of the procedure.

Before continuing it should perhaps be pointed out that the determination of the state dimension of a hidden Markov chain is one of the key problems of interest that must be solved in order for signal processing methods based on hidden Markov chain models to be of use in biophysical applications. The fluctuation of currents resulting from the opening and closing of an ion channel cannot be generally modeled as a first-order binary Markov chain. Numerous types of channel show multiple current levels: instead of switching between the fully open and closed states directly, currents can dwell at intermediate levels for variable lengths of time. In some cases, these sub-levels are clearly spaced in equal steps. Such an n-state Markov process may be envisaged as a superposition of $n-1$ identical binary homogeneous Markov processes, which are neither totally independent nor fully synchronized but partially coupled. This means that, in terms of our observations, we will see n distinct current levels although, once buried in noise, it may not be immediately evident from laboratory measurements. To gain insight into the underlying physical mechanism, one needs to assess the number of elementary channels contributing to the observed process and quantify whether the constituent channels are independent or have some form of linkage. The identification of the number of states of a hidden Markov chain is thus of considerable theoretical importance in cell membrane kinetics.

Despite this significance the estimation of state dimension in hidden Markov chain processes has received little attention. The only other work that we are aware of is an unpublished PhD thesis by Finesso (1990), which uses the idea of a penalized likelihood function to construct an estimator of n, an approach that is closely related to that adopted by Marhav *et al.* (1989). Unfortunately, practical experience indicates that these methods do not work well, experience that appears to be supported by the recent results of Nadas and Mercer (1994). Moreover, in the applications that we have in mind data sets in excess of 100,000 observations are not uncommon. In such situations it is of some importance to reduce the computational burden associated with maximizing a highly nonlinear likelihood function, particularly as the likelihood function can be multi-modal and standard algorithms can appear to have converged at the wrong point, see Nadas and Mercer (1994). Hence we have chosen to base our approach on least-squares type calculations for which fast, recursive algorithms are readily available. The method relies on the representation of $\{y_t\}$ as an ARMA process. The problem of order estimation in ARMA processes has given rise to an extensive literature, much of which is due to Hannan and his co-authors (see Hannan (1980), Hannan and Rissanen (1982) and Hannan and Kavalieris (1984) for example), and it would seem that any of the currently available procedures designed to handle this problem might be appropriate for identifying n. Evidence on the performance of these techniques does not look promising, however. We have felt it necessary, therefore, to devise an approach to the determination of n that takes explicit account of the special structure inherent in a hidden Markov chain process. The technique that we propose is also of interest in its own right. It depends on a method of obtaining initial estimates described in Hannan (1970, pp. 390–391) that appears to have lain dormant for over two decades without being further investigated and the selection process is conducted via a sequence of statistical tests. Its appeal lies in the fact that it obviates the need for the user to specify external design parameters, such as the order of a preliminary autoregressive approximation or an *a priori* selection of a choice set from which the preferred specification is to be selected, in order to implement the procedure in practice.

2. Realization properties

From (1) and the basic assumptions it is clear that the moments of $\{y_t\}$ can be calculated from those of $\{x_t\}$ and $\{\zeta_t\}$. In particular, $\mathbb{E}[y_t] = \mu_y = \mu_x = \mathbb{E}[x_t]$, and $\mathbb{E}[y_t y_{t+k}] = \mathbb{E}[x_t x_{t+k}] + \delta_{0,k}\sigma_\zeta^2$, $k = 0, \pm 1, \pm 2, \cdots$, where $\delta_{j,k}$ denotes the Kronecker delta. It follows that

$$\gamma_y(k) = \gamma_x(k) + \delta_{0,k}\sigma_\zeta^2, \qquad k = 0, \pm 1, \pm 2, \cdots \tag{2}$$

wherein we have employed an obvious notation for the autocovariances of a process. Using (2) we can establish the following basic results.

Theorem 2.1. *Let $\{y_t\}$ denote a hidden Markov chain satisfying assumptions (A1)–(A3). Then there exists a zero mean white noise process $\{\varepsilon_t\}$ with variance σ_ε^2, defined on the same probability space as $\{y_t\}$, such that $y_t - \mu_y = \varepsilon_t + \sum_{j=1}^{\infty} K_j \varepsilon_{t-j}$, where $K(z) = 1 + \sum_{j\geq 1} K_j z^{-j}$ is rational. The order of $K(z)$ is $n-1$ and there exist coprime polynomials of degree $n-1$, $\alpha(z) = 1 + \sum_{j=1}^{n-1} \alpha_j z^{-j}$ and $\mu(z) = 1 + \sum_{j=1}^{n-1} \mu_j z^{-j}$, with zeroes inside the unit circle $|z| = 1$, such that $K(z) = \mu(z)/\alpha(z)$.*

Moreover, $\alpha(z)=\prod_{j=2}^{n}(1-\lambda_j z^{-1})$, *where* λ_j, $j=2,\cdots,n$, *are the sub-dominant eigenvalues of the transition probability matrix, each counted with its algebraic multiplicity.*

The implication of this theorem is, of course, that realizations of the process $\{y_t\}$ are equivalent to observations on a stable and invertible mixed autoregressive moving-average scheme of order $(n-1,\ n-1)$. The proof of the theorem arises naturally from the lemmas presented immediately below.

Before we proceed, let us introduce the following symbolism and definitions. Set $\boldsymbol{P}_\infty=[\boldsymbol{p}:\boldsymbol{p}:\cdots:\boldsymbol{p}]$. Since $\{x_t\}$ is regular, $\boldsymbol{P}_\infty=\lim_{k\to\infty}\boldsymbol{P}^k$, see Feller (1966) or Karlin (1966) once again, and it follows that $\boldsymbol{P}_\infty=\boldsymbol{P}\boldsymbol{P}_\infty=\boldsymbol{P}_\infty\boldsymbol{P}$ and $\boldsymbol{P}_\infty$ is idempotent. Now let $\boldsymbol{F}=\boldsymbol{P}-\boldsymbol{P}_\infty$. Then using these properties it is straightforward to establish the relationships $(\boldsymbol{P}^k-\boldsymbol{P}_\infty)=(\boldsymbol{P}^{k-1}-\boldsymbol{P}_\infty)\boldsymbol{F}=\boldsymbol{F}^{k-1}\boldsymbol{F}=\boldsymbol{F}^k$, $k=1,2,\cdots$. Set $\boldsymbol{S}=\operatorname{diag}(s_1,\cdots,s_n)$, $\boldsymbol{R}=\operatorname{diag}(p_{1\cdot},\cdots,p_{n\cdot})$ and let $\mathbf{1}=(1,\cdots,1)'$ denote the n element sum vector.

Lemma 2.2. *If* $\{x_t\}$ *is a Markov chain process satisfying assumption* (*A*1) *then* $\mu_x=\boldsymbol{p}'\boldsymbol{s}$, $\gamma_x(0)=\mathbf{1}'\boldsymbol{S}(\boldsymbol{I}-\boldsymbol{P}_\infty)\boldsymbol{S}\boldsymbol{p}$ *and* $\gamma_x(k)=\mathbf{1}'\boldsymbol{S}\boldsymbol{F}^{|k|}\boldsymbol{S}\boldsymbol{p}$, $k=\pm1,\pm2,\cdots$.

Proof. By definition $\mu_x=\sum_{i=1}^{n}s_i\Pr(x_t=s_i)=\boldsymbol{p}'\boldsymbol{s}$ and

$$\mathbb{E}[x_tx_{t+k}]=\sum_{i=1}^{n}\sum_{j=1}^{n}s_is_j\Pr(x_{t+k}=s_i\mid x_t=s_j)\Pr(x_t=s_j)=\mathbf{1}'\boldsymbol{S}\boldsymbol{P}^k\boldsymbol{S}\boldsymbol{p},\quad k=0,1,2,\cdots,$$

where we have used the fact that the entries of $\boldsymbol{P}^k$ are the k-step transition probabilities. It is easy to verify that $(\boldsymbol{p}'\boldsymbol{s})^2=\sum_{i=1}^{n}\sum_{j=1}^{n}s_is_jp_{i\cdot}p_{j\cdot}=\mathbf{1}'\boldsymbol{S}\boldsymbol{P}_\infty\boldsymbol{S}\boldsymbol{P}$ and hence it follows that $\gamma_x(0)=\mathbf{1}'\boldsymbol{S}(\boldsymbol{P}^0-\boldsymbol{P}_\infty)\boldsymbol{S}\boldsymbol{p}$ and $\gamma_x(k)=\mathbf{1}'\boldsymbol{S}(\boldsymbol{P}^k-\boldsymbol{P}_\infty)\boldsymbol{S}\boldsymbol{p}=\mathbf{1}'\boldsymbol{S}\boldsymbol{F}^k\boldsymbol{S}\boldsymbol{P}$, $k=1,2,\cdots$.

Similarly, if $q_{ij}=\Pr(x_{t-1}=s_i\mid x_t=s_j)$, $i,j=1,\cdots,n$, then $\mathbb{E}[x_{t-k}x_t]=\mathbf{1}'\boldsymbol{S}\boldsymbol{Q}^k\boldsymbol{S}\boldsymbol{p}$, but $\boldsymbol{Q}=\boldsymbol{R}\boldsymbol{P}'\boldsymbol{R}^{-1}$ and $\mathbf{1}'\boldsymbol{S}\boldsymbol{R}(\boldsymbol{P}')^k\boldsymbol{R}^{-1}\boldsymbol{S}\boldsymbol{p}=\boldsymbol{p}'\boldsymbol{S}(\boldsymbol{P}')^k\boldsymbol{S}\mathbf{1}$ because $\boldsymbol{S}\boldsymbol{p}=\boldsymbol{R}\boldsymbol{S}\mathbf{1}$. Hence $\gamma_x(-k)=\mathbf{1}'\boldsymbol{S}\boldsymbol{F}^k\boldsymbol{S}\boldsymbol{p}$, $k=1,2,\cdots$, and $\gamma_x(k)=\mathbf{1}'\boldsymbol{S}\boldsymbol{F}^{|k|}\boldsymbol{S}\boldsymbol{p}$, $k=\pm1,\pm2,\cdots$, as required.

If $S_x(\omega)$ denotes the spectral density of $\{x_t\}$, then

$$\begin{aligned}2\pi S_x(\omega)&=\sum_{j=-\infty}^{\infty}\gamma_x(j)e^{(-i\omega j)},\ -\pi\leqq\omega\leqq\pi,\\&=\mathbf{1}'\boldsymbol{S}\left[(\boldsymbol{I}-\boldsymbol{P}_\infty)+2\mathbb{R}\sum_{j=1}^{\infty}\boldsymbol{F}^je^{(-i\omega j)}\right]\boldsymbol{S}\boldsymbol{p}\\&=\mathbf{1}'\boldsymbol{S}[(\boldsymbol{I}-\boldsymbol{P}_\infty)+2\mathbb{R}e^{(-i\omega)}\boldsymbol{F}(\boldsymbol{I}-e^{(-i\omega)}\boldsymbol{F})^{-1}]\boldsymbol{S}\boldsymbol{p},\end{aligned}\tag{3}$$

the last line being a consequence of the fact that for any matrix $\boldsymbol{A}$ all of whose eigenvalues are less than one in modulus the Laurent expansion $\sum_{j\geqq0}\boldsymbol{A}^j=(\boldsymbol{I}-\boldsymbol{A})^{-1}$ is valid and the eigenvalues of $\boldsymbol{F}$ are zero and λ_i, $|\lambda_i|<1$, $i=2,\cdots,n$, the sub-dominant eigenvalues of $\boldsymbol{P}$ counted with their algebraic multiplicity. Note that the characteristic values of $\boldsymbol{P}-\boldsymbol{P}^k$ are $\lambda_i-\lambda_i^k$, $i=1,\cdots,n$, and since $\boldsymbol{F}=\lim_{k\to\infty}\boldsymbol{P}-\boldsymbol{P}^k$ and the eigenvalues of a matrix are holomorphic functions of the elements of the matrix the eigenvalues of $\boldsymbol{F}$ equal $\lim_{k\to\infty}\lambda_i-\lambda_i^k$, $i=1,\cdots,n$.

To investigate the structure of (3) further, observe that $z^{-1}\boldsymbol{F}(\boldsymbol{I}-z^{-1}\boldsymbol{F})^{-1}=\boldsymbol{F}(\boldsymbol{I}z-\boldsymbol{F})^{-1}=\boldsymbol{QJ}(\boldsymbol{I}z-\boldsymbol{J})^{-1}\boldsymbol{Q}^{-1}$ where $\boldsymbol{J}=\boldsymbol{Q}^{-1}\boldsymbol{FQ}$ is the Jordan canonical form of $\boldsymbol{F}$. Given that $\boldsymbol{F}$ is non-derogatory the Jordan matrix $\boldsymbol{J}=\operatorname{diag}(\boldsymbol{J}_1:\cdots:\boldsymbol{J}_u)$ where each diagonal block $\boldsymbol{J}_i$ is a $v_i\times v_i$ matrix of the form

$$\begin{pmatrix} \varphi_i & 1 & 0 & \cdots & 0 \\ 0 & \varphi_i & 1 & \cdots & 0 \\ \vdots & & & & \vdots \\ & & & & 1 \\ 0 & & & & \varphi_i \end{pmatrix},$$

$i=1,\cdots,u$, the φ_i, $i=1,\cdots,u$, being the *unique* eigenvalues of $\boldsymbol{F}$ and $v_i\geqq 1$, $v_1+\cdots+v_u=n$, their respective algebraic multiplicities. Set $\boldsymbol{Q}=[\boldsymbol{Q}_1:\cdots:\boldsymbol{Q}_u]$ and $\boldsymbol{Q}^{-1}=[\boldsymbol{Q}^1:\cdots:\boldsymbol{Q}^u]'$ where $\boldsymbol{Q}_i$ and $\boldsymbol{Q}^i$ are $n\times v_i$ matrices, $i=1,\cdots,u$, that determine partitions of $\boldsymbol{Q}$ and its inverse conformable with the block diagonality of $\boldsymbol{J}$. A trivial calculation shows that $\boldsymbol{J}(\boldsymbol{I}z-\boldsymbol{J})^{-1}=\operatorname{diag}(\boldsymbol{J}_1(\boldsymbol{I}z-\boldsymbol{J}_1)^{-1}:\cdots:\boldsymbol{J}_n(\boldsymbol{I}z-\boldsymbol{J}_u)^{-1})$ and hence that

$$\boldsymbol{QJ}(\boldsymbol{I}z-\boldsymbol{J})^{-1}\boldsymbol{Q}^{-1}=\sum_{i=1}^{u}\boldsymbol{Q}_i\boldsymbol{J}_i(\boldsymbol{I}z-\boldsymbol{J}_i)^{-1}(\boldsymbol{Q}^i)'. \tag{4}$$

Now let $q_{ij},\cdots,q_{iv_i}$ and $q^{i1},\ldots,q^{iv_i}$ denote the columns of $\boldsymbol{Q}_i$ and $\boldsymbol{Q}^i$ respectively. Via some straightforward, if somewhat tedious, algebra we find that $\boldsymbol{J}_i(\boldsymbol{I}z-\boldsymbol{J}_i)^{-1}$ equals an upper triangular Toeplitz matrix with first row $[\varphi_i(z-\varphi_i)^{-1}, (z-2\varphi_i)(z-\varphi_i)^{-2},\cdots,(z-2\varphi_i)(z-\varphi_i)^{-v_i}(-1)^{v_i}]$, from which it can be deduced that each of the summands in (4) gives rise to an expansion of the form

$$\boldsymbol{R}_{i1}\frac{\varphi_i}{(z-\varphi_i)}+\sum_{j=2}^{v_i}\boldsymbol{R}_{ij}\frac{(z-2\varphi_i)(-1)^j}{(z-\varphi_i)^j} \tag{5}$$

where the second and subsequent terms only appear if $v_i\geqq 2$ and

$$\boldsymbol{R}_{ij}=\sum_{l=1}^{v_i-j+1}q_{il}(q^{il+j-1})',\qquad j=1,\cdots,v_i. \tag{6}$$

Collecting the components of the expansion together in terms of the lowest common denominator yields the result that $\boldsymbol{1}'\boldsymbol{SQ}_i\boldsymbol{J}_i(\boldsymbol{I}z-\boldsymbol{J}_i)^{-1}(\boldsymbol{Q}^i)'\boldsymbol{Sp}=b_i(z)/\beta_i(z)$ where

$$b_i(z)=\boldsymbol{1}'\boldsymbol{S}\left[\boldsymbol{R}_{i1}\varphi_i(z-\varphi_i)^{v_i-1}+\sum_{j=2}^{v_i}\boldsymbol{R}_{ij}(z-2\varphi_i)(z-\varphi_i)^{v_i-j}(-1)^j\right]\boldsymbol{Sp} \tag{7}$$

and

$$\beta_i(z)=(z-\varphi_i)^{v_i},\qquad i=1,\cdots,u. \tag{8}$$

Substituting into (4) it follows that $\boldsymbol{1}'\boldsymbol{S}(\boldsymbol{QJ}(\boldsymbol{I}z-\boldsymbol{J})^{-1}\boldsymbol{Q}^{-1})\boldsymbol{Sp}=\sum_{i=2}^{u}b_i(z)/\beta_i(z)$, the index of summation beginning at two because $\varphi_1 = 0$ and $\nu_1 = 1$. This leads us to the conclusion that

$$
\begin{aligned}
2\pi S_x(\omega) &= \gamma_x(0) + \sum_{i=2}^{u} \frac{b_i(e^{i\omega})}{\beta_i(e^{i\omega})} + \frac{b_i(e^{-i\omega})}{\beta_i(e^{-i\omega})} \\
&= \gamma_x(0) + \frac{a(e^{i\omega})}{\alpha(e^{i\omega})} + \frac{a(e^{-i\omega})}{\alpha(e^{-i\omega})}
\end{aligned} \tag{9}
$$

where

$$
\alpha(z) = z^{1-n} \prod_{i=2}^{u} \beta_i(z) = \prod_{j=2}^{n} (1 - \lambda_j z^{-1}) \tag{10}
$$

and

$$
a(z) = z^{1-n} \sum_{i=2}^{u} \left\{ b_i(z) \prod_{j=2, j\neq i}^{u} \beta_j(z) \right\}. \tag{11}
$$

From (10) and (11) it is clear that the degree of $z^{n-1}\alpha(z)$ is $\nu_2 + \cdots + \nu_u = n - 1$. From expression (7) it is readily verified that the degree of $b_i(z)$ is $\nu_i - 1$, at most, and that $b_i(\varphi_i) = \mathbf{1}'\boldsymbol{S}\boldsymbol{R}_{i\nu_i}\boldsymbol{S}\boldsymbol{P}(-\varphi_i)(-1)^{\nu_i} \neq 0$. This implies, via (11), that the degree of $z^{n-1}a(z)$ is $\nu_2 + \cdots + \nu_u - 1 = n - 2$ and, since $b_i(z)$ and $\beta_i(z)$ are relatively prime and $\beta_i(z)$ and $\beta_j(z)$ are obviously coprime $i \neq j$, that the ratio $a(z)/\alpha(z)$ is irreducible.

Placing the three terms on the right-hand side of (9) over a common denominator yields the result that

$$
\begin{aligned}
S_x(\omega) &= \frac{1}{2\pi} \frac{\rho(e^{i\omega})}{\alpha(e^{i\omega})\alpha(e^{-i\omega})}, \\
\rho(z) &= \sum_{j=-n+1}^{n-1} \rho_j z^{-j} \\
&= a(z)\alpha(z^{-1}) + \gamma_x(0)\alpha(z^{-1}) + a(z^{-1})\alpha(z).
\end{aligned}
$$

The polynomial $z^{n-1}\rho(z)$ has $2(n-1)$ roots because $\rho_{n-1} = \rho_{1-n} = \gamma_x(0)\alpha_{n-1} \neq 0$ and we may number and group these into two sets $\{\nu_1, \cdots, \nu_{n-1}\}$ and $\{\bar{\nu}_1^{-1}, \cdots, \bar{\nu}_{n-1}^{-1}\}$ such that $|\nu_j| \leqq 1$, $j = 1, \cdots, n-1$, because $\rho_j = \rho_{-j}$, $j = 1, \cdots, n-1$ and $\rho(z)$ is real and non-negative on the unit circle $|z| = 1$. Since the coefficients of $\rho(z)$ are real the roots are real or occur in complex conjugate pairs. Moreover, $\rho(\lambda_j) = \rho(\lambda_j^{-1}) = a(\lambda_j)\alpha(\lambda_j^{-1}) \neq 0$, $j = 2, \cdots, n$, because $a(z)$ and $\alpha(z)$ are coprime. Hence we can construct the operator $m(z) = m_0 \prod_{j=1}^{n-1} (1 - \nu_j z^{-1})$ where $m_0 = \{\prod(-\nu_j)^{-1}\rho_{n-1}\}^{1/2}$ such that

$$
\begin{aligned}
\rho(z) &= \rho_{n-1} z^{-n+1} \prod_{j=1}^{n-1} (1 - \nu_j z)(1 - \nu_j^{-1} z) \\
&= m(z)m(z^{-1})
\end{aligned}
$$

and $m(z)$ and $\alpha(z)$ will be relatively prime. We have, therefore, established the following property.

Lemma 2.3. The spectral density of a Markov chain process $\{x_t\}$ *satisfying assumption* (A1) *can be expressed in the rational form*

$$S_x(\omega) = \frac{1}{2\pi}\frac{|m(e^{i\omega})|^2}{|\alpha(e^{i\omega})|^2}, \qquad -\pi < \omega < \pi,$$

where $m(z) = m_0 + m_1 z^{-1} + \cdots + m_{n-1}z^{-n+1}$ *and* $\alpha(z) = 1 + \alpha_1 z^{-1} + \cdots + \alpha_{n-1}z^{-n+1}$ *are relatively prime and* n *is the state dimension. Furthermore, if* λ_j, $j = 2, \cdots, n$, *are the sub-dominant eigenvalues of* $\boldsymbol{P}$, *then* $\alpha(z) = \prod_{j=2}^{n}(1 - \lambda_j z^{-1})$.

We remark that alternative formulations for the moments and power spectrum of a finite state Markov chain have appeared in the literature. The derivations of μ_x and $\gamma_x(k)$, $k = 0, \pm 1, \cdots$ given above parallel those employed by Fredkin and Rice (1987) in the continuous time case and expansions of $S_x(\omega)$ in terms of partial fractions, analogous to (9), have been presented in the communications literature, see Lee and Messerschmitt (1988, Appendix B3), for example. In communications, however, the emphasis is on Markov chains that exhibit special structure, coded signals or shift register processes for instance, and interest focuses on the production of closed form expressions for the power spectrum suitable for direct numerical evaluation. The formulae presented here are perfectly general and apply to any discrete time Markov chain process. We are also concerned with the analytic properties of the power spectrum rather than computation and it is the result given in Lemma 2.3, that $S_x(\omega)$ is rational, that is critical to our theoretical developments. Lemma 2.4 given immediately below may be viewed as providing a generalization of results presented in Pagano (1974, Section 1), see also Engel (1984).

Lemma 2.4. Define the process $\{v_t\}$ *via the equation* $v_t = y_t + \alpha_1 y_{t-1} + \cdots + \alpha_{n-1}y_{t-n+1}$ *where* $\{y_t\}$ *is a hidden Markov chain process satisfying assumptions* (A1)–(A3) *and* $\alpha(z)$ *is determined as in* (10). *Then* $\{v_t\}$ *is a moving-average process of order* $n-1$ *and there exists a zero-mean white noise* $\{\varepsilon_t\}$ *with variance* σ_ε^2, *defined on the same probability space as* $\{y_t\}$, *such that* $v_t = \varepsilon_t + \mu_1\varepsilon_{t-1} + \cdots + \mu_{n-1}\varepsilon_{t-n+1}$, *the coefficients* $\mu_1, \cdots, \mu_{n-1}$ *having been chosen so that the autocovariance generating function of* $\{v_t\}$ *equals* $\sigma_\varepsilon^2\mu(z)\mu(z^{-1})$ *and* $\mu(z) \neq 0$, $\|z\| \geqq 1$.

Proof. From Lemma 2.3 and standard results on linear filtering we know that the spectral density of $\{v_t\}$ satisfies $2\pi S_v(\omega) = |m(e^{i\omega})|^2 + \sigma_\zeta^2|\alpha(e^{i\omega})|^2$, which implies that the autocovariance generating function

$$\begin{aligned} r(z) &= \sum_{j=-(n-1)}^{n-1} r_j z^{-j}, r_j = r_{-j}, j = 1, \cdots, n-1, \\ &= m(z)m(z^{-1}) + \sigma_\zeta^2\alpha(z)\alpha(z^{-1}). \end{aligned}$$

As in the derivation of Lemma 2.3 we can now select $n-1$ roots of $r(z)$ to construct $\mu(z) = 1 + \mu_1 z^{-1} + \cdots + \mu_{n-1}z^{-n+1}$ such that $r(z) = \sigma_\varepsilon^2\mu(z)\mu(z^{-1})$. Since $r(e^{i\theta}) = r(e^{-i\theta}) = |m(e^{i\theta})|^2 + \sigma_\zeta^2|\alpha(e^{i\theta})|^2 > 0$, $-\pi < \theta \leqq \pi$, $r(z)$ has no zeros on the unit circle so the roots may be chosen in such a way that $\mu(z) \neq 0$, $|z| \geqq 1$. The existence of the white noise process $\{\varepsilon_t\}$ providing a moving-average representation of $\{v_t\}$ now follows from the spectral factorization theorem, Rozanov (1967, Theorem 9.1, p. 41).

Armed with the above lemmas it is readily established that $S_y(\omega) = \sigma_\varepsilon^2 |K(e^{i\omega})|^2/2\pi$ where $K(z) = \mu(z)/\alpha(z)$. All that remains in order to complete the proof of Theorem 2.1 is to confirm that $\mu(z)$ and $\alpha(z)$ are coprime. Recall that the roots of $\alpha(z)$ are λ_j, $j = 2, \cdots, n$, the sub-dominant eigenvalues of $\boldsymbol{P}$, and $\alpha(z)$ and $m(z)$ are relatively prime. It follows that $\sigma_\varepsilon^2 \mu(\lambda_j)\mu(\lambda_j^{-1}) = m(\lambda_j)m(\lambda_j^{-1}) \neq 0$, $j = 2, \cdots, n$, as required. Thus $K(z) = \mu(z)/\alpha(z)$ is irreducible and the order of $K(z)$, the maximum of the degrees of $\mu(z)$ and $\alpha(z)$, is clearly $n-1$.

At this point let us note that we will require some kind of ergodicity assumption in order to analyse the properties of the statistical procedure described in the following section. We will formulate this now in terms of the innovations $\{\varepsilon_t\}$.

(A4) The innovation $\{\varepsilon_t\}$ is a stationary and ergodic process satisfying sufficient regularity conditions to ensure that for $H_T = (\log T)^a$, $1 \leqq a < \infty$, and $Q_T^2 = \log\log T/T$, where T is the sample size, the quantity

$$T^{-1} \sum_{t=1}^{T-|r|} y_t y_{t+|r|} = \mathbb{E}[y_t y_{t+|r|}] + O(Q_T)$$

for $r = 0, \pm 1, \cdots, \pm H_T$ and

$$T^{-1} \sum_{t=1}^{T-r} y_t \varepsilon_{t+r} = O(Q_T)$$

for $r = 1, \cdots, H_T$. Furthermore, the variables $T^{-1/2} \sum \varepsilon_t \varepsilon_{t-r}$, $r = 0, 1, \cdots, H_T$, converge in distribution to zero mean Gaussian random variates with covariance $\delta_{i,j}\sigma_\varepsilon^4$.

A sufficient condition that will ensure that (A4) holds commonly used in time series analysis is to suppose that $\{\varepsilon_t\}$ is a martingale difference process, see Theorems 5.3.1 and 5.3.2 of Hannan and Deistler (1988) for example. Although such an assumption seems quite natural in the current context, it is in fact inappropriate. The basic difficulty lies in the fact that although $\{\varepsilon_t\}$ is uncorrelated it is not independent, so the optimal mean squared error predictor is no longer linear and it is possible to gain information from higher-order moments than the first and second. Nevertheless, assumption (A4) still seems not unreasonable. It is possible to show that a Markov chain process is mixing, see Blum *et al.* (1963), and if $\{\zeta_t\}$ is also mixing then $\{y_t\}$ will be mixing. Inverting the representation given in Theorem 2.1 and expressing ε_t as a linear combination of y_{t-j}, $j = 0, 1, \cdots$, with geometrically declining weights leads to the conclusion that $\{\varepsilon_t\}$ is near epoch dependent. Thus (A4) may follow as a consequence of the law of large numbers and the central limit theorem applicable to near epoch dependent processes. Gallant and White (1988) provide an illuminating discussion of such processes.

3. Statistical algorithm

Let y_t, $t=1,\cdots,T$ denote a realization of T observations on a hidden Markov chain $\{y_t\}$ satisfying assumptions (A1)–A(4). The algorithm for identifying the state dimension n is described in the following steps.

Step 0. Evaluate the sample average $\bar{y}=T^{-1}\sum_{i=1}^{T} y_t$ and calculate the mean corrected values $\bar{y}_t=y_t-\bar{y}$, $t=1,\cdots,T$. Set $n=1$ and $\tilde{\sigma}^2_{\varepsilon,n}=T^{-1}\sum_{t=1}^{T}\bar{y}_t^2$.

Step 1. Put $n=n+1$.

Stage (i): Evaluate the initial estimates $\tilde{\alpha}_j^{(0)}$, $j=1,\cdots,n-1$, by solving the equations

$$\sum_{j=0}^{n-1}\tilde{\alpha}_j^{(0)}C_y(r-j)=0,\qquad r=n,\cdots,2(n-1),$$

where $C_y(r)=C_y(-r)=T^{-1}\sum_{t=1}^{T-|r|}\bar{y}_t\bar{y}_{t+|r|}=(2\pi/T')\sum_{j=1}^{T'}I_{T,y}(e^{i\omega_j'})e^{i\omega_j' r}$, $r=1,\cdots,T-1$, and $I_{T,y}(z)=(2\pi T)^{-1}|\sum_{t=1}^{T}y_t z^{-t}|^2$ with $\omega_j'=2\pi j/T'$, $T'\geqq 2T$, $j=1,\cdots,T'$.

Stage (ii): Now form for $r=1,\cdots,n-1$,

$$\begin{aligned}\tilde{C}_v(r)=\tilde{C}_v(-r)&=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\tilde{\alpha}_i^{(0)}\tilde{\alpha}_j^{(0)}C_y(r+i-j)\\&=(2\pi/T')\sum_{j=1}^{T'}|\tilde{\alpha}^{(0)}(e^{i\omega_j'})|^2 I_{T,y}(e^{i\omega_j'})e^{i\omega_j' r}\end{aligned}$$

where $\tilde{\alpha}^{(0)}(z)=1+\tilde{\alpha}_1^{(0)}z^{-1}+\cdots+\tilde{\alpha}_{n-1}^{(0)}z^{1-n}$ and set $\tilde{S}_v^{(0)}(z)=(2\pi)^{-1}\sum_{r=-n+1}^{n-1}\tilde{C}_v(r)z^{-r}$.

Compute the initial estimates $\tilde{\mu}_j^{(0)}$, $j=1,\cdots,n-1$, from the equation system $\sum_{j=0}^{n-1}\tilde{\mu}_j^{(0)}(\sum_{r=-T+1}^{T-1}C_y(r)\tilde{h}^{(0)}(l-j-r))=0$, $l=1,\cdots,n-1$, where

$$\tilde{h}^{(0)}(u)=\frac{2\pi}{T'}\sum_{j=0}^{T'-1}\frac{|\tilde{\alpha}^{(0)}(e^{i\omega_j'})|^2}{\tilde{S}_v^{(0)}(e^{i\omega_j'})^2}e^{i\omega_j' u}.$$

Step 2. For $i=1,2,\cdots$ generate the processes $\tilde{\varepsilon}_t^{(i)}$, $\tilde{\eta}_t^{(i)}$ and $\tilde{\varphi}_t^{(i)}$, $t=1,\cdots,T$ using the recursions

$$\tilde{\varepsilon}_t^{(i)}=\bar{y}_t+\sum_{j=1}^{n-1}(\tilde{\alpha}_j^{(i-1)}\bar{y}_{t-j}-\tilde{\mu}_j^{(i-1)}\tilde{\varepsilon}_{t-j}^{(i)})$$

$$\tilde{\eta}_t^{(i)}=\tilde{\varepsilon}_t^{(i)}-\sum_{j=1}^{n-1}\tilde{\alpha}_j^{(i-1)}\,\tilde{\eta}_{t-j}^{(i)}$$

$$\tilde{\varphi}_t^{(i)}=\tilde{\varepsilon}_t^{(i)}-\sum_{j=1}^{n-1}\tilde{\mu}_j^{(i-1)}\,\tilde{\varphi}_{t-j}^{(i)}$$

and calculate the residual mean square

$$(\tilde{\sigma}_{\varepsilon,n}^{(i)})^2=T^{-1}\sum_{t=1}^{T}\left(\tilde{\varepsilon}_t^{(i)2}+\sum_{j=1}^{n-1}(\delta\tilde{\alpha}_j^{(i)}\tilde{\eta}_{t-j}^{(i)}-\delta\tilde{\mu}_j^{(i)}\tilde{\varphi}_{t-j}^{(i)})\tilde{\varepsilon}_t^{(i)}\right)$$

where the parameter adjustments $\delta\tilde{\alpha}_j^{(i)}$ and $\delta\tilde{\mu}_j^{(i)}$, $j=1,\cdots,n-1$ are obtained from the regression of $\tilde{\varepsilon}_t^{(i)}$ on $-\tilde{\eta}_{t-j}^{(i)}$ and $\tilde{\varphi}_{t-j}^{(i)}$, $j=1,\cdots,n-1$. Construct the revised estimates $\tilde{\alpha}_j^{(i)}=\tilde{\alpha}_j^{(i-1)}+\delta\tilde{\alpha}_j^{(i)}$ and $\tilde{\mu}_j^{(i)}=\tilde{\mu}_j^{(i-1)}+\delta\tilde{\mu}_j^{(i)}$, $j=1,\cdots,n-1$, and repeat the iterations until $(\tilde{\sigma}_{\varepsilon,n}^{(i)})^2$ converges. Let $\tilde{\sigma}^2_{\varepsilon,n}=\lim_{i\to\infty}(\tilde{\sigma}_{\varepsilon,n}^{(i)})^2$.

Step 3. Evaluate the statistic $\Lambda_T(n-1)=T(\tilde{\sigma}^2_{\varepsilon,n-1}-\tilde{\sigma}^2_{\varepsilon,n})/\tilde{\sigma}^2_{\varepsilon,n}$ and compare its value to $C_T(n-1)$ where $C_T(n-1)>0$ is a real valued function, non-decreasing in T. If $\Lambda_T(n-1) \geqq C_T(n-1)$ then return to Step 1. If $\Lambda_T(n-1) < C_T(n-1)$ set $\tilde{\eta}_T = n-1$.

Although not stated explicitly, it is assumed that when implementing the algorithm advantage will be taken of the numerical efficiency of the fast Fourier transform (FFT), as described in Bingham (1974), when computing the covariances and convolutions required to determine the initial estimates in Step 1. The FFT can also be used to evaluate the mean squares and cross product terms needed in subsequent calculations, although the use of the QR method in conjunction with fast Givens transformations to solve the least squares problem recursively as the regressand and regressor variables are generated, see Golub and Van Loan (1989), could be competitive at Step 2.

4. Consistency of the estimates

In this section of the paper we will follow the format of Section 2 by first stating the main theorem and then presenting a set of lemmas that form the basis of the proof of the theorem. Our main result presents conditions on the critical value $C_T(n)$ to be applied in conjunction with the statistic $\Lambda_T(n)$ that will ensure that the value $\tilde{n}_T$ determined via the procedure just presented will yield a consistent estimate of the true dimension n_o of the hidden Markov chain $\{y_t\}$.

Theorem 4.1. *Suppose that y_t, $t=1,\cdots,T$, is a realization of a hidden Markov chain process with state dimension n_o satisfying (A1)–A(4). If $\tilde{n}_T$ is obtained by implementing the above algorithm with $C_T(n)$ a possibly stochastic function of n and T, then: (i) if $C_T(n)/T \to 0$ almost surely as $T\to\infty$ then $\tilde{n}_T \geqq n_o$ with arbitrarily large probability, as $T\to\infty$; (ii) if as $T\to\infty$, $\liminf C_T(n)/L(T)>0$ almost surely, where $L(T)$ is a real valued, increasing function of T such that $\log\log T/L(T) \to 0$, then* $\Pr(\lim_{T\to\infty} \tilde{n}_T \leqq n_o) = 1$.

In what follows we will append a subscript o to quantities of interest to indicate those values corresponding to the actual data generating mechanism giving rise to the observations, as we have already done for the state dimension. Thus, $\sigma^2_{\varepsilon o}$ and $K_o(z)$ will represent the true innovation variance and transfer function of the observed process $\{y_t\}$ and $\alpha_o(z)$ and $\mu_o(z)$ will denote the true autoregressive and moving-average operators with coefficients α_{oj} and μ_{oj}, $j=1,\cdots,n_o-1$, respectively.

Lemma 4.2. *Suppose that (A1)–(A4) obtain and set $\tilde{\alpha}^{(0)}(z) = 1+\tilde{\alpha}^{(0)}_1 z^{-1}+\cdots+\tilde{\alpha}^{(0)}_{n-1}z^{1-n}$ and $\tilde{\mu}^{(0)}(z) = 1+\tilde{\mu}^{(0)}_1 z^{-1}+\cdots+\tilde{\mu}^{(0)}_{n-1}z^{1-n}$.*

(i) *If $n<n_o$ define $\alpha^*(z)$ in an obvious manner via the solution to the equations*

$$\frac{\sigma^2_{\varepsilon o}}{2\pi}\int_{-\pi}^{\pi}\sum_{j=0}^{n-1}\alpha^*_j\,|K_o(e^{i\omega})|^2\,e^{i\omega(r-j)}d\omega = 0, \qquad r=n,\cdots,2(n+1). \tag{12}$$

Then $\tilde{\alpha}^{(0)}(z) = \alpha^*(z) + O(Q_T)$ *uniformly in* $|z| \geqq 1$, $Q_T = (\log\log T/T)^{\frac{1}{2}}$. *Given* $\alpha^*(z)$, *let* $\mu^*(z)$ *be formed from*

$$\frac{\sigma_{\varepsilon o}^2}{2\pi}\int_{-\pi}^{\pi}\sum_{j=0}^{n-1}\mu_j^*\frac{|\alpha^*(e^{i\omega})|^2\,|K_o(e^{i\omega})|^2}{S_v^*(e^{i\omega})^2}e^{i\omega(r-j)}\,d\omega = 0, \qquad r = 1, \cdots, n \tag{13}$$

where

$$S_v^*(e^{i\omega}) = \frac{\sigma_{\varepsilon o}^2}{4\pi^2}\int_{-\pi}^{\pi}\sum_{r=-n+1}^{n-1}|\alpha^*(e^{i\theta})|^2\,|K_o(e^{i\theta})|^2 e^{i(\theta-\omega)r}\,d\theta. \tag{14}$$

Then uniformly in $|z| \geqq 1$, $\tilde{\mu}^{(0)}(z) = \mu^*(z) + O(Q_T)$.

(ii) *If* $n > n_o$ *then* $\tilde{\alpha}^{(0)}(z) = \tilde{\phi}^{(0)}(z)\alpha_o(z) + O(Q_T)$ *and* $\tilde{\mu}^{(0)}(z) = \tilde{\phi}^{(0)}(z)\mu_o(z) + O(Q_T)$ *uniformly in* z, $|z| \geqq 1$, *where* $\tilde{\phi}^{(0)}(z) = 1 + \tilde{\phi}_1^{(0)}z^{-1} + \cdots + \tilde{\phi}_r^{(0)}z^{-r}$, $r = n - n_0$, $\tilde{\phi}^{(0)}(z) \neq 0$, $|z| \geqq 1$.

Proof. (i) From assumption (A4) we know that $|C_y(r) - \gamma_y(r)| = O(Q_T)$, $r = 0, \cdots, H_T \leqq (\log T)^a$, $a < \infty$. Re-expressing the autocovariance of $\{y_t\}$ in terms of the Fourier transform of the power spectrum in the usual way we recognize that $\tilde{\alpha}_j^{(o)}$ and α_j^*, $j = 1, \cdots, n-1$, correspond to the solutions of systems of linear equations in which the augmented matrices of the two systems differ by $O(Q_T)$. The coefficient matrix of the system of equations is non-singular, since

$$\sum_{s=1}^{n-1}\sum_{r=1}^{n-1}\beta_s\gamma_y(n+s-r-1)\beta_r = (\sigma_{\varepsilon o}^2/2\pi)\int_{-\pi}^{\pi}\left|\sum_{r=1}^{n-1}\beta_r e^{i\omega r}\right|^2|K_o(e^{i\omega})|^2\,e^{i\omega(n-1)}\,d\omega$$

and $(\sum_{r=1}^{n-1}\beta_r z^{-r})K_o(z)$ is rational of order at least $n_o - 1 \geqq n$ by the fundamental theorem of algebra and Theorem 2.1; see also An *et al.* (1983, Theorem 1). We are therefore led to the conclusion that $\sum_{j=1}^{n-1}|\tilde{\alpha}_j^{(0)} - \alpha_j^*| = O(Q_T)$ and the first statement of the lemma follows since $|\tilde{\alpha}^{(0)}(z) - \alpha^*(z)| \leqq \sum_{j=1}^{n-1}|\tilde{\alpha}_j^{(0)} - \alpha_j^*|$, $|z| \geqq 1$.

A completely parallel argument to that just employed will show that $\tilde{\mu}^{(0)}(z) = \mu^*(z) + O(Q_T)$, $|z| \geqq 1$, if it can be verified that the augmented matrices of the two equation systems defining these two operators are, likewise, $O(Q_T)$. To this end, let $\tilde{H}^{(0)}(\omega) = |\tilde{\alpha}^{(0)}(e^{i\omega})|^2/\tilde{S}_v^{(0)}(e^{i\omega})^2$ and set $h^*(u) = \int_{-\pi}^{\pi}H^*(\omega)e^{i\omega u}\,d\omega$, $u = 0, \pm 1, \cdots$, where $H^*(\omega) = |\alpha^*(e^{i\omega})|^2/S_v^*(e^{i\omega})^2$. Since $S_v^*(e^{i\omega})$ corresponds to the power spectrum of a finite moving average $|h^*(u)| < \kappa\lambda^{|u|}$, for some λ, $0 < \lambda < 1$, and $|h^*(u) - \sum_j h^*(u + jT')| < 2\kappa\exp(T'\log\lambda)/(1 - \lambda^{T'})$ where, here as elsewhere, κ is used to denote a fixed constant. We can also infer that $\tilde{H}^{(0)}(\omega) = H^*(\omega) + O(Q_T)$ uniformly in $\omega \in [-\pi, \pi]$. Now observe that

$$\sum_{r=-(T-1)}^{(T-1)}C_y(r)\tilde{h}^{(0)}(u-r) = \frac{2\pi}{T'}\sum_{j=0}^{T'-1}I_{T,y}(e^{i\omega_j'})\tilde{H}^{(0)}(\omega_j')e^{i\omega_j'u},$$

and that an equivalent expression obtains with $\tilde{h}^{(0)}(u-r)$ replaced by $\sum_j h^*(u - r + jT')$ and $\tilde{H}^{(0)}(\omega)$ replaced by $H^*(\omega)$. But

$$\left|\frac{2\pi}{T'}\sum_{j=0}^{T'-1}I_{T,y}(e^{i\omega_j'})[\tilde{H}^{(0)}(\omega_j') - H^*(\omega_j')]e^{i\omega_j'u}\right| \leqq \sup_\omega|\tilde{H}^{(0)}(\omega) - H^*(\omega)|\,C_y(0)$$

and

$$\left|\sum_{r=-(T-1)}^{T-1} C_y(r)\left[h^*(u-r)-\sum_j h^*(u-r+jT')\right]\right| \leqq 4\kappa C_y(0)T\exp(T'\log\lambda)/(1-\lambda^T)$$

and the right-hand side of both of these expressions are $O(Q_T)$ or smaller. Thus the remaining term that needs to be considered is

$$\int_{-\pi}^{\pi}\left[I_{T,y}(e^{i\omega})-\frac{\sigma_{\varepsilon o}^2}{2\pi}|K_o(e^{i\omega})|^2\right]H^*(\omega)e^{i\omega u}\,d\omega=\sum_{r=-\infty}^{\infty}[C_y(r)-\gamma_y(r)]h^*(u-r),$$

wherein we have set $C_y(r)=0$, $|r|\geqq T$. The modulus of the right-hand side of this equation is less than or equal to $\kappa\sum_{|r|<c\log T}|C_y(r)-\gamma_y(r)|\lambda^{|r|}+\kappa\sum_{|r|\geqq c\log T}|C_y(r)-\gamma_y(r)|\,\lambda^{|r|}$ and the second term in this expression is bounded by $2\kappa[C_y(0)+\gamma_y(0)]\lambda^{c\log T}/(1-\lambda)$. Taking $c\geqq -1/(2\log\lambda)$ and using (A4) once again we obtain the desired result.

(ii) Assume that $n=n_o$. Then it is readily verified that $\alpha_j^*=\alpha_{o,j}$, $j=1,\cdots,n-1$, provides the solution to (12) and hence (see Lemma 2.3) that $S_v^*(\omega)=\sigma_{\varepsilon o}^2|\mu_o(e^{i\omega})|^2/2\pi$. Substituting into the expression for $H^*(\omega)$ we find that the equation system (13) corresponds directly to the 'Yule–Walker' equations constructed from the power spectrum $\sigma_{\varepsilon o}^2/2\pi\,|\mu_o(e^{i\omega})|^2$ and hence that $\mu_j^*=\mu_{oj}$, $j=1,\cdots,n_o-1$. Using the same argument as employed in (i) above we find, therefore that $\tilde{\alpha}^{(0)}(z)=\alpha_o(z)+O(Q_T)$ and $\tilde{\mu}^{(0)}(z)=\mu_o(z)+O(Q_T)$ as required.

Now consider the case $n>n_o$. From Theorem 2.1 we know that the product $\alpha_o(e^{i\omega})\,|K_o(e^{i\omega})|^2=|\mu_o(e^{i\omega})|^2/\alpha_o(e^{-i\omega})$, from which it follows that $\int_{-\pi}^{\pi}\alpha_o(e^{i\omega})\,|K_o(e^{i\omega})|^2\,e^{i\omega u}\,d\omega=0$, $u>n_o-1$. Using this relationship it is straightforward to show that the solutions to (12) are characterized by operators of the form $\alpha^*(z)=\phi(z)\alpha_o(z)$ where $\phi(z)=1+\phi_1z^{-1}+\cdots+\phi_rz^{-r}$, $r=n-n_o$. Moreover, since $\sum\alpha_j^*\gamma_y(s-j)=0$, $s\geqq n$, $|\sum_{j=0}^{n-1}\alpha_j^*C_y(s-j)|\leqq\sum_{j=0}^{n-1}|\alpha_j^*|\,|C_y(s-j)-\gamma_y(s-j)|=O(Q_T)\sum_{j=0}^{n-1}|\alpha_j^*|$. Thus for T sufficiently large we can, with arbitrarily large probability, determine a measurable solution $\tilde{\alpha}_j^{(0)}$, $j=1,\cdots,n-1$, such that $\tilde{\alpha}^{(0)}(z)=\tilde{\phi}^{(0)}(z)\alpha_o(z)+O(Q_T)$ where $\tilde{\phi}^{(0)}(z)\neq 0$, $|z|\geqq 1$. We therefore find that

$$\begin{aligned}\tilde{S}_v^{(0)}(e^{i\omega})&=\frac{\sigma_{\varepsilon o}}{4\pi^2}\int_{-\pi}^{\pi}\sum_{r=1-n}^{n-1}|\tilde{\phi}^{(0)}(e^{i\theta})\alpha_o(e^{i\theta})K_o(e^{i\theta})|^2\,e^{i(\theta-\omega)r}\,d\theta+O(Q_T)\\&=\frac{\sigma_{\varepsilon o}^2}{2\pi}|\tilde{\phi}^{(0)}(e^{i\omega})\mu_o(e^{i\omega})|^2+O(Q_T)\end{aligned}$$

and hence that $\tilde{H}^{(0)}(\omega)=|\alpha_o(e^{i\omega})|^2/|\tilde{\phi}^{(0)}(e^{i\omega})|^2\,|\mu_o(e^{i\omega})|^4+O(Q_T)$.

As before, it follows that $\tilde{\mu}^{(0)}(z)=\mu^*(z)+O(Q_T)$ where $\mu^*(z)$ is formed from the 'Yule–Walker' equations

$$\frac{\sigma_{\varepsilon o}^2}{2\pi}\int_{-\pi}^{\pi}\sum_{j=0}^{n-1}\frac{\mu_j^*e^{i\omega(r-j)}}{|\tilde{\phi}^{(0)}(e^{i\omega})\mu_o(e^{i\omega})|^2}\,d\omega=0,\qquad r=1,\cdots,n-1,$$

giving the desired conclusion.

Lemma 4.3. *Set*

$$F(\alpha(z),\,\mu(z))=\frac{\sigma_{\varepsilon o}^2}{2\pi}\int_{-\pi}^{\pi}\frac{|\alpha(e^{i\omega})|^2}{|\mu(e^{i\omega})|^2}\,|K_o(e^{i\omega})|^2\,d\omega$$
$$=\sigma_{\varepsilon o}^2\left[1+\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|\alpha(e^{i\omega})K_o(e^{i\omega})-\mu(e^{i\omega})|^2}{|\mu(e^{i\omega})|^2}d\omega\right].$$

If assumptions $(A1)$–$(A4)$ *hold then*:

(i) *If* $n<n_o$, $\tilde{\sigma}_{\varepsilon n}^2=\min_{\alpha(z),\mu(z)}F(\alpha(z),\,\mu(z))+O(Q_T)$.

(ii) *If* $n\geqq n_o$, $\tilde{\sigma}_{\varepsilon n}^2=T^{-1}\sum_{t=1}^T \varepsilon_t^2+O(Q_T)$.

Proof. (i) Let $F_T(\alpha(z),\,\mu(z))=T^{-1}\sum e_t^2$, where $e_t=\bar{y}_t+\sum(\alpha_j\bar{y}_{t-j}-\mu_j e_{t-j})$, $e_t=0$, $t\leqq 0$. Following the method of proof employed in Poskitt (1987) it can be shown that $F_T(\alpha(z),\,\mu(z))=F(\alpha(z),\,\mu(z))+O(Q_T)$ uniformly in $\{\alpha(z),\,\mu(z)\}$. Thus we can determine a $\delta_T=O(Q_T)$ such that for T sufficiently large $|F_T(\cdot,\cdot)-F(\cdot,\cdot)|<\delta_T$ almost surely. Set $\{\tilde{\alpha}(z),\,\tilde{\mu}(z)\}=\arg\min F_T(\alpha(z),\,\mu(z))$. Then $F_T(\tilde{\alpha}(z),\,\tilde{\mu}(z))\leqq F_T(\alpha(z),\,\mu(z))\leqq F(\alpha(z),\,\mu(z))+\delta_T$ for all $\{\alpha(z),\,\mu(z)\}$ and

$$\min_{\alpha(z)\mu(z)}F(\alpha(z),\,\mu(z))-\delta_T\leqq F(\tilde{\alpha}(z),\,\tilde{\mu}(z))-\delta_T<F_T(\tilde{\alpha}(z),\,\tilde{\mu}(z)).$$

This implies that $|F_T(\tilde{\alpha}(z),\,\tilde{\mu}(z))-\min_{\alpha(z)\mu(z)}F(\alpha(z),\,\mu(z))|<\delta_T$. Thus the result will follow if we can show that $\tilde{\sigma}_{\varepsilon n}^2=F_T(\tilde{\alpha}(z),\,\tilde{\mu}(z))$. To achieve this note that the iterations defined in Step 2 correspond to a Gauss–Newton scheme designed to minimize $F_T(\alpha(z),\,\mu(z))$ since, to use an obvious generic notation, $\partial F_T/\partial\mu_j=2T^{-1}\sum\eta_{t-j}e_t$ and $\partial F_T/\partial\mu_j=-2T^{-1}\sum\varphi_{t-j}e_t$, $j=1,\cdots,n-1$. By an argument that parallels the proof of Theorem 3 of Kohn (1978), using the strong version of Kohn's (1978) Lemma 8, it can be shown that the iterates $\{\tilde{\alpha}^{(i)}(z),\,\tilde{\mu}^{(i)}(z)\}\to\{\tilde{\alpha}(z),\,\tilde{\mu}(z)\}$ as $i\to\infty$. But by definition,

$$\tilde{\sigma}_{\varepsilon,n}^{(i)2}=F_T(\tilde{\alpha}^{(i-1)}(z),\,\tilde{\mu}^{(i-1)}(z))+\sum_{t=1}^{T}\sum_{j=1}^{n-1}(\delta\tilde{\alpha}_j^{(i)}\tilde{\eta}_{t-j}^{(i)}-\delta\tilde{\mu}_j^{(i)}\tilde{\varphi}_{t-j}^{(i)})\tilde{\varepsilon}_t^{(i)}$$

and $T^{-1}\sum\tilde{\eta}_{t-j}^{(i)}\tilde{\varepsilon}_t^{(i)}$ and $T^{-1}\sum\tilde{\varphi}_{t-j}^{(i)}\tilde{\varepsilon}_t^{(i)}$, and therefore also $\delta\tilde{\alpha}_j^{(i)}$ and $\delta\tilde{\mu}_j^{(i)}$, $j=1,\cdots,n-1$, will converge to zero as $i\to\infty$. Applying the Cauchy–Schwartz inequality to the second term leads to the conclusion that $|\tilde{\sigma}_\varepsilon^{(i)2}-F_T(\tilde{\alpha}(z),\,\tilde{\mu}(z))|\to 0$ as $i\to\infty$ because this quantity is bounded above by $|\tilde{\sigma}_\varepsilon^{(i)2}-F_T(\tilde{\alpha}^{(i-1)}(z),\,\tilde{\mu}^{(i-1)}(z))|+|F_T(\tilde{\alpha}^{(i-1)}(z),\tilde{\mu}^{(i-1)}(z))-F_T(\tilde{\alpha}(z),\,\tilde{\mu}(z))|$.

(ii) We will establish this by proving by induction $\tilde{\sigma}_{\varepsilon n}^{(i)2}=T^{-1}\sum\varepsilon_t^2+O(Q_T^2)$ for all $i\geqq 1$, from which the stated equality follows directly. Consider $\tilde{\sigma}_{\varepsilon n}^{(i)2}$ and suppose that $\{\tilde{\alpha}^{(i-1)}(z),\,\tilde{\mu}^{(i-1)}(z)\}=\{\alpha_0(z)(1+\tilde{\phi}^{(i-1)}z^{-1}),\,\mu_0(z)(1+\tilde{\phi}^{(i-1)}z^{-1})\}+O(Q_T)$, $|\tilde{\phi}^{(i-1)}|<1$. Then

$$\frac{\alpha_o(z)}{\mu_o(z)}-\frac{\tilde{\alpha}^{(i-1)}(z)}{\tilde{\mu}^{(i-1)}(z)}=\sum_{j\geq 1}\psi_j z^{-1}$$

where the ψ_j are $O(Q_T)$ and decline at a geometric rate, so that $|\psi_j|<O(Q_T)\lambda^j$, $j=1,2,\cdots$, for some λ, $0<\lambda<1$. This implies that $T^{-1}\sum(\varepsilon_t-\tilde{\varepsilon}_t^{(i)})^2\leqq O(Q_T^2)C_y(0)/(1-\lambda)^2$ and $T^{-1}\sum_{t=1}^T(\varepsilon_t-\tilde{\varepsilon}_t^{(i)})\varepsilon_t=\sum_{j=1}^{c\log T}\psi_j[T^{-1}\sum_{t=1}^T y_{t-j}\varepsilon_t]+R_T$, where the remainder R_T is dominated by $O(Q_T)(C_y(0)T^{-1}\sum \varepsilon_t^2)^{1/2}\lambda^{c \log T}/(1-\lambda)$. Taking c sufficiently large and appealing to assumption (A4) gives the result that $T^{-1}\sum \bar{\varepsilon}_t^{(i)2} = T^{-1}\sum \varepsilon_t^2 + O(Q_T^2)$. Using a similar argument it can also be established that $T^{-1}\sum \tilde{\eta}_{t-j}^{(i)}\bar{\varepsilon}_t^{(i)}$ and $T^{-1}\sum \tilde{\varphi}_{t-j}^{(i)}\bar{\varepsilon}_t^{(i)}$, $j=1,\cdots,n-1$, are $O(Q_T)$. For example, $T^{-1}\sum \tilde{\eta}_{t-j}^{(i)}\bar{\varepsilon}_t^{(i)}$ can be expressed as a weighted sum of $T^{-1}\sum y_{t-j-k}\varepsilon_t$ and $T^{-1}\sum y_{t-j-k}[\varepsilon_t - \bar{\varepsilon}_t^{(i)}]$, $k=0,\cdots,c\log T$, with weights that decline geometrically, plus a remainder that is bounded by a constant times $\lambda^{c\log T}/(1-\lambda)$. The first two components contribute terms that are $O(Q_T)$ and the remainder is $o(T^{-\frac{1}{2}})$ for $c > -1/2 \log \lambda$. This implies, however, that $\delta\tilde{\alpha}_j^{(i)}$ and $\delta\tilde{\mu}_j^{(i)}$, $j=1,\cdots,n-1$, are $O(Q_T)$ and applying the Cauchy–Schwartz inequality to $\tilde{\sigma}_{\varepsilon,n}^{(i)2}$ as was done before we find that $\tilde{\sigma}_{\varepsilon,n}^{(i)2} = T^{-1}\sum \bar{\varepsilon}_t^{(i)2} + O(Q_T^2)$. Putting this result together with the first one produces the desired expression and starting the induction at $i=1$ using Lemma 4.1 (ii) completes the proof.

It is apparent from Lemma 4.3, though not explicitly stated, that if $n < n_o$, $\tilde{\sigma}_{\varepsilon,n}^2$ will converge to a value greater than $\sigma_{\varepsilon o}^2$ as T increases whereas, if $n \geqq n_o$, $\tilde{\sigma}_{\varepsilon,n}^2$ will converge to $\sigma_{\varepsilon o}^2$. It is also implicit in the proof that when $n \geq n_o$, $\tilde{\sigma}_{\varepsilon,n}^2$ will, with probability arbitrarily close to one, equal $\tilde{\sigma}_{\varepsilon,n}^{(1)2}$ for T large, so that only one iteration of Step 2 will be required.

Corollary 4.4. *If the conditions of Lemmas* 4.2 *and* 4.3 *obtain then for T sufficiently large*
(*i*) $\tilde{\sigma}_{\varepsilon,n}^2 > \tilde{\sigma}_{\varepsilon,n+1}^2$ *if* $n < n_o$, *and* (*ii*) $\tilde{\sigma}_{\varepsilon,n}^2 - \tilde{\sigma}_{\varepsilon,n+1}^2 = O(Q_T^2)$ *if* $n \geqq n_o$, *almost surely.*

Proof. (i) To begin we follow the proof of Theorem 5.2 in Pötscher (1983). Let $F^n = \min_{\alpha(z)\mu(z)} F(\alpha(z), \mu(z))$ where $\alpha(z)$ and $\mu(z)$ are of degree $n-1$. Clearly, $F^{n+1} \leqq F^n$. Now suppose that $F^{n+1} = F^n$. This implies that the function

$$\mathscr{F}(a, b) = \frac{\tilde{\sigma}_{\varepsilon o}^2}{2\pi}\int_{-\pi}^{\pi}\frac{(1-a\mathrm{e}^{-i\omega})}{(1-b\mathrm{e}^{-i\omega})}\frac{|\alpha(\mathrm{e}^{i\omega})|^2}{|\mu(\mathrm{e}^{i\omega})|^2}|K_o(\mathrm{e}^{i\omega})|^2\,d\omega$$

is minimised at $a=b$, $|b|<1$, leading to the conclusion that $K_o(z) = \alpha(z)/\mu(z)$, $|z| \geqq 1$, as is shown by Pötscher (1983). Thus we have a contradiction, since by Theorem 2.1 $K_o(z)$ is of order $n_o - 1$ and $\alpha(z)$ and $\mu(z)$ are of degree $n-1$, $n < n_o$. Hence we can infer that $F^{n+1} < F^n$. Now set $0 < \delta_n < (F^n - F^{n+1})/4$. Lemma 4.3 ensures that when T is sufficiently large both the events $\tilde{\sigma}_{\varepsilon,n}^2 > F^n - \delta_n$ and $\tilde{\sigma}_{\varepsilon,n+1}^2 < F^{n+1} + \delta_n$ will occur with probability arbitrarily close to one. Thus $\tilde{\sigma}_{\varepsilon,n}^2 - \tilde{\sigma}_{\varepsilon,n+1}^2 > (F^n - F^{n+1}) - 2\delta_n > \frac{1}{2}(F^n - F^{n+1}) > 0$. Statement (ii) is a trivial consequence of Lemma 4.3 (ii).

Note that it is in fact possible to show that for T large the sequence $\tilde{\sigma}_{\varepsilon,n}^2$, $n=1,2,\cdots$ will, with probability close to one, be monotonically decreasing. This can be established by first observing that the difference between $F_T(\alpha(z), \mu(z))$ and $\int_{-\pi}^{\pi}|\alpha(\mathrm{e}^{i\omega})|^2 I_{T,y}(\mathrm{e}^{i\omega})/|\mu(\mathrm{e}^{i\omega})|^2\,d\omega$ will converge to zero almost surely as $T\to\infty$. Replacing $\tilde{\sigma}_{\varepsilon o}^2|K_o(\mathrm{e}^{i\omega})|^2/2\pi$ by $I_{T,y}(\mathrm{e}^{i\omega})$ in $\mathscr{F}(a, b)$, let us label this $\mathscr{F}_T(a, b)$, and repeating the same logic as was used in the proof of Corollary 4.4 in conjunction with $\mathscr{F}_T(a, b)$, recognizing that $\sum y_t e^{i\omega t}$ is of order T with probability one, then leads us to the conclusion that $\tilde{\sigma}^2_{\varepsilon,n} > \tilde{\sigma}^2_{\varepsilon,n+1}$. Thus the statistics $T(\tilde{\sigma}^2_{\varepsilon,n} - \tilde{\sigma}^2_{\varepsilon,n+1})/\tilde{\sigma}^2_{\varepsilon,n+1}$ will be positive definite for all n almost surely. The importance of the corollary is that it tells us that the numerator remains bounded away from zero when $n < n_o$ but will be artibrarily small for $n \geqq n_o$. Thus $(\tilde{\sigma}^2_{\varepsilon,n} - \tilde{\sigma}^2_{\varepsilon,n+1})/\tilde{\sigma}^2_{\varepsilon,n+1} > 0$ whilst $C_T(n)/T \to 0$ as $T \to \infty$ when $n < n_o$, which yields part (i) of Theorem 4.1, whereas when $n \geqq n_o$, $[T(\tilde{\sigma}^2_{\varepsilon,n} - \tilde{\sigma}^2_{\varepsilon,n+1})/\tilde{\sigma}^2_{\varepsilon,n+1}]/L(T) \to 0$ if $\log\log T/L(T) \to 0$ whilst $\liminf C_T(n)/L(T) > 0$ as $T \to \infty$, giving part (ii).

5. Asymptotic extreme values

Although Theorem 4.1 specifies the precise order of magnitude of $C_T(n)$ required to obtain a strongly consistent estimate of n_o the upper and lower bounds on the rate of increase of $C_T(n)$ with T implied by the theorem are of limited use as a guide to the actual value to be employed in practice. The computations outlined in Section 3 can be viewed, however, as a method of implementing a sequence of statistical tests and the following theorem provides the required asymptotic null distribution. This means, of course, that the standard apparatus associated with hypothesis testing procedures and the choice of significance levels can be used to determine an appropriate magnitude for the critical, or extreme, value $C_T(n)$.

Theorem 5.1. *If* $\{y_t\}$ *is a hidden Markov chain process with state dimension* n_o *satisfying* (A1)–(A4) *then*

$$\lim_{T\to\infty} \Pr(\Lambda_T(n_o) \leqq 2(\log\log T - \log 8\pi + x)) = \exp(-2e^{-x}).$$

Before providing the proof of this theorem let us note that the statistic $\Lambda_T(n_o)$ has an unusual asymptotic distribution. The distribution function in fact corresponds to that of the extreme value of a continuous, zero mean, stationary Gaussian process on the real line with autocovariance function $(\cosh\{(t-s)/2\})^{-1}$. The relevance of this probability distribution in this context is due to Hannan (1982), who first obtained the distribution in the special case of testing white noise against a first order autoregressive moving-average alternative using the likelihood ratio principle. The results presented by Hannan were subsequently extended to more general autoregressive moving-average processes by Veres (1987) and the last part of the proof given below is modeled on the derivations used by these two authors. Observe also that an immediate corollary of Theorem 5.1 is that if $C_T(n)$ is chosen to correspond to the $(1-p_s)$th quantile point of $\exp(-2e^{-x})$ where $\log(1-p_s) = -2T^{-\delta_s}$, $\delta_s > 0$, then $x = \log T^{\delta_s}$ and the conditions of Theorem 4.1 will be satisfied. For example, if the probability-value p_s is less than the conventional 0.01 significance level and $T = 10{,}000$, a sample size not uncommon with cell membrane ion channel data, $x > \log T^{1/2}$.

Proof. The statistic $\Lambda_T(n_o) = T(\tilde{\sigma}^2_{\varepsilon,n_o} - \tilde{\sigma}^2_{\varepsilon,n_o+1})/\tilde{\sigma}_{\varepsilon,n_o+1}$ and from Lemma 4.3 (ii) it follows that the denominator equals $\sigma^2_{\varepsilon o} + o(1)$. Moreover, from the proof of the same result it is straightforward to verify via Markov's inequality that for $n \geqq n_o$

$$\tilde{\sigma}^{(i)2}_{\varepsilon,n} = T^{-1}\sum_{t=1}^{T} \varepsilon_t^2 + T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n-1} (\delta\tilde{\alpha}_j^{(i)}\tilde{\eta}^{(i)}_{t-j} - \delta\tilde{\mu}_j^{(i)}\tilde{\varphi}^{(i)}_{t-j})\tilde{\varepsilon}_t^{(i)} + o_p(T^{-1}),$$

where $\delta\tilde{\alpha}_j^{(i)}$ and $\delta\tilde{\mu}_j^{(i)}$, $j = 1, \cdots, n-1$, are determined by solving the equation system

$$\sum_{j=1}^{n-1}\sum_{t=1}^{T} (\delta\tilde{\alpha}_j^{(i)}\tilde{\eta}^{(i)}_{t-j}\tilde{\eta}^{(i)}_{t-k} - \delta\tilde{\mu}_j^{(i)}\tilde{\varphi}^{(i)}_{t-j}\tilde{\eta}^{(i)}_{t-k}) = -\sum_{t=1}^{T} \tilde{\eta}^{(i)}_{t-k}\tilde{\varepsilon}_t^{(i)} \qquad k = 1, \cdots, n-1,$$

$$\sum_{j=1}^{n-1}\sum_{t=1}^{T} (-\delta\tilde{\alpha}_j^{(i)}\tilde{\eta}^{(i)}_{t-j}\tilde{\varphi}^{(i)}_{t-k} + \delta\tilde{\mu}_j^{(i)}\tilde{\varphi}^{(i)}_{t-j}\tilde{\varphi}^{(i)}_{t-k}) = \sum_{t=1}^{T} \tilde{\varphi}^{(i)}_{t-k}\tilde{\varepsilon}_t^{(i)} \qquad k = 1, \cdots, n-1.$$

Re-expressing these relationships using matrix vector notation we find that the numerator of $\Lambda_T(n_o)$ equals $T^{-1}(RSS_{n_o+1} - RSS_{n_o}) + o_p(1)$ where $RSS_n = \lim_{i\to\infty} \boldsymbol{g}_n'\boldsymbol{M}_n^-\boldsymbol{g}_n$. The gradient vector $\boldsymbol{g}_n$ contains $\sum \tilde{\eta}^{(i)}_{t-j}\tilde{\varepsilon}_t^{(i)}$ in location j and $\sum \tilde{\varphi}^{(i)}_{t-j}\tilde{\varepsilon}_t^{(i)}$ in location $n-1+j$, $j = 1, \cdots, n-1$, and $\boldsymbol{M}_n^-$ denotes a generalized inverse of the matrix of mean squares and cross products of the associated derivative processes $\tilde{\eta}^{(i)}_{t-j}$ and $\tilde{\varphi}^{(i)}_{t-j}$, $j = 1, \cdots, n-1$. From Poskitt (1987), see also Poskitt and Tremayne (1981), we know that $\lim_{T\to\infty} \boldsymbol{M}_{n_o}$ is non-singular but $\boldsymbol{M}_{n_o+1}$ converges to a matrix of rank $2n_o - 1$. For the case $n = n_o + 1$ let us therefore make the transformation from $\alpha_1, \cdots, \alpha_{n_o}$ to $a_1, \cdots, a_{n_o-1}$, ϕ_a defined by $\alpha(z) = a(z) \times (1 + \phi_a z^{-1})$ and from $\mu_1, \cdots, \mu_{n_o}$ to $m_1, \cdots, m_{n_o-1}$, ϕ_m with $\mu(z) = m(z)(1 + \phi_m z^{-1})$, once again reverting to generic notation. The Jacobean $\boldsymbol{J} = \mathrm{diag}\,(\boldsymbol{J}_\alpha, \boldsymbol{J}_\mu)$ where

$$\boldsymbol{J}_\alpha = \begin{bmatrix} 1 & \phi_a & & & & & \\ & 1 & \phi_a & & & & \\ & & \cdot & \cdot & & & \\ & & & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \\ & & & & & \cdot & \cdot \\ & & & & & 1 & \phi_a \\ 1 & a_1 & \cdot & \cdot & \cdot & \cdot & a_{n_o-1} \end{bmatrix}$$

and $\boldsymbol{J}_\mu$ has the same structure with $a_1, \cdots, a_{n_o-1}$ and ϕ_a replaced by $m_1, \cdots, m_{n_o-1}$ and ϕ_m respectively. (The use of a common notation for the Jacobean and Jordan matrices should cause no confusion as they are employed in totally different situations.)

Now consider evaluating the quadratic form $\boldsymbol{g}_{n_o}'\boldsymbol{M}_{n_o}^-\boldsymbol{g}_{n_o}$ at $\{\alpha(z), \mu(z)\}$ and

$g'_{n_o+1}M^{-}_{n_o+1}g_{n_o+1} = (Jg_{n_o+1})'(JM_{n_o+1}J')^{-}Jg_{n_o+1}$ at $\{\alpha(z)(1+\phi z^{-1}), \mu(z)(1+\phi z^{-1})\}$. A fairly routine calculation shows that the elements of Jg_{n_o+1} correspond to those of g_{n_o} with the addition of the term $\sum \xi_{t-1}(\phi)e_t$, where $\xi_t(\phi)+\phi\xi_{t-1}(\phi)=e_t$, for the two extra entries associated with ϕ_a and ϕ_m. Similarly, $JM_{n_o+1}J'$ contains the elements of M_{n_o} supplemented with the mean squares and cross products $T^{-1}\sum \eta_{t-j}\xi_{t-1}(\phi)$, $T^{-1}\sum \varphi_{t-j}\xi_{t-1}(\phi)$ and $T^{-1}\sum \xi_{t-1}(\phi)^2$ so as to add two further rows and columns. A generalised inverse of $JM_{n_o+1}J'$ can now be constructed by replacing one of the duplicated rows and columns by zeroes and then inverting the remaining $2n_o-1$ linearly independent rows and columns. The value of the quadratic form so obtained will be invariant with respect to which row and column is implicitly eliminated.

Let $\Gamma_{n_o}=\Gamma'_{n_o}$ denote the $2(n_o-1)$ order matrix with elements

$$\frac{\sigma^2_{\varepsilon o}}{2\pi}\int_{-\pi}^{\pi}\frac{e^{i\omega(r-s)}}{|\alpha_o(e^{i\omega})|^2}\,d\omega,$$

$$\frac{\sigma^2_{\varepsilon o}}{2\pi}\int_{-\pi}^{\pi}\frac{e^{i\omega(r-s)}}{\alpha_o(e^{i\omega})\mu_o(e^{-i\omega})}\,d\omega$$

and

$$\frac{\sigma^2_{\varepsilon o}}{2\pi}\int_{-\pi}^{\pi}\frac{e^{i\omega(r-s)}}{|\mu_o(e^{i\omega})|^2}\,d\omega$$

in locations $r, s, r, s'=n_o-1+s$ and $r', s', r'=n_o-1+r, r, s=1,\cdots,n_o-1$, respectively and let $C_{n_o}(\phi)$ be the $2(n_o-1)$ component column vector with

$$(\pm)\frac{\sigma^2_{\varepsilon o}}{2\pi}\int_{-\pi}^{\pi}\frac{e^{i\omega r}}{\alpha_o(e^{i\omega})(1+\phi e^{-i\omega})}\,d\omega$$

in the rth position and

$$(\mp)\frac{\sigma^2_{\varepsilon o}}{2\pi}\int_{-\pi}^{\pi}\frac{e^{i\omega r}}{\mu_o(e^{i\omega})(1+\phi e^{-i\omega})}\,d\omega$$

in the r'th. By Lemma 4.2 for all T sufficiently large $\{\tilde{\alpha}^{(0)}(z), \tilde{\mu}^{(0)}(z)\}$ will equal $\{\alpha_o(z), \mu_o(x)\}$ when $n=n_o$ and $\{\alpha_o(z)(1+\tilde{\phi}^{(0)}z^{-1}), \mu_o(z)(1+\tilde{\phi}^{(0)}z^{-1})\}$, $|\tilde{\phi}^{(0)}|<1$, when $n=n_o+1$ plus a constant times Q_T with probability arbitrarily close to one. Furthermore, by construction the values $\delta\tilde{\alpha}^{(i)}_j$ and $\delta\tilde{\mu}^{(i)}_j$, $j=1,\cdots,n-1$, are determined so as to make the regression sums of squares as large as possible, but in the proof of Lemma 4.3 we have established that the parameter adjustments will be $O(Q_T)$. This implies that $\{\tilde{\alpha}^{(i)}(z), \tilde{\mu}^{(i)}(z)\}=\{\alpha_o(z), \mu_o(z)\}+O(Q_T)$ when $n=n_o$ and $\{\tilde{\alpha}^{(i)}(z), \tilde{\mu}^{(i)}(z)\}=\{\alpha_o(z)(1+\tilde{\phi}^{(i)}z^{-1}), \mu_o(z)(1+\tilde{\phi}^{(i)}z^{-1})\}+O(Q_T)$, $|\tilde{\phi}^{(i)}|<1$, when $n=n_o+1$ where $|\tilde{\phi}^{(i)}-\tilde{\phi}^{(i-1)}|=O(Q_T)$. From Theorem 1 of Poskitt (1987) it follows that $M_{n_o}=\Gamma_{n_o}+O(Q_T)$ and the remaining elements of $JM_{n_o+1}J'$ will converge to the corresponding components of $C_{n_o}(\phi)$ and $\sigma^2_{\varepsilon o}/(1-\phi^2)$ where ϕ is the value that maximizes the regression sum of squares RSS_{n_o+1}. Using well known results on partitioned inversion in conjunction with the relationships outlined in the previous paragraph leads us to the conclusion that $\Lambda(n_o) = (\Xi_T/\sigma_{\varepsilon o})^2[1 + o(1)] + o_p(1)$ where $\Xi_T^2 = \max_{|\phi|<1-\delta_T} [\{T^{-\frac{1}{2}}\sum_{t=1}^T \xi_{t-1}(\phi)\varepsilon_t\}h(\phi, \phi)^{-\frac{1}{2}}]^2$, $h(\phi, \phi') = \sigma_{\varepsilon o}^2/(1-\phi\phi') - C_{n_o}(\phi)'\Gamma_{n_o}^{-1}C_{n_o}(\phi')$ and $\delta_T \leqq 2/(1+T^{\frac{1}{2}})$.

The sum $T^{-\frac{1}{2}}\sum \xi_{t-1}(\phi)\varepsilon_t$ converges to a Gaussian random variable. Indeed, from the second part of assumption (A4), see also Kohn (1978, Theorem 5), it follows that $T^{-\frac{1}{2}}\boldsymbol{J}\boldsymbol{g}_{n_o+1}$ obeys the central limit theorem and hence for any finite set of values ϕ_j, $j = 1, \cdots, N$, $|\phi_j| < 1 - \delta_T$, the quantities $T^{-\frac{1}{2}}\sum \xi_{t-1}(\phi_j)\varepsilon_t$ will have an asymptotic marginal distribution that is normal with zero mean and covariance $\sigma_{\varepsilon o}^2 h(\phi_i, \phi_j)$, $i, j = 1, \cdots, N$. Duplicating the argument presented in Hannan (1982, pp. 409–410) or Veres [1987, p. 355] we find that $T^{-\frac{1}{2}}\sum \xi_{t-1}(\phi)\varepsilon_t$ is continuous in ϕ uniformly in T and from Theorem 1 of Gihman and Skorohod (1974, Ch. VI, Section 4) it follows that Ξ_T^2 converges in distribution to $\max_{|\phi|<1} \mathscr{X}(\phi)^2$ where $\mathscr{X}(\phi)$ is a zero mean Gaussian process with autocovariance $\sigma_{\varepsilon o}^2 h(\phi, \phi)^{-\frac{1}{2}}h(\phi, \phi')h(\phi', \phi')^{-\frac{1}{2}}$. Let $t = \log(1+\phi)/(1-\phi)$ and set $\mathscr{Z}(t) = \mathscr{X}(\phi)/\sigma_{\varepsilon o}$ where $\phi = (1-e^{-t})/(1+e^{-t})$. Adapting the derivation used by Veres (1987, pp. 349–351) we find that $\mathscr{Z}(t)$ is a zero mean Gaussian process with autocovariance function $2/(e^{t/2} + e^{-t/2})$. The statement in the theorem now follows from Theorem 7.1 of Berman (1971) by recognizing that for ϕ in the interval $\{\phi : |\phi| \leqq (T^{\frac{1}{2}} - 1)/(1 + T^{\frac{1}{2}})\}$ $|t| \leqq \log T^{\frac{1}{2}}$ and that $\max(\mathscr{X}(\phi)/\sigma_{\varepsilon o})^2$ and $\max \mathscr{Z}(t)^2$ have the same distribution.

6. Conclusion

Preliminary simulation evidence, that we hope to present elsewhere, indicates that the statistical techniques described above behave as would be expected on the basis of the theoretical results that we have established. If these results are to be of practical value, however, it will be necessary to ascertain the extent to which the basic model and its associated assumptions can be thought to provide a sensible representation of the phenomenon being studied. In cell membrane kinetics there is considerable evidence to suggest that observed ion currents behave very much like Markov chain processes and that measurement and background noise obtained using the giga-seal patch-clamp technique is essentially white, despite the vagaries due to laboratory experimentation. This suggests that the method of identifying the state dimension of a hidden Markov chain process that we are advocating could work well in the context of modelling ion currents.

Acknowledgements

We are grateful to Mark Dwyer for helpful discussions on the regularity conditions associated with Assumption (A4).

This work was in an embryonic state at the time of Ted Hannan's sudden and unexpected death. Regretfully, therefore, we have not been able to benefit directly from the fine judgment and statistical expertise of our former friend and colleague. Nevertheless, the intellectual debt that this paper owes to Ted should be apparent. He will be sorely missed.

References

AN, H-Z., CHEN, Z. G. AND HANNAN, E. J. (1983) A note on ARMA estimation. *J. Time Series Anal.* **4**, 9–17.

BAUM, L. E. AND PETRIE, T. (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **37**, 1554–1563.

BERMAN, S. M. (1971) Asymptotic independence of the numbers of high and low level crossings of stationary Gaussian processes. *Ann. Math. Statist.* **42**, 927–945.

BINGHAM, E. O. (1974) *The Fast Fourier Transform*. Prentice Hall, Englewood Cliffs.

BLACKWELL, E. AND KOOPMANS, L. (1957) On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **28**, 1011–1015.

BLUM, J. R., HANSON, D. L. AND KOOPMANS, L. (1963) On the strong law of large numbers for a class of stochastic processes. *Z. Wahrscheinlichkeitsth.* **2**, 1–11.

CHUNG, S. H., KRISHNAMURTHY, V. AND MOORE, J. B. (1991) Adaptive processing techniques based on hidden Markov models for characterizing very small channel currents buried in noise and deterministic interferences. *Phil. Trans. R. Soc. Lond.* B **334**, 357–384.

DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc.* B **39**, 1–38.

ENGEL, E, M. R. A. (1984) A unified approach to the study of sums, products, time-aggregation and other functions of ARMA processes. *J. Time Series Anal.* **5**, 159–171.

FELLER, W. (1966) *An Introduction to Probability Theory and its Application.* Vol. II. Wiley, New York.

FINESSO, L. (1990) *Consistent Estimation of the Order of Markov and Hidden Markov Chains*. PhD Dissertation, University of Maryland.

FREDKIN, D. R. AND RICE, J. A. (1987) Correlation functions of a finite-state Markov process with applications to channel kinetics. *Math. Biosci.* **87**, 161–172.

GALLANT, A. R. AND WHITE, H. (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* Blackwell, Oxford.

GIHMAN, I. I. AND SKOROHOD, A. V. (1974) *The Theory of Stochastic Processes.* Springer, Berlin.

GOLUB, G. H. AND VAN LOAN, C. F. (1989) *Matrix Computations*. 2nd edn. Johns Hopkins University Press, Baltimore.

HANNAN, E. J. (1970) *Multiple Time Series.* Wiley, New York.

HANNAN, E. J. (1980) The estimation of the order of an ARMA process. *Ann. Statist.* **8**, 1071–1081.

HANNAN, E. J. (1982) Testing for autocorrelation and Akaikes criterion. In *Essays in Statistical Science.* ed. J. Gani and E. J. Hannan. pp 403–412. Applied Probability Trust, Sheffield.

HANNAN, E. J. AND DEISTLER, M. (1988) *The Statistical Theory of Linear Systems* Wiley, New York.

HANNAN, E. J. AND KAVALIERIS, L. (1984) A method of autoregressive-moving average estimation. *Biometrika* **71**, 273–280.

HANNAN, E. J. AND RISSANEN, J. (1982) Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **69**, 81–94.

HELLER, A. (1965) On stochastic processes derived from Markov chains. *Ann. Math. Statist.* **36**, 1286–1291.

KARLIN, S. A. (1966) *A First Course in Stochastic Processes*. Academic Press, New York.

KOHN, R. (1978) Asymptotic properties of time domain Gaussian estimators. *Adv. Appl. Prob.* **10**, 339–359.

LEE, E. A. AND MESSERSCHMITT, D. G. (1988) *Digital Communication.* Kluwer, Boston, MA.

LEROUX, B. G. (1992) Maximum likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.* **40**, 127–143.

Marhav, N., Gutman, M. and Ziv, J. (1989). On the estimation of the order of a Markov chain and universal data compression. *IEEE Trans. Inf. Th.* **35**, 1614–1619.

Nadas, A. and Mercer, R. L. (1994) Hidden Markov chains can be hard to train. *Paper presented at 230th IMS Meeting, Cleveland.*

Pagano, M. (1974) Estimation of models of autoregressive signal plus white noise. *Ann. Statist.* **2**, 99–108.

Pötscher, D. S. (1983) Order estimation in ARMA-models by Lagrangian multiplier tests. *Ann. Statist.* **11**, 872–885.

Poskitt, D. S. (1987) A modified Hannan–Rissanen strategy for mixed autoregressive-moving average order determination. *Biometrika* **74**, 781–790.

Poskitt, D. S. and Tremayne, A. R. (1980) Testing the specification of a fitted autoregressive-moving average model. *Biometrika* **67**, 359–363.

Qian, W. & Titterington, D. M. (1992) Estimation of parameters in hidden Markov models. *Phil. Trans. R. Soc Lond.* A **337**, 407–428.

Rozanov, Y. A. (1967) *Stationary Random Processes.* Holden Day, San Francisco.

Veres, S. (1987) Asymptotic distributions of likelihood ratios for overparameterized ARMA processes. *J. Time Series Anal.* **8**, 345–357.