A Note on the Equivalence of Two Approaches for Specifying a Markov Process

K. S. Chan†

University of Iowa, Iowa City, U.S.A.

H. Tong

London School of Economics, U.K. & University of Hong Kong, Pokfulam, Hong Kong

Summary. The probabilistic structure of a discrete-time (high-order) vector Markov process may be studied via two approaches. In the first approach, the Markov process is specified by the transition probability and the initial distribution. An alternative approach is via a stochastic difference equation. We have proved that these two approaches are equivalent under very mild conditions.

Keywords: ARMA model; equivalence of distribution functions; stationarity; stochastic difference equation; transition probability

1. INTRODUCTION

Consider a discrete-time Markov process. One can study its structure by reference to the conditional probability, i.e. the transition probability, in addition to the initial distribution. This approach will be referred to as the direct approach. An alternative approach is via a stochastic difference equation. In the linear Gaussian case, the two approaches are clearly equivalent. Now, for a real-valued Markov chain and in the order one case, the two approaches can be shown to be equivalent under very general conditions; see Rosenblatt (1971, Lemma 2 on p. 169) and Tong (1990, Lemma 3.1 on p. 97). Whether this equivalence can be generalised to the case of higher-order vector-valued Markov processes is an

†*Address for correspondence*: K.S. Chan, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242, U.S.A.
E-mail: kchan@stat.uiowa.edu

open problem as far as we are aware. This open problem deserves careful attention as the equivalence of the two approaches, if true, has the following useful consequences. First, it implies that whether we should use one approach in preference to the other becomes a matter of either taste or convenience. Second, it is then guaranteed that one can convert freely between the two approaches. While the direct approach facilitates likelihood calculations, the stochastic difference equation approach yields a direct description of the dynamics of the process, which is of relevance for forecasting purposes. Moreover, it provides a recipe of recovering the noise terms under certain conditions, which are useful for (not necessarily likelihood-based) model estimation and diagnostics. Hence, the possibility of converting between the two approaches is of some practical value.

The purpose of this note is to address this open problem on the equivalence of the direct approach and the stochastic difference equation approach. We show in §2 that under very general conditions, the two approaches are indeed equivalent.

2. MAIN RESULT

We recall the well-known device of writing a pth-order Markov chains, say, $\{Y_n\}$ as a first-order vector-valued Markov chains via the stacking operation: $X_n = (Y_n', Y_{n-1}', \cdots, Y_{n-p+1}')'$ where X_n is k-dimensional with k equal to p times the dimension of Y_n. Henceforth, it suffices to consider first-order k-dimensional Markov chains.

We first illustrate some of the basic ideas in the case of $k = 2$, i.e., two dimensional Markov chains. Let $\{X_n = (X_{n,1}, X_{n,2})'\}$ be a stationary Markov chain with the σ-algebra generated by $\{X_1, \cdots, X_n\}$ denoted as B_n. Let $\epsilon_n = (\epsilon_{n,1}, \epsilon_{n,2})'$ be the white noise driving the process. One can consider the case of generating the process at epoch $n+1$ componentwise in the order of $X_{n+1,1}, X_{n+1,2}$. Hence, we may think of $\epsilon_{n+1,1}$ as the noise (input) needed to obtain $X_{n+1,1}$ given all past X's. Then, $\epsilon_{n+1,2}$ is the independent input needed to obtain $X_{n+1,2}$ given $X_{n+1,1}$ and all past X's. Thus, it seems not restrictive to assume that ϵ_n consists of independent components that are independent of past X's. In other words, we shall consider the case that the σ-algebra B_{n+1} generated by all observations up to and including epoch $n+1$ equals $B_n \times C_{n+1,1} \times C_{n+1,2}$ where $C_{n,j}$ denotes the σ-algebra generated by the jth component of the noise at epoch n.

Now, let us introduce some notations for the general case of k-dimensional Markov

chains. Denote by $B_{n+1,j}$ the σ-algebra generated by all X's up to and including epoch n, and $X_{n+1,i}, 0 < i \leq j$, that is, the first j-components of X_{n+1}. In particular, $B_{n+1,0} = B_n$. It is natural to impose the requirement that $B_{n+1,j} = B_n \times C_{n+1,1} \times C_{n+1,2} \times \cdots \times C_{n+1,j}$. This is an explicit requirement on how the process is generated, namely, the jth component of X_{n+1} is built up by the past X's and the first j components of the noise term ϵ_{n+1}. If this requirement is satisfied, then for fixed j, the one-dimensional conditional distributions of $X_{n+1,j+1}$ given $X_n = x, X_{n+1,i} = x_{n+1,i}, i \leq j$ are equivalent functions for all $x, x_{n+1,i}, i \leq j$. [Following Rosenblatt (1971), two 1-dimensional distribution functions, say, F and G are said to be equivalent if and only if the discontinuity points of F can be mapped to those of G in a one-to-one manner with the jump size preserved.] It is well known that a distribution function admits at most countably many points of discontinuity. For any distribution function, say, H, the set of all discontinuity points of H will be denoted as $D(H)$. Hence, F and G are equivalent if and only if there exists a one-to-one map $\phi : x \in D(F) \to \phi(x) \in D(G)$ preserving the jump sizes, that is, for all $x \in D(F)$, $F(x) - F(x-) = G(\phi(x)) - G(\phi(x)-)$ where, for example, $F(x-) = \lim_{y \uparrow x} F(y)$. In particular, continuous distribution functions are, by definition, equivalent. Conversely, following the same kind of construction as in Rosenblatt (1971, p.169), it will be shown below that if the equivalence requirement on the conditional distribution functions holds, there exist $\epsilon_n = (\epsilon_{n,1}, \epsilon_{n,2}, \cdots, \epsilon_{n,k})'$ and $G_j, j = 1, \cdots, k$ such that $X_{n+1,j+1} = G_j(X_n, X_{n+1,i}, i \leq j, \epsilon_{n+1,j+1})$. In other words, we have $X_{n+1} = G(X_n, \epsilon_{n+1})$ a.e., a stochastic difference equation representation.

We can now summarize the preceding heuristic discussion in the main result below by first posing the following question

Q: Does there exist $\epsilon_n = (\epsilon_{n,1}, \cdots, \epsilon_{n,k})'$ of independent components such that for all n and all j, $\epsilon_{n+1,j+1}$ is independent of $B_{n+1,j}$ and yet $B_{n+1,j+1}$ equals the σ-algebra generated by $B_{n+1,j}$ and $\epsilon_{n+1,j+1}$?

LEMMA 2.1. *Let $\{X_n\}$ be a k-dimensional stationary Markov chain with conditional distributions $F_j(x_{j+1}|x, x_{n+1,i}, i \leq j) = P(X_{n+1,j+1} \leq x_{j+1}|X_n = x, X_{n+1,i} = x_{n+1,i}, i \leq j)$. The answer to question Q is affirmative if and only if for each j, $F_j(x_{j+1}|x, x_{n+1,i}, i \leq j)$, as functions of x_{j+1}, are equivalent for almost all $x, x_{n+1,i}, i \leq j$, with respect to the stationary probability measure of the process $\{X_n\}$.*

Before we present the proof of this lemma, three remarks are in order. First, in the vector case, the stochastic difference equation representation is ordinarily not unique. In fact, the non-uniqueness of the difference equation representation is related to the identifiability of the model. This problem is very challenging and well documented in the case of linear Auto-Regressive Moving-Average (ARMA) models for vector time series; see Tuan (1978) and Hannan (1979). Second, we have assumed a particular order of the components of X_n. Already in the linear ARMA case, different order of the components may yield different stochastic difference equation representations, some of which may be more tractable than some others in terms of statistical inference. Third, jumps may occur naturally in the conditional distribution functions. For example, consider an AR(2) model: $Y_n = \alpha_1 Y_{n-1} + \alpha_2 Y_{n-2} + e_n$ where e_n are iid. Let $X_n = (X_{n,1} = Y_{n-1}, X_{n,2} = Y_n)'$. Clearly, the conditional probability distribution of $X_{n+1,1}$ given X_n jumps at the point $X_{n,2}$ from 0 to 1.

Proof: First, we consider the proof of the necessity of the equivalence of F_j's for each j. Let $C_{n,j}$ denote the σ-algebra generated by $\epsilon_{n,j}$. For fixed j, the equality $B_{n+1,j} = B_{n+1,j-1} \times C_{n+1,j}$ implies that conditional on $X_n, X_{n+1,i}, i < j$, the σ-algebra generated by $X_{n+1,j}$ equals $C_{n+1,j}$ and hence the F_j's are equivalent functions.

Conversely, suppose that for all j, F_j's are equivalent functions. For simplicity, first assume that all F's are continuous functions with no jumps. Then, let $U_{n+1,j} = F_j(X_{n+1,j}|X_n, X_{n+1,i}, i < j)$ which are uniformly distributed over the unit interval $[0,1]$. Hence, $\{U_n = (U_{n,1}, \cdots, U_{n,k})'\}$ is a sequence of iid random variables, each component of U_n being uniformly distributed over $[0,1]$. Therefore, $X_{n+1,j} = F_j^{-1}(U_{n,j}|X_n, X_{n+1,i}, i < j)$ almost surely. Here, the inverse of a distribution function, say, F is defined as $F^{-1}(x) = \inf\{y : F(y) > x\}$. This demonstrates that there exists a G such that $X_{n+1} = G(X_n, U_{n+1})$, a stochastic difference equation representation.

If the functions F_j have jumps, we adapt below the arguments on p.169 of Rosenblatt to show that the above difference equation representation still holds, but the components of U_n need not be uniformly distributed over $[0,1]$. For fixed j, let the set of discontinuity points of $F_j(\cdot|x, x_{n+1,i}, 0 < i < j)$ be $D = \{d_1, d_2, d_3, \cdots\}$ where the d's are assumed to be labelled so that their corresponding jump sizes are non-increasing: $p_1 \geq p_2 \geq p_3 \geq \cdots$. (Note that the only accumulation point of the jump sizes must be zero, as the total probability mass equals 1; hence the jump sizes attain their maximum value.) In other words, for all k, $F_j(d_k|x, x_{n+1,i}, 0 < i < j) - F_j(d_k - |x, x_{n+1,i}, 0 < i < j) = p_k$ with the

d's labelled sequentially so that the corresponding p's are non-increasing. Moreover, the set of p's depends on j but is independent of $x, x_{n+1,i}, 0 < i < j$ by the equivalence of the F_j's. On the other hand, the set of discontinuity points does depend on both j and the conditioning values, $x, x_{n+1,i}, 0 < i < j$. We have suppressed the dependence of the p's on j for conciseness; similarly we adopt a simpler notation for the d's. Now, define $\epsilon_{n+1,j} = i$ when $X_{n+1,j} = d_i$ given $X_n = x, X_{n+1,i} = x_{n+1,i}, 0 < i < j$; otherwise define $\epsilon_{n+1,j} = F^{-1}(X_{n+1,j}|X_n = x, X_{n+1,i} = x_{n+1,i}, 0 < i < j)$. It is readily checked that $\epsilon_{n,j}$ is a probabilistic mixture such that it equals i with probability p_i and is uniformly distributed on $[0, 1]$ with probability $1 - \sum_{i=1}^{\infty} p_i$. Clearly, there exist measurable functions G_j such that $X_{n+1,j} = G_j(\epsilon_{n+1,j}, X_n, X_{n+1,i}, 0 < i < j)$. This completes the proof that there exists G such that $X_{n+1} = G(X_n, \epsilon_{n+1})$, a stochastic difference equation representation.

3. CONCLUSION

In the case of vector ARMA modeling, the non-uniqueness of the stochastic difference equation representation has given rise to much research on convenient parameterization of vector ARMA models; see, e.g., Akaike (1976), Tuan (1978), Hannan (1979), Hannan and Deistler (1988) and Tiao and Tsay (1989). This non-uniqueness problem is much more challenging in the case of nonlinear time series, partly because the functional form of the nonlinear model is often unknown. Innovative research on this challenging problem is clearly needed for advancing nonlinear multiple time series modeling.

4. ACKNOWLEDGMENT

H. Tong was partially supported by grants from the BBSRC/EPSRC of the United Kingdom and the Research Grants Council of Hong Kong.

References

Akaike, H. (1976) Canonical correlation analysis of time series and the use of an information criterion, in *System Identification: Advances in Case Studies*, Eds R. K. Mehra and D.G. Lainiotis, 27-96, New York: Academic Press.

Hannan, E.J. (1979) A note on autoregressive-moving average identification. *Biometrika*, **66**, 672-4.

Hannan, E.J. and Deistler, M. (1988) *The statistical theory of linear system*, Wiley.

Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press.

Rosenblatt, R. (1971) *Markov processes, structure, and asymptotic behavior*, New York: Springer-Verlag.

Tiao, G.C. and Tsay, R.S. (1989) Model specification in multivariate time series. *Journal of the Royal Statistical Society, Series B*, **51**, 157-195, 195-213.

Tuan, P.D. (1978) On the fitting of multivariate processes of the autoregressive- moving average type, *Biometrika*, **65**, 99-107.