# Tree Indexed Markov Processes and Long Range Dependency

Langford B. White[1] and Sylvie L. Perreau[2]

[1] *Department of Electrical and Electronic Engineering, The University of Adelaide, North Terrace, Adelaide 5005, AUSTRALIA. Email Address: lwhite@eleceng.adelaide.edu.au*
[2] *Institute for Telecommunications Research, University of South Australia. Email Address: sylvie@spri.levels.unisa.edu.au*

**Abstract.** This paper describes the second order statistics of a finite state Markov process indexed on a binary tree. Such models are the discrete state analogues of the continuous state Gauss-Markov processes as described by Basseville *et al* [1]. Such processes are termed *tree-indexed* processes. The idea is to use the leaf nodes of the tree at a specified depth, as indices for a time series, and to derive a probabilistic model for this time series. The paper shows that such processes possess covariance functions which decay as a power law thus exhibiting a long range dependent (LRD) or self- similarity property. These models are motivated in part by recent evidence that suggests some communications network traffic may exhibit such behaviour. However, the processes are highly non-stationary in nature. The paper poses as an open question whether there exists a modification of the tree structure which permits the leaf node process to be stationary but retains the LRD property.

## I INTRODUCTION

In computer communications networks [2], messages are generally transmitted in a packetised form. From the point of view of designing control mechanisms for the network, it is of interest to examine the statistical variability or burstiness of the data at several time scales. For example, bursty data at fine time scales is likely to be related to the message structure of a single transaction ; burstiness at longer time scales is more likely to arise from some more global properties of traffic on the network. In either case, different control and design considerations are likely to arise. This paper describes a novel type of model which, we argue, offers the potential to represent such behaviour. The idea of indexing Markov processes on trees is not new. Such models were addressed in the linear-Gaussian framework in a series of papers [1], [3], [4]. Also the idea of modelling a time-series by such a process's leaf nodes has appeared in [5] where the question of identification based on such measurements is addressed in an autoregressive framework.

CP511, *Unsolved Problems of Noise and Fluctuations,* edited by D. Abbott and L. B. Kish

In [6], results are presented which argue that some computer network data exhibits a kind of self-similarity of statistical properties in scale. Reference [7] also considers a class of processes which exhibit such behaviour. These processes exhibit a very large degree of variability and extremely long term correlations, both attributes which are suggested relevant by the studies reported in [6]. The work [8] also considers LRD processes but uses a different mechanism to obtain an approximately LRD process.

The purpose of this paper is to describe the second order statistics of a general class of tree-indexed finite state Markov processes and demonstrate that the data generated by such models can possess a LRD or self-similar characteristic. The problem with our existing theory is that the leaf node process (indeed all processes derived by sequential access of all nodes of a fixed depth), are highly non-stationary. The problem of matching an observed time-series to the leaf nodes of a tree-indexed process is also addressed in [5]. We pose as an open question whether there is a modification or generalisation of the above model to yield stationary processes at the leaf nodes. Proofs are not given but may be found in [9].

## II TREE-INDEXED PROCESSES

In this section, we introduce the concept of a finite state Markov process indexed on a binary tree, and derive the joint probability and the covariance of the state process between two arbitrary nodes of the tree. We then focus attention on the leaf nodes by defining a time series indexed by the leaf nodes taken in a specified order. We demonstrate that this process has the LRD property.

Let $\mathcal{T}$ denote a binary tree of depth $D$ with root node denoted by 0. Let $X_t, t \in \mathcal{T}$ denote a random process indexed on the binary tree $\mathcal{T}$. We assume that $X_t$ takes values in a finite set $Q = \{q_1, \ldots, q_N\}$ typically a subset of $\mathbb{R}$. The Markov structure is imposed on $X_t$ by specifying the downwards (ie towards the leaf nodes) transition probabilities

$$\Pr\{X_{t\alpha} = q_j, X_{t\beta} = q_k | X_t = q_i\} = A_{i,j}(t)\ A_{i,k}(t) \tag{1}$$

where, following [1], $t\alpha$ and $t\beta$ denote respectively the left and right child nodes of node $t$. The designation $t\gamma$ denotes the parent node of node $t$, whilst $t\delta$ denotes the sibling of node $t$. Thus, rather than the state occupancy probabilities of the Markov process depending on the state at the previous time together with a transition probability, in the tree case, the probability distribution at a node depends only on its parent node together with a downwards (ie parent to child) transition probability. Also these transition probabilities are assumed to be independent (for each child node) and identical. Notice that we restrict the transition probabilities to be identical for each child node, although this restriction is not a necessary feature of the most general model.

We assume the root node 0 is intitialised with stationary probability $\pi$ satisfying $A^T\pi = \pi$ with $\pi_i > 0 \ \forall\ i$. We assume the state processes all have zero mean by suitable choice of state levels, ie $\sum_{i=1}^{N} \pi_i q_i = 0$. We also assume that $A$ is non-singular.

We give some definitions based on the idea that the leaf nodes represent a time series. Proofs of these results may be found in [9].

Definition

Consider a binary tree $\mathcal{T}$ of depth $D \geq 0$ (ie there are $2^D$ leaf nodes). The depth of a node $t$ will be denoted by $d(t) \in \{0, \ldots, D\}$. Let $s, t \in \mathcal{T}$, then we denote by $s \wedge t$ the unique common ancestor of $s$ and $t$ of maximal depth. Let $\delta(s,t)$ denote the distance between two nodes $s$ and $t$, ie $\delta(s,t) = d(s) + d(t) - 2d(s \wedge t)$.

We firstly have the following result for the joint probability between any two nodes.

Theorem 1

Let $s, t \in \mathcal{T}$, then the joint probability of the states at nodes $s$ and $t$ is given by

$$\Pr\{X_t = q_i, X_s = q_j\} = \sum_{k=1}^{N} \pi_k \left[A^{\delta(s,s\wedge t)}\right]_{k,i} \ \left[A^{\delta(t,s\wedge t)}\right]_{k,j}. \tag{2}$$

Corollary

The covariance between the state at nodes $s$ and $t$ is given by

$$\mathrm{Cov}(X_t, X_s) = \sum_{k=1}^{N} \pi_k \left[A^{\delta(s,s\wedge t)} q\right]_k \ \left[A^{\delta(t,s\wedge t)} q\right]_k. \tag{3}$$

Comment

For the case when $s$ and $t$ are restricted to the same depth, (3) becomes

$$\begin{aligned}\mathrm{Cov}(X_t, X_s) &= \sum_{k=1}^{N} \pi_k \left[A^d q\right]_k^2 \\ &= q^T \ Q_d \ q, \end{aligned} \tag{4}$$

where

$$\begin{aligned} q^T &= [q_1, \dots, q_N] \\ Q_d &= \left(A^T\right)^d \Pi A^d \end{aligned} \tag{5}$$

and $d = \delta(s, s \wedge t) = \delta(t, s \wedge t)$. Here $\Pi = \mathrm{diag}(\pi_1, \dots, \pi_N)$.

Lemma 2

Let $\mathcal{L}$ denote the leaf nodes of $\mathcal{T}$. We can find a sequence $t_n, n = 0, \dots, 2^D - 1$ of the elements of $\mathcal{L}$ so that

$$\delta(t_0, t_n) = 2\left(1 + \lfloor \log_2(n) \rfloor\right) \tag{6}$$

for $n \geq 1$ with $\delta(t_0, t_0) = 0$.

Comment

Let the leaf nodes be ordered according to the sequence $0\alpha^D, 0\alpha^{D-1}\beta$, $0\alpha^{D-2}\beta\alpha, \dots, 0\beta^{D-1}\alpha, 0\beta^D$, define the sequence $\{t_n\}$. This sequence can be readily shown to have the specified property.

We can now establish a formula for the covariance between the state at the two leaf nodes $t_0$ and $t_n$ in terms of their linear distance apart $n$.

Theorem 3

For $n \geq 1$,

$$\mathrm{Cov}(t_0, t_n) = \sum_{k=1}^{N} \pi_k \left[A^{(1+\lfloor \log_2(n) \rfloor)} q\right]_k^2. \tag{7}$$

We now focus on the dynamic properties of Markov chains by examining the properties of $A$. By assumption, we may write the eigendecomposition of $A$ as

$$A = \mathbf{1}\pi^T + \sum_{k=2}^{N} \lambda_k u_k v_k^T, \tag{8}$$

where $\mathbf{1}$ denotes the vector consisting of all 1s, and the $u_k$ and $v_k$ satisfy $v_k^T \mathbf{1} = 0 \,\forall\, k$, $u_k^T \pi = 0 \,\forall\, k$, and $u_k^T v_l = \delta_{kl}$. We assume that $1 > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_N| > 0$, and that $A$ is simple (ie diagonalisable). This last restriction is not necessary and can be relaxed (see [9]). Thus for any integer $d$, we have

$$A^d = \mathbf{1}\pi^T + \sum_{k=2}^{N} \lambda_k^d \, u_k \, v_k^T. \tag{9}$$

We now state our result on the LRD property of the state process indexed by nodes $\{t_n\}$.

Theorem 4

The state process covariance between nodes $t_0$ and $t_n$ satisfies

$$\mathrm{Cov}(t_0, t_n) \geq C \, n^{-\mu}, \tag{10}$$

where $\mu = -2\log_2(|\lambda_N|)$, and $C > 0$ depends on $A$.

Comment

We generally require the Hurst parameter $H$ given by $H = 1 - \mu/2 = 1 + \log_2(|\lambda_N|)$ to satisfy $0.5 < H < 1$. Thus we require $0 < \mu < 1$ or equivalently $1 > |\lambda_N| > 2^{-1/2}$.

## III CONCLUSION AND OPEN QUESTIONS

This paper has defined the concept of a Markov process indexed on a binary tree, and has derived the covariance structure of the process. If we define a time series by an ordering of the states at the leaf nodes of the tree, such a process is shown to have a long range dependency structure ie the covariance decays with a fractional power law, rather that exponentially in conventional (time-series) Markov processes. In other work, we have extended the concept to both continuous and discrete hidden Markov models, and we have also derived optimal smoothing and parameter estimation algorithms for such processes [9]. Application to data analysis has been addressed in [10] where it is argued that the tree indexed model is useful for multiscale analysis of network traffic.

Several open questions are :

1. What about when $A$ is singular ? Do you get exponential modes or modes of finite support ?

2. What happens when $|\lambda_N| < 2^{-1/2}$ ? This is usually a restriction on a time-series to enforce stationarity, which is not a property enjoyed here.

3. To construct a stationary process, we require $\mathrm{Cov}(t_0, t_n) = \mathrm{Cov}(t_m, t_{n+m})$ for all $m \geq 0$ (or at least over some range of $m$). Is there a transformation of the state process $X_{t_n}$ which yields a stationary process having the LRD property ?

## ACKNOWLEDGEMENTS

The authors acknowledge the support of the Department of Defence, and the helpful comments made by Dr David Wilson, Communications Division, Defence Science and Technology Organisation, Australia.

## REFERENCES

1. M. Basseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukhah and A. S. Willsky, "Modeling and Estimation of Multiresolution Stochastic Processes", *IEEE Trans. Information Theory*, v. 18, no. 2, pp. 766-784, 1992.
2. A. S. Acampora, *An Introduction to Broadband Networks*, New York : Plenum Press, 1994.
3. M. Basseville, A. Benveniste, and A. S. Willsky, "Multiscale Autoregressive Processes - Part I : Schur-Levinson Parameterizations", *IEEE Trans. Signal Processing*, v. 40, pp. 1915-1934, Aug. 1992.
4. M. Basseville, A. Benveniste, and A. S. Willsky, "Multiscale Autoregressive Processes - Part II : Lattice Structures for whitening and modeling", *IEEE Trans. Signal Processing*, v. 40, pp. 1935-1954, Aug. 1992.
5. B. Claus, "Multiscale Statistical Signal Processing : Identification of a Multiscale AR Process from a sample of an Ordinary Signal", *IEEE Trans. Signal Processing*, v. 41, no. 12, pp. 3266-3274, Dec. 1993.
6. W. E. Leland, M. Taqqu, W. Willinger and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic", *Trans. SIGCOMM93*, Ithaca NY, Sept. 1993, pp. 183-193, Association for Computing Machinery.
7. D. H. Johnson and A. R. Kumar, "Modeling and analyzing fractal point processes", *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Dallas USA 1990, pp. 1353-1356.
8. A. T. Andersen and B. F. Nielsen, "A Markovian Approach for Modeling Packet Traffic with Long-Range Dependence", *IEEE J. Sel. Areas Communications*, v. 16, no. 5, pp. 719-732, June 1998.
9. L. B. White and S. L. Perreau, "Tree-indexed Hidden Markov models and Long Range Dependency", *Department of Electrical and Electronic Engineering Report*, University of Adelaide, June 1999.
10. L. B. White, "Multiscale Markov Point Processes with Application to the Analysis of Discrete Event Data", *IEEE Workshop on Statistical Signal and Array Processing*, Corfu, Greece, 1996 (unpublished).