

# Basic Estimation of Markovian Pseudo Long-Range Dependent Processes

Stephan Robert

**Abstract**—The pseudo self similar processes are quite attractive due to their simplicity but the question we are interested in this paper concerns the basic estimation of such models. How do the standard estimators (sample mean and variance) converge with time? This will give us an indication about the time we have to collect data in order to accurately model them. With no surprise we notice that this is dependant of the Hurst parameter of course and on the number of states the model has (which defines the domain in which the behavior is self-similar). One has to collect more data with higher Hurst parameters and with more states in the Markov chain to accurately estimate the mean and variance of the process. Outside the domain where the process is self similar, standard statistics methods apply.

**Index Terms**—Pseudo self similar process, traffic engineering, modelling.



## 1 INTRODUCTION

SINCE the measurements done at Bellcore in the nineties [1] it has been generally admitted that network traffic exhibits self-similarities and long range dependance. It has also been seen that the pseudo self similar models [2], [3] are especially attractive to represent data traffic [4] although there are still queueing problems which have to be well explained (future work). There is also a nice work which have been done [5], [6] which explains how TCP generates pseudo-self-similarity. More generally than traffic engineering the pseudo self similar processes have the potentiel to model different kind of data. To generate data having a pseudo self similar behavior, it has quite recently been found [7] that the generator based on multiplexing ON/OFF sources may perform better than generators based on chaotic maps. Such statistics have already been conducted for self similar processes [8] but certainly not for pseudo self similar processes.

## 2 MARKOV MODEL

Here, we investigate the use of a simple, discrete time Markov modulated model for representing self-similar data traffic.  $X_t$  is assumed to be 0 or 1 during the  $t^{\text{th}}$  time slot, namely during time interval  $[t-1, t)$ . The Markov modulated chain is assumed to be stationary and homogeneous. Let  $Y_t = i$  be the modulator's state  $i$ ,  $i \in 1, 2, 3, \dots, k$  at time  $t$ . The arrivals of cells are modulated by a  $k$ -state discrete time Markov chain with transition probabilities  $a_{ij}(t_1, t_2) = Pr(Y_{t_2} = j | Y_{t_1} = i)$ ,  $a_{ij}(t, t+1) = a_{ij}$ . Let  $\phi_{ij}$  denote the probability of having  $j$  cells given that the modulator's state is  $i$ . More specifically  $\phi_{ij}(t) = Pr(X_t = j | Y_t = i)$ . The Markov modulated chain state probabilities are noted as  $\pi_i(t) = Pr(Y_t = i)$ ,  $i$  is referred to as the modulator's state and  $t$  to the time. All the moments are equal and no longer time dependent.

$$E\{X_t^k\} = E\{X\} = \vec{\pi} \Lambda \vec{e}$$

with  $\vec{\pi} = (\pi_1, \pi_2, \dots, \pi_n)$  and  $\vec{e}$  the unity vector.  $\Lambda$  is defined as

$$diag(E\{X|Y=1\}, E\{X|Y=2\}, \dots, E\{X|Y=k\}).$$

If  $N_m$  represents the number of arrivals in a window of  $m$  time intervals, the variance of  $N_m$  can be written ([2], [3])

---

S. Robert is with the Institute for Information and Communication Technologies, HEIG-Vd, University of Applied Sciences of Western Switzerland, 1401 Yverdon-les-Bains, Switzerland. E-mail: stephan.robert@heig-vd.ch. This work has been supported by the "Fonds National Suisse de la Recherche Scientifique".

$$\text{Var}\{N_m\} = mE\{X\} - m^2(E\{X\})^2 + 2(\sum_{i=1}^{m-1}(m-i)(\vec{\pi}\Lambda\mathbf{A}^{i+1}\Lambda\vec{e}))$$

where, according to the family of models proposed,  $\mathbf{A}$  is equal to

$$\begin{pmatrix} 1 - 1/a - \dots - 1/a^{k-1} & 1/a & \dots & 1/a^{k-1} \\ b/a & 1 - b/a & \dots & 0 \\ \dots & \dots & \dots & \dots \\ (b/a)^{k-1} & 0 & \dots & 1 - (b/a)^{k-1} \end{pmatrix}$$

with  $a < 1$  and

$$\Lambda = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

It can be shown that this process exhibits self-similarities over a finite timescale [2]. The Markov chain is driven by only 3 parameters:  $a$ ,  $b$  plus the number of states in the Markov chain  $k$ . Here, the expectation is given by

$$E\{X\} = \frac{1 - 1/b}{1 - 1/b^k}$$

with  $\lim_{n \rightarrow \infty} E\{X\} = 1 - 1/b$  and  $1/b < 1$ . In this case, Markov chains have a self-similar behavior over all timescales.

### 3 CONVERGENCE OF THE MEAN AND VARIANCE ESTIMATORS

This section is concerned with the question of how long one must observe the traffic in order to accurately estimate the mean and the variance (in order to determine  $a$  and  $b$  in our model). We will see that the self-similarity (long-range dependences) has important consequences for parameter estimation. Many estimators are more biased and converge much more slowly than in the usual short range case.

#### 3.1 Variance of the sample mean

Although the mean of the sample mean is trivial, one can be interested by the variance of the sample mean. Let  $X_t$  a discrete stationary process, pseudo self-similar, defined in Section 2,  $X = (X_t : t \geq 0)$ . By definition, the variance of the sample mean is given by

$\text{Var}\{\widehat{X}\} = E\{\widehat{X}^2\} - E^2\{\widehat{X}\}E\{\widehat{X}^2\}$  can be written

$$E\{\widehat{X}^2\} = E\left[\frac{(X_1 + X_2 + \dots + X_n)^2}{n^2}\right]$$

$$E\{\widehat{X}^2\} = \frac{E\{\sum_{i=1}^n X_i + 2\sum_{i=1}^n \sum_{j=i+1}^{n-1} X_i X_j\}}{n^2}$$

$$E\{\widehat{X}^2\} = \frac{1}{n^2} \sum_{i=1}^n E\{X_i^2\} + \frac{2}{n^2} \sum_{i=1}^n \sum_{j=i+1}^{n-1} E\{X_i X_j\}$$

where  $E\{X_i X_j\}$  are the terms due to the correlation. These terms can introduce both a strong bias and a very low convergence of the estimator. The double sum can be written

$$\sum_{i=1}^n \sum_{j=i+1}^{n-1} E\{X_i X_j\} = \sum_{i=1}^{n-1} (n-i) E\{X_0 X_i\}$$

if the process is stationary in the wide sense (WSS), in other words if  $E\{X_0 X_i\} = E\{X_j X_{i+j}\}$ . Finally,

$$E\{\widehat{X}^2\} = \frac{\vec{\pi}\Lambda^2\vec{e}}{n} + \frac{2}{n^2} \sum_{i=1}^{n-1} (n-i)(\vec{\pi}\Lambda\mathbf{A}^i\Lambda\vec{e})$$

The variance can be written

$$\begin{aligned} \text{Var}\{\widehat{X}\} &= E\{\widehat{X}^2\} - E^2\{\widehat{X}\} = \sigma^2\{\widehat{X}\} = \\ &= \frac{\vec{\pi}\Lambda^2\vec{e}}{n} - (\vec{\pi}\Lambda\vec{e})^2 + \frac{2}{n^2} \sum_{i=1}^{n-1} (n-i)(\vec{\pi}\Lambda\mathbf{A}^i\Lambda\vec{e}) \end{aligned}$$

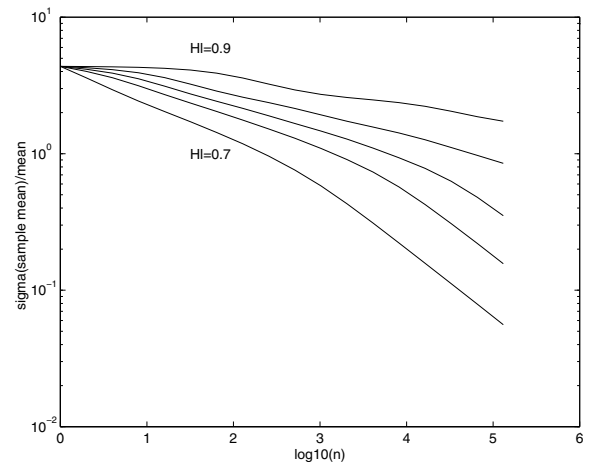


Fig. 1.  $\sigma^2\{\widehat{X}\}/\overline{X}$  as a function of  $\log_{10}(n)$  for a 5-state Markov chain with  $H_i=0.7/0.75/0.8/0.85/0.9$ ,  $E\{X\}=0.05$

This formula is valid for each discrete modulated Markov process described in Section 2.

The figure 1 shows  $\sigma\{\widehat{X}\}/\overline{X}$  as a function of  $\log_{10}(n)$  for a 5-state Markov chain. If we compare the same quantity for a pseudo self-similar process, one can show in figure 1 that the decaying of the variance of the estimator is the same after a certain number of samples. This number varies with the local Hurst parameter of the process. For  $H_l = 0.7$ , the knee is very apparent. At the beginning, for small  $n$  ( $n \leq 500$ ), the slope of the curve is  $-0.6$  and for large  $n$  ( $n > 500$ ), the slope of the curve is  $-1$ . Intuitively, we see that the knee separates the self-similar and the Poisson behavior of the process. Notice here that the sample mean can be computed on-line with the following formula:

$$\widehat{X}_n = \frac{1}{n} \left( (n-1)\widehat{X}_{n-1} + X_n \right)$$

and the sample variance  $Var\{\widehat{X}_n\}$  by

$$\frac{1}{n-1} \left( (n-2)Var\{\widehat{X}_{n-1}\} + (X_n - \widehat{X})^2 \right)$$

If we consider the aggregation now, one can observe that if  $X_t^{tot} = \sum_{i=1}^n X_t^{(i)}$  and  $X_t^{(i)}$  are independant,  $Var(X_t^{tot}) = nVar(X_t^{(i)})$ . So

$$\frac{\sqrt{Var\{X_t^{tot}\}}}{E\{X_t^{tot}\}} = \frac{\sqrt{Var\{X_t^{(i)}\}}}{\sqrt{n}E\{X_t^{(i)}\}}$$

### 3.2 Mean of the sample variance

Here again, we will consider the general case of Markov modulated processes. The structure we studied depends only on 3 parameters:  $a$ ,  $b$  and  $n$ . The local Hurst parameter  $H_l$  can be obtained from the variances but the estimation of it can be difficult, especially for large  $H_l$ . By definition

$$Var\{\widehat{X}\} = \frac{1}{n-1} \sum_{i=1}^n \left( X_i - \frac{1}{n} \sum_{j=1}^n X_j \right)^2$$

After a straightforward calculation, we find that

$$(n-1)E\{Var\{\widehat{X}\}\} = nE\{X^2\} - \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^n E\{X_i X_j\} + E\left\{ \sum_{j=1}^n X_j \right\}^2$$

but

$$\left( \sum_{i=1}^n X_i \right)^2 = nE\{X_t^2\} + 2E\left\{ \sum_{i=1}^n X_i \sum_{j=1}^{n-1} X_j \right\}$$

and

$$\sum_{i=1}^n \sum_{j=1}^n E\{X_i X_j\} = nE\{X^2\} + 2E\left\{ \sum_{i=1}^n n X_i \sum_{j=i+1}^{n-1} X_j \right\}.$$

One can find that

$$\begin{aligned} (n-1)E\{Var\{\widehat{X}\}\} &= \\ (n+1)E\{X^2\} + \frac{2}{n} \left\{ \sum_{i=1}^n \sum_{j=i+1}^{n-1} E\{X_i X_j\} \right. \\ &\quad \left. - nE\{X^2\} + 2E\left\{ \sum_{i=1}^n \sum_{j=i+1}^{n-1} X_j \right\} \right\} = \\ (n-1)E\{X^2\} - \frac{2}{n} \sum_{i=1}^n \sum_{j=i+1}^{n-1} E\{X_i X_j\} \end{aligned}$$

Assume now that the process is WSS. In this case, the above expression becomes

$$\begin{aligned} (n-1)E\{Var\{\widehat{X}\}\} &= \\ (n-1)Var\{X\} + (n-1)E^2\{X\} \\ &\quad - \frac{2}{n} \sum_{i=1}^n \sum_{j=i+1}^{n-1} (Cov\{X_i X_j\} + E^2\{X\}). \end{aligned}$$

It can be shown that

$$\sum_{i=1}^n \sum_{j=i+1}^{n-1} E^2\{X\} = \frac{n(n-1)}{2} E^2\{X\}$$

and

$$\begin{aligned} \sum_{i=1}^n \sum_{j=i+1}^{n-1} Cov\{X_i X_j\} &= \\ \sum_{i=1}^{n-1} (n-1)Cov\{X_0 X_i\} &= \\ \sum_{i=1}^{n-1} (n-i)C(i) & \end{aligned}$$

So, we find that

$$\begin{aligned} E\{Var\{\widehat{X}\}\} &= \\ Var\{X\} \left[ 1 - \frac{2}{n-1} \sum_{i=1}^{n-1} \left( 1 - \frac{i}{n} \frac{C(i)}{Var\{X\}} \right) \right] & \end{aligned}$$

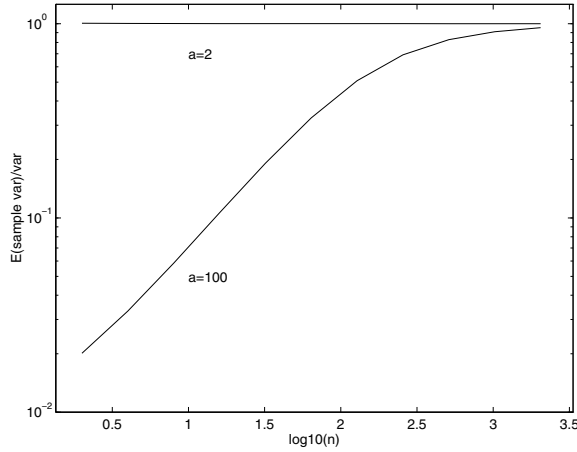


Fig. 2.  $E\{\widehat{Var}\{X\}\}/Var\{X\}$  as a function of  $\log_{10}(n)$  for a 2-state Markov chain with  $b=1$  and  $a=2/100$

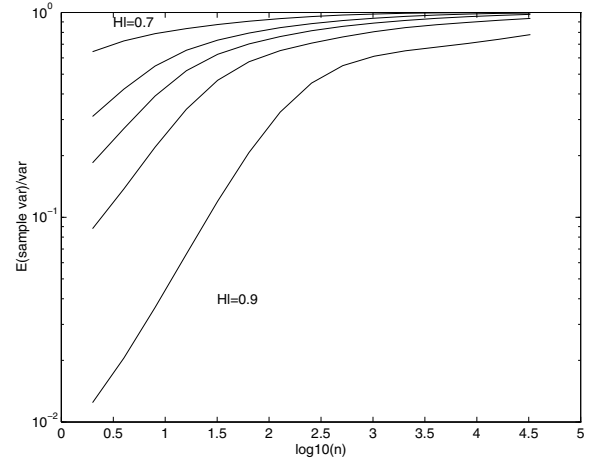


Fig. 3.  $E\{\widehat{Var}\{X\}\}/Var\{X\}$  as a function of  $\log_{10}(n)$  for a 2-state Markov chain with  $b=1$  and  $a=2/100$

with

$$C(i) = \vec{\pi}\Lambda(\mathbf{A}^i - \vec{e}\vec{\pi})\Lambda\vec{e}$$

and

$$Var\{X\} = \vec{\pi}\Lambda^2\vec{e} - (\vec{\pi}\Lambda\vec{e})^2$$

Now, look at the figure 2. We consider two 2-state Markov chains, when  $a = 2$  and when  $a = 100$ ,  $b = 1$ . In the first case, the process is uncorrelated and the estimator of the variance unbiased. In the second case, the process is positively correlated and need to converge.  $E\{\widehat{Var}X\}$  converges if  $\sum_{i=1}^{n-1} C(i)(1 - i/n)$  converges when  $n \rightarrow \infty$ .  $(1 - i/n) \leq 1$  then it is sufficient to examine the convergence of  $\sum_{i=1}^{n-1} C(i)$ . By hypothesis this condition is true for each finite modulated Markov chain as described in section 2. We notice that  $E\{\widehat{Var}\{X\}\}/Var\{X\} \uparrow 1$  when  $C(i) > 0, \forall i$  because  $2/(n-1)\sum_{i=1}^{n-1}(1 - (i/n)(C(i)/Var\{X\})) > 0, \forall n > 1$ . This is not the case when the process is negatively correlated. Figure 3 shows  $E\{\widehat{Var}\{X\}\}/Var\{X\}$  for a 5-state Markov chain exhibiting self-similarities on a finite timescale. As before, the curve converges to 1 but much slower, especially when  $H_l$  is large. If we consider the aggregation now, one can observe that if  $X_t^{tot} = \sum_{i=1}^n X_t^{(i)}$  and  $X_t^{(i)}$  are independant,  $Var(X_t^{tot}) = nVar(X_t^{(i)})$ . So

$$\frac{E\{Var\{X_t^{tot}\}\}}{Var\{X_t^{tot}\}} = \frac{E\{Var\{X_t^{(i)}\}\}}{Var\{X_t^{(i)}\}}$$

### 3.3 Variance of the sample variance

In this subsection, we are interested in observing the variance of the sample variance. First, we must compute  $E\{\widehat{Var}^2 X\}$ , note that  $E^2\{\widehat{Var}X\}$  is already known. By definition

$$Var\{\widehat{X}\}^2 = \frac{1}{(n-1)^2} \left( \sum_{i=1}^n (X_i - \widehat{X})^2 \right)^2$$

$$Var\{\widehat{X}\}^2 = \frac{1}{(n-1)^2} \sum_{i=1}^n \sum_{j=1}^n (X_i - \widehat{X})(X_j - \widehat{X})$$

This can be written

$$\begin{aligned} Var\{\widehat{X}\}^2 = & \frac{1}{(n-1)^2} \left( \sum_{i,j} X_i^2 X_j^2 - \frac{2}{n} \sum_{i,j,k} X_i^2 X_j X_k + \right. \\ & \left. \sum_{i,k,l} X_i^2 X_k X_l - \frac{2}{n} \sum_{i,j,k} X_i X_j^2 X_k + \frac{4}{n^2} \sum_{i,j,k,l} X_i X_j X_k X_l - \right. \\ & \left. \frac{2}{n^3} \sum_{i,k,l,m} X_i X_k X_l X_m + \frac{1}{n^2} \sum_{j,k,l} X_j^2 X_k X_l - \right. \\ & \left. \frac{2}{n^3} \sum_{j,k,l,m} X_j X_k X_l X_m + \frac{1}{n^4} \sum_{k,l,m,p} X_k X_l X_m X_p \right) \end{aligned}$$

Finally

$$\begin{aligned} Var\{\widehat{X}\}^2 = & \frac{1}{(n-1)^2} \left( \sum_{i=1}^n \sum_{j=1}^n X_i^2 X_j^2 - \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n X_i^2 X_j X_k + \right. \end{aligned}$$

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n X_i X_j X_k X_l$$

One can simply verify this formula in setting  $X = X_1 + X_2$ . Now, it is necessary to compute  $E\{X_{t_1}^{s_1} X_{t_2}^{s_2} X_{t_3}^{s_3} X_{t_4}^{s_4}\}$  in the general case. Without loss of generality, assume that  $t_1 < t_2 < t_3 < t_4$  and  $s_i \in N, i \in N$ .

$$E\{X_{t_1}^{s_1} X_{t_2}^{s_2} X_{t_3}^{s_3} X_{t_4}^{s_4}\} = \sum_{x_1, x_2, x_3, x_4 > 0} x_1^{s_1} x_2^{s_2} x_3^{s_3} x_4^{s_4}$$

$$Pr(X_{t_1} = x_1 \cap X_{t_2} = x_2 \cap X_{t_3} = x_3 \cap X_{t_4} = x_4)$$

but

$$Pr(X_{t_1} = x_1 \cap X_{t_2} = x_2 \cap X_{t_3} = x_3 \cap X_{t_4} = x_4) =$$

$$\sum_{i,j,k,l} Pr(X_{t_4} = x_4 \cap X_{t_3} = x_3 \cap X_{t_2} = x_2 \cap$$

$$X_{t_1} = x_1 | Y_{t_4} = l \cap Y_{t_3} = k \cap Y_{t_2} = j \cap Y_{t_1} = i)$$

$$Pr(Y_{t_4} = l \cap Y_{t_3} = k \cap Y_{t_2} = j \cap Y_{t_1} = i)$$

$$= \sum_{i,j,k,l} Pr((X_{t_4} = x_4 \cap X_{t_3} = x_3 \cap X_{t_2} = x_2 \cap$$

$$X_{t_1} = x_1 | Y_{t_4} = l) \cap$$

$$(X_{t_4} = x_4 \cap X_{t_3} = x_3 \cap X_{t_2} = x_2 \cap X_{t_1} = x_1 | Y_{t_3} = k) \cap$$

$$(X_{t_4} = x_4 \cap X_{t_3} = x_3 \cap X_{t_2} = x_2 \cap X_{t_1} = x_1 | Y_{t_2} = j) \cap$$

$$(X_{t_4} = x_4 \cap X_{t_3} = x_3 \cap X_{t_2} = x_2 \cap X_{t_1} = x_1 | Y_{t_1} = i))$$

$$Pr(Y_{t_4} = l \cap Y_{t_3} = k \cap Y_{t_2} = j \cap Y_{t_1} = i)$$

$$= \sum_{i,j,k,l} Pr(X_{t_4} = x_4 | Y_{t_4} = l) Pr(X_{t_3} = x_3 | Y_{t_3} = k)$$

$$Pr(X_{t_2} = x_2 | Y_{t_2} = j) Pr(X_{t_1} = x_1 | Y_{t_1} = i)$$

$$Pr(Y_{t_4} = l | Y_{t_3} = k) Pr(Y_{t_3} = k | Y_{t_2} = j)$$

$$Pr(Y_{t_2} = j | Y_{t_1} = i) Pr(Y_{t_1} = i)$$

$$E\{X_{t_1}^{s_1} X_{t_2}^{s_2} X_{t_3}^{s_3} X_{t_4}^{s_4}\} =$$

$$\sum_{x_1, x_2, x_3, x_4 > 0} E\{X_{t_4} = x_4 | Y_{t_4} = l\} Pr(Y_{t_4} = l | Y_{t_3} = k)$$

$$E\{X_{t_3} = x_3 | Y_{t_3} = k\} Pr(Y_{t_3} = k | Y_{t_2} = j)$$

$$E\{X_{t_2} = x_2 | Y_{t_2} = j\} Pr(Y_{t_2} = j | Y_{t_1} = i)$$

$$E\{X_{t_1} = x_1 | Y_{t_1} = i\} Pr(Y_{t_1} = i)$$

and finally

$$E\{X_{t_1}^{s_1} X_{t_2}^{s_2} X_{t_3}^{s_3} X_{t_4}^{s_4}\} = \vec{\pi} \Lambda^{s_1} \mathbf{A}_{12} \Lambda^{s_2} \mathbf{A}_{23} \Lambda^{s_3} \mathbf{A}_{34} \Lambda^{s_4} \vec{e}$$

A particular case is given by  $E\{X_{t_1} X_{t_2}\} = \vec{\pi} \Lambda \mathbf{A}_{12} \Lambda \vec{e}$ . The above formula is easily generalizable. Taking the mean of the sample variance, one obtains

$$E\{\widehat{Var}\{X\}^2\} = \frac{1}{(n-1)^2} \left( \sum_{i=1}^n \sum_{j=1}^n E\{X_i^2 X_j^2\} -$$

$$\frac{2}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n E\{X_i^2 X_j X_k\} +$$

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n E\{X_i X_j X_k X_l\} \right)$$

with

$$E\{X_i^2 X_j^2\} = \sum_{i=1}^n E\{X_i\}^4 + \sum_{i=1}^n 2(n-i) E\{X_0^2 X_i^2\} =$$

$$\vec{\pi} \Lambda^2 \mathbf{A}_{ij} \Lambda^2 \vec{e}$$

$$E\{X_i^2 X_j X_k\} = \vec{\pi} \Lambda^2 \mathbf{A}_{ij} \Lambda \mathbf{A}_{jk} \Lambda \vec{e}$$

$$E\{X_i X_j X_k X_l\} = \vec{\pi} \Lambda \mathbf{A}_{ij} \Lambda \mathbf{A}_{jk} \Lambda \mathbf{A}_{kl} \Lambda \vec{e}$$

with  $\mathbf{A}_{ij} = \mathbf{A}^{|i-j|}$  and  $\mathbf{A}^0 =$  identity matrix. Now if we know in which state the chain is at the beginning of the sequence, one can notice that  $E\{X_i^s | Y_0\} = (\vec{\pi}_0 \mathbf{A}^{i-1}) \Lambda^s \vec{e}$ . It is obvious that, for  $i \rightarrow \infty, \vec{\pi}_0 \mathbf{A}^{i-1} \rightarrow \vec{\pi}$  stationary.

## REFERENCES

- [1] W.E. Leland and D.V. Wilson. High Time-Resolution Measurement and Analysis of LAN Traffic: Implications for LAN Interconnection. In *IEEE Infocom*, Bal Harbor, FL, April 1991.
- [2] S. Robert and J.-Y. Le Boudec. On a markov modulated chain exhibiting self-similarities over finite timescale. *Performance Evaluation, Elsevier*, 27 and 28:159-173, October 1996.
- [3] S. Robert and J.-Y. Le Boudec. New models for self-similar traffic. *Performance Evaluation, Elsevier*, 30:57-68, July 1997.
- [4] M. Kahane; Y. Ben-Shimol; D. Sadot. Performance evaluation of pseudo self-similar traffic. In *ICECS 2004*, December 2004.
- [5] I. Matta L. Guo, M. Crovelly. How does tcp generate pseudo-self-similarity? *MASCOTS*, August 2001.
- [6] I. Matta L. Guo, M. Crovelly. Corrections to 'how does tcp generate pseudo-self-similarity?'. *ACM SIGCOMM Computer Communication Review*, 32(2), April 2002.
- [7] J. Blomskld G. Horn, A. Kvalbein and E. Nilsen. An empirical comparison of generators for self similar simulated traffic. *Performance Evaluation, Elsevier*, 64:162-190, February 2007.
- [8] P. Abry M. Roughan, D. Veitch. Real-time estimation of the parameters of long-range dependence. *IEEE/ACM Transactions on Networking*, 8:467-478, August 2000.