On the entropy of wide Markov chains

Valerie Girardin

Laboratoire de Mathématiques N. Oresme, UMR 6139, Campus II, BP5186, 14032 Caen, France, girardin@math.unicaen.fr

Abstract. Burg entropy concepts are here introduced in the field of wide Markov chains. These random sequences are the second-order equivalent of Markov chains: their future evolution, in terms of second order properties, conditional on the past and present, depends only on the present. Either periodically correlated or multivariate stationary, they can be characterized in terms of autoregressive models of order one.

Keywords: Auto-regressive processes; Burg entropy; Multivariate stationary processes; Periodically correlated processes; Wide Markov processes.
PACS: 02.50.Ga, 89.70.Cf, 02.50.Cw

INTRODUCTION

The future evolution of a (strong) Markov chain conditional to its past and present is known to depend only on its present. In terms of second order (that is L^2) properties, it is sufficient to consider the projection onto the linear subspaces spanned by the sequence, which leads to the notion of wide sense Markov chains. Specifically, a square integrable random sequence is a wide Markov (WM) chain if its second-order projection (that is in the sense of the L^2-norm) onto its past and present depends only on its present.

A square integrable scalar process is periodically correlated if its covariance function is periodic. A multivariate square integrable process is multivariate (weakly) stationary if its covariance function is invariant by translation of time. A one-to-one relationship exists between periodically correlated (PC) sequences and multivariate stationnary (MS) sequences; this duality allows one to study jointly the second-order structure of periodically correlated wide Markov (PCWM) chains and of multivariate stationary wide Markov (MSWM) chains, in terms of covariance, correlation and reflection coefficients. The definitive characterization of the subclass of MSWM chains dual to PCWM chains is given in Castro and Girardin [4] in terms of autoregressive processes of order one, using generalized reflection coefficients matrices introduced in Castro and Girardin [3].

The convenient entropy for studying weakly stationary random sequences is known to be Burg entropy applied to spectral densities. The maximum of entropy among multivariate stationary sequences is proven in Castro and Girardin [3] to be obtained for a multivariate autoregressive process (MAR). Burg entropy of WM chains will be determined explicitly below. The maximum of entropy is then discussed under various constraints. A closed form expression for the variance of the innovation of the studied sequence also derives from the computation of Burg entropy.

The paper is organized as follows. Necessary basics on WM chains, MS random sequences and PC scalar random sequences are given in the next section, with a focus on their duality. In the following section, autoregressive models are defined, with a particular attention given to their spectral densities, and the WM chains are characterized in terms of autoregressive processes of order one. Burg entropy is applied to WM chains in the last section, with explicit computation.

MULTIVARIATE STATIONARY AND PERIODICALLY CORRELATED WIDE MARKOV CHAINS

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $L^2(\Omega)$ denote as usual the space of zero-mean second order random variables. Let $L^2_d(\Omega)$ for $d \in \mathbb{N}^*$ denote the space of all d-variate random variables $V = (V_1, \ldots, V_d)$ such that $V_i \in L^2(\Omega)$, equipped with the Euclidean norm and inner product. A real-valued d-variate sequence $\mathbf{Z} = (Z(n))_{n\in\mathbb{Z}}$ is a second-order multivariate stochastic process if $Z(n) \in L^2_d(\Omega)$ for all $n \in \mathbb{Z}$. A random sequence $\mathbf{Z}$ is a wide sense Markov chain if (with probability 1)

$$\widehat{\mathbb{E}}[(Z(n_k))|Z(m), n_1 \leq m \leq n_{k-1}] = \widehat{\mathbb{E}}[Z(n_k)|Z(n_{k-1})], \quad n_1 < \cdots < n_k, \tag{1}$$

where $\widehat{\mathbb{E}}[\ \cdot \mid Z(m), k \leq m \leq l]$ denotes the (second-order) projection onto the linear subspace $Sp\{Z(n) : k \leq n \leq l\}$ of $L^2_d(\Omega)$. Note that if $\mathbf{Z}$ is Gaussian, then it is a Markov chain in the usual sense.

The coefficients of the covariance matrices $R_Z(m,n)$ of $\mathbf{Z}$ are defined as

$$R_Z(m,n)_{kl} = \mathbb{E}[Z_k(m)Z_l(n)], \quad m,n \in \mathbb{Z},\ 0 \leq k,l \leq d-1;$$

thus the covariance function R_Z is a positive definite matrix-valued function of two variables. For a so-called basic process, these coefficients are never null. The process is stationary (in the weak or second-order sense) if

$$R_Z(m,n) = \mathscr{R}_Z(n-m), \quad m,n \in \mathbb{Z},$$

for a positive definite matrix valued-function $\mathscr{R}_Z$ of one variable.

A straightforward application of the definition shows that a second order scalar random sequence is a WM chain if and only if its covariance function R_Z is triangular, that is satisfies

$$R_Z(n,n)R_Z(m,u) = R_Z(m,n)R_Z(n,u), \quad m \leq n \leq u \in \mathbb{Z}.$$

The correlation function ρ_Z of $\mathbf{Z}$ is defined by

$$\rho_Z(m,n) = R_Z(m,n)R_Z(m,m)^{-1}, \quad m,n \in \mathbb{Z}.$$

If the process is stationary, then $\rho_Z(m,n) = \mathscr{R}_Z(n-m)\mathscr{R}_Z(0)^{-1}$. A triangular characterization of WM chains in terms of correlations also exists, due to Doob [2] for the scalar continuous-time case and to Beutler [1] for the multivariate case. It applies directly to WM chains. As shown in Castro and Girardin [4], properties of the reflection coefficients can also be used to characterize WM chains.

A scalar second order process $\mathbf{Y}$ is periodically correlated if its covariance function is periodic, that is if some $d \in \mathbb{N}^*$ exists such that $\mathbb{E}Y(n+d)Y(m+d) = \mathbb{E}Y(n)Y(m)$, for $n,m \in \mathbb{Z}$. Theoretically, $d \geq 1$, even if, in all meaningful applications, $d > 1$; see Franses [5] for application in econometrics. A one-to-one relationship is defined between the class of scalar non stationary PC processes $\mathbf{Y}$ and a subclass of MS processes $\mathbf{Z}$ by setting $Z_k(n) = Y(k+dn)$ for the k-th component of the d-variate process $\mathbf{Z}$. Gladyshev [6] proved that $\mathbf{Y}$ is periodically correlated if and only if $\mathbf{Z}$ is weakly stationary. For $m,n \in \mathbb{Z}$ and $0 \leq k,l < d$, we have

$$\begin{aligned} R_Y(k+dn,l) &= \mathbb{E}[Y(k+dn)Y(l)] = \mathbb{E}[Y(k+d(n+m))Y(l+dm)] \\ &= \mathbb{E}Z_k(n)Z_l(0) = R_Z(m,m+n)_{kl} = \mathscr{R}_Z(n)_{kl}. \end{aligned}$$

Nematohellahi and Soltani [10] gave an explicit expression for the coefficients of the covariance function of the PCWM chains, namely

$$R_Y(k+dn,l) = \widetilde{g}(d-1)^n \frac{\widetilde{g}(k-1)}{\widetilde{g}(l-1)]} R_Y(l,l), \quad 0 \leq k,l < d,$$

where

$$\widetilde{g}(-1) = 1 \quad \text{and} \quad \widetilde{g}(j) = \prod_{i=0}^{j} \frac{R_Y(i,i)}{R_Y(i,i+1)}, \quad j \in \mathbb{N}. \tag{2}$$

This yields the next characterization of these sequences in terms of covariance matrices, proven to hold in Castro and Girardin [4].

Theorem 1 *There is a one-to-one correspondence between the PCWM chains and the MSWM chains such that $\mathscr{R}(n) = c^n AB'$, for the constant $c \in \mathbb{R}$ and column vectors $A = (a_i)$ and $B = (b_i)$ defined as follows:*

$$c = \widetilde{g}(d-1), \quad a_i = \widetilde{g}(i-1) \quad \textit{and} \quad b_i = \frac{R_Y(i,i)}{\widetilde{g}(i-1)}, \quad 0 \leq i \leq d-1. \tag{3}$$

In the following, we will refer to these special MSWM chains as to MSD chains.

AUTOREGRESSIVE MODELS AND SPECTRAL DENSITY

An MS process $\mathbf{Z}$ is an autoregressive process, or MAR(N), if it has a representation

$$\sum_{k=0}^{N} A(k)Z(n-k) = \varepsilon(n), \quad n \in \mathbb{Z}, \tag{4}$$

where the coefficients $A(k)$ are $d \times d$ matrices, $A(0)$ is a unit lower triangular matrix and ε is a multivariate white noise process with diagonal covariance matrix Σ. Similarly, a PC process $\mathbf{Y}$ with period d is a periodic autoregressive process, or PAR$(d,(N_1,\ldots,N_d))$, if it has a representation

$$Y(n) + \sum_{j=1}^{N_n} \alpha_n(j)Y(n-j) = w(n), \quad n \in \mathbb{Z},$$

where $N_n = N_{n+d}$, $\alpha_n(j) = \alpha_{n+d}(j)$ and w is a white noise process with periodic variance $\sigma_n^2 = \sigma_{n+d}^2$, for $n \in \mathbb{Z}$. This relation can be written

$$Y(k+ld) + \sum_{j=1}^{N_k} \alpha_k(j) Y(k+d(l-j)) = w(k+ld), \quad l \in \mathbb{Z},\ k = 0, \ldots, d-1,$$

obviously related to Relation (4) so that $\mathbf{Y}$ is a PAR$(d,(N_1,\ldots,N_d))$ if and only if the dual $\mathbf{Z}$ is a MAR(N) with $N = \max_k [(N_k - k)/d] + 1$, where $[\cdot]$ denotes the integer part of a real number.

The following two structural characterizations of PCWM and MSWM chains in terms of autoregressive models are essential. They are proven to hold in Castro and Girardin [4] by using reflection coefficients.

Theorem 2 *The class of MSWM chains is exactly the class of stationary MAR(1) processes, with general representation*

$$A(0)Z(n) + A(1)Z(n-1) = \varepsilon(n), \quad n \in \mathbb{Z}.$$

They are dual with the PAR$(d,(N_0,\ldots,N_{d-1}))$ processes with $1 \leq N_i \leq 2d - i$.

The class of PCWM chains is exactly the class of PAR$(d,(1,\ldots,1))$ processes, with representation

$$Y(n) + \alpha_n Y(n-1) = w(n), \quad n \in \mathbb{Z}, \tag{5}$$

with $\alpha_n = \alpha_{n+d}$. The class of their dual MSD chains is exactly the class of stationary MAR(1) processes $\mathbf{Z}$, with representation

$$A(0)Z(n) + A(1)Z(n-1) = \varepsilon(n), \quad n \in \mathbb{Z}, \tag{6}$$

where ε is a white noise with diagonal covariance matrix Σ, the matrix $A(0)$ is a unit upper triangular matrix with only $2d-1$ non zero entries,

$$A(0)_{l,l} = 1,\ 0 \leq l \leq d-1, \quad \textit{and} \quad A(0)_{l,l-1} = -\rho_{\mathbf{Y}}(l-1,l),\ 1 \leq l \leq d-1, \tag{7}$$

and the matrix $A(1)$ has a unique non zero entry,

$$A(1)_{0,d-1} = -\rho_{\mathbf{Y}}(d-1,d). \tag{8}$$

Exemple 1 The MAR(1) process $\mathbf{Z}$ with representation

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -0.1 & 1 & 0 & 0 \\ 0 & -0.2 & 1 & 0 \\ 0 & 0 & -0.3 & 1 \end{pmatrix} Z(n) + \begin{pmatrix} 0 & 0 & 0 & -0.4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} Z(n-1) = \begin{pmatrix} w(4n) \\ w(4n+1) \\ w(4n+2) \\ w(4n+3) \end{pmatrix}$$

is an MSD process dual to the PAR(2,(1,1,1,1)) process $\mathbf{Y}$ with representation

$$\begin{cases} Y(4n) & - & 0.4Y(4n-1) & = & w(4n) \\ Y(4n+1) & - & 0.1Y(4n) & = & w(4n+1) \\ Y(4n+2) & - & 0.2Y(4n+1) & = & w(4n+2) \\ Y(4n+3) & - & 0.3Y(4n+2) & = & w(4n+3). \end{cases}$$

Both $\mathbf{Z}$ and $\mathbf{Y}$ are WM chains. $\square$

The spectral density $H = (h_{kl})$ of an MS sequence $\mathbf{Z}$ is a positive-definite Hermitian $d \times d$-matrix valued function such that

$$\mathscr{R}(n)_{kl} = \int_{[0,2\pi]} h_{kl}(\lambda) e^{in\lambda} d\lambda, \quad n \in \mathbb{Z},\ 0 \leq k,l \leq d-1.$$

The cross-specrtral densities h_{kl} for $k \neq l$ are genrally complex-valued, while the auto-spectral densities h_{kk} are real-valued and nonnegative; see Priestley [11] for examples and more on spectral analysis of multivariate processes. The matrix function H can be considered as a spectral density also for the dual PC sequence $\mathbf{Y}$ which, being non stationary, does not have a natural spectral density.

If $\mathbf{Z}$ is an MSD chain, then, by Theorem 2, $\mathbf{Z}$ is a MAR(1) process with representation (6). Set

$$P(\lambda) = A(0) + A(1)e^{ik\lambda}, \quad \lambda \in [0,2\pi].$$

If the polynomial Det$P$ has all its zeros outside the unit circle, the spectral density of $\mathbf{Z}$ is well defined and takes the form

$$H(\lambda) = P^{-1}(\lambda)\Sigma[P^{-1}(\lambda)]^*, \tag{9}$$

where Σ is the diagonal covariance matrix of ε and the star denotes conjugate and transpose; see Troutman [12] for details. Reciprocally, an MS sequence with spectral density given by (9) is a MAR sequence with representation (6); see Castro and Girardin [3] for details. The matrix structure of the spectral density of any wide Markov chain is thus completely known. For MSD (or PCWM) chains, due to the form of P induced by relations (7) and (8), this structure is particularly simple.

Nematollahi and Soltani [10] have specifically studied the spectral density of an MSD sequence through its coefficients, proving straightforwardly from (9) that

$$h_{jk}(\lambda) = \frac{\alpha_{jk}e^{i\lambda} + \beta_{jk}}{|1 + \widetilde{g}(d-1)e^{i\lambda}|^2}, \tag{10}$$

where $\widetilde{g}$ is defined in (2) and

$$\begin{aligned}
\alpha_{jk} &= \widetilde{g}(d-1)\left[\frac{\widetilde{g}(k-1)R_{\mathbf{Y}}(j,j)}{\widetilde{g}(j-1)} - \frac{\widetilde{g}(j-1)R_{\mathbf{Y}}(k,k)]}{\widetilde{g}(k-1)}\right], \\
\beta_{jk} &= \frac{\widetilde{g}(j-1)R_{\mathbf{Y}}(k,k)}{\widetilde{g}(k-1)} - \frac{\widetilde{g}(k-1)R_{\mathbf{Y}}(j,j)\widetilde{g}(d-1)^2}{\widetilde{g}(j-1)}.
\end{aligned}$$

WIDE MARKOV CHAINS AND ENTROPY

The classical Burg entropy can be applied to MS sequences $\mathbf{Z}$, and hence by duality to PC sequences $\mathbf{Y}$, under the form $\mathscr{I}_{\mathbf{Z}} = \mathscr{I}_{\mathbf{Y}} = \mathscr{I}[H]$, where

$$\mathscr{I}[H] = \int_{[0,2\pi]} \ln \mathrm{Det}\, H(\lambda) d\lambda,$$

with $\mathscr{I}[H] = -\infty$ if the integral is not defined.

For WM chains, this entropy takes the following simple form.

Proposition 1 *The Burg entropy of an MSD chain or PCWM chain is*

$$\mathscr{I}_{\mathbf{Z}} = \mathscr{I}_{\mathbf{Y}} = d \ln \left(\frac{b^2 - a^2}{a^{2\pi} b^{2\pi - 1}} \right), \qquad (11)$$

where

$$b = \prod_{i=0}^{d-1} R(i,i) \quad \textit{and} \quad a = \prod_{i=0}^{d-1} R(i+1,i). \qquad (12)$$

Note that $a < b$.

Proof Indeed, since $H(\lambda)$ is positive-definite for any λ, we know that

$$\mathscr{I}(\mathbf{Z}) = \int \ln \mathrm{Det} H(\lambda) d\lambda = \int \mathrm{Tr}[\ln H(\lambda)] d\lambda,$$

where Tr denote the trace operator. Since by definition $\widetilde{g}(d-1) = a/b$ in (3), we deduce from (10) that

$$\begin{aligned} \mathscr{I}(\mathbf{Z}) &= \sum_{k=0}^{d-1} \int_{[0,2\pi]} \ln h_{kk}(\lambda) d\lambda = \sum_{k=0}^{d-1} \int_{[0,2\pi]} \ln \frac{R(k,k)(1-\widetilde{g}(d-1)^2)}{|1-\widetilde{g}(d-1)e^{i\lambda}|^2} d\lambda \\ &= \sum_{k=0}^{d-1} \ln R(k,k) \\ &\quad + d \int_{[0,2\pi]} \ln(1-\widetilde{g}(d-1)^2) d\lambda - d \int_{[0,2\pi]} \ln |1-\widetilde{g}(d-1)e^{i\lambda}|^2 d\lambda \\ &= \ln b + d \ln \left(1 - \frac{a^2}{b^2}\right) - d \int_{[0,2\pi]} \ln \left(1 - 2\frac{a}{b}\cos\lambda + \frac{a^2}{b^2}\right) d\lambda, \end{aligned}$$

and the result follows using the change of variable $\cos\lambda = (1-t^2)/(1+t^2)$. □

The entropy tends to infinity when either a or b tends to zero. When $b = 0$, the variance of $Y(i)$ is null for some i, meaning that $Y(i)$ is deterministic. When $a = 0$, the covariance of $Y(i)$ and $Y(i+1)$ is null for some i; due to (5), $\mathbb{V}\mathrm{ar} Y(i+1) = -\alpha_i R_Y(i,i+1)$, and hence $b = 0$ and the variable $Y(i)$ is deterministic.

If both $\mathscr{R}_Z(0)$ and $\mathscr{R}_Z(1)$ (that is $R_Y(k,l)$ and $R_Y(k+1,l)$ for $0 \le k,l < d$) are fixed, the maximum of Burg entropy among general MS sequences is shown to exist and to be obtained for a MAR(1) process in Theorem 3 of Castro and Girardin [3] by using spectral densities arguments. Since $\mathscr{R}_Z(0)$ and $\mathscr{R}_Z(1)$ characterize a unique MSWM chain, which is a MAR(1) process with representation (5) and spectral density given by (10), the maximum entropy among MS sequences is obtained for this MAR(1) process whose dual PCWM chain is a PAR($d,(1,\ldots,1)$) process with representation (6).

If only $\mathscr{R}_Z(0)$ (that is $R_Y(k,l)$ for $0 \le k,l < d$) is fixed, we can study entropy among WM chains. Clearly, due to (11), the maximum entropy is obtained for all MSD sequences such that the covariance between two successive steps is null for al least one coordinate, that is such that $Y(i)$ is deterministic for some $0 \leq i < d$.

For any random sequence $\mathbf{Z}$, let e_n denote the error of the best linear prediction of $Z(t)$ knowing the finite past $Z(t-1), \ldots, Z(t-n)$. The innovation process e of $\mathbf{Z}$ is the error of the best linear prediction of $Z(t)$ knowing the infinite past. It represents the information on $Z(n)$ brought by the knowledge of the whole past. Let σ_n^2 denote the variance of e_n and σ^2 the variance of innovation, that is the variance of the white noise e.

Proposition 2 *The variance of innovation of any WM chain is*

$$\sigma^2 = \left[\frac{b^2 - a^2}{a^{2\pi} b^{2\pi - 1}}\right]^d = \sigma_n^2, \quad n \in \mathbb{Z},$$

where a and b are defined in (12).

Proof Precisely,

$$\sigma_n^2 = \mathbb{V}\mathrm{ar}\left(Z(m) - \widehat{\mathbb{E}}[Z(m) \mid Z(m-1), \ldots, Z(m-n)]\right).$$

On the one hand, due to projection properties, σ_n^2 converges to σ^2 when n tends to infinity. On the other hand, for WM chains, due to (1), $\sigma_n^2 = \sigma_1^2$.

Finally, Helson and Lowdenslager [7] proved that $\sigma^2 = \exp \mathscr{I}(\mathbf{Z})$. The result follows from Proposition 1. □

When $a = b$, the variance of innovation is minimum, equal to zero, and the entropy is infinite. This happens especially when $R(i, i+1) = R(i, i)$ for all i; then, in representation (5), we get $\alpha_n = -1$, and hence $Y(n) = Y(n-1) + w(n)$ for any $n \in \mathbb{Z}$.

REFERENCES

1. Beutler, F. J. *Ann. Math. Stat.* **34**, 424–38 (1963).
2. Doob, J. L. *Stochastic Processes.* Wiley, New-York (1953).
3. Castro, G. and Girardin, V. *Stat. Probab. Letters* **59**, 37–52 (2002).
4. Castro, G. and Girardin, V. *Stat. and Probab. Letters* **78**, 158–164 (2008).
5. Franses, P. H. *Periodicity and stochastic trends in economic time series.* Advanced texts in econometrics, Oxford Univ. Press, Oxford (1996).
6. Gladyshev, E. *Sov. Math. Dokl.* **2**,385–88 (1961).
7. Helson, H. and Lowdenslager, D. *Acta Math.* **99**, 165-212 (1958).
8. Mandrekar, V. *Nagoya Math. J.* **33**, 7–19 (1968).
9. Mehr, C. and McFadden, J. *JRSS Ser. B,* **27**, 505–22 (1965).
10. Nematollahi, A. and Soltani, A. *Probab. Math. Stat.* **20**, 127–40 (2000).
11. Priestley, M. *Spectral Analysis and Time Series* Volume 2, Academic Press, London (1981).
12. Troutman, B. *Biometrika*, **66**, 219-228 (1979).