Annals of University of Craiova, Math. Comp. Sci. Ser.
Volume 30, 2003, Pages 106–115
ISSN: 1223-6934

A new approach to Markov processes of order 2

Ulrich Herkenrath

Abstract. We propose an approach to study Markov processes of order 2 which is based on their "natural" transition probability and differs from a recommendation of Doob how to transform Markov processes of order 2 to such of order 1, i.e. the usual ones. We extend the concept of uniform ergodicity from Markov processes of order 1 to such of higher order. This property makes them accessible for statistics. Making use of their natural transition probability sufficient conditions for their uniform ergodicity may be derived. We apply our method of analysis to two types of Markov processes of order 2, which arise in processing a given Markov process of order 1.

2000 Mathematics Subject Classification. Primary 60J05; Secondary 60J20.
Key words and phrases. Markov process of order 2, uniform ergodicity, non-linear time series, exponential smoothing, Markovian noise sequence, Markovian innovations.

1. Introduction

Markov processes may be considered as stochastic processes with a short memory or a very simple dependence structure. The subject of generalization of such processes towards a longer memory or a more complex dependence structure is a classical field of the Romanian probability school and some French mathematicians, as can be seen e.g. in the books of Iosifescu and Theodorescu (1969), Iosifescu and Grigorescu (1990).

The classical questions concerning such processes refer to their asymptotic stability and related to that to the validity of the important limit theorems of probability theory to make such processes accessible for statistics. In this context the property of (uniform) ergodicity is essential.

The reasons for treating processes with more complex dependence structures are on one hand theoretical considerations in handling dependencies within stochastic processes and on the other hand are given by applications.

Here we deal with a class of processes in discrete time in between the usual Markov processes and the "chains of infinite order" which were created by the Romanian mathematicians Onicescu and Mihoc (see Iosifescu and Theodorescu (1969)), namely Markov processes of order 2. Although they were defined in the literature long ago, they were scarcely treated in detail. Doob (1953, p. 89) remarks that they could be reduced to Markov processes of order 1, i.e. the usual ones, by a simple expansion of the state space and a transition to vector-valued variables. But this advice is not constructive in so far as for a given Markov process of order 2 the question remains, which conditions on its characteristics ensure e.g. its ergodicity.

In the following we discuss the concept of uniform ergodicity for Markov processes of order 2. We propose a different transformation than Doob to study these processes. An important class of examples of such processes is considered. The effiency of our approach is demonstrated by two theorems which give sufficient conditions for the

Received: 11 December 2002.

uniform ergodicity of two types of Markov processes of order 2. Our method can be extended to Markov processes of order higher than 2.

2. Definitions and Examples

Definition 2.1. *A stochastic process* $(Z_n, n \geq 1)$ *on the probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *with values in a measurable space* $(Z, \mathcal{Z})$ *is a Markov process of order p,* $p \in \mathbb{N}$*, if and only if*

$$\mathbb{P}(Z_{n+1} \in A | Z_1, \ldots, Z_{n-1}, Z_n) = \mathbb{P}(Z_{n+1} \in A | Z_{n-p+1}, \ldots, Z_{n-1}, Z_n) \text{ a.s.}$$

for all $A \in \mathcal{Z}, n \in \mathbb{N}$ *and* $n \geq p$.

In the following we restrict ourselves to Markov processes of order $p \leq 2$. Those of order $p = 1$ are simply called Markov processes and special cases of Markov processes of order 2.

As well as for Markov processes of order 1, in the case of a Polish state space $(Z, \mathcal{Z})$ endowed with the Borel-σ-algebra $\mathcal{Z}$, "regular" versions of the conditional probabilities can be chosen, i.e. $\mathbb{P}(Z_{n+1} \in A | Z_{n-1}, Z_n)$ can be expressed as $Q(Z_{n-1}, Z_n; A)$ by means of a transition probability Q. For a Markov process of order 2 we call that Q "its transition probability".

Often instead of the process there is given a transition probability Q, i.e. the process is defined by means of Q and a starting pair (z_1, z_2) by the Ionescu Tulcea theorem: For a given transition probability Q from $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$ to $(Z, \mathcal{Z})$ and each starting pair $(z_1, z_2) \in (Z \times Z)$ there exist a probability space $(\Omega, \mathcal{A}, \mathbb{P}_{z_1,z_2})$ and random variables $(Z_n, n \geq 1)$ with values in Z, such that

$\mathbb{P}_{z_1,z_2}(Z_1 = z_1, Z_2 = z_2) = 1$ and
$\mathbb{P}_{z_1,z_2}(Z_{n+1} \in A | Z_1 = z_1, Z_2 = z_2, \ldots, Z_{n-1} = z_{n-1}, Z_n = z_n) = Q(z_{n-1}, z_n; A)$ a.s.
for all $A \in \mathcal{Z}, n \geq 2$,

that is, $(Z_n, n \geq 1)$ is a Markov process of order 2 with transition probability Q.

In particular we consider the following class of examples of Markov processes of order 2:

Example 2.1. *Let* $(W, \mathcal{W}), (X, \mathcal{X})$ *be measurable spaces and u a measurable function*

$$u : (W \times X, \mathcal{W} \otimes \mathcal{X}) \to (W, \mathcal{W}),$$

for which moreover all mappings $u_w : X \to W$ *are injective,* $u_w(x) = u(w, x) \; \forall w \in W$, $x \in X$.

Let $(X_n, n \geq 1)$ *be a Markov process on* $(X, \mathcal{X})$ *with transition probability P and* $(W_n, n \geq 1)$ *a recursively generated sequence in terms of* $(X_n, n \geq 1)$ *and u:*

$$W_{n+1} = u(W_n, X_n), \; n \in \mathbb{N},$$

for a given $W_1 = w_1$.

(A) The function u is "given by nature" or induced by a "system". In the first case, $(W_n, n \geq 1)$ *is a (non-linear) time series with a Markov process as noise sequence. In the second case,* $(W_n, n \geq 1)$ *can be regarded as a sequence of states of a system for which the inputs* $(X_n, n \geq 1)$ *are a Markov process.*
(B) The function u is chosen by a "statistician", e.g. with the purpose to smoothen sequentially the Markovian sequence $(X_n, n \geq 1)$ *of observations or innovations. In this context the special function* $u(w, x) = (1 - \lambda)w + \lambda x, 0 < \lambda < 1$, *represents the well-known procedure of "exponential smoothing", see Brown (1963), Bowerman and O'Connell (1993) as general references for this procedure.*

By the Ionescu Tulcea theorem, the existence of the processes $(X_n, n \geq 1)$ and $(W_n, n \geq 1)$ is ensured, given P and u: For each $w_1 \in W, x_1 \in X$ there exist a probability space $(\Omega, \mathcal{A}, \mathbb{P}_{w_1,x_1}) = ((W \times X)^{\mathbb{N}}, (\mathcal{W} \otimes \mathcal{X})^{\mathbb{N}}, \mathbb{P}_{w_1,x_1})$ and two stochastic processes $(W_n, n \geq 1)$ and $(X_n, n \geq 1)$ with values in $(W, \mathcal{W})$, respectively in $(X, \mathcal{X})$, such that

$$\begin{aligned} \mathbb{P}_{w_1,x_1}(W_1 = w_1, X_1 = x_1) &= 1, \\ \mathbb{P}_{w_1,x_1}(W_{n+1} \in A | W_1 = w_1, X_1 = x_1, \ldots, W_n = w_n, X_n = x_n) &= \delta_{u(w_n,x_n)}(A), \\ \mathbb{P}_{w_1,x_1}(X_{n+1} \in B | W_1 = w_1, X_1 = x_1, \ldots, W_n = w_n, X_n = x_n, & \\ W_{n+1} = w_{n+1}) &= P(x_n, B) \end{aligned}$$

for all $A \in \mathcal{W}, B \in \mathcal{X}, n \in \mathbb{N}$, where δ_w denotes the probability measure concentrated at $\{w\}$.
Of course, $(X_n, n \geq 1)$ is a Markov process with transition probability P, while $(W_n, n \geq 1)$ is a Markov process of order 2 since

$$\begin{aligned} & \mathbb{P}_{w_1,x_1}(W_{n+1} \in A | W_1 = w_1, \ldots, W_{n-1} = w_{n-1}, W_n = w_n) \\ = \; & \mathbb{P}_{w_1,x_1}(W_{n+1} \in A | W_1 = w_1, \ldots, W_{n-1} = w_{n-1}, X_{n-1} = u^{-1}_{w_{n-1}}(w_n), W_n = w_n) \\ = \; & P(u^{-1}_{w_{n-1}}(w_n), u^{-1}_{w_n}(A)) \; = \; \mathbb{P}_{w_1,x_1}(W_{n+1} \in A | W_{n-1} = w_{n-1}, W_n = w_n) \end{aligned}$$

for all $A \in \mathcal{W}, n \geq 2$. This conclusion is valid by the injectivity of the mappings $u_w(\cdot), w \in W$. Note that $\mathbb{P}_{w_1,x_1}$ can be written as $\mathbb{P}_{w_1,w_2}$ with $w_2 = u(w_1, x_1)$.

In the case when W is a Borel-set of a Polish space, the Markov process $(W_n, n \geq 1)$ of order 2 has a transition probability Q given by

$$Q(w, w'; A) = P(u^{-1}_w(w'), u^{-1}_{w'}(A))$$

for $(w, w') \in (Z \times Z), A \in \mathcal{W}$.

As for the analysis of Markov processes of order 2 in general, we come back to Doob's advice to reduce such processes to Markov processes of order 1. Doob and numerous other authors recommend the following transformation:
If $(Z_n, n \geq 1)$ is a Markov process of order 2 on the state space $(Z, \mathcal{Z})$ with the transition probability Q, then $(\tilde{Y}_n = (Z_n, Z_{n+1}), n \geq 1)$ is a Markov process of order 1 on the state space $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$, as can be easily checked.

The sequence $(\tilde{Y}_n, n \geq 1)$ with terms $\tilde{Y}_n = (Z_n, Z_{n+1})$ reveals a remarkable redundance, as each Z_n appears twice for $n \geq 2$. Therefore the question arises whether the above transformation is really apted to obtain results for $(Z_n, n \geq 1)$, e.g. its ergodicity.

Tong (1995, Appendix 1) deals with non-linear autoregressive time series which are Markov processes of order p. He studies them by making use of Doob's transformation.

Here we propose and deal with a different transformation of the Markov process $(Z_n, n \geq 1)$ of order 2 to one of order 1. We simply omit the variables $\tilde{Y}_{2n}, n \geq 1$, from the above sequence thus removing its redundance. Consequently we define the sequence of random variables $(Y_n, n \geq 1)$ as

$$Y_n = (Z_{2n-1}, Z_{2n}), \quad n \in \mathbb{N}.$$

One easily can prove:

Lemma 2.1. *Let $(Z_n, n \geq 1)$ be a Markov process of order 2 on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(Z, \mathcal{Z})$. Then $(Y_n = (Z_{2n-1}, Z_{2n}), n \geq 1)$ is a Markov process of order 1 with values in $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$.*
If moreover $(Z, \mathcal{Z})$ is a Polish space endowed with the Borel-σ-algebra and Q is the transition probability of $(Z_n, n \geq 1)$ from $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$ to $(Z, \mathcal{Z})$, then the Markov

process $(Y_n, n \geq 1)$ *has a transition probability* R *on* $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$ *given by*

$$R(z, z'; A_1 \times A_2) = \int_{A_1} Q(z, z'; dz'')Q(z', z''; A_2) \tag{1}$$

for $(z, z') \in (Z \times Z), A_1, A_2 \in \mathcal{Z}$.

The above method of transformation of a Markov process of higher order to one of order 1 is possible for orders $p > 2$ as well and the Lemma is valid with the obvious modifications.

For a Markov process $(Z_n, n \geq 1)$ of order 2 its transition probability Q, in case it exists, is a natural characteristic. Therefore the study of such a process may be based on Q.

If one wants to apply the rich theory of Markov processes of order 1, that amounts to deal with the transition probability R. This R again is some kind of iterated Q's. Thus it is natural to ask whether "good properties" for Q are inherited by R.

We shall show in the next Sections that this method leads to fruitful results.

3. Uniform ergodicity

With regard to the literature, see e.g. Iosifescu and Theodorescu (1969), Iosifescu and Grigorescu (1990), we define uniform ergodicity for Markov processes of order 2 as follows:

The Markov process $(Z_n, n \geq 1)$ of order 2 with transition probability Q is transformed to the Markov process $(Y_n = (Z_{2n-1}, Z_{2n}), n \geq 1)$ of order 1 with transition probability R. Uniform ergodicity of $(Z_n, n \geq 1)$ is defined by the well-known uniform ergodicity of the Markov process $(Y_n, n \geq 1)$, see the above references and Meyn and Tweedie (1996) for the definition of uniform ergodicity of a Markov process of order 1. This device is applicable to Markov processes of order $p > 2$ as well. For Markov processes $(Z_n, n \geq 1)$ of order 1 no transformation is necessary.

Definition 3.1. *Let* $(Z_n, n \geq 1)$ *be a Markov process of order 2 on the probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *with values in* $(Z, \mathcal{Z})$ *and with transition probability* Q *and let* $(Y_n = (Z_{2n-1}, Z_{2n}), n \geq 1)$ *be the associated Markov process of order 1 with transition probability* R *defined by (1).*
The process $(Z_n, n \geq 1)$ *is called uniformly ergodic if and only if there exists a probability measure* π *on* $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$ *such that*

$$\lim_{n \to \infty} \sup_{(z,z') \in (Z \times Z)} \| R^n(z, z'; \cdot) - \pi(\cdot) \| = 0,$$

where R^n *denotes the* n*-step transition probability associated with* R *and* $\| \cdot \|$ *the norm of total variation.*

With respect to the unique invariant probability measure π associated to R we have the following result.

Lemma 3.1. *If* π *is the unique invariant probability measure on* $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$ *of a uniformly ergodic Markov process* $(Z_n, n \geq 1)$ *of order 2, then the two marginal measures of* π *are identical, i.e.*

$$\pi(A \times Z) = \pi(Z \times A) =: \rho(A), \ A \in \mathcal{Z}.$$

If the process $(Z_n, n \geq 1)$ *is defined on* $(\Omega, \mathcal{A}, \mathbb{P}_{z_1,z_2})$, *i.e. it starts from* (z_1, z_2), *then*

$$\lim_{n\to\infty} \sup_{A_1,A_2\in\mathcal{Z}} |\mathbb{P}_{z_1,z_2}(Z_{n+1}\in A_1, Z_{n+2}\in A_2) - \pi(A_1\times A_2)| = 0,$$
$$\lim_{n\to\infty} \sup_{A\in\mathcal{Z}} |\mathbb{P}_{z_1,z_2}(Z_n\in A) - \rho(A)| = 0.$$

Proof. According to Definition 3.1 and Lemma 2.1 we have

$$\begin{aligned}
\pi(A_1\times Z) &= \lim_{n\to\infty} R^{n-1}(z_1,z_2;A_1\times Z) = \lim_{n\to\infty} \mathbb{P}_{z_1,z_2}(Z_{2n-1}\in A_1)\\
&= \lim_{n\to\infty} \textstyle\int_Z R^{n-1}(z_2,z_3;A_1\times Z)Q(z_1,z_2;dz_3)\\
&= \lim_{n\to\infty} \textstyle\int_Z \mathbb{P}_{z_1,z_2}(Z_{2n}\in A_1|Z_3=z_3)\mathbb{P}_{z_1,z_2}(dz_3) = \lim_{n\to\infty} \mathbb{P}_{z_1,z_2}(Z_{2n}\in A_1)\\
&= \lim_{n\to\infty} R^{n-1}(z_1,z_2;Z\times A_1) = \pi(Z\times A_1) = \rho(A_1)
\end{aligned}$$

for all $A_1\in\mathcal{Z}, n\in\mathbb{N}$, which proves all statements. □

For a Markov process of order 2 the joint distribution of $r>2$ subsequent variables $(Z_{n+1},\dots,Z_{n+r}), n\in\mathbb{N}$, can be composed of the distribution of the first 2 variables Z_{n+1}, Z_{n+2} and Q:

$$\begin{aligned}
&\mathbb{P}_{z_1,z_2}(Z_{n+1}\in A_1,\dots,Z_{n+r}\in A_r)\\
&=\int_{A_1}\int_{A_2} \mathbb{P}_{z_1,z_2}(Z_{n+3}\in A_3,\dots,Z_{n+r}\in A_r|Z_{n+1}=z', Z_{n+2}=z'')\,\mathbb{P}_{z_1,z_2}(Z_{n+1}\in dz', Z_{n+2}\in dz'')\\
&=\int_{A_1}\int_{A_2}\int_{A_3}\dots\int_{A_r} Q(z',z'';dy_3)Q(z'',y_3;dy_4)\dots Q(y_{r-2},y_{r-1};dy_r)R^{n/2}(z_1,z_2;dz',dz'')
\end{aligned}$$

for $A_1,\dots,A_r\in\mathcal{Z}$. Thus for any $r\in\mathbb{N}$ the limit $\lim_{n\to\infty}\mathbb{P}_{z_1,z_2}(Z_{n+1}\in A_1,\dots,Z_{n+r}\in A_r)$ is determined by $\lim_{n\to\infty} R^n(z_1,z_2;\cdot)$.

Therefore one can state:

Corollary 3.1. *Uniform ergodicity of a Markov process of order 2 is equivalent to the property:*
For any $r\in\mathbb{N}$ there exists a probability measure $\mathbb{P}_r^\infty$ on $(Z^r,\mathcal{Z}^r)$, where Z^r is the r-fold Cartesian product of Z and $\mathcal{Z}^r$ the corresponding Borel-σ-algebra, such that

$$\lim_{n\to\infty}\mathbb{P}_{z_1,z_2}((Z_{n+1},\dots,Z_{n+r})\in A) = \mathbb{P}_r^\infty(A)$$

uniformly with respect to $z_1,z_2\in Z, A\in\mathcal{Z}^r$ and $r\in\mathbb{N}$.
The probabilities $\mathbb{P}_r^\infty$ are determined by the unique invariant probability measure π and Q:

$$\mathbb{P}_r^\infty(A_1\times\dots\times A_r) = \int_{A_1}\dots\int_{A_r}\pi(dy_1,dy_2)Q(y_1,y_2;dy_3)\dots Q(y_{r-2},y_{r-1};dy_r)$$

for $A_1,\dots,A_r\in\mathcal{Z}$.

The law of large numbers and the central limit theorem which are valid for the uniformly ergodic process $(Y_n=(Z_{2n-1},Z_{2n}), n\geq 1)$ are transferred to the process $(Z_n, n\geq 1)$ by setting

$$f(Z_{2n-1})+f(Z_{2n}) = \overline{f}(Y_n)$$

for appropriate functions f and $\overline{f}$ on Z and $(Z\times Z)$, respectively.

Again the extension of those concepts and results to Markov processes of order $p>2$ is obvious: π is a probability on $(Z^p,\mathcal{Z}^p)$ and "gives weight" to the first p sets $A_1,\dots,A_p$ out of $A_1,\dots,A_r$. In case of order 1 nothing changes to the well-known results, see e.g. Iosifescu and Grigorescu (1990, p. 42).

To prove the uniform ergodicity of a Markov process of order 2 one asks for sufficient conditions for this property in case of a Markov process $(Z_n, n\geq 1)$ of order 1 on $(Z,\mathcal{Z})$ with transition probability P.

We mention:

Condition (MIX): *P is called mixing if and only if there exists $n \in \mathbb{N}$ such that the ergodicity coefficient $\alpha(P^n)$ of the iterate P^n satisfies*

$$\alpha(P^n) = 1 - \sup_{\substack{z',z'' \in Z \\ A \in \mathcal{Z}}} |P^n(z', A) - P^n(z'', A)| > 0.$$

A condition on P which is generally important for Markov processes is the so-called Doeblin-condition, see e.g. Doob (1953):

Condition (D): *P satisfies the Doeblin-condition (D) if and only if there exist a finite measure φ on $(Z, \mathcal{Z})$ with $\varphi(Z) > 0$ and $n \in \mathbb{N}, \varepsilon > 0, \eta > 0$ such that*

$$\forall A \in \mathcal{Z} : \ \varphi(A) \le \varepsilon \Rightarrow P^n(\cdot, A) \le 1 - \eta.$$

An obvious sufficient condition for (D) again is given, if φ "majorizes" all $P^n(z, \cdot)$ in the sense:

Condition (DM): *P satisfies the strenghtened Doeblin-condition (DM) if and only if there exist a finite measure φ on $(Z, \mathcal{Z})$ with $\varphi(Z) > 0$ and $n \in \mathbb{N}$ such that*

$$\forall z \in Z, A \in \mathcal{Z} : \ P^n(z, A) \le \varphi(A).$$

Following Doeblin's original work already Doob (1953, p. 192 ff) extensively studied the consequences of the validity of Condition (D), in particular the existence of a finite number of ergodic kernels $E_1, \dots, E_k$, each of which may contain a finite number $p_j, j = 1, \dots, k$, of subergodic kernels $F_j^1, \dots, F_j^{p_j}$. Moreover $E_i \cap E_j = \emptyset$ for $i \neq j$, $F_j^r \cap F_j^s = \emptyset$ for $r \neq s$ and $j = 1, \dots, k$, the E_j are stochastically closed sets, the process will move exponentially fast to one of the E_j's and cyclically move around the $F_j^r, r = 1, \dots, p_j$. For a comprehensive description, see Iosifescu and Grigorescu (1990, p. 72 f). In case "$k = 1$ and $p_1 = 1$" the corresponding Markov process is called "regular" by some authors, see e.g. Iosifescu and Grigorescu (1990, p. 72). Doob (1953, p. 221) denotes that as "Condition (D_0)":

Condition (D_0): *P satisfies the "regularity" condition (D_0) if and only if P satisfies (D) and in addition there is exactly one ergodic kernel $E_1 = E$ which does not contain subergodic kernels, i.e. $p_1 = 1$.*

As is well-known, (D_0) as well as (MIX) are equivalent to uniform ergodicity, see e.g. Meyn and Tweedie (1996, pp. 384, 391, 392).

Next the question arises whether "good properties" of Q are transfered to the transition probability R in order to ensure the uniform ergodicity of a corresponding Markov process of order 2.

Lemma 3.2. *Let Q be a transition probability from $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$ to $(Z, \mathcal{Z})$ and R the corresponding "iterated" transition probability on $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$:*

$$R(z, z'; A_1 \times A_2) = \int_{A_1} Q(z, z'; dz'') Q(z', z''; A_2)$$

for $(z, z') \in (Z \times Z), A_1, A_2 \in \mathcal{Z}$.

(a) If Q satisfies (DM), then R also does.

In the case when $Z \subset \mathbb{R}$ and Q has a "positive" Lebesgue density, we get the mixing property for R:

(b) For $Z \subset \mathbb{R}$ let Q have a Lebesgue density q, which is bounded away from zero on a finite interval $[\alpha, \beta] \subset Z$:

$$\exists \gamma > 0 \quad \forall z_1, z_2 \in Z, \, y \in [\alpha, \beta] : \quad 0 < \gamma \le q(z_1, z_2; y).$$

Then the transition probability R is mixing.

Proof. (a) Let Q satisfy (DM) with a finite measure φ. Then R satisfies (DM) with the product measure φ^2 on $(Z \times Z, \mathcal{Z} \otimes \mathcal{Z})$, since

$$R(z, z'; A_1 \times A_2) = \int_{A_1} Q(z, z'; dz'')Q(z', z''; A_2) \leq \varphi(A_1)\varphi(A_2) = \varphi^2(A_1 \times A_2),$$

$(z, z') \in (Z \times Z), A_1, A_2 \in \mathcal{Z}$. The above computation suffices, as $\mathcal{Z} \times \mathcal{Z}$ is a $\cap$-stable generator of $(\mathcal{Z} \otimes \mathcal{Z})$.

(b) If Q has the density q w.r.t. the Lebesgue measure l, by its very definition R has $l^2 = (l \otimes l)$-density

$$r(z_1, z_2; y_1, y_2) = q(z_1, z_2; y_1)q(z_2, y_1; y_2)$$

for $(z_1, z_2), (y_1, y_2) \in (Z \times Z)$. The assumption on q implies that

$$\forall (z_1, z_2) \in (Z \times Z), y_1, y_2 \in [\alpha, \beta] : \ r(z_1, z_2; y_1, y_2) \geq \gamma^2 > 0.$$

For $C \in (\mathcal{Z} \otimes \mathcal{Z})$ let us denote

$$C_I = C \cap [\alpha, \beta]^2, \ \ C_C = [\alpha, \beta]^2 \setminus C_I.$$

Now, if $l^2(C_C) > \frac{1}{2} l^2([\alpha, \beta]^2)$, then we have

$$R(z_1, z_2; C) \leq 1 - \gamma^2 l^2(C_C) \leq 1 - \tfrac{1}{2}\gamma^2 l^2([\alpha, \beta]^2),$$

which for $\delta = \frac{1}{2}\gamma^2 l^2([\alpha, \beta]^2)$ yields the estimate

$$\sup_{\substack{(z_1, z_2) \in (Z \times Z) \\ (z_1', z_2') \in (Z \times Z)}} |R(z_1, z_2; C) - R(z_1', z_2'; C)| \leq 1 - \delta.$$

If $l^2(C_C) \leq \frac{1}{2} l^2([\alpha, \beta]^2)$, then we have

$$\begin{array}{rcccl} R(z_1, z_2; C) & \geq & R(z_1, z_2; C_I) & = & R(z_1, z_2; [\alpha, \beta]^2 \setminus C_C) \\ & \geq & \gamma^2 l^2([\alpha, \beta]^2 \setminus C_C) & = & \gamma^2 \{l^2([\alpha, \beta]^2) - l^2(C_C)\} \\ & \geq & \frac{1}{2}\gamma^2 l^2([\alpha, \beta]^2) & = & \delta, \end{array}$$

which also yields

$$\sup_{\substack{(z_1, z_2) \in (Z \times Z) \\ (z_1', z_2') \in (Z \times Z)}} |R(z_1, z_2; C) - R(z_1', z_2'; C)| \leq 1 - \delta.$$

Therefore $\alpha(R) \geq \delta > 0$, i.e. R is mixing. ☐

In the following Section we show that our method of studying Markov processes $(Z_n, n \geq 1)$ of order 2, i.e. the use of their transition probability Q and their transformation to $(Y_n, n \geq 1)$, is useful. For two types of such processes out of the class of examples discussed we give sufficient conditions for their uniform ergodicity.

4. Applications

First, we consider a non-linear autoregressive time series $(W_n, n \geq 1)$ of order 1 with a Markov process $(X_n, n \geq 1)$ as noise sequence, i.e. a special case of Example 2.1(A). The time series $(W_n, n \geq 1)$ is generated by the recursion

$$W_{n+1} = f(W_n) + \sigma(W_n)X_n, \quad n \in \mathbb{N}, \tag{2}$$

with known functions f, σ for given $W_1 = w_1, X_1 = x_1, \ w_1, x_1 \in \mathbb{R}$.

Theorem 4.1. *Let the noise sequence $(X_n, n \geq 1)$ in (2) be given as a Markov process of order 1 on the state space $(\mathbb{R}, \mathcal{B})$ with transition probability P defined by*

$$P(X_n, \cdot) = N(e(X_n), v(X_n)),\ n \geq 1,$$

with bounded, measurable functions, i.e.

$$e : \mathbb{R} \to [a, A] \quad \text{and} \quad v : \mathbb{R} \to [b, B]$$

with $-\infty < a \leq A < \infty$, $0 < b \leq B < \infty$. Assume in addition that the functions f and σ in (2) are bounded and Borel-measurable:

$$f : \mathbb{R} \to [m, M] \quad \text{and} \quad \sigma : \mathbb{R} \to [s, S]$$

with $-\infty < m \leq M < \infty, 0 < s \leq S < \infty$.

Then the Markov process $(W_n, n \geq 1)$ of order 2 is uniformly ergodic.

Proof. As was shown in Section 2, $(W_n, n \geq 1)$ is a Markov process of order 2 on the state space $(\mathbb{R}, \mathcal{B})$ with respect to a probability measure $\mathbb{P}_{w_1,x_1}$. Its transition probability is given by

$$Q(z, z'; A) = P(\tfrac{z'-f(z)}{\sigma(z)}, \tfrac{1}{\sigma(z')}(A - f(z'))), \quad (z, z') \in \mathbb{R} \times \mathbb{R},\ A \in \mathcal{B},$$

where $(A - r)$ denotes $\{a' \in \mathbb{R} | a' = a - r, a \in A\}$ for $r \in \mathbb{R}$.
We compute the conditional distribution of W_{n+1} given W_{n-1}, W_n:

$$\begin{aligned}\mathbb{P}_{w_1,x_1}(W_{n+1} \in A | W_{n-1}, W_n) &= \mathbb{P}_{w_1,x_1}(X_n \in \tfrac{1}{\sigma(W_n)}(A - f(W_n)) | W_{n-1}, W_n) \\ &= \mathbb{P}_{w_1,x_1}(\sigma(W_n)X_n + f(W_n) \in A | W_{n-1}, W_n).\end{aligned}$$

Since W_{n-1}, W_n determine $X_{n-1} = \frac{1}{\sigma(W_{n-1})}(W_n - f(W_{n-1}))$ and

$$\begin{aligned}\mathbb{P}_{w_1,x_1}(X_n \in B | W_{n-1}, W_n) &= \mathbb{P}_{w_1,x_1}(X_n \in B | X_{n-1}) = P(X_{n-1}, B) \\ &= N(e(X_{n-1}), v(X_{n-1}))(B),\end{aligned}$$

it follows that

$$\mathbb{P}_{w_1,x_1}(W_{n+1} \in A | W_{n-1}, W_n) = N(\sigma(W_n)e(X_{n-1}) + f(W_n), \sigma(W_n)v(X_{n-1}))(A).$$

This means that the transition probability Q of the process $(W_n, n \geq 1)$ is given by

$$Q(w, w'; A) = N(\sigma(w')e(\tfrac{w'-f(w)}{\sigma(w)}) + f(w'); \sigma(w')v(\tfrac{w'-f(w)}{\sigma(w)}))(A).$$

Due to Lemma 3.2(b) the transition probability R which is constructed by means of Q is mixing:
Take as interval $[\alpha, \beta]$ in the Lemma e.g. $[-1, +1]$. By the boundedness assumptions on the functions e, v, f and σ and the shape of the normal density functions the density q of Q w.r.t. the Lebesgue measure satisfies $q(z, z'; y) \geq \gamma > 0$ for all $(z, z') \in \mathbb{R} \times \mathbb{R}$, $y \in [-1, +1]$.

Since R is mixing, $(Z_n, n \geq 1)$ is uniformly ergodic. □

Now, we turn to the procedure of sequential adaptive exponential smoothing of a Markovian observation sequence, i.e. Example 2.1(B).

Consider a sequence of observations $(X_n, n \geq 1)$ with values in $(X, \mathcal{X}) = ([a, b], \mathcal{B}_{[a,b]}), -\infty < a < b < \infty$, which is a Markov process with transition probability P. This sequence is sequentially and recursively smoothed by means of a smoothing function u with values in $(W, \mathcal{W}) = (X, \mathcal{X})$, i.e. if at time period n the smoothed value W_n is available and X_n is the actual observation, then the new or updated value for time period $(n + 1)$ is given by

$$W_{n+1} = u(W_n, X_n), \quad n \in \mathbb{N},$$

where $W_1 = w_1, X_1 = x_1, w_1 \in W, x_1 \in X$, are given as starting points. In the classical procedure of exponential smoothing, W_{n+1} is a weighted average of W_n and X_n with a constant weight $\lambda, 0 < \lambda < 1$, i.e.

$$W_{n+1} = (1-\lambda)W_n + \lambda X_n, \quad n \in \mathbb{N}.$$

In the procedure of adaptive exponential smoothing the weight λ may be a function of W_n, i.e.

$$W_{n+1} = (1-\lambda(W_n))W_n + \lambda(W_n)X_n, \quad n \in \mathbb{N},$$

see e.g. Bonsdorff (1989), Herkenrath (1994) for a further discussion in case that $(X_n, n \geq 1)$ are i.i.d. observations. In (adaptive) exponential smoothing a sequence $(X_n, n \geq 1)$ of i.i.d. observations induces as sequence of smoothed values $(W_n, n \geq 1)$ a Markov process of order 1, a Markovian observation sequence $(X_n, n \geq 1)$ induces as $(W_n, n \geq 1)$ a Markov process of order 2. Under mild assumptions one can ensure its uniform ergodicity:

Theorem 4.2. *(Adaptive exponential smoothing). Let* $(W, \mathcal{W}) = (X, \mathcal{X}) = ([a,b], \mathcal{B}_{[a,b]}), -\infty < a < b < \infty$, *and the transition probability* P *have a bounded and positive density* p *w.r.t. the Lebesgue measure, i.e.*

$$\forall x \in X, B \in \mathcal{X}: \; P(x,B) = \int_B p(x,y)\, l(dy) \quad \text{with } 0 < p(x,y) \leq C < \infty.$$

Moreover, the smoothing function $u: W \times X \to W$ *has the form*

$$u(w,x) = (1-\lambda(w))w + \lambda(w)x,$$

where $\lambda: W \to [p,1], 0 < p < 1$, *is continuous.*
Then the sequence of smoothed values $(W_n, n \geq 1)$ *is uniformly ergodic.*

Proof. Since all mappings $u_w(\cdot)$ are injective, $(W_n, n \geq 1)$ is a Markov process of order 2 with transition probability Q given by

$$\begin{aligned} Q(w,w';A) &= P(u_w^{-1}(w'); u_{w'}^{-1}(A)) \\ &= P\left(\tfrac{w'-(1-\lambda(w))w}{\lambda(w)}, \tfrac{1}{\lambda(w')}(A-(1-\lambda(w'))w')\right) \\ &\leq C\, l\left(\tfrac{1}{\lambda(w')}(A-(1-\lambda(w'))w')\right) \;=\; C\,\tfrac{1}{\lambda(w')}\, l(A) \\ &\leq \tfrac{C}{p}\, l(A) \;=:\; \varphi(A), \end{aligned}$$

$(w,w') \in (W \times W), A \in \mathcal{W}$. Here $\varphi(\cdot)$ denotes $\frac{C}{p} l(\cdot)$, whence Q satisfies (DM). To prove the "regularity" or Condition (D_0) we conclude:
For given $W_{2n-1} = w, W_{2n} = w'$ the range of $W_{2n+1} = u(W_{2n}, X_{2n})$ depends on $W_{2n-1} = w$ only implicitly via X_{2n-1}. Moreover, since $supp\ P(x,\cdot) = X = [a,b]$ for all $x \in X$ and because of the assumptions on u, the range of W_{2n+1} equals $u(W_{2n}, X) = \{w'' \in W | \exists x \in X : u(W_{2n}, x) = w''\}$ and $supp\ \mathbb{P}_{w_1,x_1}(W_{2n+1} \in \cdot)$ $= supp\ R^n(w_1, u(w_1,x_1); \cdot \times W) = [u(W_{2n}, a), u(W_{2n}, b)]$. By the same argument $supp\ \mathbb{P}_{w_1,x_1}(W_{2n+2} \in \cdot) = supp\ R^n(w_1, u(w_1,x_1); W \times \cdot) = [u(u(W_{2n},a),a), u(u(W_{2n},b),b)]$. Starting with $W_1 = w_1, W_2 = u(w_1,x_1)$ we get recursively

$$\begin{aligned} &supp\ R^n(w_1, u(w_1,x_1); \cdot) \\ &\quad = [u(w_1,x_1,a^{2n-1}), u(w_1,x_1,b^{2n-1})] \times [u(w_1,x_1,a^{2n}), u(w_1,x_1,b^{2n})], \end{aligned}$$

where e.g. $u(w_1,x_1,a^{2n-1})$ abbreviates the value $u(\ldots u(u(w_1,x_1),a)\ldots a)$ in which a appears $(2n-1)$ times.

Now, as in the proof of Lemma 2 in Herkenrath (1994, p. 680), one can show:

$$\lim_{n\to\infty} u(w_1,x_1,a^n) = a, \quad \lim_{n\to\infty} u(w_1,x_1,b^n) = b,$$

and as a consequence it holds

$$\lim_{n\to\infty} supp\ R^n(w_1, u(w_1, x_1); \cdot) = (W \times W).$$

Since $\varphi(\cdot) = C'l(\cdot)$, there cannot exist different respectively additional ergodic or subergodic kernels, because such kernels had to have positive mass under φ and therefore under the Lebesgue measure. □

Remark 4.1. *1. The assumptions on P in Theorems 4.1 and 4.2 in particular imply the uniform ergodicity of the Markov process $(X_n, n \geq 1)$.*
2. Since X is compact, the assumption of a continuous density p on $(X \times X)$ is sufficient for its boundedness.
3. The above assumptions on u cover the important case of a constant weight function λ, i.e. $\lambda(w) \equiv \lambda$, $0 < \lambda < 1$.

For a further discussion of more general suitable smoothing functions u and their properties we refer to Herkenrath (1994).

5. Concluding Remarks

We think that time series with a Markov process as noise sequence are quite suitable to model certain dependencies within stochastic processes, like e.g. cycles of high variances. As well, the statistical procedure of adaptive exponential smoothing of a Markovian observation sequence is of interest, in order to make applicable this procedure to more general sequences than i.i.d. ones. It seems that both subjects were not studied in the past, because that amounts to deal with Markov processes of order 2. The methods which are presented here for such a study are constructive and even open the treatment of Markov processes of order higher than 2.

References

[1] H. Bonsdorff, A comparison of the ordinary and a varying parameter exponential smoothing, *J. Appl. Prob.*, **26**, 784-792 (1989).
[2] B.L. Bowerman, R.T. O'Connell, *Forecasting and Time Series: An Applied Approach*, Belmont, CA , Duxbury Press, 1993.
[3] R.G. Brown, *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, NJ, Prentice-Hall, 1963.
[4] J.L. Doob, *Stochastic Processes*, New York, Wiley, 1953.
[5] U. Herkenrath, Generalized adaptive exponential smoothing procedures, *J. Appl. Prob.*, **31**, 673-690 (1994).
[6] M. Iosifescu, S. Grigorescu, *Dependence with Complete Connections and its Applications*, Cambridge University Press, 1990.
[7] M. Iosifescu, R. Theodorescu, *Random Processes and Learning*, Berlin, Springer, 1969.
[8] S.P. Meyn, R.L. Tweedie, *Markov Chains and Stochastic Stability*, London, Springer, 1996.
[9] H. Tong, *Non-Linear Time Series*, Oxford, University Press, 1995.

(Ulrich Herkenrath) Fakultät 4 – Institut für Mathematik,
Gerhard-Mercator-Universität Duisburg, Lotharstr. 65, D–47048, Duisburg, Germany,
Tel./Fax: +49 (203) 379-2690/-2689
E-mail address: herkenrath@math.uni-duisburg.de