ON ITERATIVE SOLUTION OF A LINEAR EQUATION WITH MARKOV OPERATOR

ON EXPONENTIAL MIXING FOR STATIONARY PROCESSES

Alexander Yu. Veretennikov

Preprint No 215
ON ITERATIVE SOLUTION OF A LINEAR EQUATION
WITH MARKOV OPERATOR
ON EXPONENTIAL MIXING
FOR STATIONARY PROCESSES

Alexander Yu. Veretennikov

Preprint No 215

UNIVERSITÄT KAIERSLAUTERN
Fachbereich Mathematik
Erwin–Schrödinger–Straße
6750 Kaiserslautern

Januar 1992
On iterative solution of a linear equation with Markov operator

Veretennikov, A.Yu.
Institute for Information Transmission Sci.,
Ermolovoy 19, GSP 101447, Moscow, Russia *

January 28, 1992

Abstract

The convergence of an iterative procedure for approximate solution of an equation \( \nu A = \mu \) is studied where \( A \) is a Markov operator, \( \nu \) and \( \mu \) are probability measures. The iterative procedure is based on Bayes formula. The convergence is studied with the help of the coupling method.

1 Introduction

Let \( A = A(x, dy) \) be a transition probability measure from some measurable space \( (E, \mathcal{E}) \) to another measurable space \( (E', \mathcal{E}') \). Assume that \( A(x, \cdot) \) is a probability measure for any \( x \in E \) and \( \int A(x, dy)I(y \in \Gamma) \) is measurable for any \( \Gamma \in \mathcal{E}' \). Consider the equation

\[
\nu A = \mu,
\]

where \( \mu \) is a given element of the space \( V(E') \) of all probability measures on \( E' \), \( \nu \) is unknown element of \( V(E) \) and

\[
(\nu A)(\Gamma) = \int \int \nu(dx)A(x, dy)I(y \in \Gamma).
\]

*the author is grateful to Department of Mathematics of Kaiserslautern University which partially supported this investigation
It is an empirical observation that the following procedure based on Bayes formula converges to the exact solution \( \nu \) (proved earlier for finite \( E \) and \( E' \)):

**Procedure**
1) Let \( \nu_0 \in V(E) \).
2) Let \( \mu_0 = \nu_0 A \).
3) Let

\[
A_0(dx, \Gamma) = \int \nu_0(dx)A(x, dy)f(y \in \Gamma), \quad \Gamma \in \mathcal{E}',
\]

\[
\hat{A}_0(dx | y) = \frac{A_0(dx, dy)}{\mu_0(dy)}.
\]

4) Let \( \nu_1 = \mu \hat{A}_0 \).

5) By induction, let

\[
\mu_{n+1} = \nu_{n+1} A;
\]

\[
A_{n+1}(dx, dy) = \nu_{n+1}(dx)A(x, dy),
\]

\[
\hat{A}_{n+1}(dx, dy) = \frac{A_{n+1}(dx, dy)}{\mu_{n+1}(dy)};
\]

\[
\nu_{n+2} = \mu \hat{A}_{n+1}, \ldots
\]

2 **Main result**

We'd like to investigate conditions under which \( \nu_n \to \nu \) in some sense. In fact, we'd obtain the convergence in variance.

Our main (and unique) assumption is the following:

**Assumption.** There exists such such a measure \( \lambda(dy) \in V(E') \) that

\[
0 < C^{-1} = \inf_{x,\gamma} \frac{A(x, dy)}{\mu(dy)} \leq \sup_{x,\gamma} \frac{A(x, dy)}{\mu(dy)} = C < \infty.
\] (2)
Theorem 1. Let a solution $\nu_*$ of (1) exists. Then under the Assumption for any "initial data" $v_0 \in V(E)$ the convergence in variance holds:

$$\text{var}(\nu_n - \nu_*) \to 0.$$  \hspace{1cm} (3)

Moreover, there exist such constants $C, \lambda > 0$ that

$$\text{var}(\nu_n - \nu_*) \leq C \exp(-\lambda n), \quad n \geq 0,$$  \hspace{1cm} (4)

uniformly in initial data $v_0$.

3 Proof

We use the coupling method or the method of unique probability space, see Nummelin (1984), Veretennikov (1991). When one use this method it is preferable to consider processes instead of measures. So we consider the sequences $(X_n)$ and $(Y_n)$ with

$$\mathcal{L}(X_n) = \nu_n, \quad \mathcal{L}(Y_n) = \mu_n, \quad n \geq 0$$

(here $\mathcal{L}$ is a distribution). Random values with given distributions may be constructed, for example, with the help of Kolmogorov’s theorem. We may consider all $X_n$ and $Y_n$ on the unique probability space, say, $(\Omega', F', P')$. At the same time we consider the stationary Markov sequences $(X_n^* = X^*)$ and $(Y_n^* = Y^*)$ with distributions

$$\mathcal{L}(X_n^*) = \nu_*, \quad \mathcal{L}(Y_n^*) = \mu, \quad n \geq 0$$

on another probability space $(\Omega'', F'', P'')$.

Now we shall try to construct such new sequence $(\tilde{X}_n)$ on some extension $(\tilde{\Omega}, F, P) = (\Omega', F', P') \times (\Omega'', F'', P'')$ with the following properties:

1. $\mathcal{L}(\tilde{X}_n) = \mathcal{L}(X_n)$,

2. $\tilde{X}_n = X_n$ with the maximal possible probability.

(2) is really the idea of the coupling method and in all known to the author cases it is based on the following simple "folklore" fact:

Lemma 1. Let $\xi$ and $\eta$ be two random vectors on some probability space. Then there exist such extension of the probability space and such random
value $\zeta$ on this extension that

$$\mathcal{L}(\zeta) = \mathcal{L}(\xi)$$

and

$$P(\zeta = \eta) = \int \min(1, d\mathcal{L}(\xi)/d\mathcal{L}(\eta)) d\mathcal{L}(\eta).$$

Under the Assumption the right-hand side $\mu$ in (1) has a bounded derivative $f$ with respect to $\lambda$:

$$f = d\mu/d\lambda.$$Moreover,

$$C^{-1} \leq f \leq C.$$ (5)

Note that for any $\mu_n$ this is also true: $d\mu_n \ll d\lambda$, and

$$C^{-1} \leq f_n := d\mu_n/d\lambda \leq C.$$ (6)

Thus, it follows from the Procedure (see (5)) that distributions of $X_n^*$ and $X_n$ for $n \geq 1$ are equivalent and, moreover,

$$0 < C^{-2} \leq d\mathcal{L}(X_n)/d\mathcal{L}(X_n^*) \leq C^2.$$ (7)

So we can apply our lemma to construct a new process $\tilde{X}_n$ we spoken earlier:

A) if $\tilde{X}_n = X_n^*$ then $\tilde{X}_{n+1} = \tilde{X}_{n+2} = \cdots = X^*$.

B) if $\tilde{X}_n \neq X_n^*$ then we go on this construction and apply our lemma at the next step $n + 1$.

By virtue of (7) we have

$$P(\tilde{X}_n = X_n^*) \geq C^{-1} > 0$$ (8)

on each step of our construction. Denote $L := \inf(n \geq 0 : \tilde{X}_n = X_n^*)$. Then (8) implies the inequality

$$P(L > n) \leq (1 - C^{-1})^n.$$ (9)

Thus,

$$\text{var}(\mathcal{L}(X_n) - \mathcal{L}(X^*)) \leq 2P(L > n) \leq 2(1 - C^{-1})^n.$$ (10)

The first obvious inequality in (10) is known as the main inequality of the coupling method. Theorem is proved.

The author is thanks Prof. Dr. H. von Weizsäcker who attracted his attention to the problem.
References


On exponential mixing for stationary processes

Veretennikov, A.Yu.
Institute for Information Transmission Sci.,
Ermolovoy 19, 101447 GSP, Moscow, Russia *

January 14, 1992

Abstract

The rate of mixing for stationary sequences is studied for some "ratio-mixing" coefficient under conditions of weak dependence of the "previous path".

1 Introduction


Earlier and in recent years results on the rate of mixing for various processes were obtained mainly for Gaussian stationary processes, see Kolmogorov and Rozanov (1960), Ibragimov (1961, 1962), and for Markov processes, see Doob (1953), Davydov (1973), Veretennikov (1988, 1991). Various examples and properties of mixing for stationary processes were established by Bradley (1980), Berbee and Bradley (1984) et al. In this paper we extend the approach from Veretennikov (1988, 1991) to nonmarkov and nongaussian case.

* this investigation was partially supported by Department of Mathematics of Kaiser-slautern University
2 Main results

We consider the stationary process \((X_n, -\infty < n < \infty)\) with values in the torus \([0,1]\) and regular transition probability densities

\[
p(x_{n+1}; x_n) = P(X_{n+1} \in dx_{n+1} \mid X_{\infty} = x_{\infty})/dx_{n+1}
\]

which do not depend on \(n\). Here again \(x_{\infty} = (x_n, x_{n-1}, \ldots), x_i \in [0, 1]\). We study here the following mixing coefficient:

\[
\psi(s) = ess(p) \sup_{B \in F_{\ge s}} \frac{|B(B|F_0) - 1|}{P(B)},
\]

where \(F_{\ge s} = \sigma(X_i, i \ge s)\). Let us state assumptions on the density \(p\):

**Assumption 1.** For any \(X_{\infty}, Z_{\infty}\),

\[
p(X_{\infty}, y) / p(Z_{\infty}, y) < C < \infty.
\]

**Assumption 2.** There exists such \(q \in (0, 1/2)\) that

\[
|p(x_{\infty}; y) - 1| \le q^{i+1}
\]

if

\[
x_0 = z_0, \ldots, z_{-i+1} = z_{-i+1} \ldots
\]

**Theorem 1.** If assumptions 1 + 2 are satisfied with some \(q \in (0, 1/2)\) then there exist such \(C, \lambda > 0\) that

\[
\psi(s) \le C \exp(-\lambda s).
\]

Let

\[
(0 <) 1 - c = \inf \int \min(p(x_0, y), p(z_0, y)) dy.
\]

Let us state another assumptions for the following theorem.

**Assumption 3.** There exists such \(q < c^{1/2} - c\) that

\[
|p(x_{\infty}; y) - 1| \le q^i
\]

if

\[
x_0 = z_0, \ldots, x_{-i+1} = z_{-i+1} \ldots
\]
Theorem 2. If assumptions 1 + 3 holds then there exist such $C, \lambda > 0$ that
\[ \psi(s) \leq C \exp(-\lambda s). \]

Comment. Assumption 3 is natural in the following sense. It means that the dependence of the density from the "far" coordinates is exponentially weak. Assumption 2 with a small $q$ is even stronger: it means, moreover, that dependence from any coordinates of the previous path is weak.

3 Proof of theorem 2

I. Let $(X_n)$ and $(Y_n)$ be two independent copies of the same stationary process. Denote by $(\Omega, F, P)$ the direct product of correspondent probability spaces. Fix some $x_{-\infty}^0 = (x_0, x_{-1}, ...), y_{-\infty}^0 = (y_0, y_{-1}, ...)$. To prove the statement of the theorem it is sufficient to show that for any $\Gamma \in B^1 \times B^2 \times ...$

\[ \left| \frac{P(X_n \in \Gamma \mid X_{-\infty}^0 = x_{-\infty}^0)}{P(Y_n \in \Gamma \mid Y_{-\infty}^0 = y_{-\infty}^0)} - 1 \right| \leq C \exp(-\lambda n) \]

with some $C, \lambda > 0$ which do not depend on $x_{-\infty}^0, y_{-\infty}^0, \Gamma$. Let $\Gamma = \Gamma_n \times \Gamma_{n+1} \times ... , \Gamma_i \in B^1$. Firstly let us show that for any $\Gamma_1 \in B^1$

\[ | P(X_n \in \Gamma \mid X_{-\infty}^0 = x_{-\infty}^0) - P(Y_n \in \Gamma \mid Y_{-\infty}^0 = y_{-\infty}^0) | \leq C \exp(-\lambda n). \]

II. Now assume $\Gamma = \Gamma_n \in B^1$. We use the coupling method although there is no Markov process here. Nevertheless, we may try to construct such a new process $(\tilde{Y}_n)$ on the extension of the probability space $(\Omega, F, P)$ that $(\tilde{Y}_n)$ is equal to $(Y_n)$ in distribution and $(\tilde{Y}_n)$ is equal to $Y_n$ pointwise with the maximal possible probability. (This is actually the main idea of the coupling method). It follows from assumptions that for any $X_{-\infty}^0$ and $Y_{-\infty}^0$

\[ \int_0^1 \min \left( p \left( X_{-\infty}^0 ; y \right), p \left( Y_{-\infty}^0 ; y \right) \right) dy \geq 1 - c > 0. \]

Hence, as in the coupling method we may construct such a new random value $\tilde{Y}_1$ that

\[ \mathcal{L} \left( \tilde{Y}_1, Y_{-\infty}^0 \right) = \mathcal{L} \left( Y_1, Y_{-\infty}^0 \right) \]

(here $\mathcal{L}$ is a law) and

3
\[ P(\tilde{Y}_1 = X_1|X_{-\infty}^0, Y_{-\infty}^0) \]
\[ = \int_0^1 \min(p(X_{-\infty}^0; y), p(Y_{-\infty}^0; y)) \, dy \geq 1 - c > 0. \]

(see Nummelin (1984)). Here \( P \) is already a new extension of "old" probability measure, but we do not change the notation \( P \). The value \( \tilde{Y}_1 \) is thus constructed. Now, \( \tilde{Y}_2, \ldots, \tilde{Y}_n, \ldots \) are built in the same way by induction with the properties

\[ \mathcal{L}(\tilde{Y}_{n+1}, Y_{-\infty}^n) = \mathcal{L}(Y_{n+1}, Y_{-\infty}^n) \]

and

\[ P(\tilde{Y}_{n+1} = X_{n+1}|X_{-\infty}^n, Y_{-\infty}^n) \]
\[ = \int_0^1 \min((X_{-\infty}^n; y), p(Y_{-\infty}^n; y)) \, dy \geq 1 - c > 0. \]

IV. It follows that \( P(\tilde{Y}_m = X_m) \geq c > 0. \) Moreover, Lipshitz condition on \( p \) implies that if \( \tilde{Y}_{n-1} = X_{n-1} \)

\[ P(\tilde{Y}_n = X_n | \tilde{Y}_{n-1} = X_{n-1}) \]
\[ = E\left(\int_0^1 \min(p(X_{-\infty}^n; y), p(Y_{-\infty}^n; y)) \, dy | \tilde{Y}_{n-1} = X_{n-1}\right) \geq 1 - q/(1 - q). \]

If \( \tilde{Y}_{n-k}^{n-1} = X_{n-k}^{n-1} \) then the estimate is better:

\[ P(\tilde{Y}_n = X_n | \tilde{Y}_{n-k}^{n-1} = X_{n-k}^{n-1}) \]
\[ = E\left(\int_0^1 \min(p(X_{-\infty}^n; y), p(Y_{-\infty}^n; y)) \, dy | \tilde{Y}_{n-k}^{n-1} = X_{n-k}^{n-1}\right) \]
\[ \geq 1 - q^k/(1 - q). \]

V. Consider the sequence \((\alpha_1, \ldots, \alpha_n): \alpha_k = I(X_k = \tilde{Y}_k), 1 \leq k \leq n. \)

By virtue of assumptions for any sequence \((\alpha_1, \ldots, \alpha_{n-1})\) the probability
\[ P(\alpha_1 = a_1, ..., \alpha_{n-1} = a_{n-1}) P(\alpha_n = 0 \mid \alpha_1 = a_1, ..., \alpha_{n-1} = a_{n-1}) = \\
\quad = P(\alpha_1 = a_1, ..., \alpha_{n-1} = a_{n-1}, \alpha_n = 0) \]

may be estimated as follows. Let the sequence \((a_1, ..., a_{n-1})\) has the form

\[ a_1 = ... = a_{k_1} = 1, a_{k_1+1} = ... = a_{k_1+m_1} = 0, \]
\[ a_{k_1+m_1+1} = ... = a_{k_1+m_1+k_2} = 1, \]
\[ a_{k_1+m_1+k_2+1} = ... = a_{k_1+m_1+k_2+m_2+1} = 0, \text{ etc.} \]

Let \(r\) denote the number of series of zeros between 1 and \(n\). Then by virtue of assumptions we have

\[ P(\alpha_{k_1+1} = 0 \mid \alpha_1 = ... = \alpha_{k_1} = 1) \leq q^{k_1}, \]
\[ P(\alpha_{k_1+m_1+k_2+1} = 0 \mid \alpha_{k_1+m_1+1} = ... = \alpha_{k_1+m_1+k_2} = 1) \leq q^{k_2}, \ldots, \]

and, moreover,

\[ P(\alpha_{k_1+2} = ... = \alpha_{k_1+m_1} = 0) \leq c^{m_1-1}, \]
\[ P(\alpha_{k_1+m_1+k_2+2} = ... = \alpha_{k_1+m_1+k_2+m_2} = 0) \leq c^{m_2-1}, \ldots \]

Denote \(k = \sum_{i=1}^{n-1} I(\alpha_i = 1)\). Then

\[ P(\alpha_1 = a_1, ..., \alpha_{n-1} = a_{n-1}, \alpha_n = 0) \leq q^k c^{(n-k-r)}. \]

So

\[ P(\alpha_n = 0) \leq \sum_{k=0}^{n-1} C_{n-1}^k q^k c^{(n-k-r)} = \\
\quad = c^{-r} (q + c)^{n-1}. \]

Since \(r \leq n/2\) then

\[ P(\alpha_n = 0) \leq (q + c)^{-1} (c^{-1/2} (q + c))^n \]

and assumption \(q + c < c^{1/2}\) implies the estimate

\[ P(\alpha_n = 0) \leq C \exp(-\lambda n) \quad (C, \lambda > 0). \quad (4) \]

5
Now, it follows from (4) that
\[
P(\prod_{i=0}^{n-1} (1-\alpha_i) = 0) \leq C \exp(-\lambda n)
\]
\[(5)\]
with some new $C, \lambda > 0$. Denote $\prod_{i=0}^{n/2} \alpha_i = \alpha(n-1)$.

VI. We have
\[
P(X_n \in \Gamma_n | X_{-\infty}^0 = x) - P(Y_n \in \Gamma_n | Y_{-\infty}^0 = y)
= E[I(X_n \in \Gamma_n) - I(Y_n \in \Gamma_n)]\left(1 - \exp(-\lambda n/2)\right)
= E[I(X_n \in \Gamma_n) - I(Y_n \in \Gamma_n)]\left(1 - \exp(-\lambda n/2)\right)
= E[I(X_n \in \Gamma_n) - I(Y_n \in \Gamma_n)]\left(1 - \exp(-\lambda n/2)\right)
= A_1 + A_2.
\]

VII. (5) and assumption 1 imply the bound
\[
A_1 \leq C \exp(-\lambda n) E[I(Y_n \in \Gamma) | Y_{-\infty}^0 = y; F_{[n/2]_n}] 
\leq C \exp(-\lambda n) E[I(Y_n \in \Gamma) | Y_{-\infty}^0 = y]
\]
\[(6)\]
(with new $C > 0$).

VIII.
\[
A_2 = E[I(\alpha(n-1) = 1) E[I(X_n \in \Gamma)] - I(Y_n \in \Gamma)] | Y_{-\infty}^0 = y; X_{-\infty}^0 = x; F_{n-1}]
= E[I(\alpha(n-1) = 1) E[I(X_n \in \Gamma)] - I(Y_n \in \Gamma)] | Y_{-\infty}^0 = y; X_{-\infty}^0 = x; F_{n-1}]
\]
It follows from assumptions that
\[
E[I(X_n \in \Gamma) - I(Y_n \in \Gamma)] | Y_{-\infty}^0 = y; X_{-\infty}^0 = x; F_{n-1}
\leq (1 + C \exp(-\lambda n/2)) E[I(Y_n \in \Gamma)] | Y_{-\infty}^0 = y; X_{-\infty}^0 = x; F_{n-1}]
\]
Thus,
\[
P(X_n \in \Gamma_n | X_{-\infty}^0 = x) \leq 1 + C \exp(-\lambda n).
\]
\[(7)\]
IX. We have for $\Gamma = \Gamma_n \times \Gamma_{n+1} \times \ldots$ 

\[
P(X_n^\infty \in \Gamma|X_{-\infty}^0 = x) \over P(Y_{-\infty}^0 \in \Gamma|Y_{-\infty}^0 = y) = \frac{P(X_n \in \Gamma_n|X_{-\infty}^0 = x)}{P(Y_n \in \Gamma_n|Y_{-\infty}^0 = y)} \times \frac{P(X_{n+1} \in \Gamma_{n+1}|X_{-\infty}^0 = x; X_n \in \Gamma_n)}{P(Y_{n+1} \in \Gamma_{n+1}|Y_{-\infty}^0 = y; Y_n \in \Gamma - n)} \ldots
\]

Similarly to (7) we obtain estimates

\[
P(X_{n+1} \in \Gamma_{n+1}|X_{-\infty}^0 = x; X_n \in \Gamma_n) \over P(Y_{n+1} \in \Gamma_{n+1}|Y_{-\infty}^0 = y; Y_n \in \Gamma - n) \leq 1 + C \exp(-\lambda(n + 1)),
\]

and so on. So it follows from (8) that

\[
P(X_n^\infty \in \Gamma|X_{-\infty}^0 = x) \over P(Y_n^\infty \in \Gamma|Y_{-\infty}^0 = y) \leq \prod_{k=n}^\infty (1 + C \exp(-\lambda k)) \leq 1 + C \exp(-\lambda n)
\]

with some new $C > 0$. Theorem 2 is proved.

The proof of theorem 1 is similar.

REFERENCES


