On asymptotics of Vaserstein's coupling for a Markov chain

O.A.Butkovsky* & A.Yu.Veretennikov†

Abstract

In this paper rate of convergence to stationary regime for a homogeneous Markov chain is linked with a spectral radius (not a spectral gap) of a certain "associated" Markov semigroup operator (not a "natural" one); the latter is a new version of Vaserstein's coupling.

1 Introduction

A general question about equilibrium distributions for homogeneous Markov process may be posed as follows. If there is a convergence to the (unique) stationary distribution, then how fast is convergence and how does it relate to spectral characteristics of the process generator? There is a useful technical tool to estimate convergence and mixing rate, which we call the Lemma about three random variables and which relates to such names as Markov, Dobrushin and Vaserstein. A new version of this lemma is proposed in this paper. With the help of a "Markov extension" of this new version, some insight – even though not a definite answer – to the questions above will be given in this paper. Notice that all densities below are considered with respect to the Lebesgue measure, however, this may be easily relaxed to any sigma-finite reference measure.

A global version of the coefficient $1-\kappa$ in (5) was introduced in a non-homogeneous situation by Dobrushin in [1] and, hence, is often called Dobrushin's ergdic coefficient. In a homogeneous case, Markov himself used a similar contraction coefficient, which nowadays may be found practically in all textbooks on Markov chains with a proof of Ergodic theorem; for original paper see [5] or its reprint in [6, pp 339–361]. Localized versions of this coefficient were used in several papers by the second author (see, e.g., [10]) under various names. Neither Markov nor Dobrushin used localized versions; nevertheless, to our mind it would be appropriate to use for them the name "local Markov–Dobrushin's coefficients". In this paper such local coefficients are used implicitly in the Theorem 3, where some local Markov–Dobrushin's condition may provide the inequality $r(\tilde{A}) < 1$ assumed in the Theorem for brevity.

This Markov–Dobrushin's coefficient became rather popular in *coupling* techniques; [10] is just one example. One of possible constructions based on this coefficient is due to Vaserstein and, hence, is often called Vaserstein's coupling in the literature, see [9], [7](to check). This construction allows variations and one of such variations is used in this paper. In the first papers on coupling method by Doeblin [2], [3], Markov–Dobrushin's idea was not used.

*Department of Probability, Mechanics and Mathematics Faculty, Moscow State University, Vorobjevy Gory, Moscow, 119991, Russia

†School of Mathematics, University of Leeds, LS2 9JT Leeds, UK & Institute of Inforation Transmission Problems, B.Karetny 19, Moscow, 127994, Russia

Lemma about three random variables (see below) was communicated to the second author by A.D.Solovyev in a private conversation. Who is the actual author of this useful statement is unknown to us; in any case, Solovyev himself never worked with coupling and most of papers on coupling uses some likewise technique. This Lemma itself also allows variations and we will show two versions of it.

Finally, let us notice that there are other useful conditions frequently used in coupling, such as "small set" condition. In its global form, it was used, in particular, in [4] as one of intermediate conditions under the name of "condition C". The use of Markov–Dobrushin's coefficient always provides better convergence rate estimates and serves a wider class of processes.

The paper is organized as follows. We state the Lemma and give a new version of it in the section 2 and then apply the second Lemma to a Markov chain in the section 3. The main results on convergence are formulated in the section 4; for the sake of simplicity, the Theorems 2 and 3 are presented for finite or countable state spaces. We provide a minimum of references due to the restriction on the volume of this presentation.

2 Two Lemmae of three random variables

Lemma 1 *Let ξ^1, ξ^2 be two random variables with values in R^d and densities p^1, p^2, correspondingly, on some probability spaces $(\Omega^1, \mathcal{F}^1, P^1)$, $(\Omega^2, \mathcal{F}^2, P^2)$. Let $\kappa := \int p^1 \wedge p^2 dx > 0$. Let ζ be uniformly distributed on $[0,1]$ independent on ξ^1, ξ^2, η be another independent random variable with a density $p_\eta(x) := c\left(p^1(x) - p^1 \wedge p^2(x)\right)$, where c is a normalizing constant and*

$$\xi^3 := 1\left(\frac{p^1(\xi^2)}{p^1 \vee p^2(\xi^2)} \geq \zeta\right)\xi^2 + 1\left(\frac{p^1(\xi^2)}{p^1 \vee p^2(\xi^2)} < \zeta\right)\eta.$$

Then $\xi^3 \sim \xi^1$ (same distribution) and $P(\xi^3 = \xi^2) \geq (=)\kappa$.

This Lemma practically suffices for evaluating mixing and convergence rate bounds. However, some inconvenience is that it leads to a *non*-homogeneous Markov chain. The next Lemma is free from this drawback.

Without loss of generality, we may and will assume that state space of our random variables (and a Markov process in the sequel) is $K = (0,1)^d$. Topologically any open cube is equivalent to R^d, so, in fact, this is not a real restriction.

Lemma 2 *Let ξ^1, ξ^2 be two random variables with values from K with densities p^1, p^2, respectively. Let*

$$\kappa := \int p^1 \wedge p^2 dx > 0.$$

Let $\omega = (\omega^1, \omega^2, \omega^3, \omega^4)$, with $\omega^1, \omega^2, \omega^3 \in K$ and $\omega^4 = 0, 1$, and let

$$p_{\eta^1}(x) = c_1\left(p^1(x) - p^1 \wedge p^2(x)\right), \qquad p_\zeta(x) = c_0\left(p^1 \wedge p^2(x)\right),$$

and

$$p_{\eta^2}(x) = c_2\left(p^2(x) - p^1 \wedge p^2(x)\right),$$

where c_1, c_2, c_0 are normalizing constants. Further, let $(\Omega, \mathcal{F})$ be the space of all possible ω with a Borel sigma-algebra on it. Probability measure on $(\Omega, \mathcal{F})$ is defined as follows by its density $p(\omega)$,

$$p(\omega^1, \omega^2, \omega^3, 0) = (1 - c_0^{-1})\, p_{\eta^1}(\omega^1) p_{\eta^2}(\omega^2) \equiv c_1^{-1} p_{\eta^1}(\omega^1) p_{\eta^2}(\omega^2),$$

and

$$p(\omega^1, \omega^2, \omega^3, 1) = c_0^{-1}\, p_\zeta(\omega^3).$$

Further, let

$$\eta^1(\omega^1) = \omega^1, \ \eta^2(\omega^2) = \omega^2, \ \zeta(\omega^3) = \omega^3, \ \eta^4(\omega^4) = 1(\omega^4 = 0).$$

Then (clearly) η^i *given* $\omega^4 = 0$ *has a density* p_{η^i}*. Finally, let*

$$X^1(\omega) = \begin{cases} \eta^1(\omega^1), & \omega^4 = 0, \\ \zeta(\omega^3), & \omega^4 = 1, \end{cases} \qquad X^2(\omega) = \begin{cases} \eta^2(\omega^2), & \omega^4 = 0, \\ \zeta(\omega^3), & \omega^4 = 1. \end{cases}$$

Then $p_{X^i}(x) = p^i(x)$ *and*

$$P(X^1 = X^2) \geq (=)\kappa.$$

Proof. It may be easily seen that $P(X^1 = X^2) > \kappa$ is impossible; however, we are happy to show just $P(X^1 = X^2) \geq \kappa$.

1. Let us check that $p(\omega)$ is a probability density. Notice that $\kappa = c_2^{-1}$ and

$$P(\Omega) = \int c_0^{-1} p^{\zeta}(\omega^3)\, d\omega^3 + (1 - c_0^{-1}) \int p^{\eta^1}(\omega^1) p^{\eta^2}(\omega^2)\, d\omega$$
$$= \kappa + 1 - \kappa = 1, \quad \text{as required.}$$

2. Let us verify that $p_{X^1}(x) = p^1(x)$ and $p_{X^2}(x) = p^2(x)$. Indeed,

$$Ef(X^1) = \int f(y^1) p_{\eta^1}(y^1)(1-\kappa)\, dy^1 + \int f(y^3) p_{\zeta}(y^3) \kappa\, dy^3 = \int f(y) p^1(y)\, dy.$$

Similarly,

$$Ef(X^2) = \int f(y^2) p_{\eta^2}(y^2)(1-\kappa)\, dy^2 + \int f(y^3) p_{\zeta}(y^3) \kappa\, dy^3 = \int f(y) p^2(y)\, dy.$$

3. Let us verify that $P(X^1 = X^2) \geq (=)\kappa$. Indeed,

$$P(X^1 = X^2) \geq P(\omega^4 = 1) = \kappa.$$

The Lemma 2 is proved.

3 Appying the Lemma 2 to Markov chains

We are going to construct a sequence of random variables η_n^1, η_n^2, ζ_n, χ_n, X_n^1, X_n^2 by induction, so that, in particular, (X_n^i) is a Markov chain on state space $\mathcal{S} \subset R^d$ with given transition density for $i = 1, 2$ and that (X_n^2) is a stationary version of a Markov chain (assuming that it exists) and, moreover, $Z_n := (X_n^1, X_n^2)$ is a homogeneous Markov chain such that $P(X_n^1 \neq X_n^2) \to 0$, $n \to \infty$. In principle, this construction may be applied without a stationary component, too (and even to non-homogeneous Markov chains; in the latter case, of course, we may not expect that (Z_n) may be homogeneous). Emphasize that the construction presented below is not unique and that all notations are new, not related to those in the Lemma 1 above.

1. $n = 0$. We take as p^1 the density of X_0^1 (if it is a delta-measure, we may start coupling construction from $n = 1$), p^2 the stationary density and use construction of the Lemma 2. Also, we have $\omega_0 = (\omega_0^1, \omega_0^2, \omega_0^3, \omega_0^4)$. All further coordinates will be similar, i.e., $\omega_{n+1}^j \in K := (x : \ x \in R^d, 0 < \min_i x^i \leq \max_i x^i < 1)$ for $j = 1, 2, 3$, $\omega_{n+1}^4 = 0, 1$, $\omega_{n+1} = \omega_{n+1}^1, \omega_{n+1}^2, \omega_{n+1}^3, \omega_{n+1}^4$ and $\omega = (\omega_0, \omega_1, \ldots)$, with the convention that if $\omega_k^4 = 1$, then $\omega_n^4 = 1$ for all $n \geq k$. Remind – see the Lemma 2 – that as a result of this step, we will have random variables η_0^1, η_0^2, ζ_0, X_0^1, X_0^2 and $\chi_0 := 1(\omega_0^4 = 0)$. If the measures that correspond to the densities p^1 and p^2 are non-singular, then we also have

$$P(X_0^1 \neq X_0^2) = \int p^1 \wedge p^2 dx > 0.$$

2. Case $n > 0$. Given η^1_n, η^2_n, ζ_n, χ_n, X^1_n, X^2_n, let us construct η^1_{n+1}, η^2_{n+1}, ζ_{n+1}, χ_{n+1}, X^1_{n+1}, X^2_{n+1} as follows. Let $\omega_{n+1} = (\omega^1_{n+1}, \omega^2_{n+1}, \omega^3_{n+1}, \omega^4_{n+1})$. Firstly, if $\omega^4_n = 1$, then we just use transition density $p(x,x')$ to construct $X^1_{n+1} = X^2_{n+1}$. On $\omega_n = 0$, as well as and on $\omega_n = 1$, let us define the densities

$$p_{\eta^1_{n+1}}(\eta^1_n, \eta^2_n, x) := c_1(\eta^1_n, \eta^2_n)\ \left(p^1(\eta^1_n, x) - p^1(\eta^1_n, x) \wedge p^2(\eta^2_n, x)\right),$$

$$p_{\zeta_{n+1}}(\eta^1_n, \eta^2_n, x) = c_0(\eta^1_n, \eta^2_n)\ \left(p^1(\eta^1_n, x) \wedge p^2(\eta^2_n, x)\right),$$

$$p_{\eta^2_{n+1}}(\eta^1_n, \eta^2_n, x) = c_2(\eta^1_n, \eta^2_n)\ \left(p^2(\eta^2_n, x) - p^1(\eta^1_n, x) \wedge p^2(\eta^2_n, x)\right),$$

where all $c_1(\eta^1_n, \eta^2_n), c_0(\eta^1_n, \eta^2_n), c_2(\eta^1_n, \eta^2_n)$ are normalizing constants.

As earlier, let $(\Omega_{\{n+1\}}, \mathcal{F}_{\{n+1\}})$ be the space of all possible $\omega_{n+1} \in (0,1)^4$ with a Borel sigma-algebra on it. Probability measure on $(\Omega_{n+1}, \mathcal{F}_{n+1})$ is defined as follows by the *conditional* density $p(\omega_{n+1} \mid \omega_n)$,

$$p(\omega^1_{n+1}, \omega^2_{n+1}, \omega^3_{n+1}, 0 \mid \omega_n = 0) = (1 - c_0(\eta^1_n, \eta^2_n)^{-1})\, p_{\eta^1_{n+1}}(\eta^1_n, \omega^1_{n+1}) p_{\eta^2_{n+1}}(\eta^2_n, \omega^2_{n+1}),$$

$$p(\omega^1_{n+1}, \omega^2_{n+1}, \omega^3_{n+1}, 1 \mid \omega_n = 0) = c_0(\eta^1_n, \eta^2_n)^{-1}\, p_{\zeta_{n+1}}(\eta^1_{n+1}, \eta^2_{n+1}, \omega^3),$$

$$p(\omega^1_{n+1}, \omega^2_{n+1}, \omega^3_{n+1}, 0 \mid \omega_n = 1) = 0, \quad p(\omega^1_{n+1}, \omega^2_{n+1}, \omega^3_{n+1}, 1 \mid \omega_n = 1) = p(\zeta_n, \omega^3_{n+1}).$$

The latter equation,

$$p(\omega^1_{n+1}, \omega^2_{n+1}, \omega^3_{n+1}, 1 \mid \omega_n = 1) = p(\zeta_n, \omega^3_{n+1}),$$

signifies that after hitting $\omega^4_n = 1$ the main component of the process η_n becomes ζ_n, which *now* (after hitting $\omega^4_n = 1$) follows the transition probabilities of the original Markov chain, $p(x,x')$ (and both X^1_n and X^2_n will follow this component); transition $(\omega^4_n = 1) \mapsto (\omega^4_{n+1} = 0)$ is impossible; the coordinates η^1_n and η^2_n become, a bit loosely speaking, uniformly distributed independent random variables not important for our aim. Further, let us define

$$\eta^i_{n+1}(\omega^i_{n+1}) = \omega^i_{n+1},\ i = 1,2,\ \zeta_{n+1}(\omega^3_{n+1}) = \omega^3_{n+1},\ \chi_{n+1} = 1(\omega^4_{n+1} = 0).$$

Clearly, η^i_{n+1} given $\omega^4_n = 0$ has a density $p_{\eta^i_{n+1}}$ for any $i = 1,2$; also, ζ_{n+1} given $\omega^4_n = 0$ has a density $p_{\zeta_{n+1}}$. Finally, let

$$X^1_{n+1}(\omega) = \left\{ \begin{array}{ll} \eta^1_{n+1}(\omega^1_{n+1}), & \omega^4_{n+1} = 0, \\ \zeta_{n+1}(\omega^3_{n+1}), & \omega^4_{n+1} = 1, \end{array} \right. \tag{1}$$

and

$$X^2_{n+1}(\omega) = \left\{ \begin{array}{ll} \eta^2_{n+1}(\omega^2_{n+1}), & \omega^4_{n+1} = 0, \\ \zeta_{n+1}(\omega^3_{n+1}), & \omega^4_{n+1} = 1. \end{array} \right. \tag{2}$$

The meaning of η^1_n and η^2_n is that they represent the processes X^1_n and X^2_n, respectively, under the condition that coupling was not successful until time n (inclusive). The latter is encoded by the information in the outcome ω, namely, $\gamma(\omega) := \inf(k : \omega^4_k = 1) > n$, or, equivalently, $\omega^4_n = 0$.

On the outcomes with $\gamma = n$ and $\gamma < n$ – i.e. with $\omega^4_n = 1$ – the random variables η^1_n and η^2_n may be constructed in the same way, however, they will not be linked any more with X^1 and X^2.

The meaning of ζ_n is different, this random variable represents both X^1_n and X^2_n under condition that $\gamma(\omega) = n$, i.e. at the moment of successful coupling. Denote

$$\kappa(\eta^1_n, \eta^2_n) := \int p^1(\eta^1_n, x) \wedge p^2(\eta^2_n, x) dx. \tag{3}$$

The processes $(\eta_n) := (\eta_n^1, \eta_n^2, \zeta_n, \chi_n)$ and (η_n^1, η_n^2) are homogeneous Markov chains.

3. Now we have to check that

$$p_{X_{n+1}^i | X_n^i}(x) = p(X_n^i, x) \qquad (4)$$

and

$$P(X_{n+1}^1 = X_{n+1}^2 \mid \mathcal{F}_n) \mid_{\omega_n^4 = 0} \geq \; (=) \kappa(\eta_n^1, \eta_n^2). \qquad (5)$$

Moreover, most importantly,

$$1(X_n^1 \neq X_n^2) \;=\; \prod_{k=0}^{n} 1(\omega_k^4 = 0) \equiv \prod_{k=0}^{n} \chi_k. \qquad (6)$$

Hence, let us introduce the following operator on functions from $L_\infty((0,1)^4)$

$$Ah(\eta_0) := \chi_0 \, E_{\eta_0} h(\eta_1). \qquad (7)$$

Then we have, by induction and due to the semigroup property, with $h(\eta) \equiv 1$,

$$P_{\eta_0}(X_n^1 \neq X_n^2) = E_{\eta_0} \prod_{k=0}^{n} \chi_k = A^n h(\eta_0). \qquad (8)$$

This is the key formula, which links probability of no coupling by time n with a certain positive semigroup for a homogeneous Markov chain (η_n). Emphasize that this process is more involved than (X_n^1) or even (X_n^1, X_n^2); in particular, notice the dimension. Of course, the "hope" to have some rate of convergence to zero is due to the fact that $|\chi_n| \leq 1$ and on some ω, in fact, $|\chi_n| < 1$ (actually, $\chi_n = 0$, because χ_n only takes two possible values, 0 or 1); whence, it sounds plausible that under a suitable recurrence, the right hand side in (8) should tend to zero with a certain rate; moreover, in "simple cases" this rate directly relates to a spectral radius of this semigroup. Emphasize that the operator A is built upon the process $\eta_n = (\eta_n^1, \eta_n^2, \zeta_n, \chi_n)$ rather than (X_n^1, X_n^2), even though it may be reduced to the triple $(\eta_n^1, \eta_n^2, \zeta_n)$ or even the couple (η_n^1, η_n^2), which are also homogeneous Markov chains

4. Let us check (4) given $\omega_n^4 = 0$. We have,

$$\begin{aligned} Ef(X_{n+1}^1 \mid \mathcal{F}_n) \mid_{\omega_x^4 n = 0} = & \int f(y^1) p_{\eta_{n+1}^1}(\eta_n^1, \eta_n^2, y^1)(1 - \kappa(\eta_n^1, \eta_n^2)) \, dy^1 \\ + \int f(y^3) p_{\zeta_{n+1}}(\eta_n^1, \eta_n^2, y^3) \kappa(\eta_n^1, \eta_n^2) \, dy^3 = & \int f(y^1) p(\eta_n^1, y^1) \, dy^1, \end{aligned}$$

as required. It remains to notice that given $\omega_n^4 = 1$, the equation (4) holds true by definition. This also shows that (X_n^1) is a Markov chain with a given transition density $p(x, x')$ and the same is true for (X_n^2). Moreover, (X_n^2) is a stationary version of the same Markov chain, since it starts from a stationary distribution.

4 Convergence results

Let $r(A)$ denote the spectral radius of the operator A in L_∞. Notice that

$$r(A) \leq \|A\| = \sup_x A1(x) \leq 1. \qquad (9)$$

Hence, under appropriate conditions there is a hope to show that $r(A) < 1$, which would imply some exponential estimate for convergence rate. Given two initial distributions μ_0 and ν_0, we construct random variables $\eta_0^1, \eta_0^2, \zeta_0, \chi_0$ according to the recipe of the Lemma 2.

Theorem 1 *The estimate holds true,*

$$\|\mu_n - \nu_n\|_{TV} \le EA^n 1(\eta_0^1, \eta_0^2, \zeta_0, \chi_0) \le \|A^n\|, \tag{10}$$

and, hence,

$$\limsup_n \frac{1}{n} \ln \|\mu_n - \nu_n\|_{TV} \le \ln r(A). \tag{11}$$

In particular, if $r(A) < 1$, then for any $\epsilon > 0$ and n large enough,

$$\|\mu_n - \nu_n\|_{TV} \le \exp((\ln r_A + \epsilon)n), \tag{12}$$

Proof of (10) is straightforward, due to the construction in section 3. The estimates (11) and (12) follow from (10). The Theorem 1 is proved.

Remind that in the rest of the paper state space $\mathcal{S}$ is finite or countable.

Theorem 2 *If state space $\mathcal{S}$ is finite and the process (X_n^1) irreducibile, then*

$$r(A) < 1 \quad \& \quad \lim_{n\to\infty} (EA^n 1(\eta_0^1, \eta_0^2, \zeta_0, \chi_0))^{1/n} = r(A). \tag{13}$$

In particular, convergence rate due to (12) is exponential.

Remark. Exponential convergence rate for an irreducible finite state Markov chain is well-known. What is apparently new here is the estimate of this convergence rate via the value of $r(A)$ by (12).

Proof. We only show the first inequality in (13). For finite state space irreducibility is equivalent to positivity of the chain (X_n^1). So, there exists $k \ge 1$ such that the k-step transition matrix of the process (X_n^1), say, $\mathcal{P}^k$ is positive. Then, due to the construction,

$$\sup_{\eta_0} E_{\eta_0} \prod_{j=1}^k \chi_j < 1.$$

Therefore, $\|A^k\| \le q$ with this value of k and with some constant $q < 1$, which implies that also

$$r(A) \le (\|A^k\|)^{1/k} \le q^{1/k} < 1.$$

The Theorem 2 is proved.

For countable $\mathcal{S}$, spectral radius $r(A)$ usually equals one. Nevertheless, convergence may take place and even possibly be exponential. Let $C \subset \mathcal{S}^2$ be a finite set, $\tau^C \equiv \tau_1^C := \inf(t > 0 : \, (\eta_t^1, \eta_t^2) \in C)$, $\tau_{n+1}^C := \inf(t > \tau_n^C : \, (\eta_t^1, \eta_t^2) \in C)$,

$$\tilde{A}h(\eta_0) := \chi_0 E_{\eta_0} h(\eta_{\tau^C}).$$

Theorem 3 *If there exist $\kappa > 0$, $\lambda > 0$ such that*

$$P(\tau_{[\kappa n]}^C > n) \le \exp(-\lambda n) \qquad \forall\, n > 0 \tag{14}$$

and if $r(\tilde{A}) < 1$, then

$$\limsup_n \frac{1}{n} \ln \|\mu_n - \nu_n\|_{TV} < 0. \tag{15}$$

Remark. The condition (14) may be provided by exponential recurrence of the original process X^1. The condition $r(\tilde{A}) < 1$ could be derived from irreducibility of the "process on C". We leave precise statements until a full presentation.

Proof sketch. We have for any $r(\tilde{A}) < p < 1$ and m large enough,

$$\sup_{\eta_0} E_{\eta_0} \left(\prod_{j=0}^{m} \chi_j \right) 1(\tau^C_{[\kappa m]} \leq m) \leq \sup_{\eta_0} \tilde{A}^{[\kappa m]} 1(\eta_0) \leq \|\tilde{A}^{[\kappa m]}\| \leq p^{[\kappa m]}. \tag{16}$$

Now we estimate,

$$\begin{aligned} \|\mu_n - \nu_n\|_{TV} &\leq \sup_{\eta_0} E_{\eta_0} \left(\prod_{k=0}^{n} \chi_k \right) \Big(1(\tau^C_{[\kappa n]} \leq n) + 1(\tau^C_{[\kappa n]} > n) \Big) \\ &\leq \|\tilde{A}^{[\kappa n]}\| + \sup_{\eta_0} P_{\eta_0}(\tau^C_{[\kappa n]} > n) \leq \|\tilde{A}^{[\kappa n]}\| + \exp(-\lambda n). \end{aligned}$$

For large n the desired assertion follows from (16). The Theorem 3 is proved.

Remark. The Lemma 1 was used in [10] and some other papers and the use of the Lemma 2 is quite similar. Notice that the use of the "conditional processes" (η_n) in the coupling constructions is more correct and precise. However, all recurrence estimates clamied for non-conditional versions under corresponding assumptions in terms of hitting time moments remain valid for (η_n).

References

[1] R.L. Dobrushin, Central limit theorem for non-stationary Markov chains, I and II, Theory Prob. Appl., 1, 1956, 65–80 & 329–83.

[2] W.Doeblin, Exposé de la théorie des chanes simples constantes de Markov á un nombre fini d'états. Mathématique de l'Union Interbalkanique, 2, 1938, 77–105.

[3] W.Doeblin, Éléments d'une théorie générale des chaînes simples constantes de Markoff, Ann. Sci. École Norm. Sup. 57(3), 1940, 61–111.

[4] J.L.Doob, Stochastic processes, New York, Wiley, London, Chapman & Hall, 1953.

[5] A.A. Markov, Extension of the law of large numbers to dependent quantities (in Russian), Izvestiia Fiz.-Matem. Obsch. Kazan Univ., (2nd Ser.), 15(1906), pp. 135–156.

[6] A.A. Markov, Izbrannie Trudy (Selected Works), AN SSSR, Leningrad, 1951.

[7] J.W.Pitman, On coupling of Markov chains, Probability Theory and Related Fields, 35(4), 315–322.

[8] A.D.Solovyev, personal communication.

[9] L.N.Vaserstein, Markov processes on countable space products describing large systems of automata, Probl. Inform. Transm. 5(3), 1969, 47-52.

[10] A.Yu.Veretennikov, On polynomial mixing and convergence rate for stochastic difference and differential equations, Probab. Theory Appl. 1999, 44(2), 317-321.