The Annals of Probability
1990, Vol. 18, No. 4, 1583–1601

α-CONGRUENCE FOR MARKOV PROCESSES[1]

BY KARI ELORANTA

Stanford University

We prove infinite-time extensions of invariance principles for certain random walks with essentially compact state spaces. The extensions are uniform-like in time since they use the $\bar{d}$-metric of the Bernoulli theory and imply the classical results. These are then generalized to couplings involving an isomorphism between the processes. In general a Doeblin-type condition is needed to hold for the walks but relaxation of this is indicated.

Introduction. In this paper we consider the stability/approximation properties of certain Markov processes in the light of a new notion, the α-congruence. The first results are essentially extensions of (finite-time) invariance principles for these processes to infinite-time versions. The extensions are uniform in time in the sense that the future separation of paths is not discounted. The results are all of the following type: If we have a random walk converging weakly to a diffusion process then under certain ergodicity conditions the walks in fact converge in the $\bar{d}$-metric of the Bernoulli theory. Intuitively this means that almost all of the paths of the two processes can be coupled together for all positive times except on a set of times of very small density. Under mild extra assumptions this extends to α-congruence, which in addition to this closeness also incorporates an isomorphism between the processes. A theorem of our type implies the invariance principle. Moreover the notion is applicable to various deterministic dynamical systems that exhibit chaotic behavior. Hence it can be viewed as a unifying concept in the studies of random and pseudorandom phenomena (see [11], where the concept of α-congruence is introduced, and [4] for application to certain deterministic systems).

1. Outline of the results. Invariance principles or functional limit theorems in their basic form describe the weak convergence of measures on certain function spaces. These spaces accommodate the paths of random walks and the limit measure is concentrated on the paths of a diffusion process. The topology is usually either that of uniform convergence or the Skorokhod topology. In this paper we choose the former and define our random variables to have continuous paths. From now on let $C_T = C([0, T], M)$, where M is separable space with metric d and $\{P^n\}$, $P \in \mathscr{P}(C_\infty)$, the space of probability measures defined on C_∞. Given $T < \infty$, let $\{P_T^n\}$ and P_T denote the restrictions of these measures to C_T.

The basic form of an invariance principle is that of Donsker's theorem.

Received April 1989.
[1]Research supported by Stanford University, Academy of Finland and the Emil Aaltonen Foundation.
AMS 1980 *subject classifications*. Primary 60F17; secondary 28D20.
Key words and phrases. Invariance principle, $\bar{d}$-metric, Bernoulli shift, Doeblin condition.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability*. www.jstor.org

1.1 THEOREM. *Let ξ_i be $\mathbb{R}$-valued i.i.d. random variables with mean 0 and variance $\sigma^2 \in (0, \infty)$. Let*

$$X_t^n = \frac{1}{\sigma\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}$$

have law P^n. Then $P_T^n \Rightarrow P_T$, the law of the standard Brownian motion. Or equivalently if μ^n is a coupling measure with marginals P_T^n and P_T then

$$\inf_{\mu^n} \inf \left\{ \varepsilon > 0 | \mu^n \left(\left\{ (\omega_1, \omega_2) | \sup_{t \in [0, T]} |X_t^n(\omega_1) - B_t(\omega_2)| \geq \varepsilon \right\} \right) \leq \varepsilon \right\} \to 0.$$

The theorem cannot be extended to uniform couplings for all times. Given any $a, \delta > 0$, we have for almost every ω,

$$\sup_{\substack{s \in [0, \delta] \\ t \in \mathbb{R}_+}} |B_{t+s}(\omega) - B_t(\omega)| > a.$$

Hence a random walk cannot possibly shadow the Brownian motion for all t. This holds in general for nondegenerate diffusions. To accommodate infinite-time intervals certain extensions have been considered, notably that of Stone (see [5]) which involves an exponential discount factor on the future separation of the coupled paths. This does not seem to be the right uniform-like infinite-time extension and in this work we establish an alternative.

1.2. In the theory of Bernoulli processes it is useful to consider a variation of the Prohorov metric. Let $\{X_t^i, P^i\}$, $i = 1, 2$, be processes with paths in C_T. Let

$$\bar{d}_T(P^1, P^2) = \inf_{\mu} \inf \left\{ \varepsilon > 0 | \mu \left(\left\{ (X^1, X^2) | \frac{1}{T} \int_0^T d(X_t^1, X_t^2)\, dt \geq \varepsilon \right\} \right) \leq \varepsilon \right\},$$

where μ has the marginals P_T^i (of course μ is defined on the underlying measure space $\Omega^1 \times \Omega^2$ but to simplify the notation we drop the ω^i's). This can be shown to be a metric on $\mathscr{P}(C_T)$ and we call it the $\bar{d}_T$-metric. Clearly $\bar{d}_T(P^1, P^2) \leq \rho(P^1, P^2)$, where the metric ρ has been defined analogously but using the sup-norm. If the processes are defined for all times let

$$\bar{d}(P^1, P^2) = \sup_{T \geq 0} \bar{d}_T(P^1, P^2) = \lim_{T \to \infty} \bar{d}_T(P^1, P^2).$$

Here uniform closeness of paths is replaced by closeness except on a set of times of small density. Once the processes under consideration have sufficient mixing properties this turns out to be the "correct" uniform-like infinite-time strengthening of the mode of convergence in invariance principles which also subsumes earlier infinite-time extensions.

Obviously, weak convergence implies $\bar{d}_T$-convergence but the converse is in general false. Here we will prove that for some processes weak convergence implies even $\bar{d}$-convergence. To show that this is a genuine strengthening of the mode of convergence, we have the following result.

1.3 THEOREM. *If $\{P_T^n\}$ is tight then $\bar{d}(P^n, P) \to 0$ implies that $P_T^n \Rightarrow P_T$.*

PROOF. Let $\{X_t^n\}$ and X_t denote the processes. Fix $T > \varepsilon > 0$ such that $\varepsilon T < 1$. Define the modulus of continuity as usual:

$$w_x(\delta) = \sup_{\substack{0 \le s < t \le T \\ |t-s| < \delta}} |x(t) - x(s)|.$$

From the tightness of $\{P_T^n\}$ one concludes (see, e.g., [3]) that $\exists\ \delta \in (0, \varepsilon T)$ and n_0 such that

$$P_T^n\left\{x \in C_T | w_x(\delta) \ge \frac{\varepsilon}{3}\right\} \le \frac{\varepsilon}{3}, \qquad \forall\ n \ge n_0.$$

Since a single probability measure is tight, this holds for P_T as well. Let $G, G^n \subset C_T$ be the collections of $\varepsilon/3$-regular paths in this sense.

If $\bar{d}(P^n, P) < \delta^2/9T^2$ for $n \ge n_1$ then $\bar{d}_T(P^n, P) < \delta^2/9T^2$ for these n. Hence $\exists\ \mu_T^n$ on $C_T \times C_T$ and $E^n \subset C_T \times C_T$ such that on it

$$\frac{1}{T}\int_0^T d(X_t^n, X_t)\,dt < \frac{\delta^2}{9T^2}$$

and $\mu_T^n(E^n) \ge 1 - \delta^2/9T^2$, $\forall\ n \ge n_1$.

Notice that

$$\mu_T^n(E^n \cap (G^n \times G)) \ge 1 - \frac{\delta}{3T} - \frac{2\varepsilon}{3} > 1 - \varepsilon.$$

Moreover on $E^n \cap (G^n \times G)$,

$$\sup_{t \in [0,T]} d(X_t^n(\omega_1), X_t(\omega_2)) < \frac{\delta}{3T} + \frac{2\varepsilon}{3} < \varepsilon$$

so $\rho(P_T^n, P_T) < \varepsilon\ \forall\ n \ge n_0 \vee n_1$. Since M is separable so is C_T and weak convergence follows. □

To illustrate the general line of reasoning in the extensions, we present the following simple special case.

1.4 THEOREM. *Let $\{(X_t^n, P^n)\}$ be a sequence of symmetric random walks with step size $1/n$ on $\mathbb{R}/\mathbb{Z}$. Then if (B_t, P) is the Brownian motion on $\mathbb{R}/\mathbb{Z}$ we have $\bar{d}(P^n, P) \to 0$ as $n \to \infty$.*

IDEA OF THE PROOF. We make use of the well-known results that characterize the convergence of a transition probability $P_x(X_t \in A)$ to the stationary measure. Under Doeblin or quasicompactness condition this convergence is exponential and independent of x. If this holds uniformly over the tail of the transition probability sequence then we can specify a time T_u such that $|P_x^n(X_{T_u} \in A_i) - P_y(B_{T_u} \in A_i)| \le \varepsilon m(A_i)$ for all sets A_i that are "nice" and for all x, y and $n \ge n_0$. The collection $\{A_i\}$ is taken to be a regular cover of the state space. But for any finite T_c the processes with starting points on the set

A_i at time T_u can be coupled ε-close uniformly for $[T_u, T_u + T_c]$ if the diameter of A_i's is small enough (this follows immediately from Donsker's theorem). Hence we get a good coupling independent of the distance of x and y in the $\bar{d}_T$-sense ($T = T_u + T_c$). These finite-time couplings can then be concatenated using a standard ergodic theoretic argument to yield the $\bar{d}$-convergence result.

To do this in detail, we first consider the exponential convergence results and state the exact condition for the uniformity in that. In Section 3 the extension argument is presented rigorously in the absolutely continuous and discrete cases. The two invariance principles that our results are mainly aimed to extend are those for random walks in a bounded domain in $\mathbb{R}^n$ by Stroock and Varadhan ([13]) and for geodesic random walks on compact manifolds by Jørgensen ([8]). The reason is that our ergodicity conditions are most naturally satisfied in these setups. Since the theorems involve rather general but still lengthy and technical assumptions we do not reproduce their statements here but refer the reader to the original sources. In Section 4 the $\bar{d}$-convergence results are coupled with the Bernoulli theory and a further strengthening in the form of α-congruence is shown. Hence this is yet another context in which this concept applies (for other treatments of the topic see [11] and [4]). Finally we exemplify the extension of the results to processes with noncompact state spaces by investigating the Ornstein–Uhlenbeck process.

2. Convergence to the invariant measure.

In this section we present theorems that characterize the convergence of a Markov transition probability as the time parameter approaches ∞. These are extensions of the results in [2] where we refer for most of the proofs.

2.1. Let M be a compact, connected metric space, m finite measure, $\not\equiv 0$ on $\mathscr{B}(M)$. Let $p(t, x, y)$ be a measurable Markov transition density function satisfying for some $t_0 > 0$:

(i) $p\colon (0, \infty) \times M \times M \to R$;
(ii) $\int_M p(t, x, y)\, dm(y) = 1$ for $t > 0$, $x \in M$;
(iii) $\alpha = \inf_{x, y \in M} p(t_0, x, y) \in (0, 1/m(M)]$;
(iv) $K = \sup_{x, y \in M} p(t_0, x, y) < \infty$.

Definition. Let μ be a bounded signed Borel measure on M. By μP_t, $t > 0$, we denote the measure on M with density

$$\left[\frac{d(\mu P_t)}{dm}\right](y) = \int_M p(t, x, y)\, d\mu(x).$$

$\{P_t\}$ is the transition semigroup ($\mu P_{t+s} = \mu P_t P_s$, $s, t > 0$). In the following we denote P_{nt_0} by P_n.

2.2 Lemma. *Let μ be a bounded signed Borel measure on M with $\mu(M) = 0$. Then for all $n \geq 0$,*

$$\|\mu P_n\| \leq (1 - \alpha m(M))^n \|\mu\|,$$

where $\|\ \|$ is the total variation norm on M.

Lemma 2.2 implies the existence of a unique probability measure λ s.t. $\lambda P_t = \lambda$, $t > 0$. It has measurable density ϕ with respect to m and $\phi \geq \alpha$.

2.3 LEMMA. *Let μ be as in Lemma 2.2. Then for $n \geq 1$, $t > (n+1)t_0$*

$$\left| \int_M p(t,x,y)\, d\mu(x) \right| \leq C(1 - \alpha m(M))^n \|\mu\| \phi(y).$$

PROOF. Since

$$\int_M p(t,x,y)\, d\mu(x) = \int_M p(t-(n+1)t_0, x, y)\, d[(\mu P_n)P_1](x)$$

and

$$\left| \left[\frac{d(\mu P_n)P_1}{d\lambda} \right](x) \right| = \frac{1}{\phi(x)} \left| \int_M p(t_0, z, x)\, d(\mu P_n)(z) \right|$$
$$\leq \frac{K}{\alpha}(1 - \alpha m(M))^n \|\mu\|$$

by the previous lemma we get the result. □

2.4 THEOREM. *There are constants C, β such that for $t \geq t_0$, $x, y \in M$,*

$$|p(t,x,y) - \phi(y)| \leq Ce^{-\beta t}\phi(y);$$

hence

$$\|P_x(X_t \in \cdot) - \lambda\| \leq Ce^{-\beta t}.$$

PROOF. Let $\mu = \delta_x - \lambda$. By Lemma 2.3 we get

$$|p(t,x,y) - \phi(y)| \leq C(1 - \alpha m(M))^{t/t_0 - 2} \|\mu\| \phi(y),$$

which implies the result since $\alpha > 0$. □

REMARKS. 1. The original (weak) Doeblin condition for a process is: There is a finite measure m and $\varepsilon, t > 0$ such that $P(t, x, A) < 1 - \varepsilon$ for all $x \in M$, $A \in \mathscr{B}(M)$, $m(A) < \varepsilon$. This follows from a slightly stronger condition: There exists a finite measure m and $c, t > 0$ such that $P_x(X_t \in A) \geq cm(A)$ $\forall$ $x \in M$, $A \in \mathscr{B}(M)$. The latter is also called a Doeblin condition in the literature and we follow this convention. If $P_x \ll m$ $\forall$ $x \in M$, $t \in \mathbb{R}_+$ this Doeblin condition is equivalent to (iii).

2. In the case when we have a nondegenerate diffusion process on M the conditions 2.1(i)–(iv) are satisfied under mild conditions on the drift field. In fact we can choose any $t_0 \in \mathbb{R}_+$. β is the absolute value of the largest nonzero eigenvalue of the generator of the process ([6]).

3. If a random walk on M has an absolutely continuous transition probability with respect to m the conditions 2.1(iii) and (iv) are satisfied for some $t_0 > 0$ under similar assumptions as in Case 2.

2.5. We proceed to modify the results above to the case of a finite-state aperiodic, irreducible Markov chain X_n. Let $M = \{x_i\}_{i=1}^N$ and let p_{xy}^n be a n-step Markov transition probability, $x, y \in M$, $n \in \mathbb{N}$. Let X_t denote the continuous-time extension of X_n defined naturally as

$$P_x(X_0 = x) = 1,$$

$$P_x\left(X_t = X_{nh} + \frac{t - nh}{h}\left(X_{(n+1)h} - X_{nh}\right)\right) = 1,$$

$$P_x\left(X_{(n+1)h} \in A | \mathscr{M}_{nh}\right) = \sum_{y \in A} p_{xy},$$

where h is the intertransition time and $\mathscr{M}_t$ is the σ-field generated by X_s, $s \in [0, t]$. Hence P_x is a measure on C_∞. We will assume that for some $n_0 \in N$:

(iii) $\alpha = \min_{x, y \in M} P_x(X_{n_0} = y) \in (0, 1/|M|]$;
(iv) $K = \max_{x, y \in M} P_x(X_{n_0} = y) < \infty$.

DEFINITION. Let μ be a bounded signed measure on M. Then for $t \in \mathbb{R}_+$ we define a measure on M by

$$\mu P_t(y) = \sum_{x \in M} \mu(x) P_x(X_t = y).$$

We can verify the semigroup property of P_t by the Chapman–Kolmogorov equation. Again denote $P_m = P_{mn_0}$.

Lemma 2.2 holds as it is with m being the counting measure. For the proof we argue now as follows: Since

$$\|\mu P_1\| = \sum_y \left|\sum_x \mu(x) P_x\left(X_{n_0} = y\right)\right|,$$

we have

$$\|\mu\| - \|\mu P_1\|$$
$$= \sum_y \left\{\sum_x [\mu^+(x) + \mu^-(x)] P_x\left(X_{n_0} = y\right) - \left|\sum_x \mu(x) P_x\left(X_{n_0} = y\right)\right|\right\}$$
$$= \sum_y \left\{2 \sum_x \mu^+(x) P_x\left(X_{n_0} = y\right)\right\} \geq 2\alpha \sum_y \|\mu^+\| = \alpha|M|\,\|\mu\|.$$

Hence $\|\mu P_1\| \leq (1 - \alpha|M|)\|\mu\|$ and the result follows. □

Lemma 2.3 holds with the obvious changes and $t_0 = n_0$. For the proof we argue now as follows: Since

$$\sum_x \mu(x) P_x(X_t = y) = \sum_x \mu P_{n+1}(x) P_x\left(X_{t-(n+1)n_0} = y\right)$$

and

$$\frac{1}{\phi(x)}|\mu P_{n+1}(x)| = \frac{1}{\phi(x)}\left|\sum_z \mu P_n(z) P_z(X_{n_0} = x)\right| \le \frac{K}{\alpha}\|\mu P_n\| \le \frac{K}{\alpha}(1 - \alpha|M|)^n \|\mu\|,$$

we conclude that

$$\left|\sum_x \mu(x) P_x(X_t = y)\right| \le \frac{K}{\alpha}(1 - \alpha|M|)^n \|\mu\| \sum_x \phi(x) P_x(X_{t-(n+1)n_0} = y),$$

which by the invariance of ϕ implies the result. □

Again this lemma immediately implies the analog of the convergence Theorem 2.4 with $p(t, x, y)$ replaced by $P_x(X_t = y)$ and $\lambda(A) = \sum_{x \in A}\phi(x)$.

2.6. In the context of a family of Markov processes $\{X_t^h, t \ge 0\}_{h>0}$ the convergence theorems extend as follows. Here h is essentially the intertransition time of the process which should be thought as a continuous-time extension of a Markov chain. Let us first consider the case where X_t^h satisfies the assumptions in 2.1(i) and (ii) and also for some $t_0, h_0 > 0$:

(iii) $\alpha = \inf_{x, y \in M,\ h \in (0, h_0)} p^h(t_0, x, y) > 0$;
(iv) $K = \sup_{x, y \in M,\ h \in (0, h_0)} p^h(t_0, x, y) < \infty$.

We call the critical condition (iii) the *uniform Doeblin condition*.

One can easily show that Lemmas 2.2 and 2.3 hold for the family under consideration and conclude the existence of invariant measures λ^h with densities ϕ^h, and we obtain the following theorem.

2.7 THEOREM. *There are positive constants C and β, independent of h such that*

$$|p^h(t, x, y) - \phi^h(y)| \le Ce^{-\beta t}\phi^h(y)$$

and

$$\|P_x^h - \lambda^h\| \le Ce^{-\beta t}$$

hold for all $t \ge t_0$, $h \in (0, h_0)$ and $x, y \in M$.

2.8. For the uniform extension of Section 2.5 to a family of discrete-valued Markov processes $\{X_t^h, t \ge 0\}_{h>0}$ with state spaces M^h we assume in addition to 2.1(i) and (ii) that

(iii) $\alpha(h) = \min_{x, y \in M^h} P_x(X_{t_0}^h = y) \ge \alpha/|M^h|$, $\alpha > 0$;
(iv) $K(h) = \max_{x, y \in M^h} P_x(X_{t_0}^h = y) \le \overline{K}/|M^h|$, $\overline{K} < \infty$.

Using these and the discrete version of Lemma 2.3, we can establish immediately Theorem 2.7 with $p^h(t,x,y)$ replaced by $P_x(X_t^h = y)$ and M by M^h.

REMARKS. 1. In the results above the main assumptions are the Doeblin condition and its uniform generalization. (i) and (ii) are guaranteed for any conservative process but since we need our processes to have richer ergodic properties the other conditions are required. For nondegenerate diffusion processes (iv) is almost vacuous due to well-known results on the decay of the heat kernel and we do not discuss it subsequently.

2. Verifying the Doeblin condition requires additional information on the processes. A necessary condition for the diffusion is that if L' denotes the adjoint of L with respect to m the problem $L'\phi = 0$, $\int \phi \, dm = 1$ must have a strictly positive solution. If the diffusion is a Brownian motion on a compact manifold, then $\phi =$ constant. Even more can be said: By our assumption on the nondegeneracy of the diffusion we can define an equivalent metric on M using σ^2. Then the generator will be in local coordinates $\frac{1}{2}\Delta + b$ and we can easily see that if $\nabla \cdot b = 0$ then again $\phi =$ constant. More general results can be obtained using maximum principles.

3. For the random walks the uniform Doeblin condition can in some cases be established using for example a monotonicity argument. To illustrate this, we consider a family of walks on $\mathbb{R}$ with i.i.d. increments $\overline{X}_i^n \sim f_n$, f_n being the symmetric unimodal density of the nth walk. For simplicity assume that $E\overline{X}_i^n = 0$, $E(\overline{X}_i^n)^2 = 1/n^2 \; \forall \; i$. Denote the density of $\overline{S}_{n^2}^n = \sum_{i=1}^{n^2} \overline{X}_i^n$ by $\overline{F}_n$. We know that $(*_{n^2} f_n)(x) \to \phi(x)$, the density of $N(0,1)$. By the unimodality of f_n, $\overline{F}_n(x)$, $x \geq 0$ is monotone decreasing. Choose n_0 such that $\overline{F}_n(1/2) \geq \delta > 0 \; \forall \; n \geq n_0$. Then $\overline{F}_n(x) \geq \delta \; \forall \; x \in [-1/2, 1/2)$, $n \geq n_0$. If S_m^n is the induced walk on $\mathbb{R}/\mathbb{Z}$ then clearly $F_n(x) \geq \delta \; \forall \; x$, $n \geq n_0$, where F_n is the transition density of the induced random walk.

3. The $\bar{d}$-convergence. Let (X_t^h, P^h) be a random walk on a compact metric state space (M^h, d). As before the parameter h is essentially the intertransition time. We assume the existence of an invariance principle of the form: $P_{x_n}^{h_n} \Rightarrow P_x$ when $h_n \to 0$ and $x_n \to x \in M$ (laws are for paths in C_T with initial distributions δ_{x_n} and δ_x). P is the unique law of a diffusion process X on (M, d). It is convenient to view M^h's to be embedded in M and the metric to have diameter 1. Both the invariance principles of [13] and that of [8] are of this type although their proofs differ considerably. In [8] an exponentially distributed intertransition time is also considered but since this is an easy variant we do not include it. One reason is that for technical simplicity we prefer the random walks to take values on C_T instead of D_T. Nothing in our approach however prevents one from extending them to, for example, birth and death processes.

We distinguish two cases according to whether the distribution of the random walk increments is absolutely continuous or singular with respect to the uniform probability measure m on M. The latter case is argued with a slightly more complicated argument and we will indicate the variation in the proof of Theorem 3.2.

We are now ready to state the first main result of this section.

3.1 THEOREM. *Let the diffusion process* (X_t, P_λ) *and the random walk family* $\{(X_t^h, P_{\lambda^h}^h)\}$ *with absolutely continuous transition probability measure be as above. Here* λ, λ^h *are the invariant measures with densities* ϕ, ϕ^h *with respect to* m. *Suppose that the conditions in Sections* 2.1 *and* 2.6 *also hold and that* $\phi^h \to \phi$ m-*a.s. Then* $\bar{d}(P_{\lambda^h}^h, P_\lambda) \to 0$.

PROOF. Since the argument is rather long we will break it into several steps.

Step 1: Let $p(t,x,y)$ and $p^h(t,x,y)$ be the densities of the transition probabilities with respect to m. Fix small $\varepsilon > 0$, then by Theorems 2.4 and 2.7 $\exists\ T_u, h_0 > 0$ such that

$$\max\left\{\frac{|p(t,x,y) - \phi(y)|}{\phi(y)}, \frac{|p^h(t,x,y) - \phi^h(y)|}{\phi^h(y)}\right\} < \frac{\varepsilon}{400},$$

$\forall\ t \geq T_u$, $0 < h \leq h_0$, $x, y \in M$. By Egorov's theorem we can find $E \subset M$, $m(E) < \varepsilon/200K$, where K is as in Section 2 for $t_0 = T_u$ and $h_1 \leq h_0$ such that

$$\frac{|\phi^h(y) - \phi(y)|}{\phi(y)} < \frac{\varepsilon}{400}, \qquad \forall\ y \in E^c.$$

Consequently, we can find couplings ${}_{\mathbf{x}}\nu_{T_u}^h$ on $C_{T_u}^2$ such that

$$\begin{aligned}
&{}_{\mathbf{x}}\nu_{T_u}^h\left(\left\{(X_t^h, X_t)|t \in [0, T_u],\ (X_0^h, X_0) = \mathbf{x},\ X_{T_u}^h \neq X_{T_u}\right\}\right)\\
&\quad \leq 2\left[\int_{E^c} + \int_E |p^h(T_u, x_1, y) - p(T_u, x_2, y)| m(dy)\right]\\
&\quad < 2\left[\frac{\varepsilon}{64} + 2K\frac{\varepsilon}{200K}\right] < \frac{\varepsilon}{16},
\end{aligned}$$

$\forall\ 0 < h \leq h_1$, $\mathbf{x} \in M \times M$.

Step 2: Choose $T_c > 0$ such that $T_u/(T_u + T_c) < \varepsilon/8$. Let

$$\begin{aligned}
&\rho_{T_c}\left(P_{y_1}^h, P_{y_2}\right)\\
&\quad = \inf_\eta \inf\left\{\varepsilon > 0 \middle| \eta\left(\left\{(X_t^h, X_t)|t \in [0, T_c],\ \sup_{s \in [0, T_c]} d(X_s^h, X_s) \geq \varepsilon\right\}\right) \leq \varepsilon\right\}
\end{aligned}$$

denote the Prohorov distance between the two processes on $[T_u, T]$ starting at $\mathbf{y} = (y_1, y_2)$. If $y_1 = y_2 = y$ then by the previous theorem and Egorov's theorem we can find $h_2 \in (0, h_1]$ and $F \subset M$, $m(F) < \varepsilon/200K$ such that for $0 < h \le h_2$,

$$\rho_{T_c}\big(P_y^h, P_y\big) < \frac{\varepsilon}{16}, \qquad y \in F^c.$$

Hence there exist $\varepsilon/16$-good couplings ${}_{\mathbf{y}}\eta^h_{T_c}$ on $C^2_{[T_u, T]}$ $\forall\ y \in F^c$, $0 < h \le h_2$. If $y_1 \ne y_2$ or $y_i \in F$ we let ${}_{\mathbf{y}}\eta^h_{T_c}$ denote an independent coupling. Clearly $\rho_{T_c}(P^h_{y_1}, P_{y_2}) \le 1$.

Step 3: Define a family of couplings $\{{}_{\mathbf{x}}\mu^h_T\}$ on C^2_T by

$${}_{\mathbf{x}}\mu^h_T(A) = \int_{\{\mathbf{z}(t) | \mathbf{z} \in A,\ t \in [0, T_u]\}} {}_{\mathbf{x}}\nu^h_{T_u}(\mathbf{dz}) \int_{\substack{\{\mathbf{w}(t) | \mathbf{w} \in A,\ t \in [T_u, T],\\ \mathbf{w}(T_u) = \mathbf{z}(T_u)\}}} {}_{\mathbf{z}(T_u)}\eta^h_{T_c}(\mathbf{dw}),$$

$\forall\ A \subset C^2_T$. The coupling property is trivial since if $A = A_1 \times C_T$ by the coupling properties of ${}_{\mathbf{z}}\eta^h_{T_c}, {}_{\mathbf{x}}\nu^h_{T_u}$ and the Chapman–Kolmogorov equation we see that

$$\begin{aligned}
{}_{\mathbf{x}}\mu^h_T(A) &= \int_{\{\mathbf{z}(t) | \mathbf{z} \in A,\ t \in [0, T]\}} {}_{\mathbf{x}}\nu^h_{T_u}(\mathbf{dz}) \\
&\qquad \times P^h_{z_1(T_u)}\big(X^h_t \in A_1,\ t \in [T_u, T],\ X^h_{T_u} = z_1(T_u)\big) \\
&= \int_{\{z_1(t) | z_1 \in A_1,\ t \in [0, T_u]\}} P^h_{x_1}\big(X^h_t \in dz_1\big) \\
&\qquad \times P^h_{z_1(T_u)}\big(X^h_t \in A_1,\ t \in [T_u, T],\ X^h_{T_u} = z_1(T_u)\big) \\
&= P^h_{x_1}\big(X^h_t \in A_1,\ t \in [0, T]\big).
\end{aligned}$$

The other marginal is identical. Let

$$\begin{aligned}
G_{\mathbf{x}} = \Big\{ &\big(X^h_t, X_t\big) | t \in [0, T],\ \big(X^h_0, X_0\big) = \mathbf{x}, \\
&X^h_{T_u} = X_{T_u} \in F^c,\ \sup_{s \in [T_u, T]} d\big(X^h_s, X_s\big) < \frac{\varepsilon}{8} \Big\}
\end{aligned}$$

and

$$G(T_u) = \Big\{ \big(X^h_t, X_t\big) | t \in [T_u, T],\ X^h_{T_u} = X_{T_u} \in F^c,\ \sup_{s \in [T_u, T]} d\big(X^h_s, X_s\big) < \frac{\varepsilon}{8} \Big\}.$$

Then

$$\begin{aligned}
{}_{\mathbf{x}}\mu^h_T(G_{\mathbf{x}}) &= \int_{\{(X^h_t, X_t) | t \in [0, T_u],\ X^h_{T_u} = X_{T_u}\}} {}_{\mathbf{x}}\nu^h_{T_u}(\mathbf{dz}) \int_{G(T_u)} {}_{\mathbf{z}(T_u)}\eta^h_{T_c}(\mathbf{dw}) \\
&\ge \left(1 - \frac{\varepsilon}{16}\right) \int_{\{(X^h_t, X_t) | t \in [0, T_u],\ X^h_{T_u} = X_{T_u} \in F^c\}} {}_{\mathbf{x}}\nu^h_{T_u}(\mathbf{dz}) > 1 - \frac{\varepsilon}{4}.
\end{aligned}$$

Since $m(F) < \varepsilon/200K$ so the last integral exceeds $1 - \varepsilon/16 - \varepsilon/200$. But clearly

$$d_{\text{ave},T}(X^h, X) \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T d(X_t^h, X_t)\,dt \le \frac{1}{T}\int_{T_u}^T d(X_t^h, X_t)\,dt + \frac{T_u}{T}$$
$$< \sup_{t\in[T_u,T]} d(X_t^h, X_t) + \frac{\varepsilon}{8} < \frac{\varepsilon}{4}$$

on $G_{\mathbf{x}}$. Therefore

$$\bar{d}_T\big(P_{x_1}^h, P_{x_2}\big) = \inf_{{}_{\mathbf{x}}\tilde{\mu}_T^h} \inf_{\delta}\{\delta > 0 | {}_{\mathbf{x}}\tilde{\mu}_T^h(\{(X_t^h, X_t) | d_{\text{ave},T}(X^h, X) \ge \delta\}) \le \delta\}$$
$$\le \frac{\varepsilon}{4}, \qquad \forall\, 0 < h \le h_2, \mathbf{x} \in M \times M.$$

Step 4: By induction we get a family of measures on C_{NT}^2, $N \in \mathbb{N}$:

$${}_{\mathbf{x}}\mu_{NT}^h(A) = \int_{\{\mathbf{z}(t) | \mathbf{z}\in A,\, t\in[0,(N-1)T]\}} {}_{\mathbf{x}}\mu_{(N-1)T}^h(\mathbf{dz})$$
$$\times \int_{\substack{\{\mathbf{w}(t) | w\in A,\, t\in[(N-1)T,\, NT],\\ \mathbf{w}((N-1)T) = \mathbf{z}((N-1)T)\}}} {}_{\mathbf{z}((N-1)T)}\mu_T^h(\mathbf{dw}).$$

Clearly these are again Markovian couplings. Furthermore denote the coupling where the initial distributions are the stationary ones simply by μ_{NT}^h and the limit measure by μ_∞^h. Let us now consider the dynamical system $(C_\infty \times C_\infty, \mathscr{B}, \theta_T^h \times \theta_T, \mu_\infty^h)$. As usual θ_t's are the shifts along paths. Moreover the shift $\theta_T^h \times \theta_T$ can be chosen to be ergodic with respect to μ_∞^h (e.g., [9]). But then by the ergodic theorem

$$\frac{1}{NT}\int_0^{NT} d(X_t^h, X_t)\,dt = \frac{1}{N}\sum_{i=0}^{N-1} d_{\text{ave},T}\big((X^h, X)\circ(\theta_{iT}^h, \theta_{iT})\big)$$
$$\to \int_{C_\infty \times C_\infty} d_{\text{ave},T}(\mathbf{x})\,\mu_T^h(\mathbf{dx}), \qquad \mu_\infty^h\text{-a.s.}$$

By Step 3 the last expression is bounded by $\varepsilon/2$. Therefore

$$\bar{d}\big(P_{\lambda^h}^h, P_\lambda\big) = \lim_{N\to\infty} \bar{d}_{NT}\big(P_{\lambda^h}^h, P_\lambda\big) < \varepsilon. \qquad \square$$

In the discrete case the Doeblin condition again implies the existence of an invariant measure but now the absence of densities changes the argument slightly. The essential difference between the proofs of Theorem 3.1 and the next one is that in the former the projections of couplings ${}_{\mathbf{x}}\mu_{T_u}^h$ at $t = T_u$ concentrate on the diagonal of $M \times M$, whereas in the following argument we only get an approximate match of paths at time T_u due to the discreteness of the state space. The notation is as in the previous proof.

3.2 THEOREM. *Let $\{(X_t^h, P_{\lambda^h}^h)\}$ be a family of discrete random walks with invariant measures $\{\lambda^h\}$ such that $\lambda^h \Rightarrow \lambda$, where λ is the invariant measure of the diffusion limit (X_t, P_λ). Then $\bar{d}(P_{\lambda^h}^h, P_\lambda) \to 0$.*

PROOF. *Step* 1: By the results in Sections 2.4 and 2.8 $\exists\ T_u,\ h_0 > 0$ such that

$$\max\left\{\frac{|P_{x^h}(X_t^h = y^h) - \phi^h(y^h)|}{\phi^h(y^h)}, \frac{|p(t,x,y) - \phi(y)|}{\phi(y)}\right\} < \frac{\varepsilon}{200},$$

$\forall\ t \ge T_u,\ h \in (0, h_0),\ x^h, y^h \in M^h,\ x, y \in M$. Then choose $T_c > 0$ such that $T_u/(T_u + T_c) < \varepsilon/8$.

Step 2: Define

$$\rho_{T_c}(y;\delta) = \sup_{z \in B_\delta(y) \cap M} \rho_{T_c}(P_z, P_y).$$

Then $\rho_{T_c}(y;\delta) \to 0$ as $\delta \downarrow 0\ \forall\ y \in M$. By Egorov's theorem $\exists\ \delta_0 > 0$ such that $\rho_{T_c}(y;\delta) < \varepsilon/32\ \forall\ \delta \le \delta_0,\ y \in E^c,\ m(E) < \varepsilon/200K$. Let

$$\rho_{T_c}(y;h,\delta) = \max_{y^h \in B_\delta(y) \cap M} \rho_{T_c}\left(P_{y^h}^h, P_y\right) \to 0 \quad \text{as } h, \delta \downarrow 0\ \forall\ y.$$

Also $\rho_{T_c}(y;h,\delta) \to \rho_{T_c}(y;\delta)$ as $h \downarrow 0$. But then

$$|\rho_{T_c}(y;h,\delta_0) - \rho_{T_c}(y;\delta_0)| < \frac{\varepsilon}{32}, \qquad \forall\ h \le h_{1,y} \in F^c,$$

where $h_1 \le h_0$ and $m(F) < \varepsilon/200K$. Therefore

$$\rho_{T_c}(y;h,\delta_0) < \frac{\varepsilon}{16}, \qquad \forall\ y \in (E \cup F)^c,\ h \in (0, h_1).$$

Let ${}_{(y^h,y)}\eta_{T_c}^h$ be a $\varepsilon/16$-good coupling on $[T_u, T]$ for $y \in (E \cup F)^c$, $y^h \in B_{\delta_0}(y)$. If $y \in E \cup F$ or $y^h \in B_{\delta_0}^c(y)$ let the coupling be independent.

Let $\{B_i\}_1^N$ be a disjoint cover of M with tiles that are contained in balls of radii δ_0. We further assume that the tiles are P_λ-continuity sets. Then we can find couplings ${}_{\mathbf{x}}\nu_{T_u}^h$ on $C_{T_u}^2$ with error

$$\begin{aligned}
&{}_{\mathbf{x}}\nu_{T_u}^h\left(\left\{(X_t^h, X_t) | t \in [0, T_u],\ (X_0^h, X_0) = \mathbf{x},\ X_{T_u}^h \in B_i,\ X_{T_u} \in B_i^c, \text{ some } i\right\}\right) \\
&\quad = 2\sum_i |P_{x_1}(X_{T_u}^h \in B_i) - P_{x_2}(X_{T_u} \in B_i)| \\
&\quad \le 2\sum_i \left\{|P_{x_1}(X_T^h \in B_i) - \lambda^h(B_i)| + |P_{x_2}(X_{T_u} \in B_i) - \lambda(B_i)| \right. \\
&\qquad\qquad \left. + |\lambda^h(B_i) - \lambda(B_i)|\right\} \\
&\quad \le 2\left\{\|P_{x_1}(X_{T_u}^h \in \cdot) - \lambda^h\| + \|P_{x_2}(X_{T_u} \in \cdot) - \lambda\|\right\} \\
&\qquad + \sum_i |\lambda^h(B_i) - \lambda(B_i)| \le \frac{\varepsilon}{50} + 2\sum_i |\lambda^h(B_i) - \lambda(B_i)| \le \frac{\varepsilon}{16}
\end{aligned}$$

by Step 1 for small enough h for all $\mathbf{x}$. Hence the coupling error is bounded by $\varepsilon/16$.

Step 3: Define ${}_{\mathbf{x}}\mu_T^h$ as in Step 3 of the previous theorem and let

$$G_{\mathbf{x}}(i;\tau) = \Big\{\big(X_t^h, X_t\big)|t \in [\tau, T], \big(X_0^h, X_0\big) = \mathbf{x},\ X_{T_u}^h \in B_i,$$
$$X_{T_u} \in B_i \cap (E \cup F)^c, \sup_{t\in[T_u,T]} d\big(X_t^h, X_t\big) < \frac{\varepsilon}{8}\Big\},$$

where $0 \le \tau \le T_u$. Then

$$\begin{aligned}
&{}_{\mathbf{x}}\mu_T^h\Big(\bigcup_i G_{\mathbf{x}}(i;0)\Big)\\
&\quad= \sum_i \int_{\substack{\{(X_t^h, X_t)|t\in[0,T_u],\,(X_0^h, X_0)=\mathbf{x},\\ X_{T_u}^h\in B_i,\, X_{T_u}\in B_i\cap(E\cup F)^c\}}} {}_{\mathbf{x}}\nu_{T_u}^h(\mathbf{dz})\int_{G_{\mathbf{x}}(i;T_u)} {}_{\mathbf{z}(T_u)}\eta_{T_c}^h(\mathbf{dw})\\
&\quad\ge \Big(1-\frac{\varepsilon}{16}\Big)\int_{\substack{\{(X_t^h, X_t)|t\in[0,T_u],\,(X_0^h, X_0)=\mathbf{x},\\ X_{T_u}^h\in B_i,\, X_{T_u}\in B_i\cap(E\cup F)^c\}}} {}_{\mathbf{x}}\nu_{T_u}^h(\mathbf{dw})\\
&\quad\ge 1-\frac{\varepsilon}{4}.
\end{aligned}$$

The rest of the proof is identical to that of Theorem 3.1. □

4. Bernoulliness and α-congruence.

4.1. In the absolutely continuous case the Doeblin condition implies that associated with our processes there are Markov operators of kernel type, i.e., there are contractions T on $L^1(\phi)$ such that

$$Tf(x) = \int_M p(t_0, x, y)\, f(y)\, dy \tag{4.1}$$

and $ET = TE = E$, where E is the expectation. Let θ be the corresponding left shift. It is mixing iff

$$\lim_{n\to\infty}\int_M (T^n f)(x) g(x)\phi(x)\,dx = \int_M f(x)\phi(x)\,dx \int_M g(x)\phi(x)\,dx,$$

$\forall\, f \in L^1(\phi)$, $g \in L^\infty$. Since characteristic functions of closed sets are dense in $L^1(\phi)$ we only need to observe that by the bounded convergence theorem

$$\lim_{n\to\infty}\int_M p(nt_0, x, y) 1_E(y)\, dy = \lambda(E)$$

to establish that θ is mixing. Therefore by the results in [10] the shift is isomorphic to a Bernoulli shift and consequently the continuous shifts on these processes are Bernoulli flows.

In the discrete case the Bernoulliness of the random walks is a standard result and can be found for example in [12].

4.2. Let (X_t^i, P^i), $i = 1, 2$, be two stationary random processes assuming values in the same metric space (M, d). Let Ω^i be the underlying abstract measure spaces and f^i the shifts on these spaces corresponding to the processes. If π^i is the zeroth coordinate projection from Ω^i to M then $X_t^i(\omega) = \pi^i(f_t^i(\omega))$ for all $\omega \in \Omega^i$, $t \geq 0$.

DEFINITION. $(\Omega^i, f^i, \pi^i, P^i)$, $i = 1, 2$, are *α-congruent* if there exists an isomorphism ι between (Ω^i, f^i, P^i) such that $d(X_t^1(\omega_1), X_t^2(\iota(\omega_1))) < \alpha$ except for a set of $\omega_1 \in \Omega^1$ of P^1-measure less than α.

We now establish the α-congruence extensions of the $\bar{d}$-convergence results. Let (Ω, f_t, P) and (Ω^h, f_t^h, P^h) be the abstract dynamical systems of the diffusion process and the random walk respectively. Let π and π^h denote the corresponding projections. Let $(\tilde{\Omega}^h, \tilde{f}_t^h, \tilde{P}^h)$ be an abstract process such that the shift $\tilde{f}_t^h$ is an infinite-entropy Bernoulli shift. By the previous section and the isomorphism theorem of Bernoulli shifts ([9]) we know that (Ω, f_t, P) is isomorphic (via some ι) to $(\Omega^h \times \tilde{\Omega}^h, f_t^h \times \tilde{f}_t^h, P^h \times P)$. The latter is realized via $\pi^*(x, \tilde{x}) = \pi^h(x)$. But then

$$P\big(\{\omega | d\big((\pi \circ f_t)(\omega), (\pi^h \circ f_t^h)(\iota(\omega))\big) \geq \varepsilon\}\big) \leq \varepsilon$$

for small enough h and some ι. This is because otherwise $\bar{d}(P, P^h) > \varepsilon$, which by the theorems in the previous section is impossible. Note that the $\bar{d}$-coupling involves already a measure preserving map between the infinite orbits and now we just choose the process $(\tilde{\Omega}, \tilde{f}_t, \tilde{P})$ so that the map extends into an invertible one. Hence we obtain the main result.

4.3 THEOREM. *The $\bar{d}$-convergence in Theorems* 3.1 *and* 3.2 *extends to α-congruence.*

REMARK. We present this extension mainly to illustrate the applicability of α-congruence to this class of dynamical systems as well. The abstract process corresponding to the shift $\tilde{f}_t$ can be interpreted as a perturbation of a viewer that observes the process X_t^h. For elaboration of this see [11].

5. Unbounded state space. The Doeblin condition of Section 2.1 is natural only for processes with compact state space. However the exponential convergence to an invariant measure can hold outside this class of processes. Instead of being uniformly ergodic the process is then required to be geometrically ergodic. We investigate this and derive the α-congruence result in the particular case of a diffusion on $\mathbb{R}$ which already exemplifies the critical parts of the argument.

5.1. Let us consider the Ornstein–Uhlenbeck process X_t^a on $\mathbb{R}$ with equal infinitesimal variance and drift coefficients (this is just for computational convenience). The process has transition density

$$p^a(t,x,y) = \frac{1}{\sqrt{\pi(1-e^{-2at})}} e^{-[(y-xe^{-at})^2]/(1-e^{-2at})}, \qquad a \in (0,1).$$

Note that $p^a(t,x,y)$ is a summability kernel as $t \downarrow 0$. Consequently, the associated semigroup $(S_t^a f)(x)$ converges pointwise to $f(x)$ on $\mathbb{R}$, i.e., S_t^a is Feller. One can easily see from the formula that the weak Doeblin condition fails. Hence as a Harris process the quasi-compactness fails as well. The process is φ-recurrent but its state space is not φ-uniform set hence the strong ergodic theorem by Orey (e.g., [7]) does not imply exponential convergence.

The process has the generator

$$L = \frac{a}{2}\frac{d^2}{dx^2} - ax\frac{d}{dx},$$

which has the spectrum $\{-an \mid n \in \mathbb{N}_0\}$. Therefore the transition density has the representation

$$p^a(t,x,y) = \frac{e^{-y^2}}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{1}{2^n n!} e^{-ant} H_n(x) H_n(y),$$

where H_n is the nth Hermite polynomial, $t \in \mathbb{R}_+$, $x, y \in \mathbb{R}$. By [1] $|H_n(x)| < \sqrt{2n!}\, 2^{n/2} e^{x^2/2}$ so the convergence of the representation follows easily and we also obtain

$$\left| p^a(t,x,y) - \frac{1}{\sqrt{\pi}} e^{-y^2} \right| < \frac{2}{\sqrt{\pi}} e^{(x^2-y^2)/2} e^{-at} \sum_{n=0}^{\infty} e^{-ant} \le C(a) e^{x^2/2} e^{-at}.$$

Hence we establish a weaker form of the exponential convergence result by restricting x to a compact set K. If we let $\lambda(dy) = (1/\sqrt{\pi}) e^{-y^2}\, dy$, then the result implies

$$\|P_x(X_t^a \in \cdot) - \lambda\| \le C(a,K) e^{-at}, \qquad \forall\, x \in K.$$

The norm is again total variation.

5.2. Our approximating random walks are the generalized Ehrenfest processes $\{X_t^{a,n}\}_{n \ge 1}$. Let the state space of the nth one be $M^n = \{i/n \mid i = -n^2, \dots, n^2\}$, the jumps at intervals of $1/n^2$ and the transition probability be

$$p_{ij}^{a,n} = \begin{cases} \dfrac{a}{2}\left(1 + \dfrac{i}{n^2}\right), & j = i-1, \\ 1 - a, & j = i, \\ \dfrac{a}{2}\left(1 - \dfrac{i}{n^2}\right), & j = i+1, \qquad i, j = -n^2, \dots, n^2. \end{cases}$$

By the usual extension we obtain a process with paths in C. For all $a \in (0,1)$ and all n, $\{X^{a,n}_{i/n^2} | i \in \mathbb{N}\}$ is a homogeneous, irreducible, aperiodic Markov chain. The limit $a \uparrow 1$ is the classical Ehrenfest chain which has period 2.

Clearly

$$n^2\left[\frac{a}{2n^2}\left(1-\frac{[xn]}{n^2}\right)+\frac{a}{2n^2}\left(1+\frac{[xn]}{n^2}\right)\right]=a,$$

$$n^2\left[\frac{a}{2n}\left(1-\frac{[xn]}{n^2}\right)-\frac{a}{2n}\left(1+\frac{[xn]}{n^2}\right)\right]\to -ax$$

uniformly on compact sets. By standard results (see, e.g., [5]) we get the invariance principle.

LEMMA. *$P^{a,n}_0 \Rightarrow P^a_0$ implies $P^{a,n}_T \Rightarrow P^a_T$ for all $a \in (0,1]$ as $n \to \infty$.*

It is easy to verify that

$$q^n_i = \binom{2n^2}{i+n^2} 2^{-2n^2}$$

is the stationary distribution for $X^{a,n}_t$ $\forall\, a \in (0,1)$. This is $Bin(2n^2+1, 1/2)$ so if λ^n denotes the corresponding measure then by DeMoivre–Laplace theorem $\lambda^n \Rightarrow \lambda \sim N(0,1/2)$ which is the stationary distribution of the Ornstein–Uhlenbeck process.

5.3. We now indicate the analog of the exponential convergence result in Section 5.1 for the family $\{X^{a,n}_t\}_{n\in\mathbb{N}}$. Let $P^{a,n}_1 = [p^{a,n}_{ij}]$ denote the one-step transition matrix. We observe that

$$S^{a,n} = \frac{1}{2a}P^{a,n}_1 + \frac{2a-1}{2a}I$$

is a tridiagonal stochastic matrix of the Krawtchouck type. Remarkably its spectrum is uniformly distributed on $[0,1]$: $\sigma(S^{a,n}) = \{j/2n^2 | j = 0,\dots,2n^2\}$. Therefore

$$\sigma(P^{a,n}_1) = \left\{a\left(\frac{j}{n^2}-2\right)+1 \,\middle|\, j=0,\dots,2n^2\right\}.$$

Hence for $n \ge n_0 = \sqrt{a/2(1-a)}$ the second largest of the moduli of the eigenvalues is $1 - a/n^2$. This is called the coefficient of ergodicity of the chain and we denote it by $\mu_1(P^{a,n}_1)$. Therefore we have for the time-one transition matrix

$$\mu_1(P^{a,n}_{n^2}) = \left(1-\frac{a}{n^2}\right)^{n^2} \to e^{-a}.$$

This uniformity for the time-one transitions of $\{X^{a,n}_t\}_{n\ge n_0}$ immediately implies the uniformity in the exponential convergence.

PROPOSITION. *Given $a \in (0,1)$ and $K \in \mathbb{R}_+$, we have for all $n \geq n(a, K)$ and $x \in M^n \cap [-K, K]$,*

$$\left| P_x^{a,n}\left(X_t^{a,n} = \frac{i}{n} \right) - \lambda^n\left(\frac{i}{n} \right) \right| \leq C(a, K) e^{-\alpha(a)t} \lambda^n\left(\frac{i}{n} \right), \qquad i = -n^2, \dots, n^2,$$

where C and α are independent of n. As before this implies

$$\| P_x^{a,n}(X_t^{a,n} \in \cdot) - \lambda^n \| \leq Ce^{-\alpha t}.$$

REMARK. The ergodic properties of the chain obviously disappear as $a \downarrow 0$. If $a = 1$ the proposition fails due to periodicity.

The first main result of this section is the following.

5.4 THEOREM. *The processes $(X_t^{a,n}, P_{\lambda^n}^{a,n})$ converge to (X_t^a, P_λ^a) in $\bar{d}$-metric as $n \to \infty$.*

PROOF. *Step* 0: Choose $I_K = [-K, I] \subset \mathbb{R}$ such that

$$\lambda(I_K^c) \vee \sup_{n \geq n_0} \lambda^n(I_K^c) \leq \frac{\varepsilon}{200}$$

for some $n_0 > 0$.

Steps 1–3 in the proof of Theorem 3.2 are modified only to restrict the "essential" state space to be I_K. The couplings ${}_{\mathbf{y}}\eta_{T_c}^{a,n}$ and ${}_{\mathbf{x}}\nu_{T_u}^{a,n}$ are defined to be independent if $\mathbf{x}, \mathbf{y} \notin I_K \times I_K$. In the covering argument of Step 3 the choice of Step 0 is utilized as well as the fact that excursions of length T_u from I_K far outside it are rare. □

5.5. The chain $\{(X_{i/n^2}^{a,n}, P_{\lambda^n}^{a,n}) | i \in \mathbb{N}\}$ is a finite-state irreducible and aperiodic ($a < 1$) Markov chain hence by standard results (see [12]) isomorphic to a Bernoulli shift. Consequently, the process $(X_t^{a,n}, P_{\lambda^n}^{a,n})$ is a Bernoulli flow.

For some fixed t and a let T^a be defined as in (4.1) with $p^a(t,x,y)$ and $f \in L^1(\lambda)$. T^a is a Markov operator and its shift is mixing since given any interval $[b,c]$ and $g \in L^\infty(\mathbb{R})$ we have

$$\begin{aligned} &\int_{\mathbb{R}}\int_{\mathbb{R}} p^a(t,x,y) 1_{[b,c]}(y) g(x) \frac{1}{\sqrt{\pi}} e^{-x^2}\, dx\, dy \\ &\quad = \int_{-M}^{M}\int_b^c p^a(t,x,y) g(x) \frac{1}{\sqrt{\pi}} e^{-x^2}\, dy\, dx \qquad (5.5)\\ &\qquad + \int_{[-M,M]^c}\int_b^c p^a(t,x,y) g(x) \frac{1}{\sqrt{\pi}} e^{-x^2}\, dy\, dx, \end{aligned}$$

where M is such that

$$\left| \frac{1}{\sqrt{\pi}} \int_{[-M,M]} |g(x)| e^{-x^2}\, dx - \|g\|_{L^1(\lambda)} \right| < \frac{\varepsilon}{3}.$$

Since

$$\int_b^c p^a(t,x,y)\,dy \le 1, \qquad \forall\, x \in \mathbb{R},$$

the second term on the right of (5.5) cannot exceed $\varepsilon/3$. We can also choose t_0 such that for $t > t_0$,

$$\left|\int_b^c p^a(t,x,y)\,dy - \lambda([b,c])\right| < \frac{\varepsilon}{3\|g\|_{L^1(\lambda)}}, \qquad \forall\, x \in [-M,M].$$

Consequently,

$$\left|\int_{-M}^{M}\int_b^c p^a(t,x,y)g(x)\frac{1}{\sqrt{\pi}}e^{-x^2}\,dy\,dx - \lambda([b,c])\frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} g(x)e^{-x^2}\,dx\right|$$

$$\le \left|\int_{-M}^{M}\int_b^c [p^a(t,x,y) - \lambda([b,c])]g(x)\frac{1}{\sqrt{\pi}}e^{-x^2}\,dy\,dx\right|$$

$$+\frac{1}{\sqrt{\pi}}\int_{[-M,M]^c}|g(x)|e^{-x^2}\,dx < \frac{2\varepsilon}{3}$$

and the left-hand side of (5.5) converges to $\lambda([b,c])\int g\,d\lambda$. But then by [10] the shift is Bernoulli and the process (X_t^a, P_λ^a) is a Bernoulli flow. These results together with Theorem 5.4 yield the α-congruence. The argument is identical to that of Section 4 and we do not repeat it here.

6. Conclusion. In the foregoing analysis we have covered a large portion of processes for which an invariance principle is known. For both the Bernoulliness and the $\bar{d}$-coupling argument we need the processes to have good enough mixing properties and a unique stationary distribution. Uniform ergodicity (which in our context is equivalent to the Doeblin condition) or some mechanism in the process that essentially compactifies the state space (i.e., makes excursions outside a compact set rare) seems necessary to guarantee these. In some cases geometric ergodicity suffices although the uniformity is then harder to formulate. We note that the exponential convergence to the invariant measure is not necessary per se—in the arguments the uniformity of the convergence rate over the family of random walks is the key ingredient. Verification of the uniform Doeblin condition has to be performed in the special case at hand using typically spatial homogeneity and symmetry properties that the processes might have. For families of translation-invariant random walks on compact manifolds this should be fairly easy. We also point out that the $\bar{d}$-convergence results are genuine strengthenings of the known weak convergence results as indicated in Theorem 1.4. It is also worth noticing that the infinite-time generalizations of the Skorokhod metric (see [5]) involve an exponential discount factor for the distance between the paths in the future and hence is not time homogeneous. Consequently, invariance principles using this metric are not uniform infinite-time generalizations and are implied by our results (if also the ergodicity assumptions are satisfied).

Unlike in the case of billiards or geodesic flows in the context of random processes α-congruence requires somewhat artificial introduction of a "ghost" process to adjust the entropies. This might correspond to some kind of intrinsic unpredictability in a viewers mechanism but its final interpretability remains to be seen.

Finally we point out that a completely different kind of argument might yield an extension of these $\bar{d}$-convergence results to the case where the processes do not have good ergodic properties. For example, in the case of a symmetric random walk on the real line, Donsker's theorem gives a finite-time coupling with a Brownian motion. One could extend the definition of the $\bar{d}$-metric to this case but it is presently not known if an infinite-time coupling between these processes is possible and if it is, whether the degree of transience of the process is relevant.

Acknowledgment. This work is part of a Ph.D. dissertation written at Stanford University. The author wishes to thank Professor Donald Ornstein for numerous helpful conversations on these topics.

REFERENCES

[1] Abramowitz, M. and Stegun, I. (1964). *Handbook of Mathematical Functions*. National Bureau of Standards, Washington.

[2] Baxter, J. R. and Brosamler, G. A. (1976). Energy and the law of iterated logarithm. *Math. Scand.* **38** 115–136.

[3] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[4] Eloranta, K. V. (1989). Alpha-congruence for dispersive billiards. Preprint.

[5] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

[6] Freidlin, M. (1985). *Functional Integration and Partial Differential Equations*. Princeton Univ. Press.

[7] Griffeath, D. (1976). Coupling methods for Markov processes. In *Studies in Probability Theory and Ergodic Theory*. *Adv. in Math. Suppl. Studies* **2**. Academic, New York.

[8] Jørgensen, E. (1975). The central limit problem for geodesic random walks. *Z. Wahrsch. Verw. Gebiete* **32** 1–64.

[9] Ornstein, D. S. (1974). *Ergodic Theory, Randomness, and Dynamical Systems*. Yale Univ. Press, New Haven, Conn.

[10] Ornstein, D. S. and Shields, P. C. (1973). Mixing Markov shifts of kernel type are Bernoulli. *Adv. in Math.* **10** 143–146.

[11] Ornstein, D. S. and Weiss, B. (1989). Statistical properties of chaotic systems. Preprint.

[12] Peterson, K. (1983). *Ergodic Theory*. Cambridge Univ. Press, New York.

[13] Stroock, D. W. and Varadhan, S. R. S. (1971). Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.* **24** 147–225.

Institute of Mathematics
Helsinki University of Technology
02150 Espoo
Finland