LIMIT THEOREMS FOR MARKOV PROCESSES GENERATED BY ITERATIONS OF RANDOM MAPS

OESOOK LEE

1. Introduction

Let $p(x, dy)$ be a transition probability function on (S, ρ), where S is a complete separable metric space. Then a Markov process X_n which has $p(x, dy)$ as its transition probability may be generated by random iterations of the form $X_{n+1} = f(X_n, \mathcal{E}_{n+1})$, where $\mathcal{E}_n$ is a sequence of independent and identically distributed random variables (See, e.g.,Kifer(1986), Bhattacharya and Waymire(1990))

In this paper we consider the case $f(x, \epsilon) = f_\epsilon(x)$. We consider a discrete time Markov process $\{X_n\}$ on a complete separable metric space (S, ρ), represented as $X_n = \Gamma_n \Gamma_{n-1} \cdots \Gamma_1 X_0$, where X_0 is a given random variable with values on S and $\{\Gamma_n\}$ is a sequence of independent and identically distributed random maps on S into itself. Also, X_0 and $\{\Gamma_n\}$ are independent. It is assumed that there exist a positive integer m_0 and a measurable function G on $\Gamma^{(m_0)}$ such that $\forall \gamma \in \Gamma^{(m_0)}, \rho(\gamma x, \gamma y) \leq G(\gamma)\rho(x, y)$ and $E[G(\Gamma_{m_0} \cdots \Gamma_1)] < 1$. $\{X_n\}$ obtained by random affine maps on R^n can be treated as a special case of this type. It is proved that under an additional assumption, there exists a unique invariant probability π, where $p^n(x, dy)$ converges weakly to $\pi(dy)$ an $n \to \infty$ for every $x \in S$. It is also shown that a functional central limit theorem and a strong law of large numbers hold for Lipschitzian functions. Theorems are proved without mentioning irreducibility or any kind of mixing type conditions. This generalizes an earlier results of Bhattacharya and Lee(1988), in which successive compositions of random maps were assumed to be contractions eventually.

The case that Γ has only finitely many maps has been studied by Elton(1987) and Barnsley and Elton(1988). Laskot and Rudnicki(1995) considered the case of $m_0 = 1$.

2. Main Results

Let (S, ρ) be a complete separable metric space and $\mathcal{B}(S)$ its Borel sigma field. Let Γ be a set of continuous maps on S into itself. Let $\mathcal{C}$ be the sigma field on Γ such that the map $(\gamma, x) \longrightarrow \gamma(x)$ is a measurable function on $(\Gamma \times S, \mathcal{C} \otimes \mathcal{B}(S))$. Let P be a probability measure on $(\Gamma, \mathcal{C})$. On some probability space $(\Omega, \mathcal{F}, Q)$ define a sequence of independent and identically distributed random maps $\Gamma_1, \Gamma_2, \ldots$ with common distribution P and a random variable X_0 with values in S independent of the sequence $\{\Gamma_n\}$. Define $X_1 = \Gamma_1 X_0, \ldots, X_n = \Gamma_n X_{n-1} = \Gamma_n \cdots \Gamma_1 X_0$. Here, we write γx for the value of the map $\gamma \in \Gamma$ at x, and $\gamma_n \cdots \gamma_1$ for the composition of the maps $\gamma_1, \ldots, \gamma_n$. Then X_n is a Markov process with transition probability $p(x, dy)$ given by

$$p(x, B) = P(\{\gamma \in \Gamma : \gamma(x) \in B\}), \qquad x \in S, B \in \mathcal{B}(S).$$

$p^n(x, dy)$ denotes the n-step transition probability function, i.e. the distribution of $\Gamma_n \Gamma_{n-1} \cdots \Gamma_1 x$. We shall write $X_n(x)$ for X_n with $X_0 = x$. Note that $\Gamma_n \cdots \Gamma_1 x$ has the same distribution as $\Gamma_1 \cdots \Gamma_n x$.

Write Γ^m for the usual Cartesian product $\Gamma \times \Gamma \times \cdots \times \Gamma$, and let P^m denote the product probablilty on $(\Gamma^m, \mathcal{C}^m)$. Let $\Gamma^{(m)}$ be the set of all compositions $\gamma_m \cdots \gamma_1$ with $\gamma_i \in \Gamma, 1 \leq i \leq m$. For the sigma field $\mathcal{C}^{(m)}$ on $\Gamma^{(m)}$, take the class of all sets B whose inverse under the map $(\gamma_1, \ldots, \gamma_m) \longrightarrow \gamma_m \cdots \gamma_1$ are in $\mathcal{C}^m$. Let $P^{(m)}$ be the induced probability measure on $(\Gamma^{(m)}, \mathcal{C}^{(m)})$.

We make the following assumptions:

There exist $x_0 \in S$, a positive integer m_0 and a nonnegative measurable function $G : \Gamma^{(m_0)} \longrightarrow R$ such that

(A1) $\forall \gamma \in \Gamma^{(m_0)}, \rho(\gamma x, \gamma y) \leq G(\gamma) \rho(x, y), \quad \forall x, y \in S$

(A2) $\lambda := E[G(\Gamma_{m_0} \cdots \Gamma_1)] < 1$

(A3) for each $r \geq 0$, $\sup_{\rho(x, x_0) \leq r} E\rho(X_n(x), x_0) < \infty$,
$\quad n = 1, 2, \ldots, m_0$.

THEOREM 2.1. *Assume for some x_0 ,positive integer m_0 and a non-negative measurable function G on $\Gamma^{(m_0)}$, (A1)-(A3) hold. Then there exists a unique invariant probability $\pi(dy)$ for $p(x, dy)$, and $p^n(x, dy)$ converges weakly to $\pi(dy)$ for every $x \in S$.*

First let us show a lemma.

LEMMA 2.1. *If the hypotheses of the theorem 2.1 hold, then*

$$(2.1) \qquad \sup_n E\rho(X_n(x_0), x_0) < \infty$$

and

$$(2.2) \qquad \sup_{x,y \in C} E\rho(X_n(x), X_n(y)) \to 0$$

for every compact set C, as $n \to \infty$.

Proof. We define $H_n(x) := E\rho(X_n(x), x_0), n = 1, 2, \ldots$ and let $n \geq 1$ be such that $Nm_0 < n \leq (N+1)m_0$ for some nonnegative integer N. Then by using the assumption that $\{\Gamma_n\}$ is independent and identically distributed, we have

$$
\begin{aligned}
H_n(x_0) &= E\rho(\Gamma_1 \cdots \Gamma_n x_0, x_0) \\
&\leq E\left[\rho(\Gamma_1 \cdots \Gamma_n x_0, \Gamma_1 \cdots \Gamma_{m_0} x_0) + \rho(\Gamma_1 \cdots \Gamma_{m_0} x_0, x_0)\right] \\
&\leq E[G(\Gamma_1 \cdots \Gamma_{m_0})] E\rho(\Gamma_{m_0+1} \cdots \Gamma_n x_0, x_0) + K \\
&\leq \lambda \left[\lambda E\rho(\Gamma_{2m_0+1} \cdots \Gamma_n x_0, x_0) + K\right] + K \\
&\leq (\lambda^N + \lambda^{N-1} + \cdots + 1) K \\
&\leq K(1 - \lambda)^{-1}
\end{aligned}
$$

for all n, where $K := \sup_{1 \leq n \leq m_0} H_n(x_0) < \infty$, and hence (2.1) holds.

Fix a compact set C in S. For every two points $x, y \in C$

$$(2.3)
\begin{aligned}
E\rho(X_n(x), X_n(y)) &= E\rho(\Gamma_n \cdots \Gamma_1 x, \Gamma_n \cdots \Gamma_1 y) \\
&\leq \lambda E\rho(\Gamma_{n-m_0} \cdots \Gamma_1 x, \Gamma_{n-m_0} \cdots \Gamma_1 y) \\
&\leq \lambda^N (H_{n-Nm_0}(x) + H_{n-Nm_0}(y)) \\
&\leq 2K_1 \lambda^N,
\end{aligned}
$$

where $K_1 := \sup_{1 \leq n \leq m_0} \sup_{x \in C} H_n(x) < \infty$, by the assumption (A3). Letting $n \to \infty$, i.e. $N \to \infty$ in (2.3), we have (2.2).

On the set of $\mathcal{P}(S)$ of all probability measures on $(S, \mathcal{B}(S))$ define the *bounded Lipschitzian distance*

$$\|\mu - \nu\|_{BL}$$

$$:= \sup\left\{ \left| \int f \, d\mu - \int f \, d\nu \right| : \| f \|_\infty \leq 1, \| f \|_L \leq 1 \right\} \quad (\mu, \nu \in \mathcal{P}(S)),$$

where $\| f \|_\infty = \sup\{|f(x)| : x \in S\}$, $\| f \|_L = \sup\{|f(x) - f(y)|/\rho(x,y) : x \neq y \in S\}$. It is known that $(\mathcal{P}(S), \| \cdot \|_{BL})$ is a complete metric space and $\| \cdot \|_{BL}$ metrizes the weak-star topology on $\mathcal{P}(S)$ (Dudley(1968)).

Proof of theorem 2.1. For any $x_1, x_2 \in C, C$ is a compact set, we have

(2.4)

$$\| p^n(x_1, dy) - p^n(x_2, dy) \|_{BL}$$

$$= \sup\{ |Ef(X_n(x_1)) - E(f(X_n(x_2))| : \| f \|_\infty \leq 1, \| f \|_L \leq 1 \}$$

$$\leq E[\rho(X_n(x_1), X_n(x_2)) \wedge 2] \to 0 \qquad \text{as} \qquad n \to \infty$$

by (2.2). Now writing $B(x_0, M)$ for the ball of radius M centered at x_0,

$$\| p^{n+m}(x_0, dy) - p^n(x_0, dy) \|_{BL}$$

$$\leq E[\rho(\Gamma_1 \cdots \Gamma_n \Gamma_{n+1} \cdots \Gamma_{n+m} x_0, \Gamma_1 \dots \Gamma_n x_0) \wedge 2]$$

$$\leq E[\{\rho(\Gamma_1 \cdots \Gamma_n \Gamma_{n+1} \cdots \Gamma_{n+m} x_0, \Gamma_1 \cdots \Gamma_n x_0) \wedge 2\}$$

$$I_{\{\rho(\Gamma_{n+1} \cdots \Gamma_{n+m} x_0, x_0) \geq M\}}]$$

(2.5)

$$+ E[\{\rho(\Gamma_1 \cdots \Gamma_n \Gamma_{n+1} \cdots \Gamma_{n+m} x_0, \Gamma_1 \cdots \Gamma_n x_0) \wedge 2\}$$

$$I_{\{\rho(\Gamma_{n+1} \cdots \Gamma_{n+m} x_0, x_0) \leq M\}}]$$

$$\leq 2Q(\rho(\Gamma_{n+1} \dots \Gamma_{n+m} x_0, x_0) \geq M)$$

$$+ \sup_{y \in B(x_0, M)} E[\rho(\Gamma_1 \cdots \Gamma_n y, \Gamma_1 \cdots \Gamma_n x_0) \wedge 2]$$

for every $M > 0$. By (2.1) and Chebyshev's inequality, for given $\epsilon > 0$, we may choose $M = M_\epsilon$ such that,

(2.6) $\qquad Q(\rho(\Gamma_{n+1} \cdots \Gamma_{n+m} x_0, x_0) \geq M_\epsilon) < \epsilon/4, \quad \forall m = 1, 2, \dots.$

From (2.2), $\sup_{y \in B(x_0, M_\epsilon)} E\rho(\Gamma_1 \cdots \Gamma_n y, \Gamma_1 \cdots \Gamma_n x_0) \to 0$ as $n \to \infty$, which together with (2.5),(2.6) implies for all sufficient large n,

$$\| p^{n+m}(x_0, dy) - p^n(x_0, dy) \|_{BL} < \epsilon, \qquad \forall m = 1, 2, \ldots$$

Completeness of $(\mathcal{P}(S), \| . \|_{BL})$ ensures the existence of a probablilty measure π such that

$$(2.7) \qquad \| p^n(x_0, dy) - \pi(dy) \|_{BL} \to 0 \quad \text{as} \quad n \to \infty.$$

(2.4) and (2.7) imply that $p^n(x, dy) \to \pi(dy)$ weakly for every $x \in S$. Since $\{X_n\}$ is Feller, π is the unique invariant probablilty.

REMARK. Suppose (A1)-(A3) are satisfied for $m_0 = 1$. Then (A3) holds if $E\rho(\Gamma_1 x_0, x_0) < \infty$, since $E\rho(X_1(x), x_0) \le \lambda\rho(x, x_0) + E\rho(\Gamma_1 x_0, x_0)$.

THEOREM 2.2. *Let the hypotheses of theorem 2.1 hold and let π be the unique invariant probability. Then whatever the initial distribution is,*

$$(2.8) \qquad \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \longrightarrow \int f \, d\pi \qquad \text{a.e.} \quad , \quad n \to \infty$$

for any Lipschitz function f on S.

Proof. Let f be Lipschitzian, i.e. $|f(x) - f(y)| \le M\rho(x, y) \, \forall x, y \in S$ for some $M > 0$. Then $\int |f| d\pi < \infty$, since $\int \rho(x, x_0) d\pi(x) < \infty$. Let $\{\widetilde{X}_n\}$ be the process with initial invariant distribution π, then by Birkhoff's Ergodic theorem,

$$\frac{1}{n} \sum_{i=0}^{n-1} f(\widetilde{X}_i) \longrightarrow \int f d\pi \qquad \text{a.e.}.$$

Let μ be any probability measure on S, and let $\{X_n : n \ge 0\}$ be the process with initial distribution μ and define

$$X_0^r(\omega) = \begin{cases} X_0(\omega) & \text{if } \omega \in \Omega_r \\ x_0 & \text{if } \omega \in \Omega - \Omega_r \end{cases}$$

where $\Omega_r = \{\omega : \rho(X_0(\omega), x_0) \leq r\}$. Let the distribution of X_0^r be μ^r and let $\{X_n^r : n \geq 0\}$ be the process with initial distribution μ^r. Now given $\epsilon > 0$, set $A_n = \{\omega : |f(X_n^r) - f(\widetilde{X}_n)| > \epsilon\}$. Then by similar manner used in (2.3),

$$P(A_n) \leq \frac{M}{\epsilon} E\rho(X_n^r, \widetilde{X}_n)$$
$$\leq \frac{M}{\epsilon} \lambda^N E\rho(X_{n-Nm_0}^r, \widetilde{X}_{n-Nm_0}),$$

where N is the integer part of $\frac{n}{m_0}$. Moreover for $0 \leq i \leq m_0 - 1$,

$$E\rho(X_i^r, \widetilde{X}_i) \leq \int E\rho(X_i(x), x_0)\mu^r(dx) + \int E\rho(\widetilde{X}_i(y), x_0)\pi(dy) \leq K_2$$

where

$$K_2 = \sup_{\rho(x, x_0) \leq r} \sup_{0 \leq i \leq m_0 - 1} E\rho(X_i(x), x_0) + \int \rho(y, x_0)\pi(dy) < \infty.$$

Hence

(2.9)
$$\sum_{n=0}^{\infty} P(A_n) \leq \frac{M}{\epsilon} \sum_{N=0}^{\infty} \lambda^N \sum_{i=0}^{m_0-1} E\rho(X_i^r, \widetilde{X}_i)$$
$$\leq \frac{M}{\epsilon} m_0 K_2 (1 - \lambda)^{-1} < \infty.$$

By Borel-Cantelli lemma,

(2.10)
$$\frac{1}{n} \sum_{i=0}^{n-1} f(X_i^r) \longrightarrow \int f d\pi \quad \text{a.s.}.$$

But $\frac{1}{n} \sum_{i=0}^{n-1} f(X_i^r) = \frac{1}{n} \sum_{i=0}^{n-1} f(X_i)$ on Ω_r and $P(\Omega_r) \to 1$ as $r \to \infty$, which, together with (2.10), completes the proof.

Assume the hypotheses of theorem 2.1 hold and π is the unique invariant probability measure for $p(x, dy)$. Let T be the transition operator on $L^2(S, \pi)$,

$$Tf(x) := \int f(y)p(x, dy), \quad f \in L^2(S, \pi).$$

Then $(T^n f)(x) = \int f(y) p^n(x, dy)$. Let I denote the identity operator. We denote the L^2-norm on $L^2(S, \pi)$ by $\| \cdot \|_2$. Write $\bar{f} = \int f \, d\pi$. Fix $f \in L^2(S, \pi)$. For each positive integer n, write
(2.11)

$$Y_n(t) := n^{-\frac{1}{2}} \sum_{i=0}^{[nt]} \left[(f(X_i) - \bar{f}) + (t - \frac{[nt]}{n})(f(X_{[nt]+1}) - \bar{f}) \right] \quad (t \geq 0),$$

where $[nt]$ is the integer part of nt.

If, for fixed f, $Y_n(.)$ converges in distribution to a Brownian motion, then we say that the functional central limit theorem(FCLT) holds for the function f.

THEOREM 2.3. *Under the hypotheses of the theorem 2.1 ,*

(1) *if the initial distribution of X_0 is π, then for every bounded Lipschitzian f, FCLT holds with mean zero and variance parameter $\| g \|_2^2 - \| Tg \|_2^2$, where $(T - I)g = f - \bar{f}$.*
(2) *if ,in addition, $\int \rho^2(x, r_0) \, d\pi < \infty$, then for every Lipschitzian, FCLT holds.*
(3) *convergences in (1) and (2) hold regardless of the initial distribution.*

Proof. Let f be Lipschitzian on S, $|f(x) - f(y)| \leq M \rho(x, y), \forall x, y \in S$. Note that

$$E\rho(X_{n m_0}(x), X_{n m_0}(y)) \leq \lambda^n \rho(x, y).$$

To prove (1) and (2),it is enough to show that for each case, $f \in L^2(S, \pi)$ and $\sum_{n=0}^{\infty} \| T^n(f - \bar{f}) \|_2 < \infty$ (See Bhattacharya and Lee(1988)).

(1) Assume that for some $L > 0$, $|f(x)| \leq L, \forall x \in S$. Then clearly $f \in L^2(S, \pi)$ and

$$|T^{n m_0}(f - \bar{f})(x)|^2 \leq 2L \Big(\int |Ef(X_{n m_0}(x)) - Ef(X_{n m_0}(y))| \pi(dy) \Big)$$

$$\leq 2LM \int E\rho(X_{n m_0}(x), X_{n m_0}(y)) \pi(dy)$$

$$\leq 2LM \lambda^n \int \rho(x, y) \pi(dy).$$

If we let $a := \int \rho(x, x_0) d\pi(x)$, then $a < \infty$ and

$$\| T^{nm_0}(f - \bar{f}) \|_2^2 \leq 2LM\lambda^n \int\int \rho(x, y)\pi(dy)\, \pi(dx)$$
$$= (4aLM)\lambda^n.$$

Since T is a contraction, we get

$$\| T^{nm_0+k}(f - \bar{f}) \|_2 \leq \| T^{nm_0}(f - \bar{f}) \|_2, \ k = 0, 1, 2, \ldots, m_0 - 1,$$

and therefore

$$\sum_{n=0}^{\infty} \| T^n(f - \bar{f}) \|_2 \leq \sum_{n=0}^{\infty} m_0 \| T^{nm_0}(f - \bar{f}) \|_2$$
$$\leq \frac{(4aLM)^{1/2}m_0}{(1 - \sqrt{\lambda})} < \infty.$$

(2) $f \in L^2(S, \pi)$ follows from the assumption $b := \int \rho^2(x, x_0)\pi(dx) < \infty$.

$$\| T^{nm_0}(f - \bar{f}) \|_2^2 \leq M^2 \int (\int E\rho(X_{nm_0}(x), X_{nm_0}(y))\pi(dy))^2 \pi(dx)$$
$$\leq M^2\lambda^{2n} \int (\int \rho(x, y)\pi(dy))^2 \pi(dx)$$
$$\leq M^2(b + 3a^2)\lambda^{2n}.$$

Therefore we have

$$\sum_{n=0}^{\infty} \| T^n(f - f) \|_2 \leq M\sqrt{b + 3a^2}\, m_0(1 - \lambda)^{-1} < \infty.$$

(3) Let $\{X_n\}$ be the process whose initial distribution is μ, μ is an arbitrary probability measure on $(S, \mathcal{B}(S))$ and let $Y_n(.)$ be the process defined by (2.11). Let $Y_n^r(.)$ and $\widetilde{Y}_n(.)$ be the corresponding processes with $\{X_n\}$ replaced by $\{X_n^r\}$ and $\{\widetilde{X}_n\}$ respectively. Here $\{X_n^r\}$ and

$\{\widetilde{X}_n\}$ are defined in the proof of theorem 2.2. Then we have, for fixed Lipschitzian f,

$$E(\max_{0 \le t \le 1} |Y_n^r(t) - \widetilde{Y}_n(t)|)$$

$$\le Mn^{-\frac{1}{2}} \sum_{i=0}^{n} E\rho(X_i^r, \widetilde{X}_i) \le Mn^{-\frac{1}{2}} m_0 K_2 (1 - \lambda)^{-1},$$

which goes to 0 as $n \to \infty$. From the fact that $Y_n^r(t) = Y_n(t)$ on Ω_r, and $P(\Omega_r) \to 1$ as $r \to \infty$, the conclusion follows.

COROPPARY 2.1. *In addition to the assumptions of theorem 2.1, assume $E[G^2(\Gamma_{m_0} \cdots \Gamma_1)] < 1$ and $\sup_{1 \le n \le m_0} E[\rho^2(X_n(x_0), x_0)] < \infty$. Then FCLT holds for every Lipschitzian function.*

Proof. By theorem 2.3, it remains to show that $\limsup E[\rho^2(X_n(x_0), x_0)] < \infty$, which is obtained by the following inequality:

$$\left(E[\rho^2(X_n(x_0), x_0)] \right)^{1/2}$$

$$\le E[\rho^2(\Gamma_1 \ldots \Gamma_n x_0, \Gamma_1 \ldots \Gamma_{m_0} x_0)])^{1/2}$$

$$+ \left(E[\rho^2(\Gamma_1 \ldots \Gamma_{m_0} x_0, x_0)] \right)^{1/2}$$

$$\le \eta (E[\rho^2(\Gamma_{m_0+1} \ldots \Gamma_n x_0, x_0)])^{1/2} + K_3$$

$$\le (\eta^{\lfloor \frac{n}{m_0} \rfloor} + \cdots + \eta + 1) K_3 \le K_3 (1 - \eta)^{-1},$$

where $\eta^2 = E[G^2(\Gamma_{m_0} \cdots \Gamma_1)] < 1$, and $K_3^2 = \sup_{1 \le n \le m_0} E[\rho^2(X_n(x_0), x_0)] < \infty$.

References

1. Barnsley M. F. and Elton, J. H., *A new class of Markov processes for image encoding*, Advanced Applied Probability **20** (1988), 14-32.
2. Bhattacharya R. N. and Lee, O., *Ergodicity and central limit theorem for a class of Markov processes.*, Journal of Mutivariate Analysis **27** (1988), 80-90.
3. Bhattacharya R. N. and Waymire, E. C., *Stochastic Processes and Applications*, Wiley, New York., 1990.
4. Billingsley P., *Convergence of Probability Measures*, Wiley, New York, 1968.
5. Dudley, R. M., *Distance of probability measures and random variables*, Annals of Mathematical Statistics **39** (1968), 1563-1572.

6. Elton, J. H., *An Ergodic theorem for iterated maps*, Ergodic Theory and Dynam. System **7** (1987), 481-488.

7. Elton, J. H. and Piccioni, M., *Iterated function systems arising from recursive estimation problems*, Probability Theory and Related Fields **91** (1992), 103-114.

8. Kifer, Y., *Ergodic Theory of Random Transformations*, Birkhäuser,Boston., 1986.

9. Lasota, A. and Mackey, M., *Stochastic perturbation of dynamical systems: The weak convergence of measures*, Journal of Mathematical Analysis and Applications **138** (1989), 232-248.

10. Loskot, K. and Rudnicki, R., *Limit theorems for stochastically perturbed dynamical systems*, Journal of Applied Probability **32** (1995), 459-469.

11. Nummelin, E., *General Irreducible Markov Chains and Nonnegative Operators*, Cambridge Univ. Press, Cambridge (1984).

Department of Statistics
Ewha Womans University
Korea