

# Estimating the Stationary Distribution of a Markov Chain\*

by

Krishna B. Athreya  
Department of Operations Research and Industrial Engineering  
Rhodes Hall  
Cornell University  
Ithaca, New York 14853  
[athreya@orie.cornell.edu](mailto:athreya@orie.cornell.edu)

and

Mukul Majumdar  
Department of Economics  
Uris Hall, Cornell University  
Ithaca, New York 14853  
[mkm5@cornell.edu](mailto:mkm5@cornell.edu)

May 2001

**Abstract:** Let  $\{X_n\}_{n=1}^{\infty}$  be a Markov Chain with a unique stationary distribution  $\pi$ . Let  $h$  be a function integrable w.r.t.  $\pi$ . Let  $\bar{h}_n = h d\pi$  and  $\bar{h}_n = \frac{1}{(n+1)} \sum_{k=0}^n h(X_k)$ . This paper explores the conditions for the  $\sqrt{n}$  consistency and asymptotic normality of the estimate  $\bar{h}_n$  of  $\bar{h}$  assuming the existence of  $h$  solution to the Poisson equation  $h - \bar{h} = g - Pg$ . This framework covers the case of nonirreducible Markov chains arising in many growth models in economics.

**Keywords:** Markov chains, stationary distribution,  $\sqrt{n}$  consistency, asymptotic normality, Poisson equation, martingale central limit theorem.

---

\*Thanks are due to Professors Rabi Bhattacharya, Nicholas Kiefer and Timothy Vogelsang for helpful conversations.

## I. Introduction

Here is an informal statement of the problem that is addressed to in this expository paper.

Consider a Markov Chain  $\{X_n\}$  with a unique stationary distribution  $\delta$  which is not easy to compute analytically. A natural alternative is to estimate  $\delta(A)$  for any subset  $A$  of the state space from observing the chain  $\{X_j\}$  over a finite number of periods, say  $0 \leq j \leq n$  and using the sample proportion of visits to  $A$  (defined as  $\hat{\pi}_\phi(A) \equiv \frac{1}{n} \sum_{i=1}^n I_A(X_i)$  where  $I_A$  is the indicator function of  $A$ ) as the estimate of  $\delta(A)$ . The asymptotic consistency property of such an estimate is based on laws of large numbers<sup>1</sup> that assert that, under certain conditions, *for  $\delta$ -almost all initial conditions  $x$ ,  $\hat{\pi}_\phi(A)$  converges to  $\delta(A)$  in a suitable sense*. But one may say that this is somewhat like "begging the question": since  $\delta$  is not known, and since the support of  $\delta$  is often difficult to determine, or is a small set (see Section 4.4 for an example in which the support of the stationary distribution is a subset of  $(0,1)$ , the state space, and has Lebesgue measure zero).

In the context of dynamic economic models, the initial condition is *historically given*, and cannot be chosen by the observer<sup>2</sup> (to belong to the support of  $\delta$ ).

This paper addresses the question of finding conditions under which  $\hat{\pi}_\phi(A)$  is a consistent estimator of  $\delta(A)$  for *any initial condition*, and further, when it is  $\sqrt{n}$ -consistent (i.e.  $\sqrt{n}[\hat{\pi}_\phi(A) - \pi_\phi(A)]$  is stochastically bounded) and when it is asymptotically normal. More generally, we consider a *reward function*  $h$  on the state space, and would like to estimate  $\lambda_h = \delta h$ , the expectation of  $h$  with respect to the stationary distribution. A natural candidate is the empirical average  $\hat{\lambda}_{\phi\phi} = \left[ \sum_{i=1}^n h(X_i) \right] / n + 1$ . One would like to assess the accuracy of  $\hat{\lambda}_{\phi\phi}$ , i.e., the order of magnitude of the error  $|\hat{\lambda}_{\phi\phi} - \lambda_\phi|$ . We provide sufficient conditions under which  $|\hat{\lambda}_{\phi\phi} - \lambda_\phi|$  is of the order of  $(\sqrt{n})^{-\theta}$ . When the Markov Chain  $\{X_n\}$  is *Harris irreducible* (see Orey [1971]), with respect to some nontrivial  $\sigma$ -finite measure, (this includes irreducible

countable state space Markov Chains) the techniques of regeneration due to Athreya and Ney [1978] and Nummelin [1978] can be exploited to find such conditions (see the books by Nummelin [1984], and Meyn and Tweedie [1993]). However, there are many Markov Chains that are generated by iterations of independent identically distributed random maps (also known as Iterated Function Systems (IFS)) that are in general *not irreducible* (see Bhattacharya and Lee (1988), Athreya and Stenflo (2000)). This is especially true when the IFS consists of a finite or countable number of maps and the stationary distribution turns out to be a nonatomic one. Some of the best known stochastic dynamic models in economics - both descriptive and normative - fall into this category. We note that the literature on the "inverse optimal problem" identifies conditions under which a given IFS is "generated" by a stochastic dynamic programming model (see Mitra, Montrucchio and Privileggi (2001) and the list of references). This line of research owes much to the pioneering efforts of Mordecai Kurz (1969).

The present paper is devoted to establishing  $\sqrt{n}$  consistency and asymptotic normality of the estimate  $\hat{\theta}_{hn}$  under the key assumption that there is a function  $g$  such that i)  $h - \hat{\theta}_{hn} = g - Pg$  (the so-called Poisson equation) and ii) some finiteness condition holds. There are many nonirreducible chains that satisfy these conditions. Also in some irreducible cases the present approach is an alternative to the regeneration approach referred to earlier. Kipris and Varadhan (1986) treat the case when the chain is reversible with respect to the stationary distribution.

After some preliminaries in Section 2, a variety of sufficient conditions for  $\sqrt{n}$ -consistency and asymptotic normality are presented in Section 3. In Section 4 we turn to Markov Chains generated by iterations of independent identically distributed *monotone* maps on an interval  $[c,d]$ . When the chain satisfies the splitting condition introduced by Dubins and

Friedman (1966), it has a unique invariant distribution  $\delta$ , to which the distributions of  $X_n$  converge in the Kolmogorov distance geometrically from any initial condition. This turns out to be crucial to derive the  $\sqrt{n}$ -consistency of the estimates. We also sketch in this section some applications of the results to example of growth and cycles under uncertainty [see Stokey and Lucas (1989), Ljungqvist and Sargent (2000) and Bhattacharya and Majumdar (2001)]. All the proofs are relegated to the last section.

## 2. Estimating the Stationary Distribution of a Markov Chain

Let  $\{X_n\}_0^\infty$  be a Markov Chain with state space  $(S, \mathcal{A})$ , transition function  $p(\cdot|\cdot)$  and a given initial distribution  $\pi_0$ . [which may assign mass one to a single point]. Recall that the transition function  $p(\cdot|\cdot)$  satisfies: (a) for any  $A \in \mathcal{A}$ ,  $p(\cdot|A)$  is  $\sigma$ -measurable; and, (b) for any  $x \in S$ ,  $p(x, \cdot)$  is a probability distribution over  $\mathcal{A}$ . Assume that there is a stationary (or, an *invariant*) probability distribution  $\delta$  on  $(S, \mathcal{A})$ , such that

$$\pi(A) = \int_S p(x, A) \pi(dx) \quad (2.1)$$

for every  $A$  in  $\mathcal{A}$ .

Suppose that we wish to estimate  $\delta$  from observing the Markov chain  $\{X_j\}$  for  $0 \leq j \leq n$  starting from a historically given initial condition. A "natural" estimate of  $\delta(A)$  for any  $A$  in  $\mathcal{A}$  is the *sample proportion of visits to A*, i.e.

$$\hat{\pi}_n(A) \equiv \frac{1}{n+1} \sum_0^n I_A(X_j) \quad (2.2)$$

where  $I_A(x) = 1$  if  $x \in A$  and 0 if  $x \notin A$ . We say that  $\hat{\pi}_n(A)$  is a *weakly consistent estimator* of  $\pi(A)$  under  $\hat{\pi}_n(A) \rightarrow \pi(A)$  as  $n \rightarrow \infty$  (in probability, i.e.,

$$\forall \epsilon > 0, \quad \mathbb{P}_\mu(|\hat{\pi}_n(A) - \pi(A)| > \epsilon) \rightarrow 0 \quad (2.3)$$

$\hat{\pi}_n(A)$  is a *strongly consistent estimator* of  $\pi(A)$  if (2.3) is strengthened to:

$$\mathbb{P}_\mu(\hat{\pi}_n(A) \rightarrow \pi(A) \text{ as } n \rightarrow \infty) = 1 \quad (2.4)$$

Assuming that  $\hat{\pi}_n$  is such a consistent estimator, it will be useful to know the *accuracy* of the estimate, i.e., the order of the magnitude of the *error*  $|\hat{\pi}_n(A) - \pi(A)|$ . Under fairly general second moment conditions this turns out to be of the order  $(\sqrt{n})^{-1}$  (i.e., the estimator is  $\sqrt{n}$  *consistent*) and under some further conditions a central limit theorem of the sort asserting that

$$\sqrt{n}(\hat{\pi}_n(A) - \pi(A)) \xrightarrow{d} N(0, \sigma_A^2) \quad (2.5)$$

for  $0 < \sigma_A^2 < \infty$  depending on  $A$  and possibly the initial distribution. This can then be used to give confidence intervals for  $\pi(A)$  having observed  $\{X_j\}_0^n$ .

The above issues can be considered in a more general framework where the goal is to estimate

$$\lambda_h = \int h d\pi \quad (2.6)$$

the integral of a *reward function*  $h$  with respect to  $\pi$  for a real valued measurable function

$h : (S, \mathcal{A}) \times \mathbb{R}$  that is integrable with respect to  $\pi$ . A natural estimate for  $\lambda_h$  is the *empirical average*:

$$\hat{\lambda}_{h,n} \equiv \frac{1}{(n+1)} \sum_0^n h(X_j), \quad (2.7)$$

As before it would be useful to find conditions to assess the accuracy of  $\hat{\lambda}_{h,n}$ , i.e. the order of the magnitude of  $|\hat{\lambda}_{h,n} - \lambda_h|$ . In particular, it is of interest to know whether this estimate is  $\sqrt{n}$  *consistent*, i.e., whether  $|\hat{\lambda}_{h,n} - \lambda_h|$  is of the order  $(\sqrt{n})^{-1}$  and further whether  $\sqrt{n}(\hat{\lambda}_{h,n} - \lambda_h)$  converges in distribution to  $N(0, \sigma_h^2)$  for  $0 < \sigma_h^2 < \infty$  depending on  $h$ . In the next section we state some results that provide precise conditions for the validity of the  $\sqrt{n}$  consistency of and asymptotic normality of the estimators  $\hat{\pi}_n(A)$  and  $\hat{\lambda}_{h,n}$ .

### 3. Sufficient Conditions for $\sqrt{n}$ Consistency and Asymptotic Normality

A *reward function*  $h$  is a  $\mathcal{A}$ -measurable real valued function on  $S$ . Let  $\pi$  be a stationary distribution for the transition function  $p$ ; and let

$$\int_S |h(y)| \pi(dy) < \infty \quad (3.1)$$

Let

$$\lambda_h \equiv \int h \, d\pi \quad (3.1')$$

be the mean of the reward function with respect to  $\pi$ .

In this section we present a variety of sufficient conditions on  $h$  and  $p$  that ensure

$\sqrt{n}$ -consistency and asymptotic normality of  $\hat{\lambda}_{h,n}$  defined in (2.7).

In what follows, if  $g$  is a  $\mathcal{A}$ -measurable real valued function on  $S$  such that for all  $x \in S$

$$\int_S |g(y)| p(x, dy) \quad \text{is finite,} \quad (3.2)$$

then  $Pg$  is a function defined by:

$$Pg(x) \equiv \int g(y) p(x, dy) = E[g(X_1) | X_0 = x], \quad (3.2')$$

the *conditional expectation* of  $g(X_1)$  given  $X_0 = x$ . The *conditional variance* of  $g(X_1)$  given  $X_0 = x$  is defined by

$$\begin{aligned} Vg(x) &\equiv \text{variance } [g(X_1) | X_0 = x] = E[(g(X_1) - Pg(x))^2 | X_0 = x] \\ &\equiv P(g^2)(x) - (Pg(x))^2 \end{aligned} \quad (3.3)$$

provided

$$P(g^2)(x) < \infty \quad (3.3')$$

### 3.1 $\sqrt{n}$ Consistency

The following proposition provides a sufficient condition for the estimate  $\hat{\lambda}_{h,n}$  of  $\lambda_h$  [(2.6) - (2.7)] to be  $\sqrt{n}$ -consistent:

**Proposition 1.** Let  $\delta$  be a stationary distribution for  $p$  (44), and  $h$  be a reward function [satisfying (3.1)]. Let there exist a function  $g$  [satisfying (3.2) and (3.3)] such that the *Poisson equation*

holds, i.e.,

$$h(x) - \lambda_h = g(x) - Pg(x) \quad \text{for all } x \text{ in } S \quad (3.4)$$

Let the initial distribution  $\pi$  be such that the following "finiteness condition" holds:

$$\sup_n \frac{1}{n} \sum_0^{n-1} E_\mu Vg(X_j) < \infty, \quad \sup_n \frac{1}{n} E_\mu (g(X_n))^2 < \infty, \quad E_\mu (Pg(X_0))^2 < \infty \quad (\text{FC})$$

Then,

$$E_\mu (\hat{\lambda}_{h,n} - \lambda_h)^2 \equiv E_\mu \left( \frac{1}{n+1} \sum_{j=0}^n h(X_j) - \lambda_h \right)^2 = O\left(\frac{1}{n}\right) \quad (3.5)$$

**Remark 1.** Since  $\pi$  is an equilibrium distribution for the Markov chain

$$\begin{aligned} \int Pg \, d\pi &= \int (\int g(y)p(x, dy)) \pi(dx) = \int g(y) (\int p(x, dy) \pi(dx)) \\ &= \int g(y) \pi(dy) \quad \text{and so} \quad \int (g - Pg) d\pi = 0 \end{aligned}$$

as it should be since (3.4) holds and  $\lambda_h = \int h d\pi$ .

**Corollary 1.** Under the hypothesis of Proposition 1  $\hat{\lambda}_{h,n}$  is a  $\sqrt{n}$  consistent estimate of  $\lambda_h$ ; that is,  $\sqrt{n}(\hat{\lambda}_{h,n} - \lambda_h)$  is stochastically bounded, i.e., for  $\epsilon > 0$ ,  $K_\epsilon$  such that for all  $n$

$$P_\mu(\sqrt{n} \left| \frac{1}{n+1} \sum_0^n h(X_j) - \lambda_h \right| > K_\epsilon) \leq \epsilon \quad (3.6)$$

In addition,  $\bullet \check{z} > 0$ , as  $n^* ($

$$P_\mu\left(\left|\frac{1}{n+1} \sum_0^n h(X_j) - \lambda_h\right| > \epsilon\right) \rightarrow 0. \quad (3.7)$$

**Corollary 2.** Suppose  $h$  and  $g$  in the above proposition are such that  $g$  is bounded. Then (FC) is satisfied for *any initial distribution*. and hence (3.5) - (3.7) hold.

A natural question is this: given a reward function  $h$  how does one find a function  $g$  such that the Poisson equation (3.4) holds? The following comments address this issue. Let  $h$  be such that for all  $x \in S$

$$\sum_0^\infty |E_x h(X_m) - \lambda_h| < \infty \quad (3.8)$$

Let

$$g(x) \equiv \sum_0^\infty (E_x h(X_m) - \lambda_h) \quad (3.9)$$

satisfy (3.2), i.e.  $E_x |g(X_1)| < \infty$  for all  $x$ . Let

$$(Pg)(x) \equiv E_x g(X_1) = \sum_0^\infty E_x (h(X_{m+1}) - \lambda_h), \quad (3.9')$$

i.e., the expectation and summation can be interchanged. Comparing (3.9) and (3.9') we see that  $g(x) = E_x h(X_0) - \lambda_h + (Pg)(x) = h(x) - \lambda_h + (Pg)(x)$ , i.e., (3.4) holds.

**Corollary 3.** If the function  $g$  defined in (3.9) is bounded then (3.5) - (3.7) hold for any initial distribution . .

**Corollary 4.** Let  $A \in \mathcal{S}$  be such that for all  $x \in S$

$$\sum_0^{\infty} |\mathbb{P}_x(X_m \in A) - \pi(A)| < \infty \quad (3.10)$$

Let

$$g(x) \equiv \sum_0^{\infty} (\mathbb{P}_x(X_m \in A) - \pi(A)) \quad (3.11)$$

be bounded and satisfy

$$E_x g(X_1) = \sum_0^{\infty} (\mathbb{P}_x(X_{m+1} \in A) - \pi(A)) \quad (3.11')$$

Then (3.5) - (3.7) hold with  $h(x) \# I_A(x)$ , for any initial distribution . .

**Corollary 5.** Let  $A \in \mathcal{S}$  be such that

$$\sup_{x \in S} |\mathbb{P}_x(X_m \in A) - \pi(A)| \leq a_m \quad (3.12)$$

where  $\sum_0^{\infty} a_m < \infty$ . Then  $g$  defined in (3.11) is bounded and satisfies (3.11') and hence (3.4) - (3.7) hold with  $h(x) = I_A(x)$  for any initial distribution . .

**Remark 1.** In the case of Markov processes generated by iterations of iid monotone maps on a

finite interval  $[c,d]$  that satisfy the so-called splitting condition it is known (see Dubins & Freedman (1966)) that

$$\sup_{y,x} |P_x(X_m \leq y) - \pi[c,y]| \leq (1-\delta)^m \quad (3.13)$$

for some  $0 < \delta < 1$ . It follows from (3.12) and (3.13) that the empirical distribution function

$\frac{1}{n} \sum_{j=0}^{n-1} I(X_j \leq y)$  is a  $\sqrt{n}$  consistent estimator of  $\pi[c,y]$ . Also, for  $h(x) \neq x$  it can be shown that  $g$  in (3.9) satisfies (3.4). This point is elaborated in Section 4.

### 3.2 Asymptotic Normality

The next Proposition deals with asymptotic normality.

**Proposition 2.** Let  $h, g, P_g, Vg$  be as in Proposition 1. Let

$$\begin{aligned} (Cg)(x, \epsilon) &\equiv E_x((g(X_1) - (P_g)(x))^2 : |g(X_1) - P_g(x)| > \epsilon) \\ &= \int_{M_\epsilon} (g(y) - P_g(x))^2 p(x, dy) \end{aligned} \quad (3.14)$$

where  $M_\epsilon = \{y : |g(y) - P_g(x)| > \epsilon\}$ .

Suppose that the initial distribution  $\pi$  is such that under  $P$

$$\frac{1}{n} \sum_{j=0}^{n-1} Vg(X_j) \xrightarrow{P} \sigma^2 \quad (0 < \sigma^2 < \infty, \text{ nonrandom}) \quad (3.15)$$

$$\sum_0^n (Cg)(X_j, \sqrt{n} \epsilon) \xrightarrow{P} 0 \quad \forall \epsilon > 0 \quad (3.16)$$

$$\frac{1}{n} (P_\mu g)(X_n) \xrightarrow{P} 0 \quad (3.17)$$

Then under such a  $P$

$$\sqrt{n} \left( \frac{1}{n+1} \sum_0^n h(X_j) - \lambda_h \right) \xrightarrow{d} N(0, \sigma^2) \quad (3.18)$$

**Corollary 6.** Suppose  $g$  above is bounded and (3.15) holds. Then (3.16 - 3.17) hold and hence (3.18) holds.

**Corollary 7.** Let  $h: S \rightarrow R$  be such that for all  $x \in S$ ,  $\sum_0^\infty |E_x h(X_m) - \lambda_h| < \infty$  and  $g(x) = \sum_0^\infty (E_x h(X_m) - \lambda_h)$  satisfies (3.9') and (3.15) - (3.17) of Proposition 3. Then (3.18) holds.

**Corollary 8.** Let  $A \subset S$  be such that for all  $x \in S$

$$\sum_0^{\infty} |\mathcal{P}_x(X_m \in A) - \pi(A)| < \infty.$$

Let  $g(x) = \sum_0^{\infty} (\mathcal{P}_x(X_m \in A) - \pi(A))$  be bounded and satisfy (3.11').

Suppose the initial distribution  $\pi$  is such that

$$\frac{1}{n} \sum_0^n p(X_j, A)(1 - p(X_j, A)) \xrightarrow{p} \sigma^2, \quad 0 < \sigma^2 < \infty \text{ nonrandom}$$

Then, under  $P_\pi$ ,  $\sqrt{n} \left( \frac{1}{n+1} \sum_0^n I_A(X_j) - \pi(A) \right) \xrightarrow{d} N(0, \sigma^2)$ .

**Corollary 9.** Suppose there exists a probability distribution  $\delta$  and constants  $\{b_j\}$ , function  $k(\cdot)$  such that

i)  $\|p(\cdot, A) - \delta(A)\|_{TV} < k(x)b_j$

ii)  $\sum_j b_j < \infty$

iii)  $= k^2 d \delta < \infty$

where  $\| \cdot \|_{TV}$  stands for total variation distance. Suppose also that the initial distribution  $\pi$  is such that for any  $g$  integrable with respect to  $\delta$  the weak law of large numbers holds, i.e., under

!,

$$\frac{1}{n} \sum_0^{n-1} g(X_j) \xrightarrow{p} c \in \mathbb{R}, \text{ nonrandom}$$

Then, for such an initial distribution  $\pi$ , for any bounded measurable  $h$ ,

$$\frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} (h(X_j) - \lambda_h) \xrightarrow{d} N(0, \sigma^2)$$

for some  $0 < \sigma^2 < \infty$ , non-random but depending on  $\pi$  and  $h$ .

#### 4. Markov Chains Generated by Iterations of I.I.D. Maps

We now consider Markov chains  $\{X_n\}_0^\infty$  generated by iterations of an i.i.d. sequence of random maps. Let  $S$  be a closed bounded interval  $[c,d]$  ( $c < d$ ), and  $\mathcal{B}$  the Borel  $\sigma$ -field of  $S$ . Let  $\{f_i\}$  be a family of maps from  $S$  into itself. Let  $\{\alpha_n\}_0^\infty$  be a sequence of i.i.d. maps from  $\mathcal{B}$ . For a given initial  $x$ , write

$$X_n(x) = x \quad \text{for } n = 0$$

and

$$X_n(x) = \alpha_n(\dots\alpha_1(x)) \quad \text{for } n \geq 1$$

Then for any  $x$ ,  $X_n(x)$  is a Markov Chain with state space  $S = [c,d]$  and the transition function  $p(\cdot, \cdot)$  is given by

$$p(x, A) = \text{Prob}\{\gamma \in \Gamma : \gamma(x) \in A\} \quad (4.2)$$

The initial state  $x$  can also occur (independently of  $\pi$ ) according to some probability

distribution  $\pi$ . The distribution of  $X_n$  when the initial distribution is  $\pi$  is denoted by  $P^n$ .

#### 4.1 Monotone Maps and the Splitting Condition

Let  $f_i$  be a set of monotone maps from  $S$  into  $S$ ; i.e., each element of  $f_i$  is either a nondecreasing function on  $S$  or a nonincreasing function.

Given two probability measures  $\mu$  and  $\nu$  on  $\mathbb{I}$ , let  $d_K(\cdot, \cdot)$  be the Kolmogorov distance, i.e.,

$$d_K(\mu, \nu) \equiv \sup_{x \in S} |\mu[c, x] - \nu[c, x]| \quad (4.3)$$

It should be noted that convergence in the Kolmogorov distance implies weak convergence.

**Proposition 3.** (Dubins and Freedman (1966)). Assume that the following splitting condition **(H)** holds:

**(H)** *There exist  $z_0 \in S$ ,  $\delta > 0$  and a positive integer  $N$  such that*

$$(1) \quad \text{Prob}(\cup_{n=1}^N \cup_{x \in S} z_0 \leq x \leq z_0 + \delta) \geq 1 - \delta$$

and

$$(2) \quad \text{Prob}(\cup_{n=1}^N \cup_{x \in S} z_0 - \delta \leq x \leq z_0) \geq 1 - \delta$$

(a) Then there is a unique invariant distribution  $\delta$  on  $[c, d]$  of the Markov process  $X_n$ ;

and

(b) for any initial distribution  $\mu$ ,

$$d_K(\mu P^n, \pi) \leq (1-\delta)^{[n/N]} \quad (4.4)$$

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

The estimate (4.4) of the speed of convergence of  $P^n$  to the invariant  $\pi$  plays a crucial role in applying Proposition 1 and its corollaries to the economic models that we describe now.

## 4.2 Models of Growth and Cycles

We should stress that Markov Chains generated by *iid* maps arise "naturally" not only in descriptive dynamic economics, but also in the context of dynamic optimization under uncertainty, particularly when one wishes to study the evolution of states generated by an optimal policy function (see Majumdar, Mitra and Nyarko (1989 for an extended list of references). We briefly outline two examples of applications of Proposition 3 [see Bhattacharya and Majumdar (2001, Section III) for details, particularly for the verification of the splitting condition (H)].

**Example 1.** Let  $f_i = \{F_1, \dots, F_N\}$  where each  $F_i : \mathbb{R}_+ \times \mathbb{R}_+$  satisfies:

F.1.  $F_i$  is strictly increasing, continuous and there is some  $r_i > 0$  such that  $F_i(x) > x$  on  $(0, r_i)$  and  $F_i(x) < x$  for  $x > r_i$ .

F.2. for  $i > i'$ ,  $F_i(x) > F_{i'}(x)$  for all  $x > 0$ .

Let  $\text{Prob}(\pi_n = F_i) = p_i > 0$  ( $1 \leq i \leq N$ ).

As before, let  $X_{n+1}(x) = \pi_{n+1}(X_n(x))$ . It is possible to show that this Markov chain  $\{X_n(x)\}$  with state space  $(0, \infty)$  has the following property: for each  $x \in (0, \infty)$   $X_n(x)$  enters the interval  $[r_1, r_N]$  with probability one, and remains  $[r_1, r_N]$  forever. Hence from the perspective of long run analysis we can take  $[r_1, r_N]$  as the effective state space. The splitting condition (H) is verified in Bhattacharya and Majumdar (2001).

**Example 2.** This example is motivated by the remarks of Solow in his celebrated paper (1956) and the subsequent work of Day (1982). Consider a Markov process with the state space  $S = \mathbb{R}_+$  and two possible laws of motion denoted by  $F$  and  $G$  (i.e.,  $f_i = \{F, G\}$ ) occurring with probabilities  $\hat{p}$  and  $1 - \hat{p}$  respectively ( $0 < \hat{p} < 1$ ). The law of motion  $F$  is monotone increasing and has an attracting positive fixed point (recall Figure 1 of Solow (1956)); however, the other law  $G$  triggers cyclical forces and has a pair of locally attracting periodic points of period 2 (and a repelling fixed point: the precise assumptions are stated below). One may interpret  $F$  as the dominant long run growth law ( $\hat{p}$  is "large"), and  $G$  as the law of short run cyclical interruptions. A numerical example is given in Section 5. But, first, we state the assumptions on  $F$  and  $G$  precisely and note their implications.

The law of motion that generates the growth process is represented by a continuous increasing function  $F : [0, 1] \times [0, 1]$ . We assume that

G.1.  $F$  has a fixed point  $r > \frac{1}{2}$  such that

$$F(x) > x \text{ for } 0 < x < r,$$

$$F(x) < x \text{ for } x > r.$$

Whether or not  $F(0) = 0$  is not relevant for our subsequent analysis. Note that the trajectory from any initial  $x_0$  converges to  $r$ ; indeed, if  $0 < x_0 < r$ , the sequence  $F^{(n)}(x_0)$  increases to  $r$ ; whereas if

$x_0 > r$ , the sequence  $F^{(n)}(x_0)$  decreases to  $r$ .

The law of motion that triggers cyclical forces is denoted by a continuous map

$G: [0,1] \times [0,1]$ . We assume

- C.1.  $G$  is increasing on  $[0, \frac{1}{2}]$  and decreasing on  $[\frac{1}{2}, 1]$ .
- C.2.  $G(x) > x$  on  $[0, \frac{1}{2}]$ .
- C.3.  $G$  has two periodic points of period 2 denoted by  $\{\tilde{x}_1, \tilde{x}_2\}$ , and a fixed point  $x^*$  which is a repelling fixed point of  $G$ , and no other fixed point or periodic point. Moreover,  $\{\tilde{x}_1, \tilde{x}_2\}$  are locally stable fixed points of  $G^{(2)}$ .

Finally,

- C.4.  $G$  has an invariant interval  $[c, d] (\frac{1}{2} < c < d < 1); c < \tilde{x}_1 < x^* < \tilde{x}_2 < d$ . Also  $r \notin (c, d)$ ,  $r \in \{\tilde{x}_1, x^*, \tilde{x}_2\}$ .

Now, when we consider the evolution

$$X_{n+1} = "_{n+1}(X_n)$$

where  $"_{n+1} = F$  with probability  $\hat{p}$  and  $"_{n+1} = G$  with probability  $1-\hat{p}$ . We can proceed as follows: for any initial  $x \in (0,1)$  the process  $X_n(x)$  enters  $[c, d]$  with probability one after a finite number of steps. Also, it is easy to see that  $[c, d]$  is invariant under  $F$ . Hence, for the long run analysis of the evolution of  $X_n$ , we can take  $[c, d]$  as the effective state space. The splitting condition (**H**) is verified by a careful consideration of the structure of the model (Bhattacharya and Majumdar (2001). It should be stressed that while  $G$  is not a monotone function, on the (common) invariant interval  $[c, d]$  both  $F$  and  $G$  are monotone (increasing and decreasing respectively).

Proposition 3 is, therefore, applicable directly to both examples. From (4.7) one has

$$\sup_{x \in [c,d]} |p^{(n)}(X, J) - \pi(J)| \leq (1-\delta)^{[n/N]}$$

where  $J$  is *any* subinterval of  $[c,d]$ . We can now apply Corollary 5 and conclude that the

empirical distribution function  $\frac{1}{n+1} \sum_0^n I[X_j \leq y]$  is a  $\sqrt{n}$ -consistent estimator of  $\delta[c,y]$ .

We now turn to the problem of estimating the "equilibrium mean"  $= y\delta(dy)$ . Here the reward function  $h(x) = x$ . We present some detailed calculations that can be extended to a more general class of reward functions. Going back to (3.4) let us write  $\tilde{h}(\bullet) \equiv h(\bullet) - \lambda_h$ . Note that if we can ensure the convergence of the infinite series  $\sum_{n=0}^{\infty} P^n \tilde{h}$  and its boundedness, then by defining

$$g(\bullet) = \sum_{n=0}^{\infty} P^n \tilde{h}(\bullet)$$

we can satisfy (3.4) as well as the condition (FC).

With  $\tilde{h}(z) \equiv z - \lambda_h \equiv z - \int_c^d y \pi(dy)$ , assume, without loss of generality, that  $c > 0$  [otherwise consider  $\tilde{\tilde{h}}(z) = (z-c) - \int_c^d (y-c) \pi(dy)$ ]. Now,

$$\begin{aligned} (P^n \tilde{h})(x) &= E(\tilde{h}(X_n) | X_0 = x) = E(X_n | X_0 = x) - \int_c^d y \pi(dy) \\ &= \int_{[c,d]} \text{Prob}[X_n > u | X_0 = x] du - \int_{[c,d]} \pi((u,d)) du \\ &= \int_{[c,d]} [1 - p^{(n)}(x, [c,u]) - 1 - \pi([c,u])] du \\ &= - \int_{[c,d]} p^{(n)}(x, [c,u]) - \pi([c,u]) du \end{aligned}$$

Hence,  $\sup_x |P^n \tilde{h}(x)| \leq \int_{[c,d]} (1-\delta)^{[n/N]} du = (d-c)(1-\delta)^{[n/N]}$ . This ensures that

$$\sup_x \sum_{m=1}^{\infty} |P^n \tilde{h}(x)| \rightarrow 0 \quad \text{as } m \rightarrow \infty$$

and the uniform convergence of the series  $\sum_{n=0}^{\infty} P^n \tilde{h}$  as  $m \rightarrow \infty$ . Corollary 1 now leads to the following conclusion: *the empirical mean*  $\hat{\lambda}_{h,n}$  is a  $\sqrt{n}$ -consistent estimate of  $\lambda_h = \int y \pi(dy)$ .

### 4.3 Some Additional Results

We record first an important result applicable to the case where  $f_i$  consists of monotone nondecreasing functions on  $[c,d]$ .

**Proposition 4.** Let  $S = [c,d]$  and  $f_i$  consist of monotone nondecreasing functions from  $S$  into  $S$  and assume that the splitting condition **(H)** holds. Then for any continuous function with bounded variation  $h$ , there is some bounded  $g$  satisfying (3.4). Moreover, independent of the initial distribution . ,

$$\frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} (X_j - \int y \pi(dy)) \rightarrow N(0, \sigma^2)$$

and

$$\sqrt{n} (\hat{\pi}_n([c,y]) - \pi([c,y])) \rightarrow N(0, \sigma^2)$$

[where  $\hat{\pi}_n([c,y]) = \frac{1}{n} \#\{j : 0 \leq j \leq n-1, X_j \in [c,y]\}$  and  $\sigma^2 = g^2 d6 - (hg)^2 d6$ ].

This proposition is a special case of Theorem 3.1 of Bhattacharya and Lee (who proved a functional central limit theorem given the assumptions listed above).

We turn to a pair of sufficient conditions for solving the equation  $h - h = Pg$  when the state space  $S = [c,d]$  and  $\{\alpha_i\}_1^\infty$  is sequence of i.i.d. maps from  $S$  to  $S$  (*not necessarily monotone*). Let  $h = [c,d] \times R$  be absolutely continuous. Then  $h(x) = h(c) + \int_c^x h'(u)du$  for  $0 \leq x \leq d$ . Thus

$$E_x h(X_n) = h(c) + \int_c^d h'(u) P_x(X_n > u) du$$

and

$$\lambda_h = h(c) + \int_c^d h'(u) \pi(u, d] du$$

so

$$\begin{aligned} |E_x h(X_n) - \lambda_h| &\leq \int_c^d h'(u) (P_x(X_n > u) - \pi(u, d]) du \\ &\leq \left( \int_c^d |h'(u)| du \right) d_K(P_x(X_n \in \cdot), \pi) \end{aligned}$$

yielding the following:

**Proposition 5.** Let  $h = [c,d] \times R$  be absolutely continuous. Assume that

$$\sup_{c \leq x \leq d} \sum_{n=0}^{\infty} d_K(P_x(X_n \in \cdot), \pi(\cdot)) < \infty.$$

Then

$$g(x) = \sum_{n=0}^{\infty} (E_x h(X_n) - \lambda_h)$$

is well defined on  $[c,d]$  and solves the equation

$$h - \lambda_h = g - P_g$$

**Corollary 10.** Under the hypothesis of Proposition 3), if  $h : [c,d] \rightarrow R$  is absolutely continuous then  $g(x) = \sum_{n=0}^{\infty} (E_x h(X_n) - \lambda_h)$  is bounded and satisfies  $h - \lambda_h = g - Pg$  and for any initial distribution .

$$\sqrt{n}(\lambda_{h,n} - \lambda_h) \xrightarrow{d} N(0, \sigma^2)$$

for some  $0 < \sigma^2 < \infty$ , nonrandom but depending on  $\lambda_h$  and  $h$ .

Let  $Y_h(x) = \pi_1(\pi_2 \dots \pi_n(x))$ ,  $n \geq 1$ ,  $Y_0(x) = x$  where  $\{\alpha_i\}_1^\infty$  are as before i.i.d. maps from  $S$  to  $S$  (again not necessarily monotone). Then it is clear that for each  $x$  and  $n$ ,  $Y_n(x)$  has the same distribution as  $X_n(x)$ . Suppose  $\lim Y_h(x) = Y$  exists w.p.1 and the distribution of  $Y$  is independent of  $x$ . Then  $Y$  has distribution  $\sigma$  and

$$E_x h(X_n) - \lambda_h = E_x(h(Y_n) - h(Y)) = \int_c^d h'(u)(P_x(Y_n > u) - P(Y > u))du$$

Let  $W_Y(\delta) = \sup_u |P(Y \leq u + \delta) - P(Y \leq u - \delta)|$  be the modulus of continuity of the c.d.f. of  $Y$ . Then it can be verified that

$$\sup_u |P_x(Y_n > u) - P(Y > u)| \leq W_Y(\delta) + 2 P_x(|Y_n - Y| > \delta)$$

This leads us to

**Proposition 6.** Let  $h = [c,d] \rightarrow R$  be absolutely continuous. Let  $Y_h, Y$  be as above. Suppose there exist  $\lambda_h - 0$  such that

$$\sum \mathbb{W}_Y(\delta_n) < \infty \quad (A)$$

and

$$\sup_x \sum \mathbb{P}_x(|Y_h - Y| > \delta_n) < \infty \quad (B)$$

Then the conclusions of Corollary 10 hold.

**Corollary 11.** Let  $\{\alpha_i\}_1^\infty$  be i.i.d. Lipschitz maps from  $[c,d] \times [c,d]$  such that  $\mathbb{E} \text{Fns}(\cdot) < 0$  where  $s(\alpha) \equiv \sup \{ \frac{|\alpha(x) - \alpha(y)|}{|x-y|}, c \leq x, y \leq d, x \neq y \}$  is the Lipschitz constant of the map  $\alpha$ . Let  $Y_n$  and  $Y$  be as before. Assume  $\mathbb{W}_Y(\cdot) \leq c \cdot \ell \cdot \ell$  for some constants  $0 < c < 1, 0 < \ell < 1$ .

Then for  $e^{\ell} < r < 1, \ell_n = r^n$ , conditions (A) and (B) hold and hence the conclusions of Proposition 5 and Corollary 10 hold.

#### 4.4. The Support of an Invariant Distribution: An Example

**Example 3.** Let  $S = (0,1)$  and  $\mathcal{F}_i$  be the family of all admissible laws of motion consist of a pair of functions from the 'quadratic family', specified as follows:

$$\mathcal{F}_i = \{F_i, \hat{F}_i\} \text{ where}$$

$$\hat{F}_i(x) = 2x(1-x)$$

and (5.1)

$$F_i(x) = \mu x(1-x) \quad \text{where } -\frac{1}{2} + \frac{1}{2}\sqrt{17} < \mu < 2$$

Consider the Markov process

$$X_{n+1} = \cdot_{n+1}(X_n) \quad (5.2)$$

where  $\{\alpha_n\}_0^\infty$  is an i.i.d. sequence with the distribution  $\text{Prob}(\alpha_{n+1} = \hat{F}) = p \text{ Prob}(\alpha_{n+1} = F) = 1 - p$ . For this Markov process, it is verified by Bhattacharya and Rao (1993, p. 20, Example 2) that the splitting condition (H) holds and *the support of the invariant distribution is a Cantor set of Lebesgue measure zero*.

**Example 4.** Let  $S = [0,1]$ ;  $F(x) = (3/4)x^{1/2}$ ,  $G(x) = (3/4) \times (1-x)$ . Then all the assumptions in Example 2 of Section 4 hold. The invariant distribution can be estimated through computer simulation.

## 5. Proofs

### Proposition 1

Since  $h$  and  $g$  satisfy (3.4), i.e.

$$\begin{aligned}
 h(x) - \lambda_h &= g(x) - (Pg)(x) \\
 \frac{1}{n+1} \sum_0^n h(X_j) - \lambda_h &= \frac{1}{(n+1)} \sum_0^n (g(X_j) - Pg(X_j)) \\
 &= \frac{1}{(n+1)} \sum_1^n (g(X_j) - (Pg)(X_{j-1})) \\
 &\quad + \frac{1}{(n+1)} \sum_1^n (Pg(X_{j-1})) - Pg(X_j) + g(X_0) - Pg(X_0)
 \end{aligned}$$

Thus, 
$$\frac{1}{n+1} \sum_0^n h(X_j) - \lambda_h = \frac{1}{(n+1)} \sum_1^n Y_j + \frac{1}{(n+1)} (g(X_0) - Pg(X_n)) \quad (A)$$

where  $Y_j = g(X_j) - (Pg)(X_{j-1})$  for  $1 \leq j \leq n$ . By the Markov property of  $\{X_n\}$  and the definition of  $Pg$  it follows that  $\{Y_j\}_0^\infty$  is a *martingale difference sequence*, i.e.

$$E(Y_j | X_0, X_1, \dots, X_{j-1}) = 0.$$

Thus, under (FC),  $\{Y_j : 1 \leq j \leq n\}$  are uncorrelated,

$$E_\mu \left( \sum_1^n Y_j \right) = 0 \text{ and } E_\mu \left( \sum_1^n Y_j \right)^2 = V_\mu \left( \sum_1^n Y_j \right) = \sum_1^n V_\mu(Y_j) = \sum_1^n E_\mu V(Y_j | X_0, \dots, X_{j-1}) \text{ where } V_\mu$$

and  $E_\mu$  stand for mean and variance under the initial distribution  $\mu$  and  $V(Y_j | X_0, \dots, X_{j-1})$  is the conditional variance of  $Y_j$  given  $X_0, X_1, \dots, X_{j-1}$ . Again by the Markov property and the definition of  $(Vg)(x)$  in (3.3) of Proposition 1

$$E_\mu \left( \sum_1^n Y_j \right)^2 = \sum_1^n E_\mu (Vg)(X_{j-1}) = \sum_0^{n-1} E_\mu (Vg)(X_j) \quad (B)$$

Using  $(a+b)^2 \leq 2(a^2 + b^2)$  repeatedly, we get from (A) and (B) that

$$\begin{aligned} E_\mu \left( \frac{1}{n+1} \sum_0^n h(X_j) - \lambda_h \right)^2 &\leq \frac{2}{(n+1)^2} \sum_0^{n-1} E_\mu (Vg)(X_j) \\ &\quad + \frac{4}{(n+1)^2} (E_\mu (Pg(X_n))^2 + E_\mu (g(X_0))^2). \end{aligned}$$

Now, by the hypotheses (FC) of Proposition 1 the right side above is  $O(\frac{1}{n})$  as was to be shown.

This establishes Proposition 1.

**Corollary 1.** By Chebychev's inequality,

$$P_\mu \left( \sqrt{n} \left| \frac{1}{n+1} \sum_0^n h(X_j) - \lambda_h \right| > K_\epsilon \right) \leq \frac{n}{K_\epsilon^2} E_\mu \left( \frac{1}{n+1} \sum_0^n h(X_j) - \lambda_h \right)^2$$

and by (3.5) the right side is bounded by a constant multiple of  $K_\epsilon^{-2}$ . Hence for  $\epsilon > 0$ , there exists a  $K_\epsilon$  such that (3.6) holds. Similar calculation yields (3.7)

**Corollary 2.** If  $g$  is bounded, so is  $Pg$  and hence so is  $Vg$ . Thus (FC) holds.

**Corollary 3 and 4** follow from **Corollary 2**.

**Corollary 5.** Let  $g(x)$  be as in (3.11). By (3.12)  $\{g(x)\}_{x \in \mathbb{R}}$  is bounded. Also the bounded convergence theorem implies (3.11').

**Proposition 2.**

From (A) and (3.17) it is enough to show

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n Y_j \xrightarrow{d} N(0, \sigma^2)$$

Since  $E(Y_j | X_0, \dots, X_{j-1}) = 0, \forall j$ ,  $\{Y_j : 1 \leq j \leq n\}$  is a martingale difference sequence and B.M. Brown's Martingale central limit theorem (Bhattacharya and Waymire, [1990] pp. 508) applies. Also since  $V(Y_j | X_0, \dots, X_{j-1}) = (Vg)(X_{j-1})$  and  $E(Y_j^2 : |Y_j| > \epsilon \sqrt{n} |X_0, \dots, X_{j-1}|) = Cg(X_j, \sqrt{n} \epsilon)$  (3.16) and (3.17) imply Brown's conditions. So (3.18) follows.  $\square$

**Corollary 6.** Omitted.

**Corollary 7.** As shown in the proof of Corollary 2,  $g$  and  $h$  satisfy  $h - Pg_h = g - Pg$ . So (3.18) holds, if  $g$  satisfies (3.15 - 3.17).

**Corollary 8.** It suffices to note that

$$(Vg)(x) = \text{Var}(I_A(X_1) | X_0 = x) = p(x, A) (1 - p(x, A)).$$

Since  $g$  is bounded and if

$$\frac{1}{n} \sum_0^n p(X_j, A)(1-p(X_j, A)) \xrightarrow{P} \sigma^2, \quad 0 < \sigma^2 < \infty$$

conditions (3.15 - 3.17) hold and so (3.18) holds.

**Corollary 9.** If  $h : S \times R$  is bounded and measurable then

$$\begin{aligned} & \sum_0^\infty |(E_x h(X_m) - \lambda_h)| \\ & \leq ||h|| \sum_0^\infty ||p^j(x, \cdot) - \pi(\cdot)|| \\ & \leq ||h|| k(x) \left( \sum_0^\infty b_j \right), \quad \text{where } ||h|| = \sup_x |h(x)| \end{aligned}$$

Thus

$$g(x) = \sum_0^\infty (E_x h(X_m) - \lambda_h)$$

exists for all  $x$  and satisfies (3.4), i.e.,

$$h - \lambda_h = g - Pg$$

Also thus  $g$  is bounded by  $k(x)(\sum_0^\infty b_j)$  and hence is square integrable with respect to  $\delta$ . Also

$Vg(x)$  is integrable with respect to  $\mathcal{G}$ . Hence, by hypothesis, under  $P$

$$\frac{1}{n} \sum_{j=0}^{n-1} (Vg)(X_j) \xrightarrow{P} \sigma^2$$

$0 < \sigma^2 < \infty$  depending on  $g$  and hence on  $h$ . Thus (3.15) of Proposition 2 holds. Since  $g$  is bounded, condition (3.16) - (3.17) follow. Thus (3.18) holds.

**Corollaries 10 and 11:** omitted

## References

Athreya, K.B. and P. Ney (1978): "A New Approach to the Limit Theory of Recurrent Markov chains", *Trans. Amer. Math. Soc.*, 245: 493-501.

Athreya, K.B. and O. Stenflo (2000): "Perfect Sampling For Doeblin Chains", ORIE Tech Report, Cornell University.

Bhattacharya, R.N. and O. Lee (1988): "Asymptotics of a Class of Markov Processes, That Are Not in General Reducible", *Annals of Probability*, 16, 1333-47.

Bhattacharya, R.N. and E.C. Weymire (1990): "Stochastic Processes with Applications", John Wiley, New York.

Bhattacharya, R.N. and B.V. Rao (1993): "Random Iterations of Two Quadratic Maps" in Stochastic Processes (eds. S. Cambanis, J.K. Ghosh, R.L. Karandikar and P.K. Sen), Springer Verlag, New York, pp. 13-21.

Bhattacharya, R.N. and M. Majumdar (2001): "On a Class of Stable Random Dynamical Systems: Theory and Applications", *Journal of Economic Theory*, 96, pp. 208-229.

Chung, K.L. (1974): "A Course in Probability Theory", Academic Press, New York, 2<sup>nd</sup> Edition.

Day, R. (1982): "Irregular Growth Cycles", *American Economic Review*, 72, 406-414.

Dubins, L.E. and D. Freedman (1966): "Invariant Probabilities for Certain Markov Processes", *Annals of Mathematical Statistics*, 37, 837-858.

Elton, J.H. (1987): "An Ergodic Theorem for Iterated Maps", *Ergodic Theory and Dynamical Systems*, 7, pp. 481-488.

Granger, C.W.J. and T. Terasvirta (1993): "Modelling Nonlinear Economic Relationships", Oxford University Press, Oxford.

Kipnis, C. and Varadhan, S.R.S. (1986): "Control Limit Theorem for Additive Functionals of Reversible Markov Processes and Applications to Simple Exclusions", *Communications of Mathematical Physics*, 104, 1-19.

Kurz, M. (1969): "On the Inverse Optimal Problem", in "Lecture Notes in Operations Research and Mathematical Economics", vol. 11, Springer Verlag, Berlin, pp. 189-201.

Ljungqvist, L. and T.J. Sargent (2000): "Recursive Macroeconomic Theory", M.I.T. Press,

Cambridge, MA.

Majumdar, M. Mitra, T., and Y. Nyarko (1989): "Dynamic Optimization under Uncertainty; Non-convex Feasible Set" in "Joan Robinson and Modern Economic Theory" (ed. G.R. Feiwell), MacMillan, London, pp. 545-590.

Meyn, S.P. and R.L. Tweedie (1993): "Markov Chains and Stochastic Stability", Springer-Verlag.

Mitra, T., Montrucchio, L. and F. Privileggi (2001): "The Nature of the Steady State in Models of Optimal Growth under Uncertainty", CAE Working Paper, no. 01-04, Cornell University.

Nummelim, E. (1978): "A Splitting Technique for Harris Recurrent Chains", *Z. Walrsch. und Vern-Greb.* 43, 309-318.

Nummelim, E. (1984): "General Irreducible Markov Chains and Nonnegative Operators", Cambridge University Press, Cambridge.

Orey, S. (1971): "Limit Theorems for Markov Chain Transition Probabilities", Van Nostrand Reinhold, London.

Ray, Debraj (1998): "Development Economics", Princeton University Press, Princeton, NJ.

Solow, R.M. (1956): "A Contribution of the Theory of Economic Growth", *Quarterly Journal of Economics*, 70, 65-94.

Stokey, N.L. and R.E. Lucas (1989): "Recursive Methods in Dynamic Economics", Harvard University Press, Cambridge, MA.

1. We have in mind several fundamental results on ergodicity and the strong law of large numbers that hold when the initial distribution of the Markov process is an invariant distribution [see, for example, Bhattacharya and Waymire (1990, pp. 229-230)].
2. Standard texts on development economics emphasize the *role of history* in understanding and explaining the evolution and institutions of economies [see, for example, Ray (1998, Chapter 5)], and, in a dynamic economic model it is natural to think of the initial condition as a product of history. A similar point was made in applications of ergodic theory to computer graphics: "we have no way of choosing the starting  $x$ " according to an invariant distribution; "in fact the idea is to start at some  $x$  and let a computer generated realization of the process draw a picture of the invariant distribution" [Elton (1987, p. 482)]. The emphasis in Elton was to derive ergodic theorems starting from any state