THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 949

OCTOBER 2005

**Computing the Distributions of Economic
Models Via Simulation**

by

John Stachurksi

Department of Economics
The University of Melbourne
Melbourne   Victoria   3010
Australia.

COMPUTING THE DISTRIBUTIONS OF ECONOMIC MODELS VIA SIMULATION

JOHN STACHURSKI

ABSTRACT. This paper studies the convergence properties of a
Monte Carlo algorithm for computing distributions of state vari-
ables when the underlying model is a Markov chain with absolutely
continuous transition probabilities. We show that the L_1 error of
the estimator always converges to zero with probability one. In
addition, rates of convergence are established for L_1 and integral
mean squared errors. The algorithm is shown to have many appli-
cations in economics.

1. INTRODUCTION

Many economic models are both stochastic and dynamic. The process
for the state variables often has a Markov structure, and when shocks
are nondegenerate, or when the set of agents has positive measure,
the distribution of the state is a nondegenerate over some subset of
$\mathbb{R}^n$. This distribution may indicate the dispersion of asset holdings,
wealth, capital, wages or other such attributes across agents; or the
probabilities of future outcomes for the state.

In this paper we explore a method for computing the distributions of
state variables via simulation recently introduced by Glynn and Hen-
derson (2001). As with other simulation-based techniques, the method
can be used to examine the predictive aspects of economic models too
complex to admit analytical solution.

Our study investigates global convergence properties of Glynn and Hen-
derson's estimator. Arguably the most important global measure of
error for this estimator is the L_1 distance between the estimator and

target distribution. By applying McDiarmid's bounded difference inequality, we show that the L_1 error *always* converges to zero with probability one. Second, we establish rates of convergence for expected L_1 and integral mean squared error for a large class of models. These rates are faster than those achieved by standard nonparametric kernel density estimators.

2. FORMULATION OF THE PROBLEM

Suppose that, after setting down a model, solving all relevant decision problems and combining equilibrium constraints, one arrives at a law of motion for the state variables given by the recursive stochastic sequence

(1) $$X_t = H_t(X_{t-1}, W_t), \quad X_0 \text{ given}, \quad W_t \sim \psi.$$

Here X_t takes values in $S \subset \mathbb{R}^k$ and W_t takes values in $Z \subset \mathbb{R}^j$, while $H_t \colon S \times Z \to S$. We assume that the shocks $(W_t)_{t \geq 1}$ are independent over time and identically distributed (IID) with common distribution ψ. The IID restriction on the shocks and the fact that the state variable only enters with one lag may seem restrictive, but in fact many models can be set in this framework by adjusting the definition of the state. In Section 3 below the theory is developed for a general discrete time Markov process.

When analytical results are unavailable, one can still explore the implications of (1) by computing distributions of the state variables $(X_t)_{t \geq 0}$. The distribution φ_T of X_T provides a complete description of the probabilities implied by the model for time T events.[1]

If (1) is stationary and ergodic, another common exercise is computation of the stationary (invariant) distribution for the state. The issues here are mathematically more subtle but conceptually very similar, and we discuss them in detail below.

---

[1]In the context of density forecasting, Sarno and Valente (2004) describe one rationale for computing distributions rather than just moments as follows: "In a decision-theoretical context, the need to consider the predictive density of a time series—as opposed to considering only its conditional mean and variance—seems fairly accepted in the light of the argument that economic agents may not have loss functions that depend symmetrically on the realizations of future values of potentially non-Gaussian variables."

2.1. **Methods of Computing Distributions.** Let $T \in \mathbb{N}$ and let φ_T denote the distribution of the S-valued random variable X_T defined inductively by (1). A common technique for computing φ_T is to discretize the state space onto a grid of size n, derive for each t a Markov matrix $\mathbf{P}_t^n$ on this finite state space which approximately represents the probabilistic dynamics in (1), and then compute

$$(2) \qquad\qquad \varphi_T^n := \varphi_0^n \cdot \mathbf{P}_1^n \times \cdots \times \mathbf{P}_T^n,$$

where φ_0^n is the approximate distribution of X_0, represented on the grid.[2] For (2) computation is fast, but bounds on the deviation of φ_T^n from φ_T are difficult to obtain, as errors associated with the original discretization are propagated in a complex way at each multiplication in (2).[3]

An alternative which does not involve discretizing the state space is Monte Carlo simulation. In this procedure one starts with X_0, generates a draw $(W_1, \ldots, W_T)$ for the shock process, and then computes X_T iteratively via (1). Repeating this procedure n times gives independent observations $(X_T^1, \ldots, X_T^n)$ of the random variable X_T. By definition each X_T^m is a draw from the target distribution φ_T.

With the sample, one can construct a histogram, an empirical distribution function, or a nonparametric kernel density estimate such as

$$(3) \qquad\qquad f_T^n(y) := \frac{1}{n \cdot \delta_n} \sum_{m=1}^{n} K\left(\frac{y - X_T^m}{\delta_n}\right),$$

where K is a symmetric probability density, and the "bandwidth" parameter δ_n controls the "smoothness" of the estimate, and is chosen so that $\delta_n \to 0$ as $n \to \infty$. Regarding (3), it is well-known that—at least when φ_T is a density—we always have $|f_T^n(y) - \varphi_T(y)| \to 0$ as $n \to \infty$ with probability one for all $y \in S$. Probability one (almost sure) convergence to zero also holds for the L_1 error $\int |f_T^n - \varphi_T|$, independent of the choice of kernel K (Devroye and Györfi, 1985).

---

[2] Here the distributions φ_t^n are regarded as row vectors, and $\mathbf{P}_t^n = (p_{ij}^t)$, where p_{ij}^t is the probability of moving from state i at $t-1$ to j at t. Postmultiplication by the current Markov matrix moves the distribution one period forwards, in which case the right hand side of (2) is a measure of the time T state.

[3] For a concrete example of this problem see Johnson (2005), which uses continuous state methods to revise an earlier discretization-based study.

On the other hand, the *finite* sample properties of f_T^n are not always good. For example, $\mathbb{E}|f_T^n(y) - \varphi_T(y)|$ is known to be proportional asymptotically to $(n\delta_n)^{-1/2}$, and since $\delta_n \to 0$ with n at a rate that is sensitive to dimension of the state space S, the convergence rate is strictly slower that $O(n^{-1/2})$, and possibly much slower. Thus, while asymptotic limiting properties are excellent, convergence to the target may be slow. This is a problem common to many forms of Monte Carlo simulation.

Poor finite sample properties are problematic in a number of situations, such as when the state space is high-dimensional, when drawing random variates from the state distributions is computationally expensive, or when a large parametric class is being studied. In addition, low probability regions of the state space are rarely sampled, making it difficult to uncover features of the distribution on these sets via simulation—a situation which is particularly troubling for studies of extreme events, such as bankruptcy or market crashes.

Another issue is that poor choice of bandwidth or kernels can have significant impact on rates of convergence and finite sample properties. Making good choices depends on sufficient knowledge of the target density φ_T—in particular, being able to place φ_T in a certain restricted class with desirable features. Such knowledge is not always easy to acquire for marginal distributions of state variables when the information at hand consists only of the laws of motion given in (1).

Without additional structure one cannot easily improve on f_T^n. Suppose, then, that the conditional distribution of X_t given X_{t-1} can be represented by a density. To give an elementary example, consider the Solow model defined by

$$(4) \qquad\qquad k_t = sAk_{t-1}^\alpha W_t, \quad \ln W_t \sim N(0, \sigma^2),$$

where k is capital, $s > 0$ is the savings rate, and $\alpha, A > 0$ are productivity parameters. It is clear that when k_{t-1} is taken as given, the current state k_t is lognormally distributed: $\ln k_t \sim N(\ln(sA) + \alpha \ln k_{t-1}, \sigma^2)$. Thus, the distribution of the current state given the lagged state is indeed represented by a density. This existence of density representation turns out to provide exactly the kind of structure in question.

To see this, fix $T \in \mathbb{N}$ and suppose that the conditional distribution of X_T given X_{T-1} can be represented by density $p_T(X_{T-1}, y)dy$. Using

p_T, Glynn and Henderson (2001) proposed the following "look-ahead" estimation scheme for φ_T. First, generate n independent draws of the state variable as above, but this time generate draws of X_{T-1} rather than of X_T. Now calculate

$$(5) \qquad \varphi_T^n(y) := \frac{1}{n} \sum_{m=1}^{n} p_T(X_{T-1}^m, y).$$

To see why φ_T^n is a natural estimator of the distribution of X_T, observe the following. It can be shown (formal arguments are deferred to a later section) that when the density representation p_T exists, the distribution of X_T can be represented by a density φ_T, and φ_T satisfies

$$(6) \qquad \mathbb{E}\, p_T(X_{T-1}, y) = \varphi_T(y), \quad \forall y \in S.$$

The intuition for (6) is simple: If $\varphi_T(y)$ is thought of as the probability of observing y at T, then this should be equal to the probability $p_T(x, y)$ of going from x at $T-1$ to y at T, summed over x and weighted by the probability that $X_{T-1} = x$; and this is precisely the left hand side of (6).

From (5) and (6) we have $\mathbb{E}\varphi_T^n(y) = \frac{1}{n} n \varphi_T(y) = \varphi_T(y)$ at each point y, so that φ_T^n is pointwise unbiased. Moreover, the law of large numbers implies that, with probability one,

$$(7) \qquad \varphi_T^n(y) = \frac{1}{n} \sum_{m=1}^{n} p_T(X_{T-1}^m, y) \to \mathbb{E}p_T(X_{T-1}, y) = \varphi_T(y)$$

as $n \to \infty$. In other words, $\varphi_T^n(y)$ is a consistent estimator of $\varphi_T(y)$ at each point $y \in S$.

Notice that φ_T^n makes use of the structure of the model as embodied in p_T—a key aspect of efficient computation. In contrast to f_T^n there is no bandwidth parameter, nor any need to choose a kernel K. These two features suggest that φ_T^n will have good finite sample properties to match the asymptotic result (7). Indeed, Glynn and Henderson point out that by the Central Limit Theorem we have $\mathbb{E}|\varphi_T^n(y) - \varphi_T(y)| = O(n^{-1/2})$, independent of the dimension of the state space S.

To illustrate the speed of convergence, consider Figure 1, which compares the look-ahead estimator to the actual time T density φ_T and a kernel density estimate for the elementary Solow model (4).[4] We

---

[4]In the figure, the parameters are $\alpha = 0.3$, $A = 2$, $\sigma = 0.11$, and $s = 0.2$.

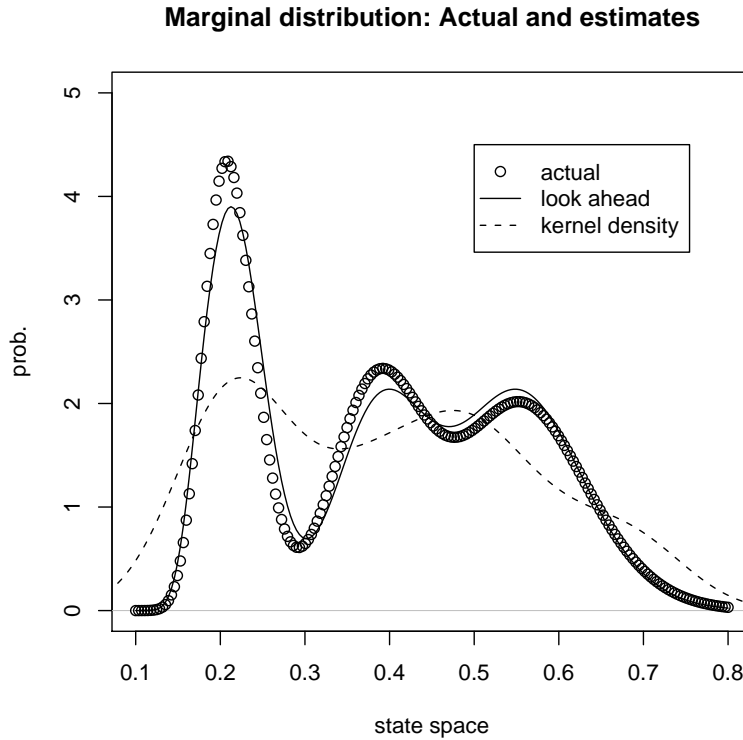**Marginal distribution: Actual and estimates**

FIGURE 1.   The Look-Ahead Estimator.

argued that the distribution for the current state given $k_{t-1} = x$ is the lognormal density $p(x, y)dy$, where

$$(8) \qquad p(x, y) = (2\pi\sigma^2)^{-1/2}\frac{1}{y}\exp\left\{\frac{-(\ln y - \ln(sA) - \alpha\ln x)^2}{2\sigma^2}\right\}.$$

Given this function p, implementation of the look-ahead estimator (5) is extremely simple. Programmed in R and using `samples` as the vector which contains the draws of the time $T - 1$ state, the look-ahead estimate $\varphi_T^n(y)$ is evaluated for Figure 1 by

```
look_ahead = function(y) {
  q = numeric(n)  # vector of length n
  for (i in 1:n) q[i] = p(samples[i],y)
  return( mean(q) )
}
```

In Figure 1 the estimates of φ_T are for $T = 2$. The initial condition φ_0 has been deliberately chosen as multi-modal, making φ_2 multi-modal and increasing the complexity of the approximation problem.[5] Despite this complexity, the combination of log-linearity and log-normality means that an analytical solution for φ_T is also available for comparison, and this is plotted using the $\circ$ symbol. The look-ahead estimate φ_T^n is the unbroken line. Although the sample size is tiny by Monte Carlo standards ($n = 100$), the estimator closely follows the actual density.

The broken line in Figure 1 is a standard kernel density estimate of the form given in (3). In this case we are using the default algorithm in R.[6] The kernel density estimate uses the same draw of shocks as the look-ahead estimate, and the same sample size ($n = 100$). At least for this default algorithm, the approximation is much poorer.[7]

In this paper we extend the analysis of Glynn and Henderson, particularly with regards to global convergence of the *function* φ_T^n to φ_T. The most important metric here is arguably the L_1 distance, which is always well-defined, and for which Scheffés identity provides a natural quantitative interpretation.[8] We prove for the first time that φ_T^n always converges to φ_T in L_1 with probability one as $n \to \infty$.

Second, we provide rates of convergence for global error measures. We prove that for a wide class of models the expected L_1 error (respectively, the integral mean square error) is $O(n^{-1/2})$ (respectively, $O(n^{-1})$). This

---

[5]We are using $\varphi_0 = (1/3)(f_1 + f_2 + f_3)$, where f_i is lognormal with parameters μ_i and σ_i; $\mu_1 = -4$, $\sigma_1 = 1$, $\mu_2 = 3$, $\sigma_2 = 1$, $\mu_3 = 7$, $\sigma_3 = 0.5$.

[6]The kernel K is Gaussian, and the bandwidth is selected according to the rule-of-thumb $\delta_n = 1.06 \min(\hat{\sigma}_n, \hat{R}_n/1.34) n^{-1/5}$, where $\hat{\sigma}_n$ is the sample standard deviation, and $\hat{R}_n$ is the inter-quartile range.

[7]This is *not* a criticism of the standard nonparametric kernel estimator, which is far more general. Further, careful choice of bandwidth and kernel will lead to improvement. The point is that the look-ahead estimator *automatically* incorporates model structure, while for the kernel estimator including enough structure to obtain similar rates of convergence is in general a nontrivial exercise.

[8]Sheffés identity states that $\int |\varphi_T^n - \varphi_T| = 2 \times \sup_B |\int_B \varphi_T^n - \int_B \varphi_T|$, where the supremum is over all Borel subsets of the state space S. It follows that if $\int |\varphi_T^n - \varphi_T| \leq \varepsilon$, then for any event B of interest the deviation in the probability assigned to B by the approximate density φ_T^n from that assigned by the true density φ_T is less than $\varepsilon/2$.

is strictly faster than rates obtained for the kernel estimator f_T^n.[9] For some common situations we provide upper bounds on the L_1 and integral mean square error in terms of the functions H_t and the distribution ψ of the shock in the benchmark model (1).

2.2. Computation of Stationary Distributions.
In some cases the model is stationary over time ($H_t = H$ for all t) and ergodic, in the sense that the distribution φ_t of X_t converges to some limiting distribution φ_∞ (usually called the *stationary* or *invariant* distribution) independent of initial conditions. For such models the stationary distribution has the interpretation of long-run stochastic equilibrium, and hence is of central interest to the modeler.

As Glynn and Henderson (2001) point out, the look-ahead estimator can often be applied. Precisely, let $p(X_{t-1}, y)dy$ again be the conditional density of X_t given X_{t-1} as implied by $X_t = H(X_{t-1}, W_t)$, and let $(X_1, \ldots, X_n)$ be a series drawn recursively from $X_t = H(X_{t-1}, W_t)$. They propose the estimator

$$(9) \qquad \varphi_\infty^n(y) := \frac{1}{n} \sum_{t=1}^n p(X_t, y).$$

Notice that we are now summing over time, rather than across independent samples of the state at a fixed point in time.

The intuition for φ_∞^n is as follows. As discussed above, a stationary density for the model $X_t = H(X_{t-1}, W_t)$ is defined as a density φ_∞ satisfying

$$(10) \qquad \int p(x, y)\varphi_\infty(x)dx = \varphi_\infty(y), \quad \forall y \in S.$$

When a stationary density exists, and moreover, $\varphi_t \to \varphi_\infty$ as $t \to \infty$, we often also have the correlated law of large numbers result

$$(11) \qquad \frac{1}{n} \sum_{t=1}^n w(X_t) \to \int w(x)\varphi_\infty(x)dx \ \text{ as } n \to \infty,$$

where w is any measurable function with $\int w(x)\varphi_\infty(x)dx$ finite. As a result,

$$(12) \qquad \varphi_\infty^n(y) = \frac{1}{n} \sum_{t=1}^n p(X_t, y) \to \int p(x, y)\varphi_\infty(x)dx = \varphi_\infty(y)$$

---

[9]See, for example, Hansen (2005) and references.

as $n \to \infty$. Thus the look-ahead estimator φ_∞^n is again seen to be a very natural estimator, and Glynn and Henderson establish strong finite sample and asymptotic properties under reasonable assumptions. We extend their analysis by establishing almost sure L_1 convergence to the true density under weaker conditions than previous results.

3. The General Model

The state space is any separable and completely metrizable topological space S. Let $\mathscr{B}_S$ denote the Borel sets of S, and let $(S, \mathscr{B}_S)$ be endowed with a σ-finite measure μ. Typically S is a Borel subset of $\mathbb{R}^k$, in which case μ will always be the Lebesgue measure. When integrating over S with respect to μ, write dx for $\mu(dx)$, dy for $\mu(dy)$, etc.; and $\int$ in place of $\int_S$. As usual, $L_1(S)$ is the set of real, $\mathscr{B}_S$ measurable functions f on S such that $|f|$ is μ-integrable.

The set of *densities* on S is the set of Borel measurable, nonnegative real functions on S that integrate to 1. A *distribution* on S is a probability measure on $(S, \mathscr{B}_S)$. A *stochastic kernel* is a family of distributions $P(x, dy)$ on S, $\forall x \in S$, with the property that $x \mapsto P(x, B)$ is Borel measurable for each $B \in \mathscr{B}_S$. A *density kernel* p on S is a measurable map $p \colon S \times S \to [0, \infty)$ such that $p(x, y)dy$ is a density on S for every $x \in S$.

We take as our primitive a Markov chain on S defined by a sequence of stochastic kernels $(P_t)_{t \geq 1}$. The interpretation is that $P_t(x, dy)$ is the probability distribution of X_t given $X_{t-1} = x \in S$; in the case of (1) we have $P_t(x, B) = \psi\{z \in Z : H_t(x, z) \in B\}$. More generally, the Markov chain $(X_t)_{t \geq 0}$ associated with initial distribution φ_0 on S and our model $(P_t)_{t \geq 1}$ is defined by:

(13) $\qquad X_0 \sim \varphi_0$ and then recursively $X_t \sim P_t(X_{t-1}, dy)$.

Formally, given an initial distribution φ_0 and a sequence of kernels $(P_t)_{t \geq 1}$, there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a sequence of S-valued random variables $(X_t)_{t \geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ with the property that $X_0 \sim \varphi_0$ and

(14) $\qquad \mathbb{P}\{X_t \in B \mid X_{t-1}\} = P_t(X_{t-1}, B), \quad \forall t \geq 1,\ \forall B \in \mathscr{B}_S.$

For more background see, for example, Durrett (1996, Chapter 5), or Lindvall (2002, Chapter 3, Section 8).

Assumption 3.1. For each stochastic kernel P_t in $(P_t)_{t\geq 1}$, there exists a density kernel p_t which represents it:

$$(15) \qquad P_t(x, B) = \int_B p_t(x, y)dy, \quad \forall B \in \mathscr{B}_S, \ \forall x \in S.$$

At this point we can verify Equation (6). To do so, take expectations of both sides of (14) to get $\mathbb{P}\{X_t \in B\} = \mathbb{E}P_t(X_{t-1}, B)$. From this expression, (15) and Fubini's Theorem we have

$$(16) \qquad \mathbb{P}\{X_t \in B\} = \int_B \mathbb{E}\, p_t(X_{t-1}, y)dy, \quad \forall B \in \mathscr{B}_S.$$

From (16) it is clear that the distribution of X_t is absolutely continuous with respect to μ for every $t \in \mathbb{N}$, with density representation $\varphi_t(y) = \mathbb{E}\, p_t(X_{t-1}, y)$.

Following (5), the *T-step look-ahead* (TSLA) estimator is the random density function φ_T^n defined by $\varphi_T^n(y) := \frac{1}{n}\sum_{m=1}^{n} p_T(X_{T-1}^m, y)$, where $X_{T-1}^1, \ldots, X_{T-1}^n$ are IID draws from φ_{T-1}. If $p_t = p$ for all t, then the stationary distribution look-ahead (SDLA) estimator is the random density function φ_∞^n defined by $\varphi_\infty^n(y) := \frac{1}{n}\sum_{t=1}^{n} p(X_t, y)$, where now we are now summing over a time series draw, rather than across independent samples of the state at a fixed point in time.[10]

4. Almost Sure Global Convergence

The measure of global convergence with most immediate quantitative interpretation is the L_1 error, given by

$$\|\varphi_t^n - \varphi_t\| := \int |\varphi_t^n(y) - \varphi_t(y)|dy, \quad t \in \mathbb{N} \cup \{\infty\}.$$

Regarding the TSLA φ_T^n, $T \in \mathbb{N}$, Glynn and Henderson (2001) establish that the L_1 error always converges to zero in expectation. They also prove almost sure convergence when p_t is uniformly continuous and bounded on $S \times S$. In fact almost sure L_1 convergence always holds:

Theorem 4.1. *The TSLA φ_T^n converges in L_1 to φ_T with probability one as $n \to \infty$.*

---

[10]In other words, $X_1, \ldots, X_n$ obeys (13), where X_0 is given.

Now consider almost sure L_1 convergence for the look-ahead estimator of the stationary distribution. We require some minimal conditions on the Markov chain to ensure that its time series satisfy the strong law of large numbers. To state them, some definitions are necessary.[11] Let $p_t = p$ for all t. A density $\varphi_\infty \in \mathscr{D}_S$ is called stationary for p if (10) holds; that is, if $\int p(x, y)\varphi_\infty(x)dx = \varphi_\infty(y)$ holds for all $y \in S$. Let $(X_t)_{t \geq 0}$ be the Markov chain generated by p and initial condition $X_0 \equiv x \in S$. For this chain define

$$L(x, A) := \mathbb{P} \cup_{t \geq 1} \{X_t \in A\}.$$

The chain is called λ-irreducible there exists a nontrivial measure λ on $(S, \mathscr{B}_S)$ such that $L(x, A) > 0$ for all $x \in S$ and all $A \in \mathscr{B}_S$ with $\lambda(A) > 0$; and Harris recurrent if $L(x, A) = 1$ for all $x \in A$ whenever $A \in \mathscr{B}_S$ and $\lambda(A) > 0$. A Harris recurrent chain with a stationary distribution is called *positive Harris*. (For Harris chains the stationary distribution is necessarily unique.)

Assumption 4.1. The model is time homogeneous: $p_t = p$ for all t. The Markov chain $(X_t)_{t \geq 0}$ generated by p is positive Harris.

This positive Harris assumption is sufficient to obtain a law of large numbers result for the series $(X_t)_{t \geq 0}$: By Meyn and Tweedie (1993, Theorem 17.1.7), if $(X_t)_{t \geq 0}$ is positive Harris with stationary distribution φ_∞, then for every function $w : S \to \mathbb{R}$ with $\int |w| d\varphi_\infty < \infty$ we have $\frac{1}{n}\sum_{t=1}^{n} w(X_t) \to \int w(x)\varphi_\infty(x)dx$ almost surely as $n \to \infty$. (In fact the converse is true, in the sense that when a stationary distribution exists and the law of large numbers holds for all such w then $(X_t)_{t \geq 0}$ is positive Harris. In this sense the positive Harris assumption is minimal for our purposes.)

For positive Harris chains, Glynn and Henderson (2001) proved almost sure L_1 convergence of the SDLA φ_∞^n to φ_∞ when p is uniformly continuous and bounded on $S \times S$. Here we show that the same result holds under the following condition, which is weaker than uniform continuity and independent of boundedness.

Assumption 4.2. Let d metrize S. The kernel p is continuous in y uniformly in x. Precisely, for all $\varepsilon > 0$ and all $y \in S$, there is a $\delta > 0$ such that $d(y', y) < \delta$ implies $\sup_{x \in S} |p(x, y) - p(x, y')| < \varepsilon$.

---

[11]See Meyn and Tweedie (1993) for further details.

Theorem 4.2. *If Assumptions 4.1 and 4.2 hold, then the SDLA φ_∞^n converges in L_1 to φ_∞ with probability one.*

5. Rates of Convergence

Asymptotic convergence results are reassuring, but without bounds on the rate of convergence they provide no guidance on finite sample properties. In this section we examine rates of convergence, and bounds on global error measures such as expected L_1 error or integral mean squared error.

Consider first the expected L_1 error. In macroeconomics it is common to deal with continuous models on compact state spaces.[12] Our first result pertains to this situation.

Theorem 5.1. *Let S be a compact subset of $\mathbb{R}^k$. If p_T is continuous on $S \times S$, then $\mathbb{E}\|\varphi_T^n - \varphi_T\| = O(n^{-1/2})$.*

To deal with state spaces which are not compact, we require that the shock is additive with exponentially decreasing tails. In addition, a mild restriction is placed on the growth rate of the law of motion:

Assumption 5.1. *Let $S = Z = \mathbb{R}^k$, and let $X_t = g_t(X_{t-1}) + W_t$, where W_t is distributed according to some density ψ on $\mathbb{R}^k$, the map $g_t \colon \mathbb{R}^k \to \mathbb{R}^k$ is measurable for all t, and, for some norm $\|\cdot\|$ on $\mathbb{R}^k$,*

   (i) *$\exists \alpha, L > 0$ s.t. $\|g_t(x)\| \leq \alpha\|x\| + L$ for all $t \in \mathbb{N}$, all $x \in \mathbb{R}^k$; and*
   (ii) *$\exists K, \varrho > 0$ s.t. $\psi(z) \leq K \exp(-\varrho\|z\|^2)$ for all $z \in \mathbb{R}^k$.*

Theorem 5.2. *Let (X_t) be the sequence in Assumption 5.1, where X_0 is a constant $x_0 \in S$, let φ_T be the density of X_T, and let φ_T^n be the TSLA of φ_T. If Assumption 5.1 holds, then $\mathbb{E}\|\varphi_T^n - \varphi_T\| = O(n^{-1/2})$.*

Another common measure of global error is the integral mean square error, defined as

$$\mathrm{IMSE}\,(\varphi_t^n) := \mathbb{E} \int [\,\varphi_t^n(y) - \varphi_t(y)\,]^2 \, dy, \quad t \in \mathbb{N} \cup \{\infty\}.$$

---

[12]See, for example, Brock and Mirman (1972), or Stokey, Lucas and Prescott (1989, Chapter 13).

We give a condition for the integral mean square error of the TSLA to be $O(n^{-1})$—a fast rate of convergence relative to standard nonparametric kernel density estimators.

Theorem 5.3. *Let $(p_t)_{t \geq 1}$ be given and let $T \in \mathbb{N}$ be fixed. If φ_T^n is the TSLA of φ_T, then $\mathrm{IMSE}\,(\varphi_T^n) = O(n^{-1})$ whenever $\int p_T(x,y)^2\,dy$ is bounded above independent of $x \in S$. In particular,*

$$(17) \qquad \mathrm{IMSE}\,(\varphi_T^n) \leq \frac{1}{n} \cdot \sup_{x \in S} \int p_T(x,y)^2\,dy.$$

Notice that the rate does not depend on the time T or the dimension of S, although the dimension of S typically influences the size of the constants in the order term. We give some applications of this result in Section 6.

6. Examples and Applications

First, let's consider Assumption 3.1, which requires that the transition probabilities have a density representation. The first lemma gives sufficient conditions for the basic model $X_t = H_t(X_{t-1}, W_t)$ to satisfy the assumption. The conditions are not necessary, but when they hold the result also provides a representation for the density kernel p_t.

Lemma 6.1. *For the model (1), let Z and S be open subsets of $\mathbb{R}^k$, and let ψ be a density on Z. Let $S_x := H(x, Z)$, the range of $z \mapsto H(x, z)$, and let $z \mapsto H_t(x, z)$ be one-to-one for each $x \in S$. Define $G_x \colon S_x \to Z$ to be the inverse mapping of this function. If G_x is a C^1 function for each $x \in S$, then Assumption 3.1 holds.[13] Moreover, if J_x denotes the Jacobian of G_x, then*

$$(18) \qquad p_t(x,y) = \begin{cases} \psi\{G_x(y)\} \cdot |\det J_x(y)| & \text{if } y \in S_x \\ 0 & \text{otherwise,} \end{cases}$$

This is just an elementary change of variable result, and the proof is omitted. It tells us how to compute the density of the random variable $H_t(x, W_t)$ when x is fixed and $W_t \sim \psi$, which is precisely what $p(x,y)dy$ represents. The following corollary helps to illustrate application of the lemma.

---

[13]A function f from one open subset of Euclidean space to another is called C^1 if it is continuously differentiable everywhere on its domain.

Corollary 6.1. *Assume the hypotheses of Lemma 6.1, and assume in addition that $Z = S = \mathbb{R}^k$, and that*

(19) $X_t = H_t(X_{t-1}, W_t) = g_t(X_{t-1}) + \Sigma_t(X_{t-1}) W_t,$

where $g_t \colon S \to S$ is any Borel measurable function, and $\Sigma_t(x)$ is an invertible $n \times n$ matrix for all t and all $x \in S$. In this case,

(20) $p_t(x, y) = \psi\{\Sigma_t(x)^{-1}[y - g_t(x)]\} \cdot |\det \Sigma_t(x)^{-1}|$

holds everywhere on $S \times S$.

Example 6.1. Let $Z = S = \mathbb{R}$, and consider the smooth transition threshold autoregression (STAR) model

(21) $X_t = (\beta_0 + \beta_1 X_{t-1})(1 - G(X_{t-1})) + (\beta_0' + \beta_1' X_{t-1})G(X_{t-1}) + \sigma W_t,$

where $(W_t)_{t \geq 1}$ is IID according to density ψ on S, $\sigma > 0$, and $G \colon S \to [0, 1]$ is a smooth transition function, such as the logistic function, satisfying $G' > 0$, $\lim_{x \to -\infty} G(x) = 0$ and $\lim_{x \to \infty} G(x) = 1$. Evidently the conditions of Corollary 6.1 are satisfied, and from (20) we get

(22) $p_t(x, y) = p(x, y) = \psi\left\{\frac{y - g(x)}{\sigma}\right\}\frac{1}{\sigma},$

where $g(x) := (\beta_0 + \beta_1 x)(1 - G(x)) + (\beta_0' + \beta_1' x)G(x).$

Example 6.2. In finance one frequently studies continuous time diffusion processes of the form

(23) $X_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t,$

where X_t is $\mathbb{R}^k$-valued, (W_t) is Brownian motion, and σ is everywhere strictly positive definite. The model (23) is used to price financial instruments such as stock options and bonds. Many pricing algorithms involve simulation, and a standard simulation method is discretization of the time parameter via Euler's scheme:

(24) $X_t = X_{t-1} + \mu(t - 1, X_{t-1}) + \sigma(t - 1, X_{t-1})W_t,$

where W_t is standard normal. Corollary 6.1 clearly applies, and the density kernel is immediate from (20).

Example 6.3. Consider the following simple one-sector optimal growth model. At t a representative household observes k_t and divides it between consumption c_t and investment x_t. The current productivity shock A_{t+1} is then observed, and production takes place, yielding output $A_{t+1}f(x_t)$ at the start of next period. Here $A_t := (1 + \gamma)^t W_t,$

where γ is the rate of productivity growth, and $(W_t)_{t \geq 1}$ are IID on $Z := (0, \infty)$ with density ψ.

Let Π be the set of all Borel measurable $h \colon [0, \infty) \to [0, \infty)$ satisfying $0 \leq h(k) \leq k$. These are the *feasible policies*, and each one defines a process

$$(25) \qquad k_t = A_t f(h(k_{t-1})) + (1 - \delta)k_{t-1},$$

where $\delta \in (0, 1]$ is the depreciation rate. The agent has period utility u and discount factor β. He or she chooses h to solve

$$(26) \qquad \max_{h \in \Pi} \mathbb{E} \left\{ \sum_{t=0}^{\infty} \beta^t u(c_t^h) \right\},$$

where $c_t^h := k_t - h(k_t)$. Let u be bounded for simplicity.[14] Let u and f both be nonnegative, differentiable, strictly increasing, with u strictly concave, $\lim_{c \to 0} u'(c) = \infty$ and $f(0) = 0$. In this case it is known that a solution h to (26) exists. Under standard conditions we also have $0 < h(k) < k$ for every $k \in S$.[15] Suppose this is the case.

Consider the optimal dynamics for k on $S := (0, \infty)$, which are given by the random sequence (25) under the optimal policy h. Since $h(k) > 0$ for all $k \in S$ and $f' > 0$ we have $f(h(k)) > 0$ for all $k \in S$. Using this fact one can verify the conditions of Lemma 6.1, and (18) gives us

$$(27) \qquad p_t(x, y) = \psi \left\{ \frac{y - (1 - \delta)x}{(1 + \gamma)^t f(h(x))} \right\} \frac{1}{(1 + \gamma)^t f(h(x))}$$

when $y > (1 - \delta)x$ and zero otherwise.

Now let's illustrate Assumption 4.1, which imposes Harris recurrence.

Example 6.4. Consider again the stochastic growth model in Examples 6.3 and 6.6. Take $\gamma = 0$, so that $p_t = p$ is stationary. It has been shown (Nishimura and Stachurski, 2005) that this model is positive

---

[14]This is assumed here only for simplicity. As is well-known, many specific models with unbounded utility can also be treated by dynamic programming on the basis of assumptions constraining maximal growth rates under the stochastic production function relative to the precise utility specification.

[15]For example, this is true when f is concave. Even when concavity fails, reasonable sufficient conditions exist. See, for example, Nishimura, Rudnicki and Stachurski (2005, Proposition 3.1).

Harris under the hypotheses of those two examples whenever the usual
Inada conditions hold. Hence Assumption 4.1 is satisfied.

Example 6.5. Consider the STAR model of Example 6.1. Recall that
a collection of random variables $(X_t)_{t \geq 0}$ taking values in S is called
tight whenever, for each $\varepsilon > 0$, there is a compact subset K of S such
that $\sup_{t \geq 0} \mathbb{P}\{X_t \notin K\} \leq \varepsilon$. Recall also that p is called (weak) Feller
if $x \mapsto \int h(y)p(x, y)dy$ is continuous and bounded on S whenever h
is. It can easily be deduced from Meyn and Tweedie (1993), Theorems
6.0.1(iii), 9.0.2 and 12.1.2(ii) that if S is a subset of $\mathbb{R}^k$ which contains
an open set, if p is Feller and irreducible with respect to the restriction
of Lebesgue measure to S, and if the Markov chain $(X_t)_{t \geq 0}$ generated
by p is tight for all the initial conditions $X_0 \equiv x_0 \in S$, then p is positive
Harris.

Returning to the STAR model, it is easy to show that if ψ is standard
normal, for example, then p defined in (22) is irreducible with respect to
Lebesgue measure on $\mathbb{R}$. Since G is assumed continuous, p is also Feller.
We now verify tightness under the hypotheses $\alpha := \max\{|\beta_1|, |\beta_1'|\} < 1$
and $\mathbb{E}|W_t| < \infty$.

Simple algebra shows that there is a finite constant L such that

(28) $$|g(x)| \leq \alpha|x| + L, \quad \forall x \in S.$$

$$\therefore \quad \mathbb{E}_{t-1}|X_t| = \mathbb{E}_{t-1}|g(X_{t-1}) + \sigma W_t| \leq \alpha|X_{t-1}| + L + \sigma \int |z|\psi(dz).$$

$$\therefore \quad \mathbb{E}|X_t| \leq \alpha\mathbb{E}|X_{t-1}| + L', \quad L' := L + \sigma \int |z|\psi(dz).$$

Iterating this inequality backwards in time to $t = 0$ we get

$$\mathbb{E}|X_t| \leq \alpha^t|x_0| + \frac{L'}{1 - \alpha}.$$

$$\therefore \quad \sup_{t \geq 0} \mathbb{E}|X_t| \leq |x_0| + \frac{L'}{1 - \alpha}.$$

Chebychev's inequality now gives

$$\mathbb{P}\{|X_t| \geq n\} \leq n^{-1}\left(|x_0| + \frac{L'}{1 - \alpha}\right), \quad \forall n \in \mathbb{N}.$$

Evidently (X_t) is tight, and the STAR model is positive Harris.

Now let's turn to Assumption 4.2. A special but important case is where S is an open subset of $\mathbb{R}$. For this case it is easy to see that Assumption 4.2 is satisfied whenever $p_t(x, y)$ is differentiable in y for each $(x, y) \in S \times S$, and

$$(29) \qquad \forall y \in S, \ \exists K_y \in \mathbb{R} \text{ s.t. } \left| \frac{\partial p_t(x, y)}{\partial y} \right| \leq K_y, \ \forall x \in S.$$

Example 6.6. Consider the stochastic growth model of Example 6.3. Let $\ln W_t \sim N(0, 1)$, and, for simplicity, let $\delta = 1$. Notice that p_t is neither bounded nor uniformly continuous on $S \times S = (0, \infty) \times (0, \infty)$.[16] However, Assumption 4.2 holds, as can easily be verified via (29). In fact, the representation (27) and some simple calculus shows that

$$\left| \frac{\partial p_t(x, y)}{\partial y} \right| \leq K_y := \frac{1}{\sqrt{2\pi y^2}}, \quad \forall x \in S.$$

Example 6.7. In the nonlinear autoregression (21), it is clear from (22) that Assumption 4.2 holds whenever ψ is differentiable on $\mathbb{R}$ and ψ' is bounded.

Next we illustrate Assumption 5.1.

Example 6.8. In the STAR model $X_t = g(X_{t-1}) + W_t$, where

$$g(x) := (\beta_0 + \beta_1 x)(1 - G(x)) + (\beta_0' + \beta_1' x)G(x), \quad W_t \sim N(0, \sigma^2),$$

Assumption 5.1 is satisfied with $\alpha = \max\{|\beta_1|, |\beta_1'|\}$, $L = \max\{|\beta_0|, |\beta_0'|\}$, $K = (2\pi\sigma^2)^{-1/2}$ and $\varrho = (2\sigma^2)^{-1}$.

Finally, an application of Theorem 5.3 is given.

Proposition 6.1. *Consider the model* (19), *where* $\Sigma_t(x)$ *is positive definite for all* t *and* x. *Let* $(p_t)_{t \geq 1}$ *be the corresponding density kernels, defined by* (20). *Let* ψ_t^x *be the density of the random term* $\Sigma_t(x)W_t$. *Let* $T \in \mathbb{N}$ *be fixed, and let* φ_T^n *be the TSLA of* φ_T. *If there exist constants* $K \geq 0$ *and* $\varrho > 0$ *such that* ψ_t^x *satisfies* $\psi_t^x(z) \leq K \exp(-\varrho\|z\|)$, *for all* $x \in S$, $t \in \mathbb{N}$ *and* $z \in Z$, *then*

$$(30) \qquad \text{IMSE}\,(\varphi_T^n) \leq \frac{K^2}{n} \frac{2\pi^{k/2}}{\Gamma(k/2)(2\varrho)^k}(k-1)!$$

---

[16] In fact, p_t may not be continuous when f is non-concave.

If ψ_t^x satisfies $\psi_t^x(z) \le K \exp(-\varrho \|z\|^2)$, for all $x \in S$, $t \in \mathbb{N}$ and $z \in Z$, then

$$(31) \qquad\qquad \mathrm{IMSE}\,(\varphi_T^n) \le \frac{K^2}{n} \left(\frac{\pi}{2\varrho} \right)^{k/2}.$$

The conditions in the proposition are just small tail assumptions for the distribution ψ of W_t. The will be satisfied if, for example, $\Sigma_t(x)$ is a constant and ψ is Gaussian.

7. Proofs

Proof of Theorem 4.1. The following proof draws on ideas in Devroye and Lugosi (2001, § 9.4) concerning concentration of measure inequalities. A discussion of McDairmid's inequality can be found there.

For the proof, fix $n \in \mathbb{N}$, and let S^n be the n-fold cartesian product of S with itself, a typical element of which is $x = (x_1, \ldots, x_n)$. Let $X_{T-1}^1, \ldots, X_{T-1}^n$ be IID draws from φ_T. By McDairmid's inequality, if g is a measurable function from S^n to $\mathbb{R}$ such that

$$\sup |g(x) - g(x')| \le c,$$

where the supremum is over all pairs x, x' in S^n which differ on at most one coordinate, then

$$\mathbb{P}\{|g(X_{T-1}) - \mathbb{E}g(X_{T-1})| \ge \varepsilon\} \le 2 \exp \left(\frac{-2\varepsilon^2}{nc^2} \right),$$

where $g(X_{T-1}) := g(X_{T-1}^1, \ldots, X_{T-1}^n)$. Setting

$$g(x) = g(x_1, \ldots, x_n) = \int \left| \frac{1}{n} \sum_{m=1}^n p_T(x_m, y) - \varphi_T(y) \right| dy$$

gives $g(X_{T-1}) = \|\varphi_T^n - \varphi_T\|$. Pick any $x, x' \in S^n$ such that x and x' differ only at the k-th coordinate. In this case $|g(x) - g(x')|$ is given by the expression

$$\left| \int \left| \frac{1}{n} \sum_{m=1}^n p_T(x_m, y) - \varphi_T(y) \right| dy - \int \left| \frac{1}{n} \sum_{m=1}^n p_T(x'_m, y) - \varphi_T(y) \right| dy \right|,$$

which is bounded above by

$$\int \left| \frac{1}{n} \sum_{m=1}^{n} p_T(x_m, y) - \frac{1}{n} \sum_{m=1}^{n} p_T(x'_m, y) \right| dy$$

$$= \frac{1}{n} \int |p_T(x_k, y) - p_T(x'_k, y)| dy.$$

$$\therefore \quad |g(x) - g(x')| \leq \frac{1}{n} \int |p_T(x_k, y) - p_T(x'_k, y)| dy \leq \frac{2}{n}.$$

$$\therefore \quad \mathbb{P}\{|g(X_{T-1}) - \mathbb{E}g(X_{T-1})| \geq \varepsilon\} \leq 2 \exp\left(\frac{-n\varepsilon^2}{2}\right).$$

$$\therefore \quad \mathbb{P}\{|\, \|\varphi_T^n - \varphi_T\| - \mathbb{E}\|\varphi_T^n - \varphi_T\|\,| \geq \varepsilon\} \leq 2 \exp\left(\frac{-n\varepsilon^2}{2}\right).$$

It now follows from the Borel-Cantelli Lemma that

$$\lim_{n \to \infty} |\, \|\varphi_T^n - \varphi_T\| - \mathbb{E}\|\varphi_T^n - \varphi_T\|\,| \to 0 \text{ almost surely.}$$

Thus, $\lim_{n \to \infty} \|\varphi_T^n - \varphi_T\| \to 0$ almost surely whenever $\mathbb{E}\|\varphi_T^n - \varphi_T\| \to 0$. In other words, convergence in expectation implies almost sure convergence. That convergence in expectation always holds was shown in Glynn and Henderson (2001, Theorem 4). $\square$

Next is the proof of Theorem 4.2. By Schéffe's Lemma, $\|\varphi_\infty^n - \varphi_\infty\| \to 0$ whenever $\varphi_\infty^n \to \varphi_\infty$ pointwise. Moreover, by the LLN in Meyn and Tweedie (1993, Theorem 17.1.7), we know that at each point $y \in S$ the look-ahead estimator $\varphi_\infty^n(y)$ converges to the true density $\varphi_\infty(y)$ on the complement of a set E_y with $\mathbb{P}(E_y) = 0$. However, since S may be uncountable, we cannot conclude that $\varphi_\infty^n \to \varphi_\infty$ pointwise with probability one. Thus, to show almost sure L_1 convergence, some degree of regularity is imposed on the density kernel p to help control the uncountable family of $\mathbb{P}$-null sets $\{E_y : y \in S\}$. This is the purpose of Assumption 4.2.

Lemma 7.1. *Let $B_\delta(y) := \{y' : d(y, y') < \delta\}$. If Assumption 4.2 holds then φ_∞ is continuous on S, and φ_∞^n is continuous on S uniformly in n, in the sense that for all $\varepsilon > 0$ and all $y \in S$ there is a $\delta > 0$ such that*

$$(32) \qquad y' \in B_\delta(y) \quad \Longrightarrow \quad \sup_{n \in \mathbb{N}} |\varphi_\infty^n(y) - \varphi_\infty^n(y')| \leq \varepsilon.$$

Proof. Regarding the first statement, fix $\varepsilon > 0$ and $y \in S$. Choose $\delta > 0$ as in Assumption 4.2. Then for $y' \in B_\delta(y)$,

$$|\varphi_\infty(y) - \varphi_\infty(y')| = \left| \int p(x,y)\varphi_\infty(x)dx - \int p(x,y')\varphi_\infty(x)dx \right|$$

$$\leq \int |p(x,y) - p(x,y')|\varphi_\infty(x)dx \leq \varepsilon.$$

Regarding (32), the same argument yields a $\delta > 0$ such that for $y' \in B_\delta(y)$ we have

$$|\varphi_\infty^n(y) - \varphi_\infty^n(y')| \leq \frac{1}{n}\sum_{t=1}^{n} |p(X_t,y) - p(X_t,y')| \leq \frac{1}{n}\sum_{t=1}^{n} \varepsilon.$$

$\square$

Proof of Theorem 4.2. As discussed above, it is sufficient to show that φ_∞^n converges to φ_∞ pointwise for all paths ω in some set $E \in \mathscr{F}$ with $\mathbb{P}(E) = 1$. So let A be a countable dense subset of S, and note by the LLN that for each $a \in A$ there is a corresponding set $E_a \subset \Omega$ with $\mathbb{P}(E_a) = 1$ and $\varphi_\infty^n(a) \to \varphi_\infty(a)$ on E_a. Let $E := \cap_{a \in A} E_a$. Clearly $\mathbb{P}(E) = 1$. We claim that for every path $\omega \in E$ we have $\varphi_\infty^n \to \varphi_\infty$ as $n \to \infty$ pointwise. To see this, fix any such path, any $y \in S$ and any $\varepsilon > 0$. By Lemma 7.1 we can take a $\delta > 0$ such that $|\varphi_\infty(y) - \varphi_\infty(y')| < \varepsilon$ for all $y' \in B_\delta(y)$, and, in addition, (32) holds. Choose $a \in A \cap B_\delta(y)$.

By the triangle inequality, $|\varphi_\infty^n(y) - \varphi_\infty(y)|$ is less than

$$|\varphi_\infty^n(y) - \varphi_\infty^n(a)| + |\varphi_\infty^n(a) - \varphi_\infty(a)| + |\varphi_\infty(a) - \varphi_\infty(y)|$$

$$\therefore \quad |\varphi_\infty^n(y) - \varphi_\infty(y)| \leq 2\varepsilon + |\varphi_\infty^n(a) - \varphi_\infty(a)|,$$

where ε does not depend on n. Because we are considering a path in E, taking limits gives

$$\lim_{n \to \infty} |\varphi_\infty^n(y) - \varphi_\infty(y)| \leq 2\varepsilon.$$

Since ε is arbitrary the proof is done. $\square$

Proof of Theorem 5.1. By Fubini's Theorem, the Cauchy-Schwartz inequality and independence of the sequence $X_{T-1}^1, \ldots, X_{T-1}^n$ we get

$$
\begin{aligned}
\mathbb{E}\|\varphi_T^n - \varphi_T\| = \int \mathbb{E}|\varphi_T^n(y) - \varphi_T(y)|dy \\
\leq \int \sqrt{\operatorname{Var}(\varphi_T^n(y))}dy \\
\leq \sqrt{\frac{1}{n}} \int \sqrt{\operatorname{Var}(p_T(X_{T-1}^m, y)}dy \\
\leq \sqrt{\frac{1}{n}} \int \sqrt{\mathbb{E}p_T(X_{T-1}^m, y)^2}dy.
\end{aligned}
$$

Since p_T is continuous and S is compact, there is a $K < \infty$ with $p_T \leq K$ everywhere on $S \times S$, and hence the bound

$$
\mathbb{E}\|\varphi_T^n - \varphi_T\| \leq \sqrt{\frac{1}{n}} K\mu(S).
$$

holds for all $n \in \mathbb{N}$. □

Next we turn to the proof of Theorem 5.2. The proof involves several lemmata.

Lemma 7.2. *If Assumption 5.1 holds and X_0 is a fixed constant $x_0 \in S$, then $\mathbb{E}\exp(r\|X_t\|) < \infty$ for all $r > 0$ and all $t \in \mathbb{N}$.*

Proof. By (i) of Assumption 5.1, we have, for all $t \in \mathbb{N}$,

$$
\|X_t\| \leq \alpha\|X_{t-1}\| + L + \|W_t\|.
$$

$$
\therefore \quad r\|X_t\| \leq r\alpha^t\|x_0\| + \sum_{i=0}^{t-1} r\alpha^i(L + \|W_{t-i}\|).
$$

$$
\therefore \quad \exp(r\|X_t\|) \leq \exp(r\alpha^t\|x_0\|) \prod_{i=0}^{t-1} \exp(r\alpha^i L) \prod_{i=0}^{t-1} \exp(r\alpha^i\|W_{t-i}\|)).
$$

$$
\therefore \quad \mathbb{E}\exp(r\|X_t\|) \leq \exp(r\alpha^t\|x_0\|) \prod_{i=0}^{t-1} \exp(r\alpha^i L) \prod_{i=0}^{t-1} \mathbb{E}\exp(r\alpha^i\|W_{t-i}\|)).
$$

From (ii) of Assumption 5.1 the expectation $\mathbb{E}\exp(a\|W_t\|))$ is finite for any $a > 0$, so the right hand side of the last inequality is finite. □

Lemma 7.3. *If (ii) of Assumption 5.1 holds, then there exists a positive constant N such that $\psi(z) \leq N\exp(-\|z\|)$ for all $z \in S$.*

Proof. Let $M := \{z : \|z\| \leq 1/\varrho\}$. For $z \notin M$ we have $\varrho\|z\| > 1$, and hence $\varrho\|z\|^2 > \|z\|$. Therefore,

$$K \exp(-\varrho\|z\|^2) \leq K \exp(-\|z\|), \quad \forall z \notin M.$$

Now set $K_0 := \sup_{z \in M} K \exp(-\varrho\|z\|^2 + \|z\|)$, so that

$$K \exp(-\varrho\|z\|^2) \leq K_0 \exp(-\|z\|), \quad \forall z \in M.$$

Now setting $N := \max\{K_0, K\}$ and applying (ii) of Assumption 5.1 provides a constant with the desired property. □

Proof of Theorem 5.2. The proof of Theorem 5.1 provides the bound

$$\mathbb{E}\|\varphi_T^n - \varphi_T\| \leq \sqrt{\frac{1}{n}} \int \sqrt{\mathbb{E}p_T(X_{T-1}^m, y)^2} dy.$$

We must verify that the integral is finite. To this end, observe that Lemma 7.3 yields an $N < \infty$ with

$$p(X_{T-1}^m, y)^2 = \psi(g(X_{T-1}^m) - y)^2 \leq N^2 \exp(-2\|g(X_{T-1}^m) - y\|).$$

But

$$\exp(-2\|g(X_{T-1}^m) - y\|) \leq \exp(-2\|y\| + 2\|g(X_{T-1}^m)\|).$$

$$\therefore \quad \sqrt{\mathbb{E}p(X_{T-1}^m, y)^2} \leq N\sqrt{\mathbb{E}\exp(-2\|y\| + 2\|g(X_{T-1}^m)\|)}$$

$$\leq N \exp(-\|y\|)\sqrt{\mathbb{E}\exp(2\|g(X_{T-1}^m)\|)}$$

$$\leq N \exp(-\|y\|)\sqrt{\mathbb{E}\exp(2\alpha\|X_{T-1}^m\| + 2L)}$$

$$\leq N \exp(-\|y\|)\exp(L)\sqrt{\mathbb{E}\exp(2\alpha\|X_{T-1}^m\|)}.$$

As a result,

$$\mathbb{E}\|\varphi_T^n - \varphi_T\| \leq \sqrt{\frac{1}{n}}N \int \exp(-\|y\|)dy \exp(L)\sqrt{\mathbb{E}\exp(2\alpha\|X_{T-1}^m\|)}.$$

Here the expectation on the right is finite from Lemma 7.2. □

Proof of Theorem 5.3. Since the TSLA is unbiased and $\{X_{T-1}^1, \ldots, X_{T-1}^n\}$ are independent, we have

$$\text{IMSE}(\varphi_T^n) = \int \mathbb{E}[\varphi_T^n(y) - \varphi_T(y)]^2 \, dy$$

$$= \int \text{Var}(\varphi_T^n(y)) \, dy = \frac{1}{n} \int \text{Var}(p_T(X_{T-1}^m, y)) \, dy.$$

But evidently

$$\int \operatorname{Var} p_T(X^m_{T-1}, y) \, dy \leq \int \mathbb{E} p_T(X^m_{T-1}, y)^2 \, dy = \mathbb{E} \int p_T(X^m_{T-1}, y)^2 \, dy.$$

$$(33) \qquad \therefore \quad \operatorname{IMSE}(\varphi^n_T) \leq \frac{1}{n} \mathbb{E} \int p_T(X^m_{T-1}, y)^2 \, dy.$$

The result (17) now follows. □

Proof of Proposition 6.1. We apply (17). Under the first condition we have

$$\int p_T(x, y)^2 \, dy = \int \psi^x_t [y - g_T(x)]^2 dy$$

$$\leq \int K^2 \exp(-2\varrho \|y - g_T(x)\|) dy = \int K^2 \exp(-2\varrho \|z\|) dz.$$

Direct integration gives

$$\int \exp(-2\varrho \|x\|) dx = \frac{2\pi^{k/2}}{\Gamma(k/2)(2\varrho)^k}(k-1)!\,,$$

from which (30) now follows. The proof for the second case is essentially identical, this time using

$$\int \exp(-2\varrho \|x\|^2) dx = \left(\frac{\pi}{2\varrho}\right)^{k/2}.$$

□

References

[1] Brock, W. A. and L. Mirman (1972): "Optimal Economic Growth and Uncertainty: The Discounted Case," *Journal of Economic Theory,* 4, 479–513.

[2] Devroye, L. and L. Györfi (1985): *Nonparametric Density Estimation: The L_1 View,* John Wiley and Sons, New York.

[3] Devroye, L. and G. Lugosi (2001): *Combinatorial Methods in Density Estimation,* Springer-Verlag, New York.

[4] Durrett, R. (1996): *Probability: Theory and Examples,* Duxbury Press, New York.

[5] Glynn, P. W. and S. G. Henderson (2001): "Computing Densities for Markov Chains via Simulation," *Mathematics of Operations Research,* 26, 375–400.

[6] Hansen, B. (2005): "Exact Mean Integrated Squared Error of Higher Order Kernel Estimators," *Econometric Theory,* forthcoming.

[7] Johnson, P.A. (2005): "A Continuous State Space Approach to Convergence by Parts," *Economics Letters,* 86, 317–21.

[8] Lindvall, T. (2002): *Lectures on the Coupling Method,* Dover Publications, New York.

[9] Nishimura, K., R. Rudnicki and J. Stachurski (2005): "Stochastic Optimal Growth with Nonconvexities," *Journal of Mathematical Economics*, in press.

[10] Nishimura, K. and J. Stachurski (2005): "Stability of Stochastic Optimal Growth Models: A New Approach," *Journal of Economic Theory*, 122 (1), 100–118.

[11] Sarno, L. and G. Valente (2004): "Comparing the Accuracy of Density Forecasts from Competing Models," *Journal of Forecasting,* 23, 541–547.

[12] Stokey, N. L., R. E. Lucas and E. C. Prescott (1989): *Recursive Methods in Economic Dynamics,* Harvard University Press, Massachusetts.

Department of Economics, University of Melbourne, VIC 3010, Australia, *j.stachurski@econ.unimelb.edu.au*