Proceedings of the 1998 Winter Simulation Conference
D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

ESTIMATION OF STATIONARY DENSITIES FOR MARKOV CHAINS

Peter W. Glynn

Department of Engineering-Economic Systems
and Operations Research
Stanford University
Stanford, CA 94305-4023, U.S.A.

Shane G. Henderson

Department of Engineering Science
University of Auckland
Private Bag 92019
Auckland, NEW ZEALAND

ABSTRACT

We describe a new estimator of the stationary density of a Markov chain on general state space. The new estimator is easier to compute, converges faster, and empirically gives visually superior estimates than more standard estimators such as nearest-neighbour or kernel density estimators.

1 INTRODUCTION

Visualization is becoming increasingly popular as a means of enhancing one's understanding of a stochastic system. In particular, rather than just reporting the mean of a distribution, one often finds that more useful conclusions may be drawn by seeing the *density* of the underlying random variable (see, e.g., Kelton 1997).

We will consider the problem of computing the stationary density of a Markov chain. For chains on a finite state space, this amounts to computing or estimating a finite number of stationary probabilities, and standard methods may be applied easily in this case. When the chain evolves on a general state space, however, the problem is not so straight-forward. General state-space Markov chains arise naturally in the simulation of discrete-event systems (Henderson and Glynn 1998). As a simple example, consider the customer waiting time (in the queue) in the single-server queue with traffic intensity $\rho < 1$ (see Section 4). The sequence of customer waiting times forms a Markov chain that evolves on the state space $[0, \infty)$. More generally, many discrete-event systems may be described by a generalized semi-Markov process, and such processes can be viewed as Markov chains on a general state space (see, e.g., Henderson and Glynn 1998).

In this paper we introduce a new methodology for stationary density estimation. For a general overview of density estimation from i.i.d. observations, see Prakasa Rao (1983), Devroye (1985) or Devroye (1987). Yakowitz (1985), (1989) has considered the stationary density estimation problem for Markov chains on state space $S \subset \mathbf{R}^d$ where the stationary distribution has a density with respect to Lebesgue measure. He showed that under certain conditions, the kernel density estimator at any point x is asymptotically normally distributed with error proportional to $(nh_n^d)^{-1/2}$, where h_n is the so-called "bandwidth", and n is the simulation runlength. One of the conditions needed to establish this result is that $h_n \to 0$ as $n \to \infty$. Hence, the rate of convergence for kernel density estimators is typically strictly slower than $n^{-1/2}$, and depends on the dimension d. In contrast, the estimator we propose converges at rate $n^{-1/2}$ independent of the dimension d.

In fact, the estimator that we propose has several appealing features.

1. It is relatively easy to compute (compared, say, to nearest-neighbour or kernel density estimators).
2. No tuning parameters need to be selected (unlike the "bandwidth" for kernel density estimators, for example).
3. Well-established steady-state simulation output analysis techniques may be applied to analyze the estimator.
4. The error in the estimator converges to 0 at rate $n^{-1/2}$ *independent of the dimension of the state space*, where n is the simulation runlength.
5. The estimator is likely to give smoother estimators than the standard kernel density estimator; see Section 4. In fact, under fairly weak assumptions, it can be shown that the estimator is k-times differentiable (Henderson and Glynn 1998b).
6. The estimator can be used to obtain a new quantile estimator. The variance estimator for the corresponding quantile estimator has a rigorous convergence theory, and converges at rate $n^{-1/2}$ (Henderson and Glynn 1998b).

7. Empirically, the estimator yields superior representations of stationary densities compared with other methods (see Section 4).

To ensure the existence of the stationary distribution, we make the following assumption. This assumption is certainly not confining. In particular, Glynn (1994) showed that in order for a steady-state simulation to be "well-posed" in a certain precise sense, assumption **A** is both necessary and sufficient. In the special case where S is discrete, assumption **A** corresponds with the usual definition of positive recurrence (with a single closed communicating class and no transient states).

Assumption A: The process $X = (X_n : n \geq 0)$ is a positive Harris recurrent Markov chain on state space S, with stationary distribution π.

Let

$$P^k(x, \cdot) = P(X_k \in \cdot | X_0 = x)$$

be the k-step transition kernel of X. The look-ahead density estimator, as we call it, relies on the following assumption.

Assumption B: For some $m > 0$, and some reference measure μ,

$$P^m(x, dy) = p(x, y)\mu(dy),$$

where $p(x, y)$ is explicitly known.

Typically, $m = 1$, but this is not necessary. The case where μ is Lebesgue measure, and S is a subset of $\mathbf{R}^d$ is perhaps most familiar, and is the context where kernel and nearest-neighbour density estimators are most easily applied. Note however, that in the definition there are no restrictions on μ, so that assumption **B** does not restrict us to this context. The example in Section 4 shows that this apparent subtlety can in fact be very useful.

Assumption **B** is critical to our approach. Therefore, our methodology cannot be applied to estimate the common density of i.i.d. r.v.'s. To see why, note that in the i.i.d. case $p(x, y) = q(y)$ for all x, and q is the unknown density. But we require explicit knowledge of p, and hence q.

The remainder of this paper is organized as follows. In Section 2 we introduce the look-ahead density estimator and derive some of its pointwise convergence properties. Then, in Section 3 we consider global convergence issues. Finally, Section 4 is devoted to a comparison of the look-ahead density estimator with a kernel density estimator through a numerical example.

2 THE LOOK-AHEAD ESTIMATOR

Before introducing the look-ahead density estimator, we prove some preliminary results.

Lemma 1 *If* **A** *and* **B** *hold, then* $\pi \ll \mu$, *i.e.,* π *is absolutely continuous with respect to* μ.

Proof: Suppose $\mu(B) = 0$. Then $P^m(x, B) = 0$ for all $x \in S$. Hence

$$\pi(B) = \int_S P^m(x, B)\pi(dx) = 0.$$

□

Hence, the Radon-Nikodym theorem implies that the stationary distribution π has a density with respect to μ, i.e., that $\pi(dx) = \pi(x)\mu(dx)$. (Note that we are using π to represent both the stationary distribution and its density with respect to μ. The appropriate interpretation should be clear from the context.) Our next result sets the scene for the definition of the look-ahead density estimator.

Lemma 2 *If* **A** *and* **B** *hold, then* $\int_S \pi(dx)p(x, \cdot)$ *is a version of the stationary density.*

Proof: Observe that for all measurable sets B (measurable with respect to the sigma-field on which μ is defined),

$$\begin{aligned} \int_B \pi(y)\mu(dy) &= \pi(B) \\ &= \int_S \pi(dx)P^m(x, B) \\ &= \int_S \pi(x) \int_B p(x, y)\mu(dy)\mu(dx) \\ &= \int_B \int_S \pi(x)p(x, y)\mu(dx)\mu(dy). \end{aligned}$$

Since this holds for all measurable sets B, it follows that

$$\pi(y) = \int_S \pi(x)p(x, y)\mu(dx) \quad \mu \text{ a.e.}$$

Hence

$$\int_S \pi(dx)p(x, y) \tag{1}$$

is a version of the stationary density. □

Remark In the case that both $\pi(\cdot)$ and $\int_S \pi(x)p(x, \cdot)$ are continuous, we can generally expect equality everywhere, and not just μ a.e. For example, this will hold when μ is Lebesgue measure on the real line.

To avoid having to state pointwise convergence properties as holding for μ-almost all y, we will henceforth *define* $\pi(y)$ to be equal to (1), even when this is infinite valued. This new definition changes $\pi(y)$ on a set of μ measure 0, and is merely a theoretical convenience. Practically speaking, this makes no difference to the estimator.

The above result suggests that we might estimate $\pi(y)$ via the look-ahead estimator

$$\pi_n(y) \triangleq \frac{1}{n} \sum_{k=0}^{n-1} p(X_k, y). \tag{2}$$

Our next result reveals certain pointwise convergence properties of this estimator, including the rate of convergence. In preparation for this result, note that Harris recurrent Markov chains are known to contain regenerative structure (see, e.g., Asmussen 1987, or Henderson and Glynn 1998). So suppose that the chain X is classically regenerative, i.e., there exist stopping times $0 = T(-1) \leq T(0) < T(1) < T(2) < \cdots$ such that the random elements

$$\tilde{X}(i) = (X_{T(i-1)+k} : 0 \leq k < T(i) - T(i-1))$$

are independent for $i \geq 0$, and identically distributed for $i \geq 1$. Now define, for a real-valued cost function $f : S \to \mathbb{R}$ and all $k \geq 1$, the random variables

$$\begin{aligned} Y_k(f) &= \sum_{j=T(k-1)}^{T(k)-1} f(X_j) \text{ and} \\ \tau_k &= T(k) - T(k-1), \end{aligned}$$

corresponding to the "cost" accumulated in regenerative cycle k, and the length of the kth cycle.

Theorem 3 *Suppose that assumptions* **A** *and* **B** *hold.*

1. *The estimator $\pi_n(y)$ is almost surely convergent, i.e., for all $y \in S$,*

   $$\pi_n(y) \to \pi(y) \quad a.s.,$$

   as $n \to \infty$.

2. *In addition, if X is classically regenerative, and $E(Y_1(p(\cdot, y))^2 + \tau_1^2) < \infty$, then*

   $$\sqrt{n}(\pi_n(y) - \pi(y)) \Rightarrow \sigma_y N(0,1), \qquad (3)$$

   where $N(0,1)$ is a standard normal r.v., $\Rightarrow$ denotes weak convergence, and

   $$\sigma_y^2 = \frac{EY_1(p(\cdot, y) - \pi(y))^2}{E\tau_1}.$$

Proof: Almost sure convergence (Part 1) follows by the strong law for Harris chains (Proposition 3.7, Asmussen (1987), p. 154). The CLT follows directly from Theorem 3.2 of Asmussen (1987), p. 136. □

Theorem 3 shows that in general, the look-ahead estimator is almost surely convergent, that the error in the estimator is normally distributed, and that the error decreases at rate $n^{-1/2}$. This rate of convergence of the error is an improvement on the rate $(nh_n^d)^{-1/2}$ obtained by Yakowitz (1989) for kernel density estimators.

To conclude this section, we discuss a generalization of the look-ahead estimator that should be of great use in applications. In the introduction we mentioned that general discrete-event simulations may be viewed as general state-space Markov chains. This is done by appending additional information to the state-space until the resulting process is Markov (see, e.g., Henderson and Glynn 1998). In such contexts, it is likely that one will only be interested in the stationary density of some functional of the process.

For example, suppose we are estimating the density of the steady-state waiting time in a single-server queue, where the arrival process is a superposition of two renewal processes. The sequence of customer waiting times is not necessarily Markov, unless we adjoin the times till the next arrival in each arrival renewal process to the state space $[0, \infty)$. But we are primarily interested in the stationary distribution of the customer waiting times, and not the r.v.'s associated with event times.

To handle this situation, let f be a function defined on the state-space S that identifies the variables of interest. In the above example, $f(w, t_1, t_2) = w$, where t_1 and t_2 are the times till the next arrivals in the 2 renewal arrival processes. Next, we replace assumption **B** with

Assumption B$'$ For some $m > 0$, and some underlying measure μ,

$$P(f(X_m) \in dy) = p(x, y)\mu(dy).$$

Note that if $f : S \to T$, for some space T, then $p : S \times T \to \mathbb{R}$ and μ is a measure on the space T. After redefining π to be the stationary distribution of $f(X_0)$, all the results of this section go through with the obvious modifications. The results of the following section also hold with similar modifications.

3 GLOBAL PROPERTIES

In the previous section we discussed the *pointwise* convergence properties of the look-ahead density estimator. In this section we turn to the estimator's *global* convergence properties. In particular, we show that π_n converges in L_1 to π (Theorem 4), and furthermore, that under reasonable conditions, $\pi_n(y)$ converges to $\pi(y)$ *uniformly* in y (Theorem 5). Convergence in L_1 ensures that for a given simulation runlength n, errors of a given size in $\pi_n(y)$ can only occur on a small (with respect to μ) set. Uniform convergence is especially important in a visualization context. If one can guarantee that the error in the estimator $\pi_n(y)$ is uniformly small in y, then graphs of π_n will be "close" to π.

Theorem 4 *Suppose that assumptions* **A** *and* **B** *hold. Then $\pi_n \to \pi$ in L_1, i.e.,*

$$\int_S |\pi_n(y) - \pi(y)|\mu(dy) \to 0$$

as $n \to \infty$. This convergence also holds in expectation, i.e.,

$$E \int_S |\pi_n(y) - \pi(y)|\mu(dy) \to 0$$

as $n \to \infty$.

Proof: First note that

$$0 \le |\pi_n(y) - \pi(y)| \le \pi_n(y) + \pi(y), \tag{4}$$

and $\pi_n(y) + \pi(y)$ integrates to the limit in the sense that

$$\begin{aligned} \lim_{n\to\infty} \int_S \pi_n(y) &+ \pi(y)\,\mu(dy) \\ &= 2 \\ &= \int_S 2\pi(y)\mu(dy) \\ &= \int_S \lim_{n\to\infty} (\pi_n(y) + \pi(y))\mu(dy). \end{aligned} \tag{5}$$

Hence (see Exercise 16.6, p. 222, Billingsley 1986), it follows from (4) and (5) that

$$\begin{aligned} \lim_{n\to\infty} \int_S |\pi_n(y) &- \pi(y)|\,\mu(dy) \\ &= \int_S \lim_{n\to\infty} |\pi_n(y) - \pi(y)|\,\mu(dy) \\ &= 0 \quad \text{a.s.} \end{aligned}$$

Finally, since $\int |\pi_n(y) - \pi(y)|\,\mu(dy) \le 2$ for all n, the second result follows by dominated convergence. □

The following theorem gives simple sufficient conditions for uniform convergence when the state space is compact. For the proof, see Henderson and Glynn (1998b).

Theorem 5 *Suppose that $S \subset \boldsymbol{R}^d$ is compact, and assumptions* **A** *and* **B** *hold. Suppose further that $p : S \times S \to [0,\infty)$ is continuous on $S \times S$. Then π_n converges uniformly to π on S, i.e.,*

$$\sup_{y\in S} |\pi_n(y) - \pi(y)| \to 0 \quad a.s.$$

as $n \to \infty$.

4 AN EXAMPLE

In this section, we compare the look-ahead density estimator with a kernel density estimator through a numerical example. Specifically, we estimate the (known) stationary distribution of the customer waiting time sequence in the M/M/1 queue. In this system, customer n ($n \ge 1$) arrives at time T_n, and requires a service time V_n. The sequence of service times $(V_n : n \ge 0)$ is i.i.d., and each V_i is exponentially distributed with mean μ^{-1}. The sequence of interarrival times $(U_n : n \ge 1)$ is i.i.d., where $U_n = T_n - T_{n-1}$ is exponentially distributed with mean λ^{-1} (note that we define $T_0 = 0$). The interarrival and service time sequences are independent. Let W_n denote the waiting time excluding service of the nth customer. Then it is well known that the sequence $W = (W_n : n \ge 0)$ satisfies the Lindley recursion

$$W_{n+1} = [W_n + X_{n+1}]^+,$$

where $X_n \triangleq V_{n-1} - U_n$ for $n \ge 1$, and for $x \in \mathbf{R}$, $[x]^+ \triangleq \max\{x, 0\}$. Observe that $(X_n : n \ge 1)$ is i.i.d., so that W is a Markov chain on $S = [0,\infty)$. If $\rho \triangleq \lambda/\mu < 1$, then W is positive Harris recurrent (see Asmussen 1987, p. 181).

The stationary distribution of W is a mixture of a unit mass at 0, and an exponentially distributed r.v. with mean $(\mu - \lambda)^{-1}$. Thus, it is convenient to take $\mu(dx) = \delta_0(dx) + I(x \ge 0)dx$, where δ_0 is the probability measure that assigns a mass of 1 to the origin. The stationary distribution then has density π with respect to μ, where $\pi(0) = 1 - \rho$, and for $x > 0$,

$$\pi(x) = \rho(\mu - \lambda)e^{-(\mu-\lambda)x}.$$

To define the look-ahead estimator we need to define $p(x,\cdot)$, the transition density of W with respect to μ. It is straightforward to show that $p(x, 0) = (1 + \rho)^{-1}e^{-\lambda x}$, and for $y > 0$,

$$p(x, y) = \frac{\lambda}{1+\rho} \exp(-\mu[y - x]^+ - \lambda[x - y]^+).$$

The look-ahead density estimator at y is then given by

$$\pi_L(y; n) \triangleq n^{-1} \sum_{k=0}^{n-1} p(W_k, y).$$

Defining the kernel density estimator is slightly more problematical, due to the presence of the point mass at 0 in the stationary distribution, and to the need to select a kernel and bandwidth. To estimate the point mass at 0 we use

$$\pi_K(0; n) \triangleq n^{-1} \sum_{k=0}^{n-1} I(W_k = 0),$$

the mean number of visits to 0 in a run of length n. For $y > 0$, we estimate $\pi(y)$ using

$$\pi_K(y; n) \triangleq (nh_n)^{-1} \sum_{k=0}^{n-1} I(W_k > 0)\varphi((y - W_k)/h_n),$$

where

$$\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

is the density of a standard normal r.v., and $h_n = n^{-1/5}$. This choice of h_n (modulo a multiplicative constant) yields the optimal rate of mean-square convergence in the case where the observations are i.i.d. (Prakasa Rao 1983, p. 182), and so it seems a reasonable choice in this context.

For this example we chose $\lambda = 0.5$ and $\mu = 1$, so that $\rho = 0.5$. To remove the effect of initialization bias (note that both estimators are affected by this), we simulated a stationary version of W by sampling W_0 from the stationary distribution.

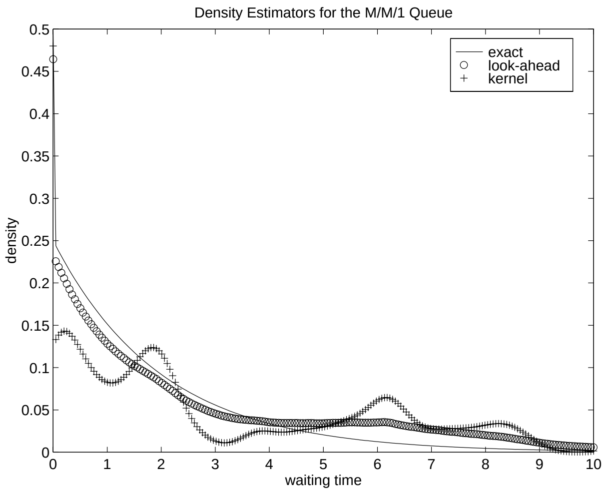

Figure 1: Density estimates from a run of length 100.

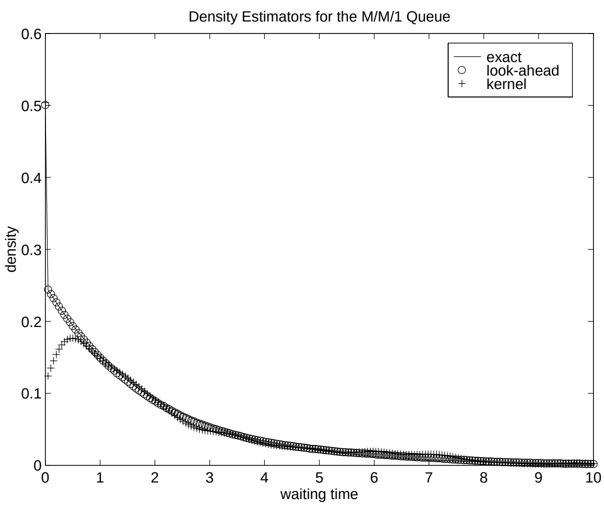

Figure 2: Density estimates from a run of length 1000.

The density estimates for $x > 0$ together with the exact density are plotted for simulation runlengths of $n = 100$ (Figure 1) and $n = 1000$ (Figure 2). We observe the following from these figures.

1. For both $n = 100$, and $n = 1000$, the look-ahead density appears to closely match the exact density.

2. For both $n = 100$, and $n = 1000$, the kernel density estimator appears to perform better away from the origin. It's performance near the origin is particularly poor.

3. Overall, the look-ahead density appears to be a far better representation of the exact density than the kernel density.

4. The look-ahead density appears to be monotone for both runlengths.

5. The kernel density has several local modes, as is strikingly evident in Figure 1.

The poor performance of the kernel density estimator relative to the look-ahead density estimator was, perhaps, expected, owing to the different convergence rates discussed in Section 2. Particularly notable is the poor performance of the kernel density estimator near the boundary. With increasingly long runlengths, and smaller bandwidths, this effect may be expected to diminish, but progress is slow as exemplified in Figure 2 for a run of length 1000.

We anticipate that similar behaviour can be expected on far more general systems, and therefore conclude that when applicable (i.e., when assumptions **A** and **B** are satisfied), the look-ahead approach discussed in this paper should be the preferred methodology for density estimation.

ACKNOWLEDGMENTS

The work of the first author was partially supported by NSF Grant DMS 9704732.

REFERENCES

Asmussen, S. 1987. *Applied Probability and Queues.* Wiley, New York.

Billingsley, P. 1986. *Probability and Measure, 2nd ed.* Wiley, New York.

Devroye, L. 1985. *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

Devroye, L. 1987. *A Course in Density Estimation.* Birkhauser, Boston.

Glynn, P. W. 1994. Some topics in regenerative steady-state simulation. *Acta Applicandae Mathematicae* 34: 225–236.

Henderson, S. G. and P. W. Glynn. 1998. Regenerative steady-state simulation of discrete-event systems. *In Preparation.*

Henderson, S. G. and P. W. Glynn. 1998b. The look-ahead estimator for estimating stationary densities of Markov chains. *In Preparation.*

Kelton, W. D. 1997. Statistical analysis of simulation output. *Proceedings of the 1997 Winter Simulation Conference.* S. Andradóttir, K. J. Healy, D. H. Withers and B. L. Nelson, eds. Institute of Electrical and Electronic Engineers., Piscataway, New Jersey.

Prakasa Rao, B. L. S. 1983. *Nonparametric Functional Estimation.* Academic Press.

Yakowitz, S. 1985. Nonparametric density estimation, prediction and regression for Markov sequences. *Journal of the American Statistical Association* 80: 215–221.

Yakowitz, S. 1989. Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *Journal of Multivariate Analysis* 30: 124–136.

AUTHOR BIOGRAPHIES

SHANE G. HENDERSON joined the Industrial and Operations Engineering Department at the University of Michigan (Ann Arbor) in 1997 after completing his Ph.D. in Operations Research at Stanford University. He is currently a lecturer in the Department of Engineering Science at the University of Auckland, while on leave from the University of Michigan. His research interests include discrete-event simulation, queueing theory and rostering problems.

PETER W. GLYNN received his Ph.D. from Stanford University, after which he joined the faculty of the Department of Industrial Engineering at the University of Wisconsin-Madison. In 1987, he returned to Stanford, where he currently holds the Thomas Ford Chair in the Department of Engineering-Economic Systems and Operations Research. He was a co-winner of the 1993 Outstanding Simulation Publication Award sponsored by the TIMS College on Simulation. His research interests include discrete-event simulation, computational probability, queueing, and general theory for stochastic systems.