MATHEMATICS OF OPERATIONS RESEARCH
Vol. 26, No. 2, May 2001, pp. 375–400
Printed in U.S.A.

COMPUTING DENSITIES FOR MARKOV CHAINS VIA SIMULATION

SHANE G. HENDERSON AND PETER W. GLYNN

We introduce a new class of density estimators, termed look-ahead density estimators, for performance measures associated with a Markov chain. Look-ahead density estimators are given for both transient and steady-state quantities. Look-ahead density estimators converge faster (especially in multidimensional problems) and empirically give visually superior results relative to more standard estimators, such as kernel density estimators. Several numerical examples that demonstrate the potential applicability of look-ahead density estimation are given.

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.

We will consider the problem of computing the densities of performance measures associated with a Markov chain. For chains on a finite state-space, this typically amounts to computing or estimating a finite number of probabilities, and standard methods may be applied easily in this case (see below). When the chain evolves on a general state-space, however, the problem is not so straightforward.

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 single-server queue with traffic intensity $\rho < 1$ (see §6). 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). General state-space Markov chains are also prevalent in the theory of control systems; see Meyn and Tweedie (1993, Chapter 2). Another important area where general state-space Markov chains arise is in Markov chain Monte Carlo simulation; see Gilks et al. (1996).

This paper is an outgrowth of, and considerably extends, Glynn and Henderson (1998), in which we introduced 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 $S \subset \mathbb{R}^d$, 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 run-length. 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 (see Remarks 5 and 7). In contrast, the estimator we propose converges at rate $n^{-1/2}$ *independent of the dimension* d.

Received February 9, 1999; revised October 9, 2000, and December 5, 2000.
MSC 2000 subject classification. Primary: 60J22.
OR/MS subject classification. Primary: Probability/Markov processes; secondary: simulation.
Key words. Markov chain, density estimator, simulation.

0364-765X/01/2602/0375/$05.00
1526-5471 electronic ISSN,

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) Under relatively mild assumptions, look-ahead density estimators consistently estimate not only the density itself, but also the derivatives of the density; see Theorem 9.

(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 $n^{-1/2}$ (§5).

(7) Empirically, the estimator yields superior representations of stationary densities compared with other methods (§6, Example 1).

We first introduce the central ideas behind look-ahead density estimation in a familiar context. Although this problem is subsumed by the treatment of §3, a separate development should prove helpful in understanding the look-ahead approach. Let $X = (X(n) : n \geq 0)$ be an irreducible positive recurrent Markov chain on finite state-space S, and $\pi(y)$ be the stationary probability of a point $y \in S$. Our goal is to estimate the stationary "density" $\pi(\cdot)$; in the finite state-space context, the stationary "density" coincides with the stationary probabilities $\pi(y)$, for $y \in S$. To estimate $\pi(y)$, the standard estimator is

$$\tilde{\pi}_n(y) \stackrel{\triangle}{=} \frac{1}{n} \sum_{i=0}^{n-1} I(X(i) = y),$$

where $I(\cdot)$ is the indicator function that is 1 if its argument is true, and 0 otherwise. The estimator $\tilde{\pi}_n(y)$ is simply the proportion of time the Markov chain X spends in the state y.

Notice however, that one could also estimate $\pi(y)$ by

$$\pi_n(y) \stackrel{\triangle}{=} \frac{1}{n} \sum_{i=0}^{n-1} P(X(i), y),$$

where $P(\cdot, \cdot)$ is the transition matrix of X. The estimator $\pi_n(y)$ is a (strongly) consistent estimator of $\pi(y)$ as can be seen by noting that

$$\pi_n(y) \to \sum_{x \in S} \pi(x) P(x, y) = \pi(y)$$

as $n \to \infty$, by the strong law for positive recurrent Markov chains on discrete state-space. Notice that $P(X(i), y) = P(X(i+1) = y | X(i))$, so that the quantity $P(X(i), y)$ is, in effect, "looking ahead" to see whether the next iterate of the Markov chain will equal y. This is the motivation for the name "look-ahead" density estimator.

In this example, we needed explicit knowledge of the transition matrix, and in general, look-ahead estimators require explicit knowledge of the transition kernel. Such knowledge is not required in standard methods such as kernel density estimation, and this observation helps explain the appealing convergence properties of look-ahead density estimators.

In the remainder of this paper we assume a general state-space (not necessarily discrete) unless otherwise specified. We refer to the density we are trying to estimate as the *target density*, and the associated distribution as the *target distribution*.

In §2, look-ahead density estimators are developed for several performance measures associated with transient simulations, and their *pointwise* asymptotic behaviour is derived.

Steady-state performance measures are similarly considered in §3. In §4, we turn to the global convergence behaviour of look-ahead density estimators. In particular, we give conditions under which the look-ahead density estimator converges to the target density in an L^q sense (Theorem 5), is uniformly convergent (Theorem 7), and is differentiable (Theorem 9).

In §5, we consider the computation of several features of the target distribution, including the mode of the target density and quantiles of the target distribution. We present three examples of look-ahead density estimation in §6, and offer some concluding remarks in §7.

2. Computing densities for transient performance measures. Let $X = (X(n) : n \geq 0)$ be a Markov chain taking values in a state-space S. Since our focus in this section is on transient performance measures, we will permit our chain to possess transition probabilities that are nonstationary.

Recall that $Q = (Q(x, dy) : x, y \in S)$ is a transition kernel if $Q(x, \cdot)$ is a probability measure on S for each $x \in S$, and if $Q(\cdot, dy)$ is suitably measurable. (If S is a discrete state space, Q corresponds to a transition matrix.) By permitting X to have nonstationary transition probabilities, we are asserting that there exists a sequence $(P(n) : n \geq 0)$ of transition kernels, such that

$$P(X(n+1) \in dy \,|\, X(j) : 0 \leq j \leq n) = P(n, X(n), dy) \quad \text{a.s.}$$

for $n \geq 0$ and $y \in S$. Our basic assumption is:

ASSUMPTION 1. *There exists a* (σ*-finite*) *measure* γ *on* S *and a function* $p : \mathbb{Z}^+ \times S \times S \to [0, \infty)$, *such that*

$$P(n, x, dy) = p(n, x, y)\gamma(dy)$$

for $n \geq 0$ *and* $x, y \in S$.

REMARK 1. Assumption 1 is automatically satisfied when S is finite or countably infinite.

REMARK 2. Given that this paper is concerned with density estimation, the case where γ is Lebesgue measure and S is a subset of $\mathbb{R}^d$ is of the most interest to us. However, it is important to note that Assumption 1 does not restrict us to this context. In fact, Example 1 in §6 shows that this apparent subtlety can in fact be very useful.

REMARK 3. If X has stationary transition probabilities, then $P(n) = P$ for some transition kernel P and $n \geq 0$. In our discussion of steady-state density estimation (see §3), we will clearly wish to restrict ourselves to such chains.

We will now describe several different computational settings to which the ideas of this paper apply. In what follows, we will adopt the generic notation $p_Z(\cdot)$ to denote the γ-density of the r.v. Z. In other words, $p_Z(\cdot)$ is a function with the property that

$$P(Z \in dy) = p_Z(y)\gamma(dy)$$

for all y in the range of Z. Also, for a given initial distribution μ on S, let $P_\mu(\cdot)$ be the probability distribution on the path-space of X under which X has initial distribution μ.

Problem 1. Compute the density of $X(r)$.

For $r \geq 1$, let $p_{X(r)}(\cdot)$ be the γ-density of $X(r)$. Note that

$$P_\mu(X(r) \in B) = \int_B \int_S P_\mu(X(r-1) \in dx) p(r-1, x, y)\gamma(dy),$$

so that

$$\begin{aligned} p_{X(r)}(y) &= \int_S P_\mu(X(r-1) \in dx) p(r-1, x, y) \\ &= E_\mu p(r-1, X(r-1), y), \end{aligned}$$

where E_μ is the expectation operator corresponding to P_μ. To compute the density $p_{X(r)}(y)$, simulate n i.i.d. replicates $X_1, X_2, \ldots, X_n$ of X under P_μ. Then, Assumption 1 and the strong law of large numbers together guarantee that

$$p_{1n}(y) \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^{n} p(r-1, X_i(r-1), y) \to p_{X(r)}(y) \quad \text{a.s.},$$

as $n \to \infty$, so that $p_{X(r)}(y)$ can indeed be computed by our look-ahead estimator $p_{1n}(y)$.

REMARK 4. Suppose that Assumption 1 is weakened to

$$P(X(n+m) \in dy | X(n) = x) = p(n, x, y)\gamma(dy)$$

for $n \geq 0$, and $x, y \in S$, so that now we are assuming the existence of a density only for the m-step transition probability distribution. Provided $r \geq m$, we can write

$$p_{X(r)}(y) = E_\mu p(r-m, X(r-m), y),$$

so that $p_{X(r)}(y)$ can again be easily computed via independent replication of X.

For a given subset $A \subseteq S$, let $T = \inf\{n \geq 0 : X(n) \in A\}$ be the first entrance time to A.

Problem 2. Compute the density of $X(T)$.

Suppose that $P_\mu(X(0) \in A^c) = 1$, so that X starts in A^c under initial distribution μ. Then, for $B \subseteq A$,

$$P_\mu(X(T) \in B) = \sum_{n=0}^{\infty} \int_B \int_S P_\mu(X(n) \in dx, T > n) p(n, x, y)\gamma(dy),$$

so that for $y \in A$,

$$\begin{aligned} p_{X(T)}(y) &= \sum_{n=0}^{\infty} \int_S P_\mu(X(n) \in dx, T > n) p(n, x, y) \\ &= E_\mu \sum_{n=0}^{T-1} p(n, X(n), y). \end{aligned}$$

Again, Assumption 1 and the strong law of large numbers ensure that

$$p_{2n}(y) \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{T_i-1} p(j, X_i(j), y) \to p_{X(T)}(y) \quad \text{a.s.},$$

as $n \to \infty$, where the X_i's are independent replicates of X under P_μ, and $T_i = \inf\{n \geq 0 : X_i(n) \in A\}$.

An important class of transient performance measures is concerned with cumulative costs. Specifically, let $\Gamma = (\Gamma(n) : n \geq 1)$ be a sequence of real-valued r.v.'s, in which $\Gamma(n)$ may be interpreted as the "cost" associated with running X over $[n-1, n)$. Then,

$$C(n) = \sum_{i=1}^{n} \Gamma(i)$$

is the cumulative cost corresponding to the time interval $[0, n)$. We assume that

$$P_\mu\big(\Gamma(1) \in dy_1, \ldots, \Gamma(n) \in dy_n \mid X\big) = \prod_{i=1}^{n} P_\mu\big(\Gamma(i) \in dy_i \mid X(i-1), X(i)\big) \tag{1}$$

so that, conditional on X, the $\Gamma(i)$'s are independent r.v.'s, and the conditional distribution of $\Gamma(i)$ depends on X only through $X(i-1)$ and $X(i)$. An important special case arises

when $\Gamma(n) = f(X(n-1))$ for $n \geq 1$, for some deterministic function $f : S \to \mathbb{R}$. In this case, (1) is automatically satisfied, and $f(x)$ may be viewed as the cost associated with spending a unit amount of time in $x \in S$. (We permit the additional generality of (1) because such cost structures are a standard ingredient in the general theory of "additive functionals" for Markov chains, and create no difficulties for our theory.)

Before proceeding to a discussion of the cumulative cost $C(n)$, we note that Problems 1 and 2 have natural analogues here. However, we will need to replace Assumption 1 with:

ASSUMPTION 2. *There exists a* (σ*-finite*) *measure* γ *on* S *and a function* $\tilde{p} : \mathbb{Z}^+ \times S \times S \times S \to [0, \infty)$ *such that*

$$P_\mu(\Gamma(n) \in dy \,|X(n-1) = x_{n-1}, X(n) = x_n) = \tilde{p}(n-1, x_{n-1}, x_n, y)\gamma(dy)$$

for $n \geq 1$ *and* $x_{n-1}, x_n, y \in S$.

Problem 3. Compute the density of $\Gamma(r)$.

For $y \in \mathbb{R}$, the density $p_{\Gamma(r)}(y)$ can be consistently estimated by

$$p_{3n}(y) \stackrel{\triangle}{=} \frac{1}{n} \sum_{i=1}^{n} \tilde{p}(r-1, X_i(r-1), X_i(r), y),$$

where $X_1, X_2, \ldots, X_n$ are independent replicates of X.

Problem 4. Compute the density of $\Gamma(T)$.

Here, the density $p_{\Gamma(T)}(y)$ can be consistently estimated via

$$p_{4n}(y) \stackrel{\triangle}{=} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{T_i - 1} \tilde{p}(j, X_i(j), X_i(j+1), y).$$

As usual, $X_1, X_2, \ldots, X_n$ are independent replicates of X under P_μ, and T_i is the first entrance time of X_i to the set A.

In addition to consistency of $p_{3n}(y)$ and $p_{4n}(y)$, Assumption 2 permits us to solve a couple of additional computational problems that relate to the density of the cumulative cost r.v. introduced earlier.

Problem 5. Compute the density of $C(r)$.

We assume here that γ is Lebesgue measure. Then, if $r \geq 1$, we may use Assumption 2 to write

$$\begin{aligned}
P_\mu(C(r) \leq y) &= E_\mu P_\mu(C(r) \leq y|X) \\
&= E_\mu \int_{\mathbb{R}} P_\mu(C(r-1) \in dz|X) P_\mu(\Gamma(r) \leq y - z|X) \\
&= E_\mu \int_{\mathbb{R}} \int_{-\infty}^{y-z} P_\mu(C(r-1) \in dz|X) \tilde{p}(r-1, X(r-1), X(r), u)\, du \\
&= E_\mu \int_{\mathbb{R}} \int_{-\infty}^{y} P_\mu(C(r-1) \in dz|X) \tilde{p}(r-1, X(r-1), X(r), t-z)\, dt \qquad (2) \\
&= \int_{-\infty}^{y} E_\mu \int_{\mathbb{R}} P_\mu(C(r-1) \in dz|X) \tilde{p}(r-1, X(r-1), X(r), t-z)\, dt, \qquad (3)
\end{aligned}$$

where (2) follows from a change of variable $t = z + u$, and (3) follows from Fubini's theorem, since the integrand is nonnegative. Hence, we may conclude that

$$\begin{aligned} P_\mu(C(r) \in dy) &= \left[E_\mu \int_{\mathbb{R}} P_\mu(C(r-1) \in dz|X) \tilde{p}(r-1, X(r-1), X(r), y-z) \right] dy \\ &= \left[E_\mu \int_{\mathbb{R}} P_\mu(C(r-1) \in dz|X) \tilde{p}(r-1, X(r-1), X(r), y-C(r-1)) \right] dy \\ &= [E_\mu \tilde{p}(r-1, X(r-1), X(r), y-C(r-1))]\, dy. \end{aligned}$$

Evidently, Assumption 2 and the strong law of large numbers together guarantee that

$$\begin{aligned} p_{5n}(y) &\stackrel{\triangle}{=} \frac{1}{n} \sum_{i=1}^{n} \tilde{p}(r-1, X_i(r-1), X_i(r), y-C_i(r-1)) \\ &\to p_{C(r)}(y) \quad \text{a.s.}, \end{aligned}$$

as $n \to \infty$, so that $p_{5n}(y)$ is a consistent estimator of $p_{C(r)}(y)$, the (Lebesgue) density of $C(r)$.

Problem 6. Compute the density of $C(T)$.

As we did earlier, we assume that $\gamma(dy) = dy$ and $P_\mu(X(0) \in A^c) = 1$. Similar arguments to those used above establish the identity

$$p_{C(T)}(y) = E_\mu \sum_{j=0}^{T-1} \tilde{p}(j, X(j), X(j+1), y-C(j)).$$

Thus Assumption 2 and the strong law prove that

$$p_{6n}(y) \stackrel{\triangle}{=} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{T-1} \tilde{p}(j, X_i(j), X_i(j+1), y-C_i(j))$$

is a consistent estimator for $p_{C(T)}(y)$.

To this point, we have constructed unbiased density estimators for each of the six density computation problems described above. We now turn to the development of asymptotically valid confidence regions for these densities. The key is to recognize that each of the six estimators may be represented as

$$p_{in}(y) = \frac{1}{n} \sum_{j=1}^{n} \chi_{ij}(y),$$

where $(\chi_{ij}(y) : y \in \Lambda)$ is i.i.d. in $j \geq 1$. (Here, the index set Λ is either S or $\mathbb{R}$, depending on which of the estimators is under consideration.) For $y \in \Lambda$, let $p(y; i) \stackrel{\triangle}{=} E_\mu \chi_{ij}(y)$. For d points, $y_1, \ldots, y_d \in \Lambda$, let $\vec{y} = (y_1, \ldots, y_d)$, and define $\vec{p}_{in}(\vec{y})$ $(\vec{p}(\vec{y}; i))$ to be a d-dimensional vector with jth component $p_{in}(y_j)$ $(p(y_j; i))$. A straightforward application of the multivariate central limit theorem (CLT) (see page 177 of Billingsley 1968) yields the following result.

PROPOSITION 1. *Let $y_1, y_2, \ldots, y_d$ be d points in Λ, selected so that $E_\mu \chi_{ij}(y_k)^2 < \infty$ for $1 \leq k \leq d$, and let $\vec{y} = (y_1, \ldots, y_d)$. Then,*

$$n^{1/2}(\vec{p}_{in}(\vec{y}) - \vec{p}(\vec{y}; i)) \Rightarrow N(0, \Sigma_i(\vec{y})),$$

as $n \to \infty$, where $N(0, \Sigma_i(\vec{y}))$ is a d-dimensional multivariate normal random vector with mean vector zero and covariance matrix $\Sigma_i(\vec{y})$ having (j,k)th element given by $\operatorname{cov}(\chi_{i1}(y_j), \chi_{i1}(y_k))$.

Proposition 1 suggests the approximation,

$$\vec{p}_{in}(\vec{y}) \overset{D}{\approx} \vec{p}(\vec{y}; i) + n^{-1/2}\Sigma_i^{1/2}(\vec{y})N(0, I), \tag{4}$$

for n large, where $\overset{D}{\approx}$ denotes the (nonrigorous) relation "has approximately the same distribution as," and $\Sigma_i^{1/2}(\vec{y})$ is a Cholesky factor of the nonnegative definite symmetric matrix $\Sigma_i(\vec{y})$. (We say that L is a Cholesky factor of the symmetric positive semidefinite matrix A if $LL' = A$, where L' denotes the transpose of L, and L is lower triangular; see §11.5.1, page 322, of Seber 1977.) Since $\Sigma_i(\vec{y})$ is easily estimated consistently from $X_1, \ldots, X_n$ by the sample covariance matrix, it follows that (4) may be used to construct asymptotically valid confidence regions for $\vec{p}(\vec{y}; i)$ in the case where $\Sigma_i(\vec{y})$ is positive definite.

REMARK 5. Equation (4) implies that the error in the look-ahead density estimator decreases at rate $n^{-1/2}$. This dimension-independent rate stands in sharp contrast to the heavily dimension-dependent rate exhibited by other density estimators, including kernel density estimators; see Prakasa Rao (1983). The convergence rate for such estimators is typically $(nh_n^d)^{-1/2}$, where h_n is the bandwidth parameter and d is the dimension of Λ. To minimize mean squared error, the bandwidth h_n is typically chosen to be of the order $n^{-1/(d+4)}$, and then the error in the kernel density estimators decreases at rate $n^{-2/(d+4)}$. Even in one dimension, this asymptotic rate is slower than that exhibited by the look-ahead density estimator, and in higher dimensions, the difference is even more apparent; see Example 2 of §6.

3. Computing densities for steady-state performance measures. We now extend our look-ahead estimation methodology to the steady-state context. In order for the concept of steady-state to be well defined, we assume that X has stationary transition probabilities, so that the transition kernels $(P(n) : n \geq 0)$ introduced in §2 are independent of n. In other words, we assume that there exists a transition kernel P such that $P(n) = P$ for $n \geq 0$. Let $p(x, y) = p(0, x, y)$ for $x, y \in S$, where $p(0, x, y)$ is defined as in Assumption 1.

PROPOSITION 2. *Under Assumption* 1, *any stationary distribution of X possesses a density* π *with respect to* γ.

PROOF. Let π be a stationary distribution of X. (Note that we are using π to represent both the stationary distribution and its density with respect to γ. The appropriate interpretation should be clear from the context.) Then,

$$P_\pi(X(1) \in B) = P_\pi(X(0) \in B) \tag{5}$$

for all (suitably) measurable $B \subseteq S$. But $P_\pi(X(0) \in B) = \pi(B)$. And

$$\begin{aligned} P_\pi(X(1) \in B) &= \int_S \pi(dx) P(X(1) \in B | X(0) = x) \\ &= \int_S \int_B \pi(dx) p(x, y)\, \gamma(dy) \\ &= \int_B \int_S \pi(dx) p(x, y)\, \gamma(dy). \end{aligned} \tag{6}$$

It follows from (5) and (6) that the stationary distribution π has a γ-density $\pi(\cdot)$ having value

$$\pi(y) = \int_S \pi(dx) p(x, y) \tag{7}$$

at $y \in S$. □

According to Proposition 2, the density $\pi(y)$ may be expressed as an expectation, namely

$$\pi(y) = E_\pi p(X(0), y); \tag{8}$$

see (7). Relation (8) suggests using the estimator,

$$\pi_n(y) = \frac{1}{n}\sum_{i=0}^{n-1} p(X(i), y),$$

to compute $\pi(y)$; $\pi_n(y)$ requires simulating X up to time $n-1$.

To establish laws of large numbers and CLT's for $\pi_n(y)$, we require that X be positive recurrent in a suitable sense.

ASSUMPTION 3. *Assume that there exists a subset $B \subseteq S$, positive scalars λ, a and b, an integer $m \geq 1$, a probability distribution $\varphi(\cdot)$ on S, and a (deterministic) function $V : S \to [1, \infty)$, such that*

(1) $P(X(m) \in \cdot \,| X(0) = x) \geq \lambda\varphi(\cdot)$, $x \in B$, *and*

(2) $E[V(X(1))|X(0) = x] \leq (1-a)V(x) + bI(x \in B)$, $x \in S$,

where $I(x \in B)$ is 1 or 0 depending on whether or not $x \in B$.

In the language of general state-space Markov chain theory, Assumption 3 ensures that X is a geometrically ergodic Harris recurrent Markov chain; see Meyn and Tweedie (1993) for details. Condition 1 of Assumption 3 is typically satisfied for reasonably behaved Markov chains by choosing B to be a compact set; λ, φ, and m are then determined so that Condition 1 is satisfied. Condition 2 is known as a Lyapunov function condition. For many chains, a potential choice for V is something of the form $V(x) = \exp(\alpha\|x\|)$ for $\alpha > 0$; for others, a great deal of ingenuity may be necessary in order to construct such a V. See Example 1 of §6 for an illustration of the verification of Assumption 3. In any case, Assumption 3 ensures that X possesses a unique stationary distribution.

REMARK 6. Assumption 3 is a stronger condition than is necessary to obtain the laws of large numbers and CLT's below. However, in most applications, Assumption 3 is a particularly straightforward sufficient condition to verify, and we offer it in that spirit.

Let $\vec{y} = (y_1, \ldots, y_d)$ consist of d points selected from S, and let $\vec{\pi}_n(\vec{y})$ ($\vec{\pi}(\vec{y})$) be a d-dimensional vector in which the jth component is $\pi_n(y_j)$ ($\pi(y_j)$).

THEOREM 3. *Assume Assumption 1, Assumption 3, and suppose that for $1 \leq i \leq d$, $p(\cdot, y_i) \leq V^{1/2}(\cdot)$. Then,*

$$\vec{\pi}_n(\vec{y}) \to \vec{\pi}(\vec{y}) \quad a.s.,$$

as $n \to \infty$. Also, there exists a nonnegative definite symmetric matrix $\Sigma = \Sigma(\vec{y})$, such that

$$n^{1/2} t\big(\vec{\pi}_{\lfloor nt \rfloor}(\vec{y}) - \vec{\pi}(\vec{y})\big) \Rightarrow \Sigma^{1/2}(\vec{y})B(t),$$

as $n \to \infty$, where $(B(t) : t \geq 0)$ is a d-dimensional standard Brownian motion and $\Rightarrow$ denotes weak convergence in $D[0, \infty)$.

PROOF. The proof follows directly from results from Meyn and Tweedie (1993). The strong law is a consequence of Theorem 17.0.1. Lemma 17.5.1 and Lemma 17.5.2 together imply the existence of a square integrable (with respect to π) solution to Poisson's equation. This then enables an application of Theorem 17.5.4 to yield the result. □

REMARK 7. The remarks on rate of convergence in Remark 5 also apply here. In particular, Theorem 3 gives conditions under which the look-ahead stationary density estimator converges at rate $n^{-1/2}$. Other existing estimators, including kernel density estimators, typically converge at rate $(nh_n^d)^{-1/2}$, where h_n is the bandwidth parameter, and d is the dimension of the (Euclidian) state space (Yakowitz 1989). Since $h_n \to 0$ as $n \to \infty$ this convergence rate is slower than $n^{-1/2}$.

Yakowitz (1989) does not give the optimal (in terms of minimizing mean squared error) choice of bandwidth h_n. However, an i.i.d. sequence is a special case of a Markov chain, and, as noted in Remark 5, the fastest possible root mean square error convergence rate in that setting is of the order $n^{-2/d+4}$. This rate is heavily dimension dependent, so that in large-dimensional problems, one might expect very slow convergence of kernel density estimators.

To obtain confidence regions for the density vector $\vec{\pi}(\vec{y})$, several different approaches are possible. If $\Sigma(\vec{y})$ is positive definite, and there exists a consistent estimator $\Sigma_n(\vec{y})$ for $\Sigma(\vec{y})$, then Theorem 3 asserts that for $D \subseteq \mathbb{R}^d$,

$$P_\mu\left(\vec{\pi}(\vec{y}) \in \vec{\pi}_n(\vec{y}) - \Sigma_n^{1/2}(\vec{y})D\right) \approx P(N(0, I) \in n^{1/2}D), \tag{9}$$

for large n, where $\Sigma_n^{1/2}(\vec{y})D$ is defined to be the set $\{x : x = \Sigma_n^{1/2}(\vec{y})\vec{w}$ for some $\vec{w} \in D\}$ etc. Approximate confidence regions for $\vec{\pi}(\vec{y})$ can then be easily obtained from (9). If X enjoys regenerative structure, the regenerative method for steady-state simulation output analysis provides one means of constructing such consistent estimators for $\Sigma(\vec{y})$; see, for example, Bratley et al. (1987).

An alternative approach exploits the functional CLT provided by Theorem 3 to ensure the asymptotic validity of the method of multivariate batch means; see Muñoz and Glynn (1998) for details.

REMARK 8. The discussion of this section generalizes to the computation of the density of $\Gamma(1)$ under the stationary distribution π. In particular, suppose that X satisfies Assumption 2 and Assumption 3. Then for $y \in \mathbb{R}$,

$$P_\pi(\Gamma(1) \in dy) = E_\pi \tilde{p}(X(0), X(1), y)\, dy,$$

where $\tilde{p}(x_0, x_1, y) \stackrel{\triangle}{=} \tilde{p}(0, x_0, x_1, y)$ for $x_0, x_1, y \in \mathbb{R}$ and $\tilde{p}(0, x_0, x_1, y)$ is defined as in Assumption 2; the methodology of this section then generalizes suitably.

4. Global behaviour of the look-ahead density estimator. In the previous section, we focused on the pointwise convergence properties of the look-ahead density estimator. Specifically, we showed that for any finite collection $y_1, y_2, \ldots, y_d$ of points in either S or $\mathbb{R}$ (depending on the estimator), the look-ahead density estimator converges a.s., and the rate of convergence is described by a CLT in which the rate is dimension independent. In this section, we turn to the estimator's global convergence properties. We assume throughout the remainder of this paper that S is a complete separable metric space. (In particular, this includes any state space that is a "reasonable" subset of $\mathbb{R}^k$.)

Let γ be as in §§2 and 3, and let Λ be either S or $\mathbb{R}$ (depending on the estimator considered). Then, for any function $f : \Lambda \to \mathbb{R}$, we may define, for $q \geq 1$, the L^q-norm

$$\|f\|_q \stackrel{\triangle}{=} \left(\int_\Lambda |f(y)|^q \gamma(dy)\right)^{1/q}.$$

For any two functions f_1 and f_2, $\|f_1 - f_2\|_q$ is a measure of the "distance" from f_1 to f_2. We first analyze the look-ahead density estimators introduced in §2.

THEOREM 4. *Suppose that* $E_\mu \int_\Lambda |\chi_{i1}(y)|^q \gamma(dy) < \infty$ *for* $q \geq 1$. *Then,*

$$\|p_{in}(\cdot) - p(\cdot; i)\|_q \Rightarrow 0$$

as $n \to \infty$.

PROOF. Evidently,

$$E_\mu|\chi_{i1}(y)|^q < \infty, \tag{10}$$

for γ a.e. y. Note that

$$\left|\frac{1}{n}\sum_{j=1}^{n}(\chi_{ij}(y) - p(y;i))\right|^q \le \frac{1}{n}\sum_{j=1}^{n}|\chi_{ij}(y) - p(y;i)|^q, \tag{11}$$

due to the convexity of $|x|^q$. For each y satisfying (10), the right-hand side of (11) converges a.s., and in expectation to $E_\mu|\chi_{i1}(y) - p(y;i)|^q$. Consequently, the right-hand side of (11) is uniformly integrable. Also, for each such y, the left-hand side of (11) converges to zero a.s. Since the left-hand side is dominated by a uniformly integrable sequence, it follows that

$$E_\mu|p_{in}(y) - p(y;i)|^q \to 0, \tag{12}$$

as $n \to \infty$ for γ a.e. y. Also, taking the expectation of both sides of (11) yields the inequality

$$\begin{aligned} E_\mu|p_{in}(y) - p(y;i)| &\le E_\mu|\chi_{i1}(y) - p(y;i)|^q \\ &\le E_\mu \max(|\chi_{i1}(y)|^q, p^q(y;i)) \\ &\le E_\mu(|\chi_{i1}(y)|^q + p^q(y;i)) \\ &= E_\mu(|\chi_{i1}(y)|^q) + |E_\mu\chi_{i1}(y)|^q \\ &\le 2E_\mu(|\chi_{i1}(y)|^q); \end{aligned}$$

the right-hand side is integrable in y, by hypothesis. The Dominated Convergence Theorem, applied to (12), then gives

$$\int_\Lambda E_\mu|p_{in}(y) - p(y;i)|^q \gamma(dy) \to 0,$$

and hence,

$$E_\mu\|p_{in}(\cdot) - p(\cdot;i)\|_q^q \to 0.$$

Consequently,

$$\|p_{in}(\cdot) - p(\cdot;i)\|_q^q \Rightarrow 0,$$

as $n \to \infty$, from which the theorem follows. □

We turn next to obtaining the analogous result for the steady-state density estimator $\pi_n(\cdot)$ of §3.

THEOREM 5. *Suppose*

$$\int_S p(x,y)^q \gamma(dy) \le V(x) \tag{13}$$

for $x \in S$, with $q \ge 1$. If Assumption 3 holds, and the initial distribution μ has a density with respect to γ, then

$$\|\pi_n(\cdot) - \pi(\cdot)\|_q \Rightarrow 0,$$

as $n \to \infty$.

PROOF. Condition (13) guarantees that

$$E_\pi \int_S p(X(0), y)^q \gamma(dy) < \infty;$$

see Theorem 14.3.7 of Meyn and Tweedie (1993). So, $E_\pi p(X(0), y)^q < \infty$ for γ a.e. y, and the proof follows the same pattern as that for Theorem 2. That argument yields the conclusion that for $\epsilon > 0$,

$$P_\pi(\|\pi_n(\cdot) - \pi(\cdot)\|_q > \epsilon) \to 0,$$

as $n \to \infty$. Hence, $P(\|\pi_n(\cdot) - \pi(\cdot)\|_q > \epsilon | X(0) = x) \to 0$ as $n \to \infty$ for π almost every x. Therefore, the Dominated Convergence Theorem allows us to conclude that

$$\begin{aligned} &P_\mu(\|\pi_n(\cdot) - \pi(\cdot)\|_q > \epsilon) \\ &\quad = \int_S \pi(dx) u(x) P\big(\|\pi_n(\cdot) - \pi(\cdot)\|_q > \epsilon | X(0) = x\big) \to 0, \end{aligned}$$

as $n \to \infty$, where $u(\cdot)$ is the π-density of μ. This is the desired conclusion. □

REMARK 9. Note that the hypotheses of both Theorems 4 and 5 are automatically satisfied when $q = 1$.

Convergence of the estimated density in L^q ensures that for a given runlength n, errors of a given size can only occur in a small (with respect to γ) set.

We now turn to the question of when the look-ahead density estimator converges to its limit uniformly. Uniform convergence is especially important in a visualization context. If one can guarantee that the error in the estimator is uniformly small, then graphs of the estimated density will be "close" to the graph of the limit.

We will focus our attention here on the steady-state density estimator π_n; similar results can be derived for our other density estimators through analogous arguments.

THEOREM 6. *Suppose that Assumption* 1 *is in force, and* $p : S \times S \to [0, \infty)$ *is continuous and bounded. If Assumption* 3 *holds, then for each compact set* K,

$$\sup_{y \in K} |\pi_n(y) - \pi(y)| \to 0 \text{ a.s.},$$

as $n \to \infty$.

PROOF. Fix $\epsilon > 0$. Since π is tight (see Billingsley 1968), there exists a compact set $K(\epsilon)$, for which π assigns at most ϵ mass to its complement. Write

$$\begin{aligned} \pi_n(y) = &\frac{1}{n} \sum_{j=1}^n p(X(j), y) I(X(j) \in K(\epsilon)) \\ &+ \frac{1}{n} \sum_{j=1}^n p(X(j), y) I(X(j) \notin K(\epsilon)). \end{aligned} \tag{14}$$

Let $\kappa = \sup\{p(x, y) : x, y \in S\} < \infty$. The second term on the right-hand side of (14) may be bounded by $\kappa n^{-1} \sum_{j=1}^n I(X(j) \notin K(\epsilon))$, which has an a.s. limit supremum of at most $\kappa\epsilon$. As for the first term, note that if K is compact, then $K(\epsilon) \times K$ is compact and p is therefore uniformly continuous there. Because of uniform continuity, there exists $\delta(\epsilon)$,

such that whenever $(x_1, y_1) \in K(\epsilon) \times K$ is within distance $\delta(\epsilon)$ of $(x_2, y_2) \in K(\epsilon) \times K$, $|p(x_1, y_1) - p(x_2, y_2)| < \epsilon$. Since K is compact, we can find a finite collection $y_1, \ldots, y_l$ of points in K such that the open balls of radius $\delta(\epsilon)$ centered at $y_1, \ldots, y_l$ cover K. Then, for each $y \in K$, there exists y_i in our collection such that $|p(X(j), y) - p(X(j), y_i)| < \epsilon$ whenever $X(j) \in K(\epsilon)$. So, for $y \in K$,

$$|\pi_n(y) - \pi_n(y_i)| \leq \epsilon + \frac{\kappa}{n} \sum_{j=1}^{n} I(X(j) \notin K(\epsilon)).$$

Letting $n \to \infty$, we conclude that $|\pi(y) - \pi(y_i)| \leq (\kappa + 1)\epsilon$. Hence, for $y \in K$, we obtain the uniform bound,

$$|\pi_n(y) - \pi(y)| \leq \max_{1 \leq i \leq l} |\pi_n(y_i) - \pi(y_i)| + \epsilon + \frac{\kappa}{n} \sum_{j=1}^{n} I(X(j) \notin K(\epsilon)) + (\kappa + 1)\epsilon.$$

By letting $n \to \infty$, applying the strong law for Harris chains to $n^{-1} \sum_{j=1}^{n} p(X(j), y_i)$ $(1 \leq i \leq l)$, and sending $\epsilon \to 0$, we obtain the desired conclusion. □

REMARK 10. If S is compact, Theorem 6 yields uniform convergence of π_n to π over S, under a continuity hypothesis on p. (The boundedness is automatic in this setting.)

Our next result establishes uniform convergence of π_n to π over all of S, provided that we assume that $p(x, \cdot)$ "vanishes at infinity."

THEOREM 7. *Suppose that Assumption* 1 *holds, and* $p : S \times S \to [0, \infty)$ *is uniformly continuous and bounded. Assume that for each* $x \in S$ *and* $\epsilon > 0$, *there exists a compact set* $K(x, \epsilon)$ *such that whenever* $y \notin K(x, \epsilon)$, $p(x, y) < \epsilon$. *If Assumption* 3 *holds, then*

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

as $n \to \infty$.

PROOF. Fix $\epsilon > 0$, and choose $\delta(\epsilon)$, so that whenever (x_1, y_1) lies within distance $\delta(\epsilon)$ of (x_2, y_2), $|p(x_1, y_1) - p(x_2, y_2)| < \epsilon$. Next, choose $K(\epsilon)$ as in the proof of Theorem 6 and let $x_1, x_2, \ldots, x_r \in K(\epsilon)$ be a finite collection of points such that the open balls of radius $\delta(\epsilon)$ centred at $x_1, \ldots, x_r$ cover $K(\epsilon)$. For each x_i, there exists $K_i(x_i, \epsilon)$, such that $p(x_i, y) < \epsilon$ whenever $y \notin K_i(x_i, \epsilon)$. Put $K = K_1(x_1, \epsilon) \cup \cdots \cup K_r(x_r, \epsilon)$ and note that K is compact. Theorem 6 establishes that

$$\sup_{y \in K} |\pi_n(y) - \pi(y)| \to 0 \quad \text{a.s.}, \tag{15}$$

as $n \to \infty$. To deal with $y \notin K$, construct the sequence $(X'(n) : n \geq 0)$ so that $X'(n) = X(n)$ whenever $X(n) \notin K(\epsilon)$, and $X'(n)$ is the closest point within the collection $\{x_1, \ldots, x_r\}$

whenever $X(n) \in K(\epsilon)$. Then, for $y \notin K$,

$$\begin{aligned}
\pi_n(y) &= \frac{1}{n}\sum_{j=1}^{n} p(X(j), y) I(X(j) \in K(\epsilon)) \\
&\quad + \frac{1}{n}\sum_{j=1}^{n} p(X(j), y) I(X(j) \notin K(\epsilon)) \\
&\leq \left| \frac{1}{n}\sum_{j=1}^{n} (p(X(j), y) - p(X'(j), y)) I(X(j) \in K(\epsilon)) \right| \\
&\quad + \frac{1}{n}\sum_{j=1}^{n} p(X'(j), y) I(X(j) \in K(\epsilon)) \\
&\quad + \frac{\kappa}{n}\sum_{j=1}^{n} I(X(j) \notin K(\epsilon)) \\
&\leq 2\epsilon \frac{1}{n}\sum_{j=1}^{n} I(X(j) \in K(\epsilon)) + \frac{\kappa}{n}\sum_{j=1}^{n} I(X(j) \notin K(\epsilon)). \qquad (16)
\end{aligned}$$

Sending $n \to \infty$ allows us to conclude that $\pi(y) \leq (2+\kappa)\epsilon$ for $y \notin K$. The Inequality (16) then yields

$$\limsup_{n\to\infty} \sup_{y \notin K} |\pi_n(y) - \pi(y)| \leq (4+2\kappa)\epsilon. \qquad (17)$$

Since ϵ was arbitrary, (15) and (17) together imply the theorem. □

The following consequence of Theorem 7 improves Theorem 5 from convergence in probability to a.s. convergence when $q = 1$, and is basically Scheffé's Theorem (see, for example, Serfling 1980, p. 17).

COROLLARY 8. *Under the conditions of Theorem* 7,

$$\int_S |\pi_n(y) - \pi(y)| \gamma(dy) \to 0 \text{ a.s.},$$

as $n \to \infty$.

PROOF. The result is immediate if γ is a finite measure (since $|\pi_n(\cdot) - \pi(\cdot)|$ is uniformly bounded and converges to zero a.s. by Theorem 7. If γ is an infinite measure (like Lebesgue measure), Theorem 7 asserts that $\pi_n(\cdot) \to \pi(\cdot)$ a.s., so that path-by-path, we may argue that

$$\begin{aligned}
\int_S |\pi_n(y) - \pi(y)| \gamma(dy) &= 2\int_S (\pi(y) - \pi_n(y)) I(\pi(y) > \pi_n(y)) \gamma(dy) \\
&\to 0 \quad \text{a.s.}
\end{aligned}$$

(since the integrand is dominated by $\pi(\cdot)$, which integrates to one, thereby permitting the application of the Dominated Convergence Theorem path-by-path). □

A very important characteristic of the look-ahead density estimator is that it "smoothly approximates" the density to be computed. To be specific, suppose that either $S = \mathbb{R}^d$, or that we are considering the density of one of the real-valued r.v.'s associated with the estimators $p_{3n}(\cdot)$, $p_{4n}(\cdot)$, $p_{5n}(\cdot)$ or $p_{6n}(\cdot)$. Since we are then working in a subset of Euclidian space, it is reasonable to measure smoothness in terms of the derivatives of the density.

Without any real loss of generality, assume $d = 1$, so that $y \in \mathbb{R}$. The look-ahead density estimators we have developed take the form

$$p_n(y) \stackrel{\Delta}{=} \frac{1}{n}\sum_{i=1}^{n} G_i(y)$$

for some sequence of random functions $(G_i(\cdot) : i \geq 1)$. (Both the estimators of §2 and §3 admit this representation.) To estimate the kth derivative of the target density to be computed, the natural estimator is therefore

$$\frac{d^k}{dy^k} p_n(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{d^k}{dy^k} G_i(y).$$

Under quite weak conditions on the problem, it can be shown that the above estimator computes the kth derivative of the target density consistently; see below for a discussion. Such a result proves that not only does look-ahead density estimation compute the density, but it also approximates the derivatives of the density in a consistent fashion. In other words, it "smoothly approximates" the target density.

As an illustration of the types of conditions needed in order to ensure that the look-ahead density estimator smoothly approximates the target density, we consider the steady-state density estimator of §3. Let $p'(x, y) = dp(x, y)/dy$.

THEOREM 9. *Suppose Assumption* 1 *holds,* $S = \mathbb{R}$, *and* $p : S \times S \to \mathbb{R}$ *is continuously differentiable with bounded derivative. If Assumption* 3 *holds, then* π *has a differentiable density* $\pi(\cdot)$, *and*

$$\sup_{y \in K} |\pi_n'(y) - \pi'(y)| \to 0 \text{ a.s.}, \tag{18}$$

as $n \to \infty$ *for each compact* $K \subseteq S$. *Furthermore, if* $|p'(x, y)| \leq V^{1/2}(x)$ *for* $x \in S$, *then there exists* $d(y)$ *such that*

$$n^{1/2}(\pi_n'(y) - \pi'(y)) \Rightarrow d(y)N(0, 1), \tag{19}$$

as $n \to \infty$.

PROOF. Note that

$$h^{-1}(\pi(y+h) - \pi(y)) = \int_S \pi(dx)[h^{-1}(p(x, y+h) - p(x, y))]. \tag{20}$$

But $h^{-1}(p(x, y+h) - p(x, y)) = p'(x, \xi)$, where ξ lies between y and $y+h$. Because the derivative is assumed to be bounded, the Bounded Convergence Theorem then ensures that the limit in (20) exists and equals $E_\pi p'(X(0), y)$. Since p' is bounded and continuous, exactly the same argument as that used in proving Theorem 6 can be used here to obtain (18).

The CLT (19) is an immediate consequence of Theorems 17.0.1 and 17.5.4 of Meyn and Tweedie (1993). □

An important implication of Theorem 9 is that the look-ahead density estimator computes the derivative accurately. In fact, the density estimator converges at rate $n^{-1/2}$, independent of the dimension of the state space, and furthermore, independent of the order of the derivative being estimated.

REMARK 11. It is also known that kernel density estimators smoothly approximate the target density; see Prakasa Rao (1983, p. 237), and Scott (1992, p. 131). The choice of bandwidth that minimizes mean squared error of the kernel density derivative estimator is larger than in the case of estimating the target density itself. The resulting rates of convergence of kernel density derivative estimators are adversely affected by both the order of the derivatives, and the dimension of the state space. For example, in one dimension, kernel density derivative estimators of an rth order derivative converge at best at rate $n^{-2/(2r+5)}$; see Scott (1992, p. 132). This rate is fastest when estimating the first derivative, and even then, is slower than the rate of convergence of the look-ahead density derivative estimator discussed above.

5. Computing special features of the target distribution using look-ahead density estimators. As discussed earlier, computation of the density is a useful element in developing visualization tools for computer simulation. In this section, we focus on the computation of certain features of the target distribution to which our look-ahead density estimator can be applied to advantage.

5.1. Computing the relative likelihood of two points. In §§2 and 3, we introduced a number of different look-ahead density estimators, each of which we can write generically as $p_n(\cdot)$. The look-ahead density estimator $p_n(\cdot)$ is an estimator for a target density $p(\cdot)$ say. For each pair of points $(y_1, y_2) \in S \times S$, $p(y_1)/p(y_2)$ represents the likelihood of the point y_1 relative to that of y_2.

The joint CLT's developed in Proposition 1 and Theorem 3 can be used to obtain a CLT (suitable for construction of large-sample confidence intervals for the relative likelihood) for the estimator $p_n(y_1)/p_n(y_2)$. Specifically, if

$$n^{1/2}(p_n(y_1) - p(y_1), p_n(y_2) - p(y_2)) \Rightarrow (N_1, N_2)$$

as $n \to \infty$, where (N_1, N_2) is bivariate Gaussian and $p(y_2) > 0$, then

$$n^{1/2}\left(\frac{p_n(y_1)}{p_n(y_2)} - \frac{p(y_1)}{p(y_2)}\right) \Rightarrow (N_1 - (p(y_1)/p(y_2))N_2)/p(y_2) \tag{21}$$

as $n \to \infty$. If the covariance matrix of (N_1, N_2) can be consistently estimated (as with, for example, the regenerative method), then confidence intervals for the relative likelihood (based on (21)) can easily be obtained. Otherwise, one can turn to the batch means method to produce such confidence intervals; see Muñoz and Glynn (1997).

5.2. Computing the mode of the density. The mode of the density provides information as to the region within which the random variable of interest attains its highest likelihood. Given that the target distribution here has density $p(\cdot)$, our goal is to compute the modal location y^* and the modal value $p(y^*)$. As discussed earlier in this section, we write our look-ahead density estimator generically as $p_n(\cdot)$. The obvious estimator of y^* is, of course, any y_n^* which maximizes $p_n(\cdot)$, and the natural estimator for $p(y^*)$ is then $p_n(y_n^*)$, (We can and will assume that the maximizer y_n^* has been selected to be measurable.) We denote the domain of $p(\cdot)$ by Λ. Because our analysis involves using a Taylor expansion, we require that $\Lambda \subseteq \mathbb{R}^d$.

THEOREM 10. *Suppose that*:

(1) $p(\cdot)$ *has a unique mode at location* y^*;

(2) $\sup_{y\in\Lambda} |p_n(y) - p(y)| \to 0$ *a.s., as* $n \to \infty$;

(3) *there exists an* ϵ*-neighbourhood of* y^* *with* $\epsilon > 0$ *such that* $p(\cdot)$ *and* $p_n(\cdot)$ *are twice continuously differentiable there a.s.*;

(4) $\sup_{\|y-y^*\|<\epsilon} |\nabla p_n(y) - \nabla p(y)| \to 0$ *a.s. as* $n \to \infty$;

(5) $\sup_{\|y-y^*\|<\epsilon} |H_n(y) - H(y)| \to 0$ *a.s. as* $n \to \infty$, *where* $H_n(y)$ *and* $H(y)$ *are the Hessians of* $p_n(\cdot)$ *and* $p(\cdot)$ *at* y, *respectively*;

(6) $H(y^*)$ *is negative definite*;

(7) $n^{1/2}(p_n(y^*) - p(y^*), \nabla p_n(y^*) - \nabla p(y^*)) \Rightarrow (\tilde{N}_1, \tilde{N}_2)$ *as* $n \to \infty$.

Then, $y_n^* \to y^*$ *a.s. as* $n \to \infty$ *and*

$$n^{1/2}(p_n(y_n^*) - p(y^*), y_n^* - y^*) \Rightarrow (\tilde{N}_1, -H(y^*)^{-1}\tilde{N}_2)$$

as $n \to \infty$.

PROOF. The almost sure convergence of y_n^* to y^* is an immediate consequence of Relations 1 and 2. For the weak convergence statement, observe that $\nabla p_n(y_n^*) = 0 = \nabla p(y^*)$, since y_n^* and y^* are local maxima of $p_n(\cdot)$ and $p(\cdot)$, respectively. (To be precise, this is valid only for n so large that y_n^* lies in the ϵ-neighbourhood specified by Relations 3–5.) So,

$$\nabla p_n(y_n^*) - \nabla p_n(y^*) = -(\nabla p_n(y^*) - \nabla p(y^*)).$$

But

$$\nabla p_n(y_n^*) - \nabla p_n(y^*) = H_n(y_n^* - y^*), \tag{22}$$

where $H_n \to H(y^*)$ a.s., as $n \to \infty$. It follows that $n^{1/2}(y_n^* - y^*) \Rightarrow -H(y^*)^{-1}\tilde{N}_2$, as $n \to \infty$. Furthermore,

$$\begin{aligned} n^{1/2}(p_n(y_n^*) - p(y^*)) &= n^{1/2}(p_n(y_n^*) - p_n(y^*)) + n^{1/2}(p_n(y^*) - p(y^*)) \\ &= \nabla p_n(\xi_n) \cdot n^{1/2}(y_n^* - y^*) + n^{1/2}(p_n(y^*) - p(y^*)), \end{aligned} \tag{23}$$

where ξ_n lies on the line segment joining y_n^* and y^*. Since $\nabla p_n(\xi_n) \to \nabla p(y^*) = 0$ a.s., and we have established weak convergence of $n^{1/2}(y_n^* - y^*)$ above, evidently the first term in (23) converges to zero in probability. Consequently, (22), (23), Relation 7, and the "converging together principle" (see Billingsley 1968, for example) imply the desired joint weak convergence result. □

REMARK 12. The uniform convergence theory of §3 for p_n, and its derivatives can be easily applied to verify Relations 2, 4, and 5.

REMARK 13. The CLT established in Theorem 10 shows that the look-ahead estimator of the mode converges at the asymptotic rate $n^{-1/2}$, independent of the dimension d of the state space. This compares very favourably with the rate of convergence of kernel estimators of the mode. A kernel estimator of the mode converges at rate $(nh_n^{d+2})^{-1/2}$ when the bandwidth h_n is chosen appropriately; see Theorem 4.5.6 of Prakasa Rao (1983, p. 284).

REMARK 14. To construct confidence intervals based on Theorem 10, there are again a couple of alternatives. Assume first that for each fixed $y \in \Lambda$, one can consistently estimate the covariance matrix that arises in the joint CLT of Relation 7. (For example, this can be done in the transient context or the setting of regenerative processes in steady-state simulation). To estimate the covariance structure at y^*, one can compute the corresponding covariance estimate evaluated at the point $y = y_n^*$. In the transient problems considered in §2, it is typically easy to verify that the covariance matrix is continuous in y, so that using the estimator associated with y_n^* in place of the covariance at y^* is asymptotically valid. In the steady-state context, it is not as straightforward to theoretically establish the continuity of the covariance, although one suspects it is valid in great generality; one potential avenue is to adopt the methods of Glynn and L'Ecuyer (1995). Henderson and Glynn (1999) give a proof in the univariate case. If consistent estimates of the covariance matrix at each fixed $y \in \Lambda$ are not available (as might occur in nonregenerative steady-state simulations), then one can potentially appeal to the method of batch means; see Muñoz (1998).

5.3. Computing quantiles of the target distribution. We focus here on the special case in which $\Lambda \subseteq \mathbb{R}$, so that the target distribution is that of a real-valued r.v. In this setting, suppose that $\gamma(dy) = dy$, and let

$$F(x) = \int_{-\infty}^{x} p(y)\, dy$$

be the target distribution. An important special feature of this distribution is the pth quantile of F. Specifically, for each $p \in (0, 1)$, we define the pth quantile of F as the quantity $q = F^{-1}(p)$.

There is a significant literature on the computation of such quantiles. Iglehart (1976) considered quantile estimation in the context of regenerative simulation. Seila (1982) introduced the batch quantile method, again for regenerative processes, that avoids some of the difficulties associated with the estimation procedure proposed by Iglehart (1976). The approach suggested by Heidelberger and Lewis (1984) is based on the so-called "maximum transformation" and mixing assumptions of the underlying process, and does not require regenerative structure. Hesterberg and Nelson (1998) and the references therein discuss the use of control variates to obtain variance reduction in quantile estimation. Avramidis and Wilson (1998) obtain variance reduction in estimating quantiles through the use of antithetic variates and Latin hypercube sampling.

Kappenman (1987) integrated and inverted a kernel density estimator for $p(\cdot)$ in the case when the observations are i.i.d. Our approach is similar to Kappenman's, in that we invert the integrated look-ahead density estimator. Let $p_n(\cdot)$ be the look-ahead density estimator, and set

$$F_n(x) = \int_{-\infty}^{x} p_n(y)\, dy.$$

The natural estimator for the quantile q is then $Q_n = F_n^{-1}(p)$.

THEOREM 11. *Suppose that*
(1) $p(q) > 0$;
(2) $p(\cdot)$ *is continuous in an* ϵ *neighbourhood of* q;
(3) $\sup_{|y-q|<\epsilon} |p_n(y) - p(y)| \to 0$ *a.s. as* $n \to \infty$;
(4) $n^{1/2}(F_n(q) - F(q)) \Rightarrow N$, *as* $n \to \infty$.
Then, $n^{1/2}(Q_n - q) \Rightarrow -N/p(q)$ *as* $n \to \infty$.

PROOF. Recall that $\|p_n(\cdot) - p(\cdot)\|_1 \Rightarrow 0$ as $n \to \infty$; see Remark 9. Hence,

$$\sup_x |F_n(x) - F(x)| \le \int_{-\infty}^{\infty} |p_n(y) - p(y)|\, dy \Rightarrow 0,$$

as $n \to \infty$. Now,

$$\begin{aligned} |F(Q_n) - F(q)| &\le |F(Q_n) - F_n(Q_n)| + |F_n(Q_n) - F(q)| \\ &= |F(Q_n) - F_n(Q_n)| \\ &\le \sup_x |F_n(x) - F(x)|, \end{aligned}$$

so that $F(Q_n) \Rightarrow F(q)$ as $n \to \infty$. Our assumptions imply that F^{-1} is continuous in a neighbourhood of $p = F(q)$. Hence, by the continuous mapping theorem (Theorem 5.1, Billingsley 1968),

$$F^{-1}(F(Q_n)) \Rightarrow F^{-1}(F(q)),$$

i.e., $Q_n \Rightarrow q$ as $n \to \infty$. Now

$$F_n(Q_n) - F_n(q) = p - F_n(q) = F(q) - F_n(q), \tag{24}$$

and

$$F_n(Q_n) - F_n(q) = p_n(\xi_n)(Q_n - q), \tag{25}$$

where ξ_n lies between Q_n and q. The result will follow from (24), (25), and Condition (4) of Theorem 11 once we establish that $p_n(\xi_n) \Rightarrow p(q)$.

Note that $\xi_n \Rightarrow q$, and

$$|p_n(\xi_n) - p(q)| \le |p_n(\xi_n) - p(\xi_n)| + |p(\xi_n) - p(q)|.$$

Condition (2) and the continuous mapping theorem together yield $|p(\xi_n) - p(q)| \Rightarrow 0$ as $n \to \infty$. Finally, let $\varepsilon > 0$ be arbitrary, and let B_n denote the event $\{|p_n(\xi_n) - p(\xi_n)| > \varepsilon\}$ (for notational convenience, we suppress the dependence of the event B_n on ε). Then Condition (3) and the fact that $\xi_n \Rightarrow q$, give

$$\begin{aligned}\limsup_{n\to\infty} P(B_n) &= \limsup_{n\to\infty}\Big[P(B_n|\,|\xi_n - q| < \epsilon)P(|\xi_n - q| < \epsilon)\\ &\qquad + P(B_n|\,|\xi_n - q| \geq \epsilon)P(|\xi_n - q| \geq \epsilon)\Big]\\ &\leq 0 \cdot 1 + 1 \cdot 0.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the proof is complete. □

REMARK 15. Similar issues to those discussed in Remark 14 arise in constructing confidence intervals based on Theorem 11. Once again, it is possible to consistently estimate the variance parameter that arises in the CLT in Condition (4) of Theorem 11 in either the transient context, or the setting of regenerative steady-state simulation. To see this, recall that the look-ahead density estimator $p_n(\cdot)$ may be expressed as $n^{-1}\sum_{i=1}^n G_i(\cdot)$, for some sequence of random functions $(G_i(\cdot) : i \geq 1)$, and then

$$\begin{aligned}F_n(x) &= \int_{-\infty}^x p_n(y)\,dy\\ &= \frac{1}{n}\sum_{i=1}^n \int_{-\infty}^x G_i(y)\,dy.\end{aligned} \tag{26}$$

Evidently, (26) is a sample mean over a sequence of real-valued r.v.'s, and the sequence is either i.i.d. or regenerative, depending on the context. Therefore, standard methods may be applied to estimate the variance parameter in the CLT in Condition (4) of Theorem 11.

The comments in Remark 14 related to continuity of the variance parameter apply directly here. In particular, one must establish that the variance of the r.v. N in Condition (4) of Theorem 11 is continuous as a function of q, so that estimating the variance at q by an estimate of the variance at q_n is asymptotically valid. See Henderson and Glynn (1999) for such a result.

It is natural to ask how the performance of the look-ahead quantile estimator compares with that of a more standard quantile estimator. For ease of exposition, in the remainder of this section we focus on the case where $X = (X_n : n \geq 0)$ is a time-homogeneous Markov chain taking values in $\mathbb{R}$. Suppose that Assumption 1 and Assumption 3 are in force with $\gamma(dx) = dx$, and $p(n, x, y) = p(x, y)$ for all n, and we are interested in computing $q = F^{-1}(p)$, where F is the distribution function of the stationary distribution π of X.

A natural approach to estimation of q is to first estimate F by the empirical distribution function $\tilde{F}_n$, where

$$\tilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leq x),$$

and then choose the estimator $\tilde{Q}_n$ of q as $\tilde{Q}_n = \tilde{F}_n^{-1}(p)$.

Alternatively, using look-ahead methodology, one could estimate q by $F_n^{-1}(p)$, where

$$F_n(x) = \frac{1}{n}\sum_{j=0}^{n-1} \int_{-\infty}^x p(X_j, y)\,dy.$$

The proof of the following proposition rests primarily on the observation that

$$\int_{-\infty}^x p(X_j, y)\,dy = E[I(X_{j+1} \leq x)|X_0, \ldots, X_j], \tag{27}$$

so that the estimators F_n and $\tilde{F}_n$ are related through the principle of extended conditional Monte Carlo (Bratley et al. 1987, p. 71; Glasserman 1993).

Let var_ν denote the variance operator associated with the path space of X, where X has initial distribution ν.

PROPOSITION 12. *Suppose that Assumption* 1 *and Assumption* 3 *hold, and Conditions* 1 *and* 2 *of Theorem* 11 *are satisfied. Then,*

$$n^{1/2}(F_n(q) - p) \Rightarrow \sigma N_1(0,1), \quad \textit{and}$$
$$n^{1/2}(\tilde{F}_n(q) - p) \Rightarrow \tilde{\sigma} N_2(0,1),$$

as $n \to \infty$, *where* $\sigma^2 = \lim_{n\to\infty} n \,\text{var}_\pi F_n(q)$, $\tilde{\sigma}^2 = \lim_{n\to\infty} n \,\text{var}_\pi \tilde{F}_n(q)$, *and* $N_1(0,1)$ *and* $N_2(0,1)$ *are standard normal r.v.'s. In addition, if* X *is stochastically monotone, then* $\sigma^2 \leq \tilde{\sigma}^2$.

PROOF. Let $p(x,y) = p(0,x,y)$ (recall that we are assuming time-homogeneity). For all $w \in \mathbb{R}$,

$$\int_{-\infty}^{q} p(w,y)\,dy = P(X_1 \leq q | X_0 = w) \leq 1 \leq V^{1/2}(w),$$

since $V(\cdot) \geq 1$, so that Theorem 17.5.3 of Meyn and Tweedie implies that

$$n^{1/2}(F_n(q) - F(q)) \Rightarrow \sigma N(0,1).$$

Similarly, Theorem 17.5.3 also gives $n^{1/2}(\tilde{F}_n(q) - F(q)) \Rightarrow \tilde{\sigma} N(0,1)$.

If X is stochastically monotone, then in view of (27), we can apply Theorem 12 of Glynn and Iglehart (1988) to achieve the result. (The required uniform integrability follows from the fact that $\sigma^2 = \lim_{n\to\infty} n \,\text{var}_\pi F_n(q)$ and $\tilde{\sigma}^2 = \lim_{n\to\infty} n \,\text{var}_\pi \tilde{F}_n(q)$.) □

Combining the results of Theorem 11 and Proposition 12, we see that under reasonable conditions,

$$\sqrt{n}(Q_n - q) \Rightarrow \frac{\sigma}{p(q)} N(0,1),$$

as $n \to \infty$. It can also be shown, again under reasonable conditions, that

$$\sqrt{n}(\tilde{Q}_n - q) \Rightarrow \frac{\tilde{\sigma}}{p(q)} N(0,1),$$

as $n \to \infty$; see Henderson and Glynn (1999). Proposition 12 asserts that $\sigma^2 \leq \tilde{\sigma}^2$, so that in the context of steady-state quantile estimation for stochastically monotone Markov chains, the look-ahead quantile estimator may typically be expected to achieve variance reduction over a more standard quantile estimator.

REMARK 16. It is well known that the waiting time sequence in the single-server queue is a stochastically monotone Markov chain, and thus the results of this section may be applied in that context.

REMARK 17. Up to this point we have examined the properties of the look-ahead quantile estimator Q_n. In some cases, one might prefer to use the standard estimator $\tilde{Q}_n$, and in such situations, look-ahead methodology might again prove useful. In view of the central limit theorem for $\tilde{Q}_n$ given above, the generation of confidence intervals for q requires the estimation of $p(q)$, the density evaluated at q. The attractive properties of look-ahead estimators outlined in the Introduction make them eminently suitable for this purpose.

6. Examples. We present three examples of the application of look-ahead density estimators. Our first example is an example of steady-state density estimation and illustrates how to establish Assumption 3.

EXAMPLE 1. It is well known that the sequence $W = (W(n) : n \geq 0)$ of customer waiting times (excluding service) in the FIFO single-server queue is a Markov chain on state space $S = [0, \infty)$. In particular, W satisfies the Lindley recursion (Asmussen 1987, p. 181),

$$W(n+1) = [W(n) + Y(n+1)]^+,$$

where $[x]^+ \stackrel{\Delta}{=} \max(x, 0)$, $Y = (Y(n) : n \geq 1)$ is an i.i.d. sequence with $Y(n+1) = V(n) - U(n+1)$, $V(n)$ is the service time of the nth customer, and $U(n+1)$ is the interarrival time between the nth and $(n+1)$st customer.

To verify Assumption 3, we proceed as follows. Define $V(x) = e^{\alpha x}$ for some yet-to-be-determined constant $\alpha > 0$. Then note that

$$\begin{aligned} &E[V(W(1))|W(0) = x] \\ &\quad = E\exp(\alpha[x + Y(1)]^+) \\ &\quad = E(e^{\alpha(x+Y(1))}; x + Y(1) > 0) + P(x + Y(1) < 0) \\ &\quad \leq Ee^{\alpha(x+Y(1))} + 1 \\ &\quad = V(x)(Ee^{\alpha Y(1)} + e^{-\alpha x}). \end{aligned} \tag{28}$$

Let us assume that the moment generating function $\phi(t) \stackrel{\Delta}{=} Ee^{tY(1)}$ of $Y(1)$ exists in a neighbourhood of zero, so that $\phi(t)$ is finite for sufficiently small t. For stability, we must have $EY(1) < 0$, which implies that $\phi'(0) < 0$. Hence, there exists an $\alpha > 0$, such that $\phi(\alpha) < 1$. Now choose $K > 0$, so that $\phi(\alpha) + e^{-\alpha K} < 1$, and then for all $x > K$, we see from (28) that $E[V(W(1))|W(0) = x] \leq (1 - a)V(x)$, where $a = 1 - (\phi(\alpha) + e^{-\alpha K})$. From (28), we also see that for $x \leq K$, $E[V(W(1))|W(0) = x] \leq b \stackrel{\Delta}{=} e^{\alpha K}\phi(\alpha) + 1$. Thus, we have verified Condition (2) of Assumption 3 for the set $B = [0, K]$.

To verify Condition (1), note that $EY(1) < 0$ implies that there exists $\beta, \delta > 0$, such that $P(Y(1) < -\delta) \geq \beta$. It follows that conditional on $W(0) = x \leq K$, after $m = \lceil K/\delta \rceil$ transitions, $P(W(m) = 0) \geq \beta^m$. Taking φ to be a point mass at 0, and $\lambda = \beta^m$, we see that Condition 3 of Assumption 3 is verified. We have therefore established the following result.

PROPOSITION 13. *If the moment generating function of $Y(1)$ exists in a neighbourhood of 0, and $EY(1) < 0$, then Assumption 3 is satisfied.*

We now focus on the M/M/1 queue with arrival rate λ, service rate μ, and traffic intensity $\rho \stackrel{\Delta}{=} \lambda/\mu < 1$. The transition kernel for W is then given by

$$P(x, dy) = p(x, y)\gamma(dy),$$

where $p(x, 0) = (1 + \rho)^{-1}e^{-\lambda x}$,

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

for $y > 0$, and $\gamma(dy) = \delta_0(y) + I(y > 0)\,dy$, where δ_0 is the probability measure that assigns unit mass to the origin. Noting that $p(\cdot, \cdot)$ is bounded by $\max(\lambda, 1)$, it follows (after possibly scaling the function V by λ^2) that the conditions of Theorem 3 are satisfied, and the look-ahead density estimator therefore converges at rate $n^{-1/2}$ to the stationary density of W.

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

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

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

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

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 the traffic intensity $\rho = 0.5$. 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.

The density estimates for $x > 0$, together with the exact density, are plotted for simulation runlengths of $n = 100$ (Figure 1) and $n = 1{,}000$ (Figure 2). We observe the following.

(1) Visually, the look-ahead density estimate appears to be a far better representation of the true density than the kernel density estimate.

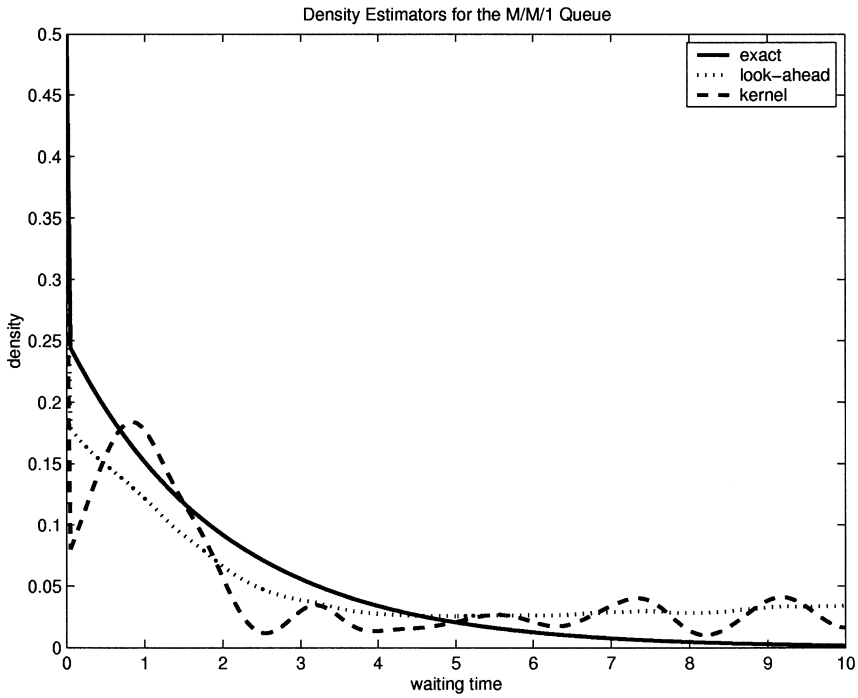

FIGURE 1. Density estimates from a run of length 100.

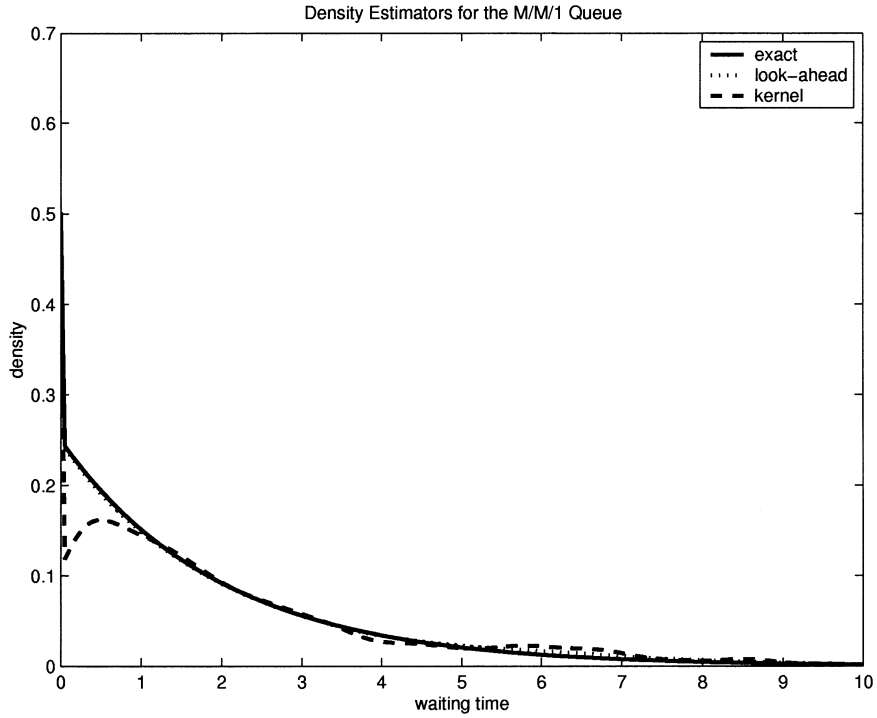

FIGURE 2. Density estimates from a run of length 1,000.

(2) The kernel density estimate has several local modes, and its performance near the origin is particularly poor, even for the run of length 1,000.

The previous example is a one-dimensional density estimation problem. Our results suggest that the rate of convergence of the look-ahead density estimators is insensitive to the underlying dimension of the problem. However, the rate of convergence of kernel density estimators is known to be adversely affected by the dimension; see Remarks 5 and 7. To assess the difference in performance in a multidimensional setting, we provide the following example.

EXAMPLE 2. Let $W = (W(k) : k \geq 1)$ be a sequence of d-dimensional i.i.d. normal random vectors with zero mean and covariance matrix the identity matrix I. Define the Markov chain $X = (X(k) : k \geq 0)$ inductively by $X(0) = 0$, and for $k \geq 0$,

$$X(k+1) = rX(k) + W(k+1),$$

where $-1 < r < 1$. The Markov chain X is a (very) special case of the linear state space model defined on p. 9 of Meyn and Tweedie (1993). We chose such a model for this example so that the steady-state density is easily computed. In particular, the stationary distribution of X is normal with mean zero and covariance matrix $(1-r^2)^{-1}I$, and thus X has stationary density

$$\pi(x) = \left(\frac{1-r^2}{2\pi}\right)^{d/2} \exp\left(-\frac{(1-r^2)x^T x}{2}\right).$$

We estimate this density at $x = 0$ for dimensions $d = 1, 2, 5$, and 10, using both a kernel density estimator and a look-ahead density estimator, with $r = 1/2$. Both estimators are constructed from simulated sample paths of length 10, 100, and 1000. We sample $X(0)$ from the stationary distribution to remove any initialization bias. To estimate the mean squared error (MSE) of the density estimators at $x = 0$, we repeat the simulations 100 times.

TABLE 1. Root MSE of estimators of $\pi(0)$ as a percentage of $\pi(0)$.

			Runlength		
d	$\pi(0)$	*Estimator*	10	100	1,000
1	0.3455	Kernel	28	14	6
1		Lookahead	7	2.4	0.8
2	0.1194	Kernel	41	22	10
2		Lookahead	9	3.5	1.1
5	$4.923e-3$	Kernel	66	51	33
5		Lookahead	16	5.3	1.7
10	$2.423e-5$	Kernel	90	82	71
10		Lookahead	22	7.8	2.5

The kernel density estimator we chose uses a multivariate standard normal distribution as the kernel, and a bandwidth $h_n = n^{-1/(d+4)}$ (see Example 1 for the rationale behind this choice of bandwidth).

Table 1 reports the root MSE for the two estimators as a percentage of the true density value $\pi(0)$.

Observe that as the dimension increases, the rate of convergence of the kernel density estimator deteriorates rapidly. In contrast, the rate of convergence of the look-ahead density estimator remains constant (for each increase in runlength by a factor of 10, relative error decreases by a factor of approximately 3), independent of the dimension of the problem.

REMARK 18. It is possible to construct look-ahead density estimators for far more complicated linear state space models than the one considered here. The critical ingredient is Assumption 1, which is easily satisfied, for example, if the innovation vectors $W(k)$ have a known density with respect to Lebesgue measure.

Our final example is an application to stochastic activity networks (SANs). This example demonstrates that the concept underlying look-ahead density estimation is basically that of conditional expectation, so that look-ahead methodology may be applied in contexts other than Markov chain simulations.

EXAMPLE 3. In this example, we estimate the density of the network completion time (the length of the longest path from the source to the sink) in a simple stochastic activity network adapted from Avramidis and Wilson (1996). Consider the SAN in Figure 3 with independent task durations, Source Node 1, and Sink Node 9. The labels on the arcs give the mean task durations. We assume that tasks (6, 9) and (8, 9) have densities (with respect to Lebesgue measure), so that the network completion time L has a density $p(\cdot)$ (with respect to Lebesgue measure).

Suppose that we sample all task durations except task (6, 9) and (8, 9), and compute the lengths $L(6)$ and $L(8)$ of the longest paths from the source node to nodes 6 and 8 respectively. Then

$$\begin{aligned} P(L \in dx) &= EP(L \in dx | L(6), L(8)) \\ &= E\left\{ f_{69}(x - L(6)) F_{89}(x - L(8)) + f_{89}(x - L(8)) F_{69}(x - L(6)) \right\} dx, \end{aligned}$$

where, for a given task ab, F_{ab} denotes the task duration distribution function, and $f_{ab}(\cdot)$ is the (Lebesgue) density. This expression is obtained using similar reasoning to that used in Example 1 of Avramidis and Wilson (1996).

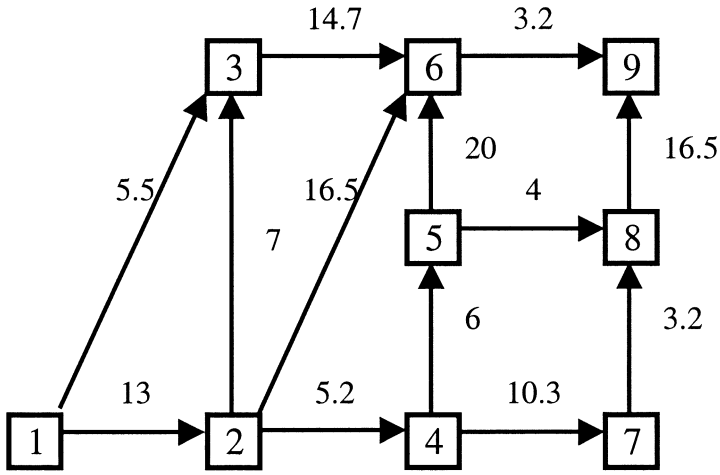

FIGURE 3. Stochastic activity network with mean task duration shown beside each task.

Now, Assumption 1 and the strong law of large numbers ensure that the look-ahead density estimator,

$$p_n(x) \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^{n} f_{69}(x - L_i(6))F_{89}(x - L_i(8)) + f_{89}(x - L_i(8))F_{69}(x - L_i(6)),$$

is a strongly consistent estimator of $p(x)$.

For the purposes of our simulation experiment, we assumed that all task durations were Weibull distributed with shape parameter 3 and with arc-specific scale parameter β chosen to give a mean as indicated on Figure 3. (So the distribution function of task durations was $F(x) = 1 - \exp(-(x/\beta)^3)$ for $x \geq 0$.) This choice of task duration distribution ensures that the look-ahead density estimator smoothly approximates the target density. The density estimate is depicted in Figure 4 for a run of length 1000.

REMARK 19. One need not base a look-ahead density estimator on the arcs that are incident on the sink. For example, one might instead focus on arcs that leave the source. In the above example, these arcs correspond to Tasks (1, 2) and (1, 3), and one would condition on the longest paths from Nodes 2 and 3 to the sink. See §4.1.2 of Avramidis and Wilson (1996) for a closely related discussion.

7. Conclusions. We have introduced look-ahead estimators for computing densities of a wide variety of performance measures within a Markov chain context. The new estimators converge (pointwise) at rate $n^{-1/2}$, where n is the sample size. This rate is dimension-independent, which stands in sharp contrast to that of more well-known density estimation methods, such as kernel density estimation, where the convergence rate is strongly dimension-dependent.

We have also explored the global convergence properties of the estimators, showing, in a certain precise sense, that under certain conditions they "smoothly approximate" the target density. This result, together with the fact that, again under certain conditions, the estimators are uniformly convergent, suggests that the estimators should prove useful for visualization purposes.

At their core, look-ahead estimators are really an example of the use of the conditional Monte Carlo method. As such, they are not confined to the Markov chain context, and one of our examples relies on this fact.

Several areas for future research immediately present themselves.

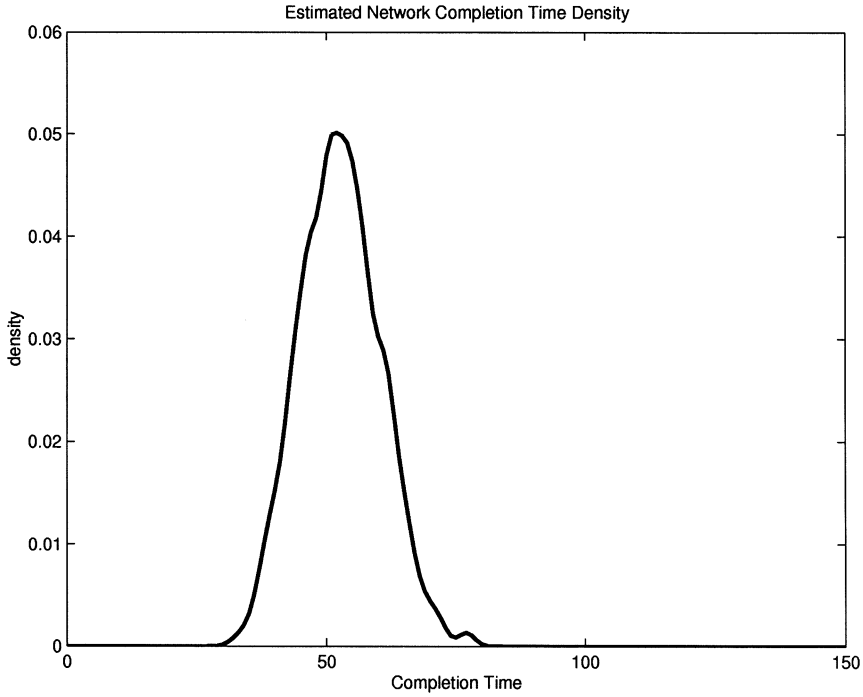

FIGURE 4. Estimate of the Network Completion Time Density.

(1) To what extent can one use similar ideas when Assumption 1 is not satisfied? This question is of some interest, given that virtually any discrete-event simulation can be described via a general state space Markov chain. Such chains are unlikely to satisfy Assumption 1; see Henderson and Glynn (1998).

(2) Can one develop density estimation methods for quantities such as steady-state customer sojourn times in general discrete-event simulation?

(3) Can we further exploit the connections between look-ahead methodology and that of conditional Monte Carlo (Fu and Hu 1997)?

(4) Where can these methods be of most use in Markov chain Monte Carlo applications?

Acknowledgments. We would like to thank the referees for suggestions that led to improvements in the exposition of the paper. The work of the first author was partially supported by New Zealand PGSF grant number UOA803 and the National Science Foundation under Grant No. DMI-9984717. The research of the second author was supported by the U.S. Army Research Office under Contract No. DAAG55-97-1-0377 and by the National Science Foundation under Grant No. DMS-9704732.

References

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

Avramidis, A. N., J. R. Wilson. 1996. Integrated variance reduction strategies for simulation. *Oper. Res.* **44** 327–346.

———, 1998. Correlation-induction techniques for estimating quantiles in simulation experiments. *Oper. Res.* **46** 574–591.

Billingsley, P. 1968. *Convergence of Probability Measures.* Wiley, New York.

Bratley, P., B. L. Fox, L. E. Schrage. 1987. *A Guide to Simulation,* 2nd ed. Springer, New York.

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

———. 1987. *A Course in Density Estimation.* Birkhauser, Boston, MA.

Fu, M., J. Q. Hu. 1997. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications.* Kluwer, Boston, MA.

Gilks, W. R., S. Richardson, D. J. Spiegelhalter, eds. 1996. *Markov Chain Monte Carlo in Practice*. Interdisciplinary Statistics. Chapman & Hall, London, U.K.

Glasserman, P. 1993. Filtered Monte Carlo. *Math. Oper. Res.* **18** 610–634.

Glynn, P. W., S. G. Henderson. 1998. Estimation of stationary densities of Markov chains. Medeiros, D., E. Watson, J. Carson, M. Manivannan, eds. *Proc. 1998 Winter Simulation Conf.* IEEE, Piscataway, NJ.

Glynn, P. W., D. L. Iglehart. 1988. Simulation methods for queues: An overview. *Queueing Syst.* **3** 221–256.

Glynn, P. W., P. L'Ecuyer. 1995. Likelihood ratio gradient estimation for stochastic recursions. *Adv. Appl. Probab.* **27** 1019–1053.

Heidelberger, P., P. A. W. Lewis. 1984. Quantile estimation in dependent sequences. *Oper. Res.* **32** 185–209.

Henderson, S. G., P. W. Glynn. 1998. Regenerative steady-state simulation of discrete-event systems. Submitted for publication. Also available from ⟨http://www-personal.umich.edu/~shaneioe/pubs.html⟩.

———. 1999. A central limit theorem for empirical quantiles in the Markov chain setting. Working paper.

Hesterberg, T., B. L. Nelson. Control variates for probability and quantile estimation. *Management Sci.* **44** 1295–1312.

Iglehart, D. L. 1976. Simulating stable stochastic systems; VI. Quantile estimation. *J. Assoc. Comput. Mach.* **23** 347–360.

Kappenman, R. F. 1987. Improved distribution quantile estimation. *Comm. Statist.* **B16** 307–320.

Meyn, S. P., R. L. Tweedie. 1993. *Markov Chains and Stochastic Stability.* Springer-Verlag, London, U.K.

Muñoz, D. F. 1998. A batch means methodology for the estimation of quantiles of the steady-state distribution. Preprint.

Muñoz, D. F., P. W. Glynn. 1998. Multivariate standarized time series for steady-state simulation output analysis. *Oper. Res.* Forthcoming.

———. 1997. Batch means methodology for estimation of a nonlinear function of a steady-state mean. *Management Sci.* **43** 1121–1135.

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

Scott, D. W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley, New York.

Seber, G. A. F. 1977. *Linear Regression Analysis.* Wiley, New York.

Seila, A. 1982. A batching approach to quantile estimation in regenerative simulations. *Management Sci.* **28** 573–581.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Yakowitz, S. 1985. Nonparametric density estimation, prediction and regression for Markov sequences. *J. Amer. Statist. Assoc.* **80** 215–221.

———, S. 1989. Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivariate Anal.* **30** 124–136.

S. G. Henderson: Department of Industrial and Operations Engineering, University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109-2117; e-mail: shane.henderson@umich.edu

P. W. Glynn: Management Science and Engineering, Terman Engineering Center, Stanford University, Stanford, CA 94305-4026; e-mail: glynn@leland.stanford.edu