The Annals of Mathematical Statistics
1969, Vol. 40, No. 4, 1386–1400

NONPARAMETRIC ESTIMATION OF THE TRANSITION DISTRIBUTION FUNCTION OF A MARKOV PROCESS[1]

By George G. Roussas

University of Wisconsin, Madison

1. Introduction and summary. In [8] the problem of nonparametric estimation in Markov processes has been considered, and estimates of the initial, two-dimensional joint, and transition densities of the process, satisfying a number of optimal properties, have been obtained. In the present paper, under the same nonparametric setup, the attention is centered primarily on the transition distribution function of the process.

It will be assumed that the underlying Markov process, defined on a probability space $(\Omega, \mathfrak{A}, P)$ and taking values in the real line R, is (strictly) stationary, and has initial, two-dimensional joint, and transition densities $p(\cdot)$, $q(\cdot\,,\,\cdot)$, and $t(\cdot \mid x)$, $x \varepsilon R$, respectively, relative to the appropriate Lebesgue measures. Let K be a probability density. On the basis of the first $n + 1$ random variables X_j, $j = 1, 2, \cdots, n + 1$ of the process, we define the random variables $p_n(x)$, $x \varepsilon R$, and $q_n(y)$, $y \varepsilon R \times R$ (by suppressing the random element ω) by the following relations

$$p_n(x) = (nh)^{-1} \textstyle\sum_{j=1}^n K((x - X_j)h^{-1}) \tag{1.1}$$

$$q_n(y) = q_n(x, x') = (nh)^{-1} \textstyle\sum_{j=1}^n K((x - X_j)h^{-\frac{1}{2}})K((x' - X_{j+1})h^{-\frac{1}{2}}), \tag{1.2}$$

where $h = h(n)$ is a sequence of positive constants satisfying also some additional conditions. We further set

$$t_n(x' \mid x) = q_n(x, x')/p_n(x). \tag{1.3}$$

Next, by means of $p_n(x)$ and $q_n(x, x')$, define the random variables

$$F_n(x) = \textstyle\int_{-\infty}^{x} p_n(z)\, dz, \qquad G_n(z \mid x) = \textstyle\int_{-\infty}^{z} t_n(dx' \mid x).$$

We finally let $F(\cdot)$ and $G(\cdot \mid x)$, $x \varepsilon R$, be the initial and transition distribution functions of the process.

Under suitable conditions on the function K, the sequence $\{h\}$, and the process, the main results of this paper are the following.

The distribution function $F_n(x)$, as an estimate of $F(x)$, obeys the Glivenko-Cantelli theorem. This is Theorem 3.1. Turning now to the estimate $G_n(\cdot \mid x)$ of $G(\cdot \mid x)$, we have been able to establish in Theorem 3.2 that sup $\{|G_n(z \mid x) - G(z \mid x)|; z \varepsilon R\}$ converges to zero, as $n \to \infty$, but in the probability sense. This is true for all $x \varepsilon R$.

In Section 4 it is assumed that the rth absolute moment of X_1 exists (for some

Received 2 October 1967; revised 28 October 1968.

[1] This research was supported by the National Science Foundation, Grant GP-8884.

$r = 1, 2, \cdots$), and the problem is that of gaining further information about $G(\cdot \mid x)$, by estimating its kth moment, to be denoted by $m(k; x)$, for $k = 1, 2, \cdots, r$. By letting the rather simple expression

$$m_n(k; x) = (nh)^{-1} p_n^{-1}(x) \sum_{j=1}^{n} X_{j+1}^{k} K((x - X_j)h^{-1})$$

stand for an estimate of $m(k; x)$, it is shown that, as $n \to \infty$, $m_n(k; x) \to m(k; x)$ in probability, for $k = 1, 2, \cdots, r$ and $x \varepsilon R$. This is the content of Theorem 4.1.

Finally in Section 5 we look into the problem of estimating the quantiles of $G(\cdot \mid x)$, and in connection with this, two results are derived. For some p in the interval $(0, 1)$, it is assumed that the pth quantile, $\xi(p, x)$, of $G(\cdot \mid x)$ is unique. By defining $\xi_n(p, x)$ as the smallest root of the equation $G_n(z \mid x) = p$ and using it as an estimate of $\xi(p, x)$, it is proved that, as $n \to \infty$,

$$\xi_n(p, x) \to \xi(p, x) \qquad \text{in probability (Theorem 5.1)},$$

and

$$(nh)^{\frac{1}{2}}[\xi_n(p, x) - \xi(p, x)] \to N(0, \tau^2(\xi, x)) \qquad \text{in law.}$$

This is Theorem 5.2, and the variance $\tau^2(\xi, x)$ is explicitly given in that theorem.

2. Notation and assumptions. It is assumed throughout this paper that the Markov process is (strictly) stationary, and it satisfies condition D_0([1] p. 221). No further mention of these assumptions will be made. It is also assumed that the process has initial, two-dimentional joint, and transition densities, with respect to appropriate Lebesgue measures, to be denoted by $p(\cdot)$, $q(\cdot\,,\cdot)$, and $t(\cdot \mid x)$, $x \varepsilon R$, respectively.

By K we denote a positive probability density defined on R into itself and satisfying the following K-assumptions.

(K1) $K(x) \leqq M_1$, $x \varepsilon R$.

(K2) $|x|K(x) \to 0$, as $|x| \to \infty$.

(K3) K is continuous.

(K4) $\int |x|^r K(x)\,dx < \infty$, for some $r = 1, 2, \cdots$ (This r is the same as the r in the condition (P2) below.)

(K5) The derivative K' exists, except possibly for a finite number of points, and is such that $x^2|K'(x)| \leqq M_2$, $x \varepsilon R$.

The quantity $h = h(n)$ stands for a sequence of positive constants which tends to zero, as $n \to \infty$, and satisfies the following H-assumptions. As $n \to \infty$,

(H1) $nh \to \infty$.

(H2) $nh \to \infty$ nondecreasingly.

(H3) $nh^2 \to 0$.

Finally the process is assumed to obey the following P-conditions.

(P1) Both $p(\cdot)$ and $q(\cdot\,,\cdot)$ are continuous and $p(x) > 0$, $x \varepsilon R$.

(P2) (i) $E(|X_1^r|) < \infty$ for some $r = 1, 2, \cdots$, and (ii) $E(|X_2^r| \mid X_1 = x) = \int |z^r| t(dz \mid x)$ is continuous.

(P3) For $p \varepsilon (0, 1)$ and $x \varepsilon R$, the pth quantile of the transition distribution function of the process, $G(\cdot \mid x)$, is unique.

(P4) The joint densities $q_{1,i}(\cdot\,,\,\cdot)$ and $q_{1,i,j}(\cdot\,,\,\cdot\,,\,\cdot)$ of the random variables X_1, X_i and X_1, X_i, X_j, respectively, are bounded; that is, $q_{1,i}(x, x') \leqq M_3$, for all $1 < i$ and $q_{1,i,j}(x, x', x'') \leqq M_3$, for all $1 < i < j$.

(P5) The density $p(\cdot)$ satisfies the Lipschitz condition $|p(x') - p(x)| \leqq M(x)|x - x'|$.

(P6) The density $q(\cdot\,,\,\cdot)$ satisfies the Lipschitz condition (i) $|q(y') - q(y)| \leqq M^*(y)\|y - y'\|$, and (ii) $\int M^*(x, x')\,dx' < \infty$, where $(x, x') = y$.

Not all of the results obtained in this paper require the validity of all assumptions listed here. It would be suggested therefore to mention at the outset which results depend on what assumptions.

In Section 3 only assumptions (K1), (K2), (K3), (H1), and (P1) are utilized.

The results of Section 4 are derived under assumptions (K1), (K2), (K4), (H1), (H2), (P1), (P2)(i), and (P2)(ii).

Finally the theorems and lemmas of Section 5 require the validity of the following assumptions: (K1), (K2), (K4) (for $r = 1$), (K5), (H1), (H3), (P1), (P3), (P4), (P5), (P6)(i), and (P6)(ii).

Probability densities satisfying the K-conditions can be found, for example, among those densities mentioned in [4].

There is no problem in choosing h so that it will satisfy the H-conditions. Assumption (H2), although somewhat restrictive, will be ordinarily fulfilled in practice.

The assumption (P2)(ii) is used only in Section 4, as it was pointed out above. It is felt that the result of that section should hold without (P2)(ii). We have not been able, however, to avoid using the assumption in question. Obviously, (P2)(ii) is implied by (P6) and (P1), if (P6)(ii) be supplemented by the condition that $\int |x'^r| M^*(x, x')\,dx' < \infty$.

Results of the same nature as those treated in Section 5, are usually derived under differentiability conditions on the densities involved. In this respect, assumptions (P5) and (P6) constitute an improvement over such differentiability conditions.

Remark. Assumption (P2)(ii) and the continuity of $p(\cdot)$ assumed in (P1) imply, in effect, that the function $\int |z^r| q(x, z)\,dz$ is continuous. It follows from this that, if A is a measurable subset of R, one has that $\int z^r I_A(z) q(x, z)\,dz$ is also a continuous function as is easily seen.

In closing this section we mention once and for all that all bounding "constants" (some of which depend on x) are positive and finite numbers, and all limits are taken as $n \to \infty$ unless otherwise explicitly stated. Integrals without limits are taken over the entire real line, and whenever suprema are taken, it is tacitly assumed that the resulting functions are measurable.

Finally it should be pointed out that in the proofs of the various lemmas and theorems only the main steps are presented here. The details can be found in the references cited.

3. Estimation of the initial and transition distribution functions. If F_n^* is the sample distribution function defined on the basis of the random variables

$X_j, j = 1, \cdots, n$ of the process, it is an easy matter to see that

$$P[\sup \{|F_n^*(x) - F(x)|; x \varepsilon R\} \to 0] = 1,$$

where F is the initial distribution function of the process and R is the real line.

Let now $p_n(x)$ be the estimate of $p(x)$ given in (1.1). Define $F_n(x)$ by $F_n(x) = \int_{-\infty}^{x} p_n(z)\, dz$. Then it is shown here that the distribution function F_n enjoys the same property as F_n^*. Namely,

THEOREM 3.1. *Let assumptions* (K2) *and* (K3) *be satisfied. Then*

$$P[\sup \{|F_n(x) - F(x)|; x \varepsilon R\} \to 0] = 1.$$

The proof of this theorem is based on the following two lemmas for the formulation of which some further notation is required.

For an arbitrary but fixed $x \varepsilon R$, define the following sequences of functions of v

$$G(x, h; v) = 1 - \int_{-\infty}^{(x-v)h^{-1}} K(z)\, dz \qquad \text{for } v < x,$$
$$g(x, h; v) = \int_{-\infty}^{(x-v)h^{-1}} K(z)\, dz \qquad \text{for } v > x,$$

and set

$$\mathcal{G} = \{G(x, h; v); n = 1, 2, \cdots\}, \qquad \mathcal{G}' = \{g(x, h; v); n = 1, 2, \cdots\}.$$

Then the first lemma is to the following effect

LEMMA 3.1. *Under the assumptions of Theorem* 3.1, *the classes* $\mathcal{G}$ *and* $\mathcal{G}'$ *consist of equicontinuous functions.*

PROOF. By means of the Mean Value theorem of integral calculus and the continuity of K ((K3)), one gets

$$|G(x, h; v) - G(x, h; v_0)| \leqq 2M_4(x - v_0)^{-1}|v - v_0|,$$

where M_4 is a bound of $|w|K(w)$ which exists by (K2) and (K3). This establishes the equicontinuity of the class $\mathcal{G}$.

The equicontinuity of the class $\mathcal{G}'$ is shown in a similar fashion.

Clearly, the members of the class $\mathcal{G}$ are uniformly bounded (by 2). This boundedness together with the equicontinuity of Lemma 3.1, and the fact that $F_n^*(x) \to F(x)$ a.s. for every x in R, imply that Theorem 3.1 in [6] applies and gives

$$\int_{-\infty}^{x} G(x, h; v)\, dF_n^*(v) - \int_{-\infty}^{x} G(x, h; v)\, dF(v) \to 0 \qquad \text{a.s.}$$

This is equivalent to

$$F_n^*(x) - \int_{-\infty}^{x} G^*(x, h; v)\, dF_n^*(v) - F(x) + \int_{-\infty}^{x} G^*(x, h; v)\, dF(v) \to 0 \qquad \text{a.s.},$$

where we set $G^*(x, h; v) = \int_{-\infty}^{(x-v)h^{-1}} K(z)\, dz$, $(v < x)$. Now $|G^*(x, h; v)| \leqq 1$, $G^*(x, h; v) \to 1\, (v < x)$, and $F_n^*(x) - F(x) \to 0$ a.s., $x \varepsilon R$. Therefore

$$\int_{-\infty}^{x} G^*(x, h; v)\, dF_n^*(v) \to F(x) \qquad \text{a.s.} \tag{3.1}$$

A similar argument applied to the members of the class $\mathcal{Q}'$ gives

$$\int_x^\infty g(x, h; v)\, dF_n^*(v) \to 0 \qquad \text{a.s.} \tag{3.2}$$

We now observe that

$$\begin{aligned} F_n(x) &= \int_{-\infty}^x p_n(z)\, dz = \int [\int_{-\infty}^{(x-v)h^{-1}} K(z)\, dz]\, dF_n^*(v) \\ &= \int_{-\infty}^x [\int_{-\infty}^{(x-v)h^{-1}} K(z)\, dz]\, dF_n^*(v) + \int_x^\infty [\int_{-\infty}^{(x-v)h^{-1}} K(z)\, dz]\, dF_n^*(v) \text{ a.s.}, \end{aligned}$$

where in the first integral v remains strictly less than x and in the second integral v remains strictly greater than x. Hence

$$F_n(x) = \int_{-\infty}^x G^*(x, h; v)\, dF_n^*(v) + \int_x^\infty g(x, h; v)\, dF_n^*(v) \text{ a.s.} \tag{3.3}$$

Utilizing the results (3.1) and (3.2), (3.3) gives then

$$F_n(x) \to F(x) \text{ a.s.} \qquad \text{for every } x \varepsilon R.$$

That is, the following lemma has been established.

LEMMA 3.2. *Under the assumptions of Theorem* 3.1, *we have*

$$F_n(x) \to F(x) \text{ a.s.} \qquad \textit{for every } x \varepsilon R.$$

This lemma provides all that is needed for the proof of Theorem 3.1. In fact,

PROOF OF THEOREM 3.1. A close examination of the proof of the Glivenko-Cantelli theorem (see, e.g., [2] p. 20) reveals that the proof rests on the fact that the distribution function which is used as an estimate of the underlying distribution function F converges a.s. to $F(x)$, for all x in R. When the sample distribution function is used as an estimate of F, this convergence follows from the strong law of large numbers for the independent case, and the Ergodic theorem in the stationary case. In the present case the a.s. convergence of $F_n(x)$, for all x in R, is provided by Lemma 3.2.

REMARKS. (1) It is to be pointed out that Theorem 3.1 remains true for (strictly) stationary sequences, since only stationarity, but not the Markovian character of our process, was used in its proof.

(2) A theorem such as Theorem 3.1 was formulated by E. A. Nadaraya in [3] for the independent case only, but no proof was presented.

A much more interesting problem in the case of Markov processes is that of (nonparametrically) estimating the transition distribution function $G(z \mid x) = \int_{-\infty}^z t(dx' \mid x)$ of the processes. This problem has not been solved as yet, to the best of our knowledge. In the present paper a solution to this problem is proposed.

What we do is to define the distribution function $G_n(z \mid x) = \int_{-\infty}^x t_n(dx' \mid x)$, where $t_n(x' \mid x)$ is the estimate of $t(x' \mid x)$ given in (1.3), and use $G_n(z \mid x)$ as an estimate of $G(z \mid x)$. Then it would be desirable to prove a theorem such as Theorem 3.1. However, it appears that a theorem of this nature would require at least that $p_n(x)$ converges a.s. to $p(x)$, while all we know so far is that this convergence is true in the probability sense only (in fact, in quadratic mean; see

[8] Theorem 3.1.) Here we confine ourselves to establishing a weaker mode of convergence of $G_n(z \mid x)$. More precisely the following theorem is proved.

THEOREM 3.2. *Let assumptions* (K1), (K2), (H1), *and* (P1) *be satisfied. Then* $\sup\{|G_n(z \mid x) - G(z \mid x)|; z \varepsilon R\} \to 0$ *in probability, for all* $x \varepsilon R$.

The following lemma will facilitate the proof of the theorem.

LEMMA 3.3. *Under the assumptions of Theorem* 3.2, *we have*

$$E \int |q_n(x, x') - q(x, x')| \, dx' \to 0, \text{ for all } x \varepsilon R,$$

where $q_n(x, x')$ *is given in* (1.2).

PROOF. This is an easy consequence of Theorems 2.2 and 3.1 in [8] and Theorem 1 in [5].

PROOF OF THEOREM 3.2. We have

$$\begin{aligned}
&\sup\{|G_n(z \mid x) - G(z \mid x)|; z \varepsilon R\} \\
&\quad = \sup\{|p_n^{-1}(x) \textstyle\int_{-\infty}^{z} q_n(x, x') \, dx' - p^{-1}(x) \int_{-\infty}^{z} q(x, x') \, dx'|; z \varepsilon R\} \\
&\quad \leqq p_n^{-1}(x) \sup\{\textstyle\int_{-\infty}^{z} |q_n(x, x') - q(x, x')| \, dx'; z \varepsilon R\} + |p^{-1}(x) - p^{-1}(x)| \\
&\qquad \cdot \sup\{\textstyle\int_{-\infty}^{z} q(x, x') \, dx'; z \varepsilon R\} \\
&\quad \leqq p_n^{-1}(x) \textstyle\int |q_n(x, x') - q(x, x')| \, dx' + |p_n^{-1}(x) - p^{-1}(x)| p(x).
\end{aligned}$$

Now from Lemma 3.3 it follows that

$$\int |q_n(x, x') - q(x, x')| \, dx' \to 0 \text{ in probability, for all } x \varepsilon R.$$

Since also $p_n(x) \to p(x)$ in probability, the theorem is proved.

4. Estimation of the moments of the transition distribution function. In this section it is assumed that for some positive integer r, $E|X_1^r|$ exists, and further information about $G(\cdot \mid x)$ will be sought through the estimates of its moments $E(X_2^k \mid X_1)$, for $k = 1, \cdots, r$. We consider again the estimate $q_n(x, x')$ of $q(x, x')$ given in (1.2), where now $h^{\ddagger}$ is replaced by h for easier writing. That is, we take

$$q_n^*(x, x') = (nh^2)^{-1} \textstyle\sum_{j=1}^{n} K((x - X_j)h^{-1}) K((x' - X_{j+1})h^{-1}).$$

By means of $q_n^*(x, x')$ and the estimate $p_n(x)$ of $p(x)$ given in (1.1), we define the estimate $t_n^*(x' \mid x) = q_n^*(x, x')/p_n(x)$ of the transition density $t(x' \mid x)$, and we further set

$$m_n^*(k; x) = \int z^k t_n^*(dz \mid x), \qquad k = 1, 2, \cdots, r.$$

The expression $m_n^*(k; x)$ will be used as an estimate of the kth conditional moment

$$m(k; x) = E(X_2^k \mid X_1 = x) = \int z^k t(dz \mid x).$$

It is readily seen that

$$\begin{aligned}
(4.1)\quad & m_n^*(k; x) \\
&= \textstyle\sum_{i=0}^{k} [(nh)^{-1} p_n^{-1}(x) \sum_{j=1}^{n} \binom{k}{i} h^{k-i} K((x - X_j)h^{-1}) X_{j+1}^{i} \int z^{k-i} K(z) \, dz].
\end{aligned}$$

This expression of $m_n^*(k; x)$ is somewhat complicated, in particular for large values of k. Thus it would be desirable to replace it by something considerably simpler. To this effect we consider

$$(4.2) \qquad m_n(k; x) = (nh)^{-1} p_n^{-1}(x) \textstyle\sum_{j=1}^{n} X_{j+1}^k K((x - X_j)h^{-1}),$$

and it will be shown that both $m_n^*(k; x)$ and $m_n(k; x)$ behave in the same way asymptotically. This is made precise by the following theorem.

THEOREM 4.1. *Let assumptions* (K1), (K2), (K4), (H1), (H2), (P1), (P2)(i), (P2)(ii) *be satisfied. Then the quantities* $m_n^*(k; x)$ *and* $m_n(k; x)$ *defined by* (4.1) *and* (4.2) *are consistent estimates of the kth conditional moment* $m(k; x)$; *that is,*

(i) $m_n^*(k; x) \to m(k; x)$ *in probability and also*

(ii) $m_n(k; x) \to m(k; x)$ *in probability,* $k = 1, 2, \cdots, r$.

The following lemma will simplify the proof of the theorem.

LEMMA 4.1. *Under the assumptions* (K1), (K2), (K4), (H1), (P1), *and* (P2)(i), *we have*

$$m_n^*(k; x) - m_n(k; x) \to 0 \textit{ in probability}, \qquad k = 1, 2, \cdots, r.$$

PROOF. Clearly, it suffices to establish the lemma for $k = r$ only, since the proof for the remaining values of k will be quite similar. We have

$$\begin{aligned} m_n^*(r; x) &\\ = \textstyle\sum_{i=0}^{r-1} [(nh)^{-1} p_n^{-1}(x) \textstyle\sum_{j=1}^{n} \tbinom{r}{i} h^{r-i} K((x - X_j)h^{-1}) X_{j+1}^i \int z^{r-i} K(z)\, dz] &\\ + m_n(r; x). \end{aligned}$$

Therefore, by the fact that $p_n(x) \to p(x)$ in probability (Theorem 3.1 [8]) and assumption (K4), it suffices to prove that

$$n^{-1} h^{r-i-1} \textstyle\sum_{j=1}^{n} X_{j+1}^i K((x - X_j)h^{-1}) \to 0 \text{ in probability.}$$

But

$$h^{r-i-1} E|X_2^i K((x - X_1)h^{-1})| = h^{r-i} \int |w^i| [h^{-1} \int K((x - v)h^{-1}) q(v, w)\, dv]\, dw,$$

and

$$h^{-1} \int K((x - v)h^{-1}) q(v, w)\, dv \to q(x, w)$$

by Theorem 1A in [4]. Hence

$$h^{r-i} h^{-1} \int K((x - v)h^{-1}) q(v, w)\, dv \to 0 \, (i < r, \text{ thus } h^{r-i} \to 0).$$

On the other hand,

$$h^{r-i-1} |w^i| \int K((x - v)h^{-1}) q(v, w)\, dv \leqq M_1 h^{r-i-1} |w^i| p(w)$$

by assumption (K1), and this is $\leqq M_5 |w^i| p(w)$, where M_5 includes M_1 and also a bound for h^{r-i-1}. The Dominated Convergence theorem applies then and gives

$$h^{r-i-1} E|X_2^i K((x - X_1)h^{-1})| \to 0.$$

The Tchebichev inequality concludes the proof of the lemma.

By Lemma 4.1 the validity of Theorem 4.1 follows from the covergence in the second part of it. Thus we would like to prove that

(4.3) $(nh)^{-1}\sum_{j=1}^{n} X_{j+1}^{r}K((x - X_j)h^{-1}) \to \int w^r q(x, w)\, dw$ in probability,

confining ourselves to the value $k = r$ only which we can do, as was pointed out in the proof of Lemma 4.1.

The convergence in (4.3) will be established by means of the familiar truncation method. To this end, we set

$$W_j = W_j(n) = h^{-1}X_{j+1}^{r}K((x - X_j)h^{-1}),$$

and, for some positive δ, define

$$\begin{aligned} V_j = V_j(n) &= W_j \text{ if } |W_j| \leqq n\delta & \qquad Z_j = Z_j(n) &= 0 \text{ if } |W_j| \leqq n\delta \\ &= 0 \text{ otherwise,} & &= W_j \text{ otherwise,} \end{aligned}$$

for $j = 1, 2, \cdots, n$. Then $W_j = V_j + Z_j$ and the left-hand side of (4.3) becomes $n^{-1}\sum_{j=1}^{n} W_j$. Next

$$\begin{aligned} (4.4)\quad n^{-1}\textstyle\sum_{j=1}^{n} W_j - c(r) &= [n^{-1}\textstyle\sum_{j=1}^{n} V_j - d_n(r)] + [c_n(r) - c(r)] \\ &\quad + [d_n(r) - c_n(r)] + n^{-1}\textstyle\sum_{j=1}^{n} Z_j, \end{aligned}$$

where

$$(4.5)\qquad c(r) = \int w^r q(x, w)\, dw,$$

$$(4.6)\qquad c_n(r) = E(W_1), \qquad \text{and}$$

$$(4.7)\qquad d_n(r) = E(V_1) = E\{W_1 I_{[|W_1| \leqq n\delta]}\}.$$

The relation (4.3) is then equivalent to proving that

$$n^{-1}\textstyle\sum_{j=1}^{n} W_j - c(r) \to 0 \text{ in probability.}$$

By (4.4), it suffices to prove that the right-hand side of that relation tends to zero in probability, and this will be done in a series of three lemmas the first of which is the following one.

LEMMA 4.2. *Let assumptions* (K1), (K2), (P2)(i), *and* (P2)(ii) *be satisfied. Then with* $c(r)$, $c_n(r)$, *and* $d_n(r)$ *defined by* (4.5), (4.6), *and* (4.7), *respectively, we have*

(i) $c_n(r) \to c(r)$ *and*

(ii) $c_n(r) - d_n(r) \to 0$.

PROOF. (i) In fact,

$$c_n(r) = h^{-1}\int K((x - v)h^{-1})[\int w^r q(v, w)\, dw]\, dv.$$

But $\int w^r q(v, w)\, dw$ is continuous, by assumption (P2)(ii), and therefore Theorem 1A [4] applies and gives the desired result. From this proof it also follows that

$$(4.8)\qquad E|W_1| = h^{-1}E[|X_2^r|\, K((x - X_1)h^{-1})] \to \int |w^r|\, q(x, w)\, dw.$$

(ii) It is seen that

$$|c_n(r) - d_n(r)| \leqq h^{-1}E[|X_2^r|\, K((x - X_1)h^{-1})] - E\{h^{-1}\, |X_2^r|\, K((x - X_1)h^{-1})I_{[|X_2^r| \leqq n\delta h M_1^{-1}]}\}.$$

The first member of the right-hand side of the last relation converges to $\int |w^r|\, q(x, w)\, dw$, by (4.8). Next

$$E\{h^{-1}\, |X_2^r|\, K((x - X_1)h^{-1})I_{[|X_2^r| \leqq n\delta h M_1^{-1}]}\} = h^{-1} \int K((x - v)h^{-1})\{\int |w^r|\, I_{[|w^r| \leqq n\delta h M_1^{-1}]}q(v, w)\, dv\, dw\}.$$

Now the integral in the square brackets is a continuous function (see Remark in Section 2), and, by the Dominated Convergence theorem, it tends (as $n \to \infty$) to the function $\int |w^r|\, q(v, w)\, dw$ which is also continuous. Then an obvious generalization of Theorem 1A in [4] applies and concludes the proof.

LEMMA 4.3. *Under the assumptions of Lemma* 4.2, *we have*

$$n^{-1} \sum_{j=1}^{n} V_j - d_n(r) \to 0 \quad \textit{in probability.}$$

PROOF. We have

$$\sigma^2(V_1) \leqq n\delta \int |W_1|\, dP = n\delta E\, |W_1|,$$

and since

$$E\, |W_1| \to \int |w^r|\, q(x, w)\, dw = p(x)E(|X_2^r| \mid X_1 = x) \quad \text{(finite)},$$

we get

$$\sigma^2(V_1) \leqq n\delta M_6(x). \tag{4.9}$$

Next

$$|\mathrm{Cov}\,(V_1, V_{j+1})| \leqq 2\gamma^{\frac{1}{2}}\rho^{\frac{1}{2}(j-1)}E(V_1^2)$$

by Lemma 7.1 [1], and hence

$$\sum_{j=1}^{n} |\mathrm{Cov}\,(V_1, V_{j+1})| \leqq M_7 E(V_1^2), \tag{4.10}$$

where $M_7 = 2\gamma^{\frac{1}{2}}(1 - \rho^{\frac{1}{2}})^{-1}$. By means of the Tchebichev inequality and the inequalities (4.8) and (4.9), one easily gets

$$P[|n^{-1} \sum_{j=1}^{n} V_j - d_n(r)| \geqq \epsilon] \leqq M_8\delta, \qquad M_8 = (1 + 2M_7)M_6(x)\epsilon^{-2}.$$

Since this last inequality is true for every $\delta > 0$, the proof of the lemma is completed.

We finally formulate and prove the following

LEMMA 4.4. *Under the assumptions* (K1), (K2), (H2), (P2)(i), *and* (P2)(ii), *we have*

$$n^{-1} \sum_{j=1}^{n} Z_j \to 0 \quad \textit{in probability.}$$

PROOF. It is easily seen that

$$P[|n^{-1} \sum_{j=1}^{n} Z_j| \geqq \epsilon] \leqq P[\sum_{j=1}^{n} Z_j \neq 0] \leqq \delta^{-1} \int |W_1| I_{[|W_1|>n\delta]} \, dP.$$

But

$$\int |W_1| I_{[|W_1|>n\delta]} \, dP \to 0. \tag{4.11}$$

In fact,

$$\int |W_1| I_{[|W_1|>n\delta]} \, dP \leqq \int h^{-1}K((x-v)h^{-1}) f_n(v) \, dv,$$

where we set

$$f_n(v) = \int |w^r| I_{[|w^r| \geqq nh\delta M_1^{-1}]} q(v, w) \, dv \, dw.$$

For each n, $f_n(\cdot)$ is a continuous function (see Remark in Section 2), and for each v, it converges to zero. If, in addition, assumption (H2) is satisfied, this convergence is nonincreasing. Then an obvious generalization of Theorem 1A in [4] applies to the expression $\int h^{-1}K((x-v)h^{-1}) f_n(v) \, dv$ and gives (4.11).

PROOF OF THEOREM 4.1. This proof follows by putting together the various facts established in this section. To summarize: by Lemma 4.1, only the second part of the theorem requires a proof. In order for this part to be true, it suffices to show the validity of (4.3), or the validity of its equivalent form (4.4). That (4.4) holds, follows from Lemma 4.2, Lemma 4.3, and Lemma 4.4.

REMARK. In this section, unlike Section 3 and Section 5, in forming the estimate $q_n^*(x, x')$, h itself rather than $h^{\frac{1}{2}}$ was used. And as was already pointed out at the beginning of the present section, this was done only for convenience in the writing. All arguments obviously go through with $h^{\frac{1}{2}}$ replacing h. One point only deserves some special attention. In proving Lemma 4.4, the nondecreasing property of nh was utilized. But this implies that $nh^{\frac{1}{2}}$ is also nondecreasing. So no problem arises there either.

5. Estimation of the quantiles of the transition distribution function. For $0 < p < 1$, the pth quantile of $G(\cdot \mid x)$ is a root of the equation $G(z \mid x) = p$. By assumption (P3), there is only one such a root which we denote by $\xi(p, x)$. The problem then is that of estimating $\xi(p, x)$. An obvious estimate for it is the sample pth quantile; that is, a root of the equation $G_n(z \mid x) = p$. For reasons of definiteness the smallest root of this equation will be taken to be the pth quantile of $G_n(\cdot \mid x)$, and it will be denoted by $\xi_n(p, x)$.

In the present section two properties of $\xi_n(p, x)$ will be established. Namely, consistency in the probability sense, and asymptotic normality. The first of these results is the following theorem.

THEOREM 5.1. *Let assumptions* (K1), (K2), (H1), (P1), *and* (P3), *be satisfied. Then, with* $\xi(p, x)$ *and* $\xi_n(p, x)$ *as defined above, we have*

$$\xi_n(p, x) \to \xi(p, x) \quad \textit{in probability}, \qquad 0 < p < 1, \qquad x \in R.$$

PROOF. In order to simplify the notation the letters p and x will be left out in the expressions $\xi_n(p, x)$ and $\xi(p, x)$ and we will simply write ξ_n and ξ. For $\epsilon > 0$, define $\delta(\epsilon)$ by

$$\delta(\epsilon) = \min\{G(\xi + \epsilon \mid x) - G(\xi \mid x), G(\xi \mid x) - G(\xi - \epsilon \mid x)\}.$$

Then $\delta(\epsilon)$ is positive because of the uniqueness of ξ. Next

$$|G(\xi_n \mid x) - G(\xi \mid x)| \leqq \sup\{|G(z \mid x) - G_n(z \mid x)|; z \varepsilon R\},$$

while from the definition of $\delta(\epsilon)$ it follows that

$$[|\xi_n - \xi| > \epsilon] \subseteq [|G(\xi_n \mid x) - G(\xi \mid x)| > \delta(\epsilon)].$$

Therefore

$$\begin{aligned} P[|\xi_n - \xi| > \epsilon] &\leqq P[|G(\xi_n \mid x) - G(\xi \mid x)| > \delta(\epsilon)] \\ &\leqq P[\sup\{|G_n(z \mid x) - G(z \mid x)|; z \varepsilon R\}], \end{aligned}$$

and this last expression tends to zero by Theorem 3.2.

The proof of the asymptotic normality of ξ_n is much more involved. Some preliminary results will facilitate it. Applying Taylor's formula to $G_n(\cdot \mid x)$, we get

$$G_n(z \mid x) = G_n(z_0 \mid x) + (z - z_0)t_n(z^* \mid x).$$

Upon replacing z by ξ_n and z_0 by ξ and observing that $G_n(\xi_n \mid x) = p = G(\xi \mid x)$, one has then

$$(5.1) \qquad (nh)^{\frac{1}{2}}(\xi_n - \xi) = -(nh)^{\frac{1}{2}}[G_n(\xi \mid x) - G(\xi \mid x)]t_n^{-1}(\zeta_n \mid x),$$

where $\zeta_n = \zeta_n(p, x)$ is a random variable whose values lie between the values of the random variables ξ_n and ξ. We intend to establish asymptotic normality for $(nh)^{\frac{1}{2}}(\xi_n - \xi)$. To this end, we first prove the following lemma.

LEMMA 5.1. *Under the assumptions of Theorem* 5.1 *and, in addition, assumption* (K5), *we have*

$$(5.2) \qquad t_n(\zeta_n \mid x) \to t(\xi \mid x) \quad \textit{in probability}, \qquad x \varepsilon R.$$

PROOF. Clearly (5.2) is equivalent to

$$\begin{aligned} &(nh)^{-1} \textstyle\sum_{j=1}^n K((x - X_j)h^{-\frac{1}{2}}) \\ &\qquad \cdot [K((\zeta_n - X_{j+1})h^{-\frac{1}{2}}) - K((\xi - X_{j+1})h^{-\frac{1}{2}})] \to 0 \quad \text{in probability.} \end{aligned}$$

Since by assumption (K5),

$$K((\zeta_n - X_{j+1})h^{-\frac{1}{2}}) - K((\xi - X_{j+1})h^{-\frac{1}{2}}) = (\xi - \zeta_n)h^{-\frac{1}{2}}K'(\zeta_{nj}),$$

where ζ_{nj} is a random variable taking values between those of the random variables $(\zeta_n - X_{j+1})h^{-\frac{1}{2}}$ and $(\xi - X_{j+1})h^{-\frac{1}{2}}$, the last relation becomes

$$(5.3) \quad (\xi - \zeta_n)(nh^{\frac{1}{2}})^{-1} \textstyle\sum_{j=1}^n [K((x - X_j)h^{-\frac{1}{2}})h^{-1}K'(\zeta_{nj})] \to 0 \quad \text{in probability.}$$

By assumption (K5) again and the fact that $h^{\frac{1}{2}}\zeta_{nj}$ is bounded in probability uniformly in $j = 1, 2, \cdots, n$, as is easily seen, it follows that $h^{-1}K'(\zeta_{nj})$ is bounded in probability uniformly in j. By means of this result and the convergence in probability of $(nh^{\frac{1}{2}})^{-1}\sum_{j=1}^{n}K((x-X_j)h^{-\frac{1}{2}})$, the entire expression on the left hand side of (5.3), except for $\xi-\zeta_n$, is bounded in probability uniformly in j. This fact together with the convergence $\zeta_n\to\xi$ in probability which follows from the fact that $\xi_n\to\xi$ in probability, concludes the proof of the lemma.

Now we will work with the expression $(nh)^{\frac{1}{2}}[G_n(\xi\mid x)-G(\xi\mid x)]$. It is readily seen that

$$G_n(\xi\mid x)=\sum_{j=1}^{n}L_n^*(Y_j)/\sum_{j=1}^{n}L_n(X_j),$$

where

$$L_n(X_j)=K((x-X_j)h^{-1})\qquad\text{and}\tag{5.4}$$

$$L_n^*(Y_j)=K((x-X_j)h^{-\frac{1}{2}})\int_{-\infty}^{\xi}K((x'-X_{j+1})h^{-\frac{1}{2}})\,dx',\tag{5.5}$$

$$Y_j=(X_j,X_{j+1}),$$

$$j=1,2,\cdots,n.$$

One verifies further that

$$\begin{aligned}(nh)^{\frac{1}{2}}&[G_n(\xi\mid x)+v_n]\\&=[(nh)^{-1}\textstyle\sum_{j=1}^{n}L_n(X_j)]^{-1}(nh)^{-\frac{1}{2}}\textstyle\sum_{j=1}^{n}[\varphi_n(Y_j)-E\varphi_n(Y_j)],\end{aligned}\tag{5.6}$$

where

$$\varphi_n(Y_j)=L_n^*(Y_j)+v_nL_n(X_j)\tag{5.7}$$

$$v_n=-[EL_n^*(Y_1)][EL_n(X_1)]^{-1}.\tag{5.8}$$

We will be interested in proving asymptotic normality for the expression $(nh)^{\frac{1}{2}}[G_n(\xi\mid x)+v_n]$. But $(nh)^{-1}\sum_{j=1}^{n}L_n(X_j)\to p(x)$ in probability, and thus it suffices to prove asymptotic normality for

$$(nh)^{-\frac{1}{2}}\sum_{j=1}^{n}[\varphi_n(Y_j)-E\varphi_n(Y_j)].$$

More precisely,

LEMMA 5.2. *Let assumptions* (K1), (K2), (H1), (P1), *and* (P4) *be satisfied. Then*

$$\mathcal{L}\{(nh)^{-\frac{1}{2}}\sum_{j=1}^{n}[\varphi_n(Y_j)-E\varphi_n(Y_j)]\}\to N(0,\sigma_0^2(\xi,x)),$$

where

$$\sigma_0^2(\xi,x)=G^2(\xi\mid x)p(x)\int K^2(z)\,dz.$$

PROOF. The proof consists in simply verifying the conditions (A1), (A2), (A3), $(A2)^{**}$, $(A3)^{*}$, and $(A4)^{**}$ in [7]. Condition (A1) is satisfied. The quantity h_n appearing in (A2) is equal to nh here which tends to infinity by (H1). As for the condition (A3), this has already been checked in [8] under the conditions (K1), (K2), (H1), (P1), and (P4). Here the $\sigma_1^2(x)$ of (A3)(iv) has the value

$$\sigma_1^2(x)=p(x)\int K^2(z)\,dz.\tag{5.9}$$

Next $(A2)^{**}$ is also true with the quantity l appearing in $(A2)^{**}$(ii) given by

$$l = p(x) \tag{5.10}$$

The verification of $(A3)^*$ and $(A4)^{**}$ is long, but rather straightforward. The interested reader is referred to [9] for the details.

Now Lemma 5.2 in conjunction with the relation (5.6) gives

$$\mathcal{L}\{(nh)^{\frac{1}{2}}[G_n(\xi \mid x) + v_n]\} \to N(0, \sigma_0^2(\xi, x)p^{-1}(x)), \tag{5.11}$$

where v_n is given by (5.8).

At this point attention will be given to the expression $(nh)^{\frac{1}{2}}[v_n + G(\xi \mid x)]$. The lemma stated and proved below relates to its convergence.

LEMMA 5.3. *Under the assumptions* (K1), (K2), (K4) (*with* $r = 1$), (H3), (P1), (P5), *and* (P6)(i), (P6)(ii), *we have*

$$(nh)^{\frac{1}{2}}[v_n + G(\xi \mid x)] = -(nh)^{\frac{1}{2}}\{[EL_n^*(Y_1)][EL_n(X_1)]^{-1} - G(\xi \mid x)\} \to 0.$$

PROOF. Clearly,

$$[EL_n^*(Y_1)][EL_n(X_1)]^{-1} - G(\xi \mid x) = [h^{-1}EK((x - X_1)h^{-1})]^{-1} \cdot h^{-1}[EL_n^*(Y_1) - G(\xi \mid x)EL_n(X_1)].$$

Hence it suffices to show that

$$h^{-1}(nh)^{\frac{1}{2}}[EL_n^*(Y_1) - G(\xi \mid x)EL_n(X_1)] \to 0.$$

This is rewritten in the following more convenient form

$$(nh)^{\frac{1}{2}}[h^{-1}EL_n^*(Y_1) - \textstyle\int_{-\infty}^{\xi} q(x, w)\, dw] - (nh)^{\frac{1}{2}}[h^{-1}G(\xi \mid x)EL_n(X_1) - \int_{-\infty}^{\xi} q(x, w)\, dw].$$

But

$$(nh)^{\frac{1}{2}}[h^{-1}G(\xi \mid x)EL_n(X_1) - \textstyle\int_{-\infty}^{\xi} q(x, w)\, dw = \int_{-\infty}^{\xi} t(z \mid x)\, dz \cdot (nh)^{\frac{1}{2}}[h^{-1}EK((x - X_1)h^{-1}) - p(x)],$$

and

$$(nh)^{\frac{1}{2}}[h^{-1}EK((x - X_1)h^{-1}) - p(x)] \to 0. \tag{5.12}$$

In fact,

$$(nh)^{\frac{1}{2}}[h^{-1}EK((x - X_1)h^{-1}) - p(x)] \leqq M(x)h(nh)^{\frac{1}{2}} \textstyle\int |z|\, K(z)\, dz$$

by assumption (P5). Finiteness of $\int |z|\, K(z)\, dz$ (assumption (K4) with $r = 1$) together with the fact that $nh^3 \to 0$, which is mplied by assumption (H3), completes the proof of (5.12). Next one has

$$(nh)^{\frac{1}{2}}[h^{-1}EL_n^*(Y_1) - \textstyle\int_{-\infty}^{\xi} q(x, w)\, dw] \leqq hn^{\frac{1}{2}} \int_{-\infty}^{\xi} M(x, x')\, dx' \int\int (v^2 + w^2)^{\frac{1}{2}}K(v)K(w)\, dv\, dw$$

by assumption (P6)(i). Since

$$\int\int (v^2 + w^2)^{\frac{1}{2}} K(v)K(w)\,dv\,dw$$
$$\leqq \int\int (|v| + |w|)K(v)K(w)\,dv\,dw = 2\int |z|\,K(z)\,dz,$$

assumptions (K4) (with $r = 1$), (H3), and (P6)(ii), imply that

$$(5.13)\qquad (nh)^{\frac{1}{2}}[h^{-1}EL_n^*(Y_1) - \textstyle\int_{-\infty}^{\xi} q(x, w)\,dw] \to 0.$$

Relations (5.13) and (5.12) taken together, conclude the proof of the lemma.

Now Lemma 5.1, Lemma 5.2, Lemma 5.3, and relation (5.11) provide all that is needed for the formulation and proof of the second main result in this section. That is,

THEOREM 5.2. *Let assumptions* (K1), (K2), (K4) (*with* $r = 1$), (K5), (H1), (H3), (P1), (P3), (P4), (P5), *and* (P6)(i), (P6)(ii) *be satisfied. Then*

$$\mathcal{L}\{(nh)^{\frac{1}{2}}[\xi_n(p, x) - \xi(p, x)]\} \to N(0, \tau^2(\xi, x)),$$

where $\tau^2(\xi, x) = t^{-2}(\xi \mid x)G^2(\xi \mid x)\int K^2(z)\,dz$, $0 < p < 1$, $x \varepsilon R$.

PROOF. By (5.1),

$$(nh)^{\frac{1}{2}}(\xi_n - \xi) = -(nh)^{\frac{1}{2}}[G_n(\xi \mid x) - G(\xi \mid x)]t_n^{-1}(\zeta_n \mid x),$$

and this is further rewritten as

$$(nh)^{\frac{1}{2}}(\xi_n - \xi) = -\{(nh)^{\frac{1}{2}}[G_n(\xi \mid x) + v_n] - (nh)^{\frac{1}{2}}[v_n + G(\xi \mid x)]\}t_n^{-1}(\zeta_n \mid x).$$

By Lemma 5.1, Lemma 5.3, and relation (5.11), this last expression converges in law to a normal law. By relation (5.11) again, Lemma 5.1, and Lemma 5.2, the mean of the limiting normal distribution is zero, and its variance $\tau^2(\xi, x)$ is given by

$$\begin{aligned}\tau^2(\xi \mid x) &= t^{-2}(\xi \mid x)\sigma_0^2(\xi, x)p^{-1}(x)\\ &= t^{-2}(\xi \mid x)G^2(\xi \mid x)p(x)\int K^2(z)\,dz p^{-1}(x)\\ &= t^{-2}(\xi \mid x)G^2(\xi \mid x)\int K^2(z)\,dz.\end{aligned}$$

The proof of the theorem is concluded.

REFERENCES

[1] DOOB, J. (1953). *Stochastic Processes.* Wiley, New York.
[2] LOÈVE, M. (1963). *Probability Theory,* (3rd ed.). Van Nostrand, N.J.
[3] NADARAYA, E. A. (1964). Some new estimates for distribution functions. *Theo. Probability Appl.* **9** 491–500.
[4] PAREN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.
[5] PRATT, J. W. (1960). On interchanging limits and integrals. *Ann. Math. Statist.* **31** 74–77.
[6] RAO, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

[7] ROUSSAS, G. G. (1967). Asymptotic normality of certain functions defined on a Markov process. Technical Report No. 109. University of Wisconsin, Madison.

[8] ROUSSAS, G. G. (1967). Nonparametric estimation in Markov processes. Technical Report No. 110. University of Wisconsin, Madison.

[9] ROUSSAS, G. G. (1968). Nonparametric estimation of the transition distribution function of a Markov process. Technical Report No. 160. University of Wisconsin, Madison.