DEPARTMENT OF STATISTICS

The University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT No. 110

April, 1967

NONPARAMETRIC ESTIMATION IN MARKOV PROCESSES

By

George G. Roussas

This paper was partially supported by the National Science Foundation, Grant GP-6242 and part by the United States Navy through the Office of Naval Research under Contract Nonr-1202,17, Project NR 042 222.

NONPARAMETRIC ESTIMATION IN MARKOV PROCESSES

By

George G. Roussas

University of Wisconsin, Madison, Wisconsin

1. Introduction and Summary. The problem os statistical inferences in Markov processes has received considerable attention during the last fifteen years. Much of the work consists in carrying over to the Markov case the maximum likelihood and chi-square methods from processes with independent identically distributed random variables. (See, for example, [1] and other references cited there.) Alternative approaches have also been adopted [10], some of which [6] refer to statistical inferences in more general processes.

It is not long ago that presumably the first paper [9] appeared on nonparametric estimation of the density in the case of independent identically distributed random variables. Soon a number of others [14], [8], [13], [3], [5] followed which by using either similar or different methods obtained further results.

The purpose of the present paper is to consider the non-parametric estimation of densities in the case of Markov processes. The methods being used and results being obtained here are similar to those in [8]. What we do specifically here is this: We first construct asymptotically unbiased estimates for the initial and (two-dimensional) joint densities. This is done in section 2. In section 3 these estimates are shown to be consistent in quadratic mean, and furthermore a consistent, in the probability sense, estimate for the transition density is obtained. Finally, it is proved in section 4 that, under suitable conditions, all three estimators mentioned, properly normalized, are asymptotically normal.

2. Asymptotically unbiased estimation of the initial and (two-dimensional) joint densities. The results of this paper, like those of [8] and [3], rely heavily on a slight variation of a theorem of Bochner [2] that we formulate and prove here. By C(f) we will denote the set of continuity points of the function f.

Theorem 2.1. Let $(\mathcal{E}_m, \mathcal{B}^{(m)})$ be the m-dimensional Euclidean space with the corresponding Borel σ-field and $(R, \mathcal{B})$ the Borel real line, and let K: $(\mathcal{E}_m, \mathcal{B}^{(m)}) \to (R, \mathcal{B})$ measurable and such that

(2.1) $|K(z)| \leqq M_1 (< \infty), \; z \in \mathcal{E}_m; \; \int |K(z)| dz < \infty.$

(2.2) $||z||^m | K(z)| \to 0, \text{ as } \|z\| \to \infty,$

where $\|\cdot\|$ is the usual norm in $\mathcal{E}_m$, and integrals without limits here and thereafter are assumed to be taken over the whole space.

Furthermore, let $g : (\mathcal{E}_m, \mathcal{B}^{(m)}) \to (R, \mathcal{B})$ measurable and such that

(2.3) $\int |g(z)| \, dz < \infty .$

Define

(2.4) $g_n(x) = h^{-m}(n) \int K(zh^{-1}(n)) g(x-z) dz ,$

where $\{h(n)\}$, $n = 1, 2, \ldots$ is a sequence of positive constants such that

(2.5) $h(n) \to 0, \text{ as } n \to \infty.$

Then for $x \in C(g)$

(2.6) $\lim g_n(x) = g(x) \int K(z) \, dz, \text{ as } n \to \infty.$

If g is continuous on $\mathcal{E}_m$, then the convergence (2.6) is uniform on compact subsets of $\mathcal{E}_m$.

(two-dimensional) joint density with respect to Lebesgue measure. It is assumed that p is strictly positive on R. Then (q/p)= t is a transition density of the process.

For i=1, 2 we consider two functions K_i such that $K_i:(\mathscr{E}_i, \mathscr{B}^{(i)}) \to (R, \mathscr{B})$ measurable and satisfying conditions (2.1), (2.2), and (2.7). On the basis then of the first n+1 random variables X_j, j=1, ..., n+1 of the Markov process we define the following random variables (suppressing the random element $\omega \varepsilon \Omega$)

(2.8) $$p_n(x)=(nh_1(n))^{-1} \Sigma_{j=1}^n K_1((x-X_j)h_1^{-1}(n)), \ x \varepsilon \mathscr{E}_1$$

(2.9) $$q_n(y)=(nh_2^2(n))^{-1} \Sigma_{j=1}^n K_2((y-Y_j)h_2^{-1}(n)), \ y \varepsilon \mathscr{E}_2,$$

where $Y_j = (X_j, X_{j+1})$, j=1, ..., n, and $h_1(n)$, $h_2(n)$ satisfy (2.5). For convenient reference we will denote by (C_i') the assumption that K_i, h_i satisfy (2.1), (2.2), (2.7), and (2.5), i = 1, 2. We intend to show that p and q are asymptotically unbiased estimates of p and q, respectively. More precisely

Theorem 2.2. Asymptotic unbiasedness. Under (C_1') and (C_2'), respectively, the random variables defined by (2.8) and (2.9) are asymptotically unbiased estimates of p and q, respectively, in the sense that

$$Ep_n(x) \to p(x), \text{ as } n \to \infty, \; x \in C(p),$$

and

$$Eq_n(y) \to q(y), \text{ as } n \to \infty, \; y \in C(q).$$

Furthermore these estimates are uniformly asymptotically unbiased on compact subsets of $\mathcal{E}_i$, i=1,2 if p and q are continuous on $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively.

Proof. The proof is an immediate application of Theorem 2.1. In fact, writing h_1 and h_2 instead of $h_1(n)$ and $h_2(n)$, we get

$$Ep_n(x) = h_1^{-1} \int K_1((x-z)h_1^{-1}) \; p(z)\, dz$$

$$= h_1^{-1} \int K_1(zh_1^{-1})\, p(x-z)\, dz,$$

and, as $n \to \infty$, this converges to p(x), provided $x \in C(p)$; this convergence is uniform on compact subsets of $\mathcal{E}_1$ if p is continuous. Similarly,

$$Eq_n(y) = h_2^{-2} \int K_2((y-z)h_2^{-1})q(z)\, dz$$

$$= h_2^{-2} \int K_2(zh_2^{-1})q(y-z)\, dz,$$

and, as $n \to \infty$, this converges to q(y), provided $y \in C(q)$; this convergence is uniform on compact subsets of $\mathcal{E}_2$ if q is continuous.

3. Consistent and uniform consistent estimation. The results of this section as well as those of the next one are derived under the additional assumption that the process satisfies hypothesis D_o([4], p. 221). Namely,

Hypothesis (D_o)(a) Condition (D) (Doeblin's condition) is satisfied,

and

(b) there is only a single ergodic set and this set contains no cyclically moving subsets.

We first prove consistency in quadratic mean. We have

$$E[p_n(x)-p(x)]^2 = \sigma^2[p_n(x)] + [Ep_n(x)-p(x)]^2 ,$$

while the second term converges to zero, as $n \to \infty$, provided $x \in C(p)$; this convergence is uniform, on compact subsets of $\mathcal{E}_1$ if p is continuous, by Theorem 2.2. Next,

$$\sigma^2[p_n(x)] = n^{-1}h_1^{-2}\,\sigma^2\,[K_1(\,(x-X_1)h_1^{-1})] + 2(nh_1)^{-2}\,\Sigma_{i<j}\,\mathrm{Cov}\,[K_1(\,(x-X_i)h_1^{-1}),\ K_1((x-X_j)\,h_1^{-1}\,] ,$$

where the summation extends over all i's and j's such that $1 \leqq i < j \leqq n$. But

$$h_1^{-1}\,\sigma^2\,[K_1(\,(x-X_1)h_1^{-1})] = h_1^{-1}\int K_1^2(\,(x-z)h_1^{-1})p(z)\,dz - h_1[h_1^{-1}\int K_1(\,(x-z)h_1^{-1})p(z)\,dz]^2 ,$$

and for $x \in C(p)$ this tends to $p(x)\int K_1^2\,(z)dz$, as $n \to \infty$, by Theorem 2.1, since $\int K_1^2\,(z)\,dz$ is finite, as is easily seen from (2.1). This convergence is uniform on compact subsets of $\mathcal{E}_1$ if p is continuous. As for the covariance we have:

Under hypothesis (D_o), Lemma 7.1, p. 222 in [4] applies and gives

$$|\mathrm{Cov}\,[K_1(\,(x-X_1)h_1^{-1},\ K_1((x-X_{j+1})h_1^{-1}]| \leqq 2\,\gamma^{\frac{1}{2}}\,\rho^{\frac{j}{2}}\;E\;|\;K_1(\,(x-X_1)h_1^{-1}\,)\,|^2$$

for some γ, ρ such that $\gamma > 0$, $0 < \rho < 1$.

Therefore

$$|(nh_1)^{-1}\;\Sigma_{i<j}\,\mathrm{Cov}\,[K_1((x-X_i)h_1^{-1}\,),\ K_1(\,(x-X_j)\,h_1^{-1})]\;| \leqq (nh_1)^{-1}\;.$$

$$\Sigma_{j=1}^{n-1}\;(n-j)2\,\gamma^{\frac{1}{2}}\rho^{\frac{j}{2}}\;E\;|\;K_1(\,(x-X_1)h_1^{-1})|^2 \leqq (nh_1)^{-1}\,n\,\rho^{\frac{1}{2}}(1-\rho^{\frac{1}{2}})^{-1}\;.$$

$$2\,\gamma^{\frac{1}{2}}\,E\;|\;K_1(\,(x-X_1)h_1^{-1})\;|^2 = 2\,\gamma^{\frac{1}{2}}\rho^{\frac{1}{2}}(1-\rho^{\frac{1}{2}})^{-1}\,h_1^{-1}\,E\;|K_1(\,(x-X_1)h_1^{-1}\,)\;|^2 ,$$

and this last expression converges, as $n \to \infty$, to

$2\,\gamma^{\frac{1}{2}}\,\rho^{\frac{1}{2}}(1-\rho^{\frac{1}{2}})^{-1}\,p(x)\int K_1^2(z)\,dz$ for $x \in C(p)$, and the convergence is uniform on compact subsets of $\mathcal{E}_1$ if p is continuous. Thus, if we assume that $h_1 = h_1(n)$ can be chosen so that

(3.1) $\quad nh_1(n) \to \infty$, as $n \to \infty$, it follows that

$\sigma^2[p_n(x)] \to 0$, as $n \to \infty$, $x \in C(p)$, and this convergence is uniform on compact subsets of $\mathcal{E}_1$ if p is continuous. Denoting, for convenience, by (C_1) the assumption that both (C_1) and (3.1) are satisfied, we get then: Under (C_1) and (D_o) $E[p_n(x)-p(x)]^2 \to 0$, as $n \to \infty$, provided $x \in C(p)$, and this convergence is uniform on compact subsets of $\mathcal{E}_1$ if p is continuous.

In a similar fashion we get that: Under (C_2) and (D_o) $E[q_n(y)-q(y)]^2 \to 0$, as $n \to \infty$, provided $y \in C(q)$, and this convergence is uniform on compact subsets of $\mathcal{E}_2$ if q is continuous. Here by (C_2) we denote the assumption that both (C_2) and (3.2) are satisfied, where

(3.2) $\quad nh_2(n) \to \infty$, as $n \to \infty$.

Putting together these results we have the following theorem:

Theorem 3.1. Consistency in quadratic mean (q.m.). Under (D_o) and (C_1), (C_2), respectively, the random variables defined by (2.8) and (2.9) are consistent in q.m. estimates of p and q, respectively, for $x \in C(p)$, $y \in C(q)$; and they are uniformly consistent in q.m. estimates on compact subsets of $\mathcal{E}_i$, i=1,2 if p and q are continuous.

Of course, consistency in q.m. (and Tchebichev inequality) implies consistency in the probability sense for $x \in C(p)$, $y \in C(q)$, and this consistency is uniform on compact subsets of $\mathcal{E}_i$, i = 1,2 if p and q are continuous.

By taking into account now that the random variables (2.8) are to be used in order to estimate the positive quantity p, one may assume that K_1 is strictly positive. Under this condition a consistent estimate of the transistion density can be constructed. More precisely

<u>Corollary</u> 3.1. Let $y = (x, x') \in C(q)$ and $x \in C(p)$.

We set

$$t_n(x'|x) = [q_n(y)/p_n(x)] \text{ and } t(x'|x) = [q(y)/p(x)].$$

Then, as $n \to \infty$,

$t_n(x'|x) \to t(x'|x)$ in probability, and this convergence is uniform on compact subsets of $\mathcal{E}_2$ if p and q are continuous.

4. <u>Asymptotic normality.</u> In this section asymptotic normality of the estimators p_n, q_n, and t_n will be obtained, under some further restrictions on the process. Actually, these results are merely an application of the results obtained in [12], and have also served as a motivation for the type of assumption being made there.

In (A2) of [12] we take $h_n = nh_1(n)$. Then (A2) is satisfied on account of (3.1) herein. Next for r=1 in (A3) of [12] and with $L_n(z)$ being replaced by $K_1((x-z)h_1^{-1})$, (A3)(i) and (A3)(iv) are automatically satisfied on the basis of Theorem 2.1 here with $\sigma_1^2(x) = p(x) \int K_1^2(z)dz$, $x \in C(p)$. As for (A3)(ii) and (A3)(iii) they clearly follow from the assumption being made below.

The joint densities of X_1, X_i and X_1, X_i, X_j

(4.1) are bounded by $M_2 (< \infty)$ for all i, j such that

$$1 < i \leqq n, \quad 1 < i < j \leqq n, \quad n = 2, 3, \ldots$$

In [12] the positive integers α, β, and μ were introduced with the property that they tended to infinity together with n and also satisfied the properties:

$$\beta \mu \alpha^{-1} \to 0 \text{ and } \alpha h_n n^{-1} \to 0, \text{ as } n \to \infty .$$

With the above choice of h_n these relations become

(4.2) $\beta \mu \alpha^{-1} \to 0$ and $\alpha h_1(n) \to 0$, as $n \to \infty$.

Theorem 2.1.1 in [12] then becomes

Theorem 4.1. Let assumptions (D_o), (C_1), (4.1), and (4.2) be satisfied. Then for $x \in C(p)$

$$\mathcal{L}\{(nh_1)^{\frac{1}{2}} [p_n(x) - Ep_n(x)]\} \to N(0, \sigma_1^2(x)), \text{ as } n \to \infty,$$

where

$$\sigma_1^2(x) = p(x) \int K_1^2(z)\, dz.$$

We next choose $h_n = nh_2^2(n)$ and then (A2) in [12] is again satisfied by (3.2) herein. For s=2 in $(A3)^*$ of [12] and with $L_n^*(z)$ being replaced by $K_2((y-z)h_2^{-1})$, $(A3)^*$ (i) and $(A3)^*$ (iv) are automatically satisfied on account of Theorem 2.1 of this paper with $\sigma_2^2(y) = q(y) \int K_2^2(z)\, dz$, $y \in C(q)$. As for $(A3)^*$ (ii) and $(A3)^*$ (iii) they follow in an obvious way from (4.3) below.

The joint densities of Y_1, Y_i and Y_i, Y_i, Y_j are bounded by $M_3(<\infty)$ for all i, j such that

(4.3) $1 < i \leqq n,\ 1 < i < j \leqq n, \quad n = 2, 3, \ldots$

We finally require α, β, and μ to tend to infinity, as $n \to \infty$, and be such that

(4.4) $\beta \mu \alpha^{-1} \to 0$ and $\alpha h_2^2(n) \to 0$, as $n \to \infty$.

Then Theorem 2.2.1 in [12] becomes

Theorem 4.2. Let assumptions (D_o), (C_2), (4.3), and (4.4) be satisfied. Then for $y \in C(q)$ and such that $q(y) > 0$ we have

$$\mathcal{L}\{(nh_2^2)^{\frac{1}{2}} [q_n(y) - Eq_n(y)]\} \to N(0, \sigma_2^2(y)), \text{ as } n \to \infty, \text{ where}$$

$$\sigma_2^2(y) = q(y) \int K_2^2(z)\, dz.$$

Finally we will examine the estimator of the transition density from the point of view of asymptotic normality.

In the first place we take $h_1(n) = h_2^2(n)=h(n)$, $n=1,2,\ldots$ for simplicity. Thus h_n in [12] is now $h_n=nh_1(n) = nh_2^2(n)$. Next $(A2)^{**}(i)$ in [12] again is clearly true, and so is $(A2)^{**}(ii)$ with ℓ being $p(x)$, $x \in C(p)$ because of Theorem 3.1 herein. Furthermore $(A4)^{**}(i)$ follows from (4.1) and (4.3), $(A4)^{**}(ii)$ is true with $v(x,y)= -[q(y)/p(x)]$ by Theorem 2.2 herein, provided $x \in C(p)$, $y \in C(q)$, and $(A4)^{**}(iii)$ is also valid with $\sigma = 0$ on account of (4.1). Therefore Theorem 2.3.1 in [12] becomes as follows

<u>Theorem</u> 4.3. Let assumptions (D_o), (C_1), (C_2), (4.1), (4.2), and (4.3) be satisfied. Then for $y=(x,x')\in C(q)$ such that $x \in C(p)$ we have that the law of

$$(nh)^{\frac{1}{2}} \{t_n(x'|x) - [EK_2((y-Y_1)h^{-\frac{1}{2}})/EK_1((x-X_1)h^{-1})]\}$$

converges to $N(0,\sigma_o^2(x,y)\,\ell^{-2}(x))$, as $n \to \infty$,

where $\sigma_o^2(x,y)= \sigma_2^2(y)+ v^2(x,y)\,\sigma_1^2(x)$ and

$$\sigma_1^2(x) = p(x) \int K_1^2(z)\,dz, \quad \sigma_2^2(y) = q(y) \int K_2^2(z)\,dz,$$

$v(x,y) = -[q(y)/p(x)]$, $\ell(x) = p(x)$, provided $q(y) > 0$.

REFERENCES

[1] BILLINGSLEY, P. (1961). Statistical Inference for Markov Processes. University of Chicago Press.

[2] BOCHNER, S. (1955). Harmonic Analysis and the Theory of Probability. University of California Press.

[3] CACOULLOS, THEO. (1966). Estimation of a Multivariate Density. Ann. of the Institute of Stat. Math. **18** 179-189.

[4] DOOB, J. (1953). Stochastic Processes. Wiley, New York.

[5] GRENANDER, U. (1965). Some Direct Estimates of the Mode. Ann. Math. Stat. **36** 131-138.

[6] LE CAM, L. (1960). Locally asymptotically normal families of distributions. University of California Publications in Statistics. **3** 37-98.

[7] LOÈVE M. (1963). Probability Theory, 3rd ed. Van Nostrand, N. J.

[8] PARZEN, E. (1962). On estimation of a probability density function and mode. Ann. Math. Stat. **33** 1065-1076.

[9] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. Ann. Math. Stat. **27** 832-837.

[10] ROUSSAS, G. (1965). Asymptotic Inference in Markov Processes. Ann. Math. Stat. **36** 978-992.

[11] ROUSSAS, G. (1965). Extension to Markov Processes of a result by Wald about the consistency of the maximum likelihood estimate. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **4** 69-73.

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

[13] WATSON, G. S. and LEADBETTER, M. R. (1963). On the estimation of the probability density, I. Ann. Math. Stat. **34** 480-491.

[14] WHITTLE, P. (1958). On the smoothing of probability density functions. J. Roy. Stat. Soc., Ser. B **20** 334-343.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

1. National Science Foundation and Office of Naval Research	2a. Unclassified
	2b. --

3. Nonparametric estimation in Markov processes.

4. --

5. George G. Roussas

6. April, 1967	7a. 11 pp.	7b. 14 references

8a. NSF-GP-6242 Nonr-1202,17	9a. Report number 110
8b. NR 042 222	9b. --

10. Distribution of this document is unlimited.

11. --	12. Office of Naval Research Washington, D.C.

13. Abstract. For i = 1, 2, let K_i be bounded, continuous probability densities defined on the i-dimensional Euclidean spaces $(\mathcal{E}_i, \mathcal{B}^{(i)})$ and satisfying the conditions: $||z||^i K_i(z) \to 0$, as $||z|| \to \infty$, $z \in \mathcal{E}_i$, and $K_1(z) > 0$, $z \in \mathcal{E}_1$. Let $\{X_n\}$, $n \geq 1$ be a Markov process having initial, 2-dimensional, and transition densities denoted by p, q, and t, respectively, and satisfying some additional regularity conditions. For two sequences of positive constants $\{h_i(n)\}$, $n \geq 1$ with the property that: $h_i(n) = h_i \to 0$, and $n^i h_i \to \infty$, as $n \to \infty$, we set: $p_n(x) = (nh_1)^{-1} \Sigma_{j=1}^n K_1[(x-X_j)h_1^{-1}]$, $q_n(y) = (nh_2^2)^{-1} \Sigma_{j=1}^n K_2[(y-Y_j)h_2^{-1}]$, $t_n(x'|x) = q_n(y)/p_n(x)$, where $x, x' \in \mathcal{E}_1$, $y = (x, x')$, $Y_j = (X_j, X_{j+1})$. Then the following theorems are proved: Theorem 1. The random variables $p_n(x)$ and $q_n(y)$ are asymptotically unbiased estimates of $p(x)$ and $q(y)$, respectively. Theorem 2. The random variables $p_n(x)$, $q_n(y)$, and $t_n(x'|x)$ are consistent estimates of $p(x)$, $q(y)$, and $t(x'|x)$, respectively, the first two in quadratic mean and the last one in probability. Theorem 3. All three estimates in Theorem 2, properly normalized, are asymptotically normal.

14. Key Words --