# ASYMPTOTIC INFERENCE IN MARKOV PROCESSES[1]

By G. G. Roussas[2]

*University of California, Berkeley*

**1. Introduction and summary.** Let $(\mathfrak{X}, \mathcal{A})$ be a measurable space and $\Theta$ be an open subset of a $k$-dimensional Euclidean space. For each $\theta \varepsilon \Theta$ let $P_\theta$ be a probability measure on $\mathcal{A}$. We assume that for every $\theta \varepsilon \Theta$, $\{X_n, n \geqq 0\}$ is a Markov process defined on $(\mathfrak{X}, \mathcal{A})$ into $(R, \mathcal{B})$, where $(R, \mathcal{B})$ denotes the Borel real line. Furthermore, we denote by $\mathcal{A}_n$ the $\sigma$-fields induced by the random variables $\{X_0, X_1, \cdots, X_n\}$, and by $P_{n,\theta}$ the restriction of $P_\theta$ to $\mathcal{A}_n$.

In the present paper we give conditions under which the sequence of families of probability measures $\{P_{n,\theta}, \theta \varepsilon \Theta\}$ has the desirable property of being differentially asymptotically normal. This implies that in some neighborhood of $\theta \varepsilon \Theta$, $\{P_{n,\theta}, \theta \varepsilon \Theta\}$, $n \geqq 0$ can be, for certain problems, approximately treated as if they were normal. For a detailed account of the notions involved in this paper the reader is referred to [5], in particular, Section 5. Also in the Appendix of the present paper one can find the definitions of the concepts most frequently used in this work, including that of the differentially asymptotically normal families of distributions.

In Section 2 the required notation is introduced and also the assumptions being made throughout the paper are listed. In Section 3 we state the main result, and in the following subsections we give the proof of it in several steps.

**2. Notation and assumptions.** In what follows we will take the measurable space $(\mathfrak{X}, \mathcal{A})$ to be of the form $(\mathfrak{X}, \mathcal{A}) = \prod_{i=0}^{\infty} (R, \mathcal{B})$. For each $\theta \varepsilon \Theta$ the probability measure $P_\theta$ will be the one induced on $\mathcal{A}$ by a set of transition probability measures $p_\theta(\cdot, \cdot)$ defined on $R \times \mathcal{B}$, and a probability distribution $p_\theta(\cdot)$ on $\mathcal{B}$, according to Kolmogorov's consistency theorem. $\{X_n, n \geqq 0\}$ will be taken to be the coordinate process, and then it will be a Markov process with initial distribution $p_\theta(\cdot)$ and (stationary) transition measures $p_\theta(\cdot, \cdot)$, We will assume in the following that the probability measures $\{P_{n,\theta}, \theta \varepsilon \Theta\}$, $n \geqq 0$ are absolutely continuous with respect to each other. Therefore for any $\theta, \theta' \varepsilon \Theta$ we will have $[dP_{0,\theta'}/dP_{0,\theta}] = q(X_0; \theta, \theta')$, $[dP_{1,\theta'}/dP_{1,\theta}] = q(X_0, X_1; \theta, \theta')$, and if we set $q(X_1 \mid X_0; \theta, \theta') = q(X_0, X_1; \theta, \theta')/q(X_0; \theta, \theta')$, we will then have for the joint measures $P_{n,\theta'}, P_{n,\theta}$: $[dP_{n,\theta'}/dP_{n,\theta}] = q(X_0; \theta, \theta') \cdot \prod_{j=1}^{n} q(X_j \mid X_{j-1}; \theta, \theta')$. It will prove convenient to set $[q(X_j \mid X_{j-1}; \theta, \theta')]^{\frac{1}{2}} = \phi_j(\theta, \theta')$, $[q(X_{j-1}, X_j; \theta, \theta')]^{\frac{1}{2}} = f_j(\theta, \theta')$, $j = 1, \cdots, n$. Then $[dP_{n,\theta'}/dP_{n,\theta}] = q(X_0; \theta, \theta') \cdot \prod_{j=1}^{n} \phi_j^2(\theta, \theta')$, while it is clear that $\int \phi_1^2(\theta, \theta')\, dP_{1,\theta}$ is finite, and, in fact, equal to 1.

Assumptions. (A1) For each $\theta \varepsilon \Theta$ the Markov process $\{X_n, n \geqq 0\}$ is stationary and metrically transitive (ergodic).

Received 23 January 1964; revised 4 January 1965.

[1] This paper was prepared with the partial support of the National Science Foundation, Grant GP-10.

[2] Now at San Jose State College.

(A2)(i) The probability measure $\{P_{n,\theta}, \theta \varepsilon \Theta\}$, $n \geqq 0$ are mutually absolutely continuous.

(ii) For any $\theta, \theta' \varepsilon \Theta$, $\theta' \neq \theta$ implies $P_{\theta'} \neq P_\theta$.

(A3)(i) For every $\theta \varepsilon \Theta$ the random function $\phi_1(\theta', \theta'')$ satisfies the following condition

$$\lambda^{-1}\{\phi_1(\theta + \lambda h, \theta + \lambda h + \lambda v) - 1\} \to v'\dot{\phi}_1(\theta) \quad \text{in quadradic mean (q.m.) } [P_\theta], \quad \text{as } \lambda \to 0,$$

uniformly on bounded sets of $h, v \varepsilon \mathcal{E}_k$.

Clearly, Condition (A3)(i) implies differentiability in q.m. $[P_\theta]$ of the random function $\phi_1(\theta, \theta')$ with respect to $\theta'$ at $(\theta, \theta)$.

(ii) If $\dot{\phi}_1(\theta)$ is the derivative of $\phi_1(\theta, \theta')$ with respect to $\theta'$ at $(\theta, \theta)$, then $\dot{\phi}_1(\theta)$ is continuous in $P_\theta$-probability at $\theta$, for every $\theta \varepsilon \Theta$, and $\mathcal{A}_2 \times \mathcal{C}$-measurable, where $\mathcal{C}$ denotes the $\sigma$-field of Borel subsets of $\Theta$.

(iii) $\mathcal{E}_\theta|h'\dot{\phi}_1(\theta)|^2 = \int |h'\dot{\phi}_1(\theta)|^2 \, dP_\theta$ is a continuous function of $\theta$, for any $h \varepsilon \mathcal{E}_k$.

Part (iii) of (A3) is replaced below by (iii)$'$ which implies (iii), as will be shown in the text, and which may be easier to verify in some instances. Namely

(iii)$'$ For each $\theta \varepsilon \Theta$ there exists a neighborhood $N_\theta$ such that for all $\theta' \varepsilon N_\theta$, $\mathcal{E}_{\theta'}|h'\dot{\phi}_1(\theta')|^{2+\delta} = \mathcal{E}_\theta\{|h'\dot{\phi}_1(\theta')|^{2+\delta} f_1^2(\theta, \theta')\} = C(\theta, h)(<\infty)$, for some $0 < \delta(< 1)$, any $h \varepsilon \mathcal{E}_k$.

Remark. All of the results, except for one, are derived under the assumption that $\phi_1(\theta, \theta')$ is differentiable with respect to $\theta'$ at $(\theta, \theta)$, with respect to $P_\theta$. There is only one occasion, Theorem 3.3.2(ii), where the full force of (A3)(i) is used.

(A4)(i) For every $\theta \varepsilon \Theta$ the random function $f_1(\theta, \theta')$ is differentiable in q.m. with respect to $\theta'$ at $(\theta, \theta)$, when $P_\theta$ is employed.

(ii) If $\dot{f}_1(\theta)$ is the derivative of $f_1(\theta, \theta')$ with respect to $\theta'$ at $(\theta, \theta)$, then $\dot{f}_1(\theta)$ is continuous in $P_\theta$-probability at $\theta$, and $|\dot{f}_1(\theta)| \neq 0$ with $P_\theta$-probability $> 0$, for every $\theta \varepsilon \Theta$.

(iii) $\mathcal{E}_\theta|h'\dot{f}_1(\theta)|^2 = \int |h'\dot{f}_1(\theta)|^2 \, dP_\theta$ is a continuous function of $\theta$, for any $h \varepsilon \mathcal{E}_k$.

Remark. The continuity assumption of $\mathcal{E}_\theta|h'\dot{f}_1(\theta)|^2$ may be replaced by a condition analogous to (A3)(iii)$'$.

This author has found that assumptions (A1)–(A4) are satisfied in a number of interesting examples. Details may be published elsewhere.

From (A2)(i) it follows that $[dP_{n,\theta'}/dP_{n,\theta}] = q(X_0; \theta, \theta') \cdot \prod_{j=1}^n \phi_j^2(\theta, \theta')$ is well defined except on $P_\theta$-null sets for all $\theta \varepsilon \Theta$. Disregarding these null sets we define the random variable $\Lambda[P_{n,\theta'}; P_{n,\theta}] = \log [dP_{n,\theta'}/dP_{n,\theta}] = \log [q(X_0; \theta, \theta') \cdot \prod_{j=1}^n \phi_j^2(\theta, \theta')]$, and from here on, unless otherwise explicitly stated, the basic probability measure to be used will be $P_{n,\theta}$.

**3. Results.** Under the above assumptions we will be able to prove the main result of this paper which is stated below.

Theorem 3.1. *Under assumptions* (A1) *to* (A4) *the sequence of families of probability measures* $\{P_{n,\theta}, \theta \varepsilon \Theta\}$ *is differentially asymptotically normal.*

The proof of Theorem 3.1, being long, will be given in several steps. First we remark that by considering any $\theta, \theta' \varepsilon \Theta$, fixing $\theta$, and letting $\theta'$ vary over $\Theta$, we do not allow it to vary in an entirely arbitrary way. In most cases we will take it to be of the form $\theta' = h \cdot n^{-\frac{1}{2}}$ with $h$ belonging to some bounded set in $\mathcal{E}_k$, or $\theta' = h_n \cdot n^{-\frac{1}{2}}$ with $h_n \varepsilon \mathcal{E}_k$ and $h_n \to h$, as $n \to \infty$. For such a choice of $\theta'$ the log-likelihood ratio of the corresponding probability measures $P_{n,\theta+h_n \cdot n^{-\frac{1}{2}}}$, $P_{n,\theta}$, i.e., $\Lambda[P_{n,\theta+h_n \cdot n^{-\frac{1}{2}}}; P_{n,\theta}]$, assumes a certain expansion in $P_{n,\theta}$-probability to be made precise by the following theorem.

3.1 THEOREM 3.1.1. *For $h_n$, $h \varepsilon \mathcal{E}_k$, $h_n \to h$, as $n \to \infty$, we have*

$$\Lambda[P_{n,\theta+h_n \cdot n^{-\frac{1}{2}}}; P_{n,\theta}] - h' \cdot \Delta_n(\theta) \to -A(h, \theta)$$

*in $P_{n,\theta}$-probability, as $n \to \infty$, where $A(h, \theta) = \frac{1}{2}h'\Gamma(\theta)h$, $h'\Gamma(\theta)h = \mathcal{E}_\theta|2h'\dot{\phi}_1(\theta)|^2$, $\Delta_n(\theta) = 2n^{-\frac{1}{2}} \cdot \sum_{j=1}^n \dot{\phi}_j(\theta)$.*

PROOF. This proof will be a consequence of a series of lemmas to be formulated and proved below. We consider the random variable $\Lambda[P_{n,\theta+h_n \cdot n^{-\frac{1}{2}}}; P_{n,\theta}] = \log q(X_0; \theta, \theta + h_n \cdot n^{-\frac{1}{2}}) + 2\sum_{j=1}^n \log \phi_j(\theta, \theta + h_n \cdot n^{-\frac{1}{2}})$, and we set $\phi_{nj}(\theta) = \phi_j(\theta, \theta + h_n \cdot n^{-\frac{1}{2}})$. Then the first lemma is to the effect that the quantities $\log \phi_{nj}(\theta)$ appearing in the definition of $\Lambda[P_{n,\theta+h_n \cdot n^{-\frac{1}{2}}}; P_{n,\theta}]$ can be replaced, asymptotically, by the quantities $\phi_{nj}(\theta) - 1$. More precisely,

LEMMA 3.1.1.

(i) $\Lambda[P_{n,\theta+h_n \cdot n^{-\frac{1}{2}}}; P_{n,\theta}] - 2\{\sum_{j=1}^n [\phi_{nj}(\theta) - 1] - \frac{1}{2}\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2\} \to 0$ *in $P_\theta$-probability, as $n \to \infty$.*

(ii) $\Lambda[P_{n,\theta+h_n \cdot n^{-\frac{1}{2}}}; P_{n,\theta}] - \{\sum_{j=1}^n [\phi_{nj}^2(\theta) - 1] - \frac{1}{2}\sum_{j=1}^n [\phi_{nj}^2(\theta) - 1]^2\} \to 0$ *in $P_\theta$-probability, as $n \to \infty$.*

PROOF. (i) The differentiability assumption of $\phi_1(\theta, \theta')$ and $f_1(\theta, \theta')$ implies their continuity in probability, and hence the continuity in probability of $q(X_0; \theta, \theta')$. Observing that $q(X_0; \theta, \theta) = 1$ we get then $\log q(X_0; \theta, \theta + h_n \cdot n^{-\frac{1}{2}}) \to 0$ in $P_\theta$-probability, as $n \to \infty$, and therefore all we have to show is $\sum_{j=1}^n \phi_{nj}^2(\theta) - 2\{\sum_{j=1}^n [\phi_{nj}(\theta) - 1] - \frac{1}{2}\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2\} \to 0$ in $P_\theta$-probability, as $n \to \infty$.

Now assumption (A3)(i) implies

$$|\lambda^{-1}| \cdot |\phi_j(\theta, \theta + \lambda h) - \phi_j(\theta, \theta) - \lambda h'\dot{\phi}_j(\theta)| \to 0 \quad \text{in q.m. } [P_\theta], \quad \text{as} \quad \lambda \to 0, \tag{3.1.1}$$

uniformly on bounded sets of $h \varepsilon \mathcal{E}_k$, for all $j$ $(1 \leqq j \leqq n)$. By taking the rate of convergence of $\lambda$ to be $n^{-\frac{1}{2}}$, observing that $\phi_j(\theta, \theta) = 1$, and picking any sequence $\{h_n\}$ with $h_n \varepsilon \mathcal{E}_k$ and $h_n \to h$, as $n \to \infty$, we have

$$n^{\frac{1}{2}}[\phi_{nj}(\theta) - 1] \to h'\dot{\phi}_j(\theta) \quad \text{in q.m. } [P_\theta], \quad \text{as} \quad n \to \infty, \quad \text{for all } j. \tag{3.1.2}$$

This last relation implies $\phi_{nj}(\theta) - 1 \to 0$ in $P_\theta$-probability, as $n \to \infty$, for all $j$. In fact, more is true; namely

$$\max\{|\phi_{nj}(\theta) - 1|; 1 \leqq j \leqq n\} \to 0 \quad \text{in } P_\theta\text{-probability}, \quad \text{as} \quad n \to \infty. \tag{3.1.3}$$

To see this we write $\phi_{nj}(\theta) - 1 = n^{-\frac{1}{2}} \cdot h'\dot{\phi}_j(\theta) + n^{-\frac{1}{2}} \cdot R_{nj}(\theta, h)$, where $\mathcal{E}_\theta|R_{nj}(\theta, h)|^2 \to 0$, as $n \to \infty$, as is seen from (3.1.2). $P_\theta[\max\{|\phi_{nj}(\theta) - 1|;$

$1 \leqq j \leqq n\} > \epsilon] \leqq P_\theta[\max\{|h'\phi_j(\theta)|; 1 \leqq j \leqq n\} > \epsilon n^{\frac{1}{2}}/2] + P_\theta[\max\{|R_{nj}(\theta, h)|; 1 \leqq j \leqq n\} > \epsilon n^{\frac{1}{2}}/2] \leqq n \cdot P_\theta[|h'\phi_1(\theta)| > \epsilon n^{\frac{1}{2}}/2] + n \cdot P_\theta[|R_{n1}(\theta, h)| > \epsilon n^{\frac{1}{2}}/2]$. But $n \cdot P_\theta[|R_{n1}(\theta, h)| > \epsilon n^{\frac{1}{2}}/2] \leqq n4 \cdot \epsilon^{-2} \cdot n^{-1} \cdot \mathcal{E}_\theta|R_{n1}(\theta, h)|^2 = 4\epsilon^{-2} \cdot \mathcal{E}_\theta|R_{n1}(\theta, h)|^2 \to 0$, as $n \to \infty$, and $n \cdot P_\theta[|h'\phi_1(\theta)| > \epsilon n^{\frac{1}{2}}/2] = n \cdot \int_{\epsilon n^{\frac{1}{2}}/2}^\infty dF(z)$, where $F(z)$ denotes the distribution function of the random variable $|h'\phi_1(\theta)| = Z$. Next

$$\tfrac{1}{4} \cdot n \cdot \epsilon^2 \int_{\epsilon n^{\frac{1}{2}}/2}^\infty dF(z) = \int_{\epsilon n^{\frac{1}{2}}/2}^\infty (\epsilon n^{\frac{1}{2}}/2)^2\, dF(z) \leqq \int_{\epsilon n^{\frac{1}{2}}/2}^\infty z^2\, dF(z) \leqq \int z^2\, dF(z) = \mathcal{E}_\theta|h'\phi_1(\theta)|^2 < \infty.$$

Therefore

$$\int_{\epsilon n^{\frac{1}{2}}/2}^\infty z^2\, dF(z) \to 0, \quad \text{as} \quad n \to \infty$$

which implies

$$n \cdot \int_{\epsilon n^{\frac{1}{2}}/2}^\infty dF(z) \to 0, \quad \text{as} \quad n \to \infty.$$

This completes the proof of (3.1.3). If we set $A_n(\theta) = [\max \cdot \{|\phi_{nj}(\theta) - 1|; 1 \leqq j \leqq n\} > \epsilon]$, then (3.1.3) implies that for $n$ sufficiently large we have $|\phi_{nj}(\theta) - 1| \leqq \epsilon$, $1 \leqq j \leqq n$ on $A_n^c(\theta)$ with $P_\theta(A_n^c(\theta)) \geqq 1 - \epsilon$, and hence the expansion $\log y = \log\{1 + (y - 1)\} = (y - 1) - \frac{1}{2}(y - 1)^2 + \alpha \cdot (y - 1)^3$, $|\alpha| < 1$, with $y = \phi_{nj}(\theta)$ holds simultaneously, for $1 \leqq j \leqq n$ on $A_n^c(\theta)$; i.e., $\log \phi_{nj}(\theta) = [\phi_{nj}(\theta) - 1] - \frac{1}{2}[\phi_{nj}(\theta) - 1]^2 + \alpha_{nj}[\phi_{nj}(\theta) - 1]^3$, $|\alpha_{nj}| < 1$, $1 \leqq j \leqq n$, and hence $\sum_{j=1}^n \log \phi_{nj}(\theta) = \sum_{j=1}^n [\phi_{nj}(\theta) - 1] - \frac{1}{2}\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2 + \sum_{j=1}^n \alpha_{nj}[\phi_{nj}(\theta) - 1]^3$. But $\sum_{j=1}^n \alpha_{nj}[\phi_{nj}(\theta) - 1]^3 \to 0$ in $P_\theta$-probability, as $n \to \infty$, because $|\sum_{j=1}^n \alpha_{nj} \cdot [\phi_{nj}(\theta) - 1]^3| \leqq \{\max \cdot |\phi_{nj}(\theta) - 1|; 1 \leqq j \leqq n\} \sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2 \to 0$ in $P_\theta$-probability, as $n \to \infty$, since $\max \cdot \{|\phi_{nj}(\theta) - 1|; 1 \leqq j \leqq n\} \to 0$ in $P_\theta$-probability, as $n \to \infty$, and $\sum_{j=1}^n |\phi_{nj}(\theta) - 1|^2 \to \mathcal{E}_\theta|h'\phi_1(\theta)|^2 (< \infty)$, in $P_\theta$-probability, as $n \to \infty$, as will be shown in the next lemma. Therefore, $\sum_{j=1}^n \log \phi_{nj}(\theta) - \{\sum_{j=1}^n [\phi_{nj}(\theta) - 1] - \frac{1}{2}\sum_{j=1}^n |\phi_{nj}(\theta) - 1|^2\} \to 0$ in $P_\theta$-probability, as $n \to \infty$, and hence $\sum_{j=1}^n \log \phi_{nj}^2(\theta) - 2\{\sum_{j=1}^n [\phi_{nj}(\theta) - 1] - \frac{1}{2}\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2\} \to 0$ in $P_\theta$-probability, as $n \to \infty$, which completes the proof of (i).

(ii) It is readily seen that (3.1.3) implies

$$(3.1.4) \quad \max \cdot \{|\phi_{nj}^2(\theta) - 1|; 1 \leqq j \leqq n\} \to 0 \quad \text{in } P_\theta\text{-probability}, \quad \text{as} \quad n \to \infty,$$

since on the set $A_n^c(\theta)$ we have $\max \cdot \{|\phi_{nj}^2(\theta) - 1|; 1 \leqq j \leqq n\} \leqq (2 + \epsilon)\epsilon$, $(\epsilon \leqq \frac{1}{3})$. Hence the expansion

$$\log \phi_{nj}^2(\theta) = [\phi_{nj}^2(\theta) - 1] - \tfrac{1}{2}[\phi_{nj}^2(\theta) - 1]^2 + \beta_{nj}[\phi_{nj}(\theta) - 1]^3, \quad |\beta_{nj}| < 1,$$

on the set $A_n^c(\theta)$ is justified as above, and we get

$$\sum_{j=1}^n \log \phi_{nj}^2(\theta) = \sum_{j=1}^n [\phi_{nj}^2(\theta) - 1] - \tfrac{1}{2}\sum_{j=1}^n [\phi_{nj}^2(\theta) - 1]^2 + \sum_{j=1}^n \beta_{nj}[\phi_{nj}^2(\theta) - 1]^3.$$

We now observe that $|[\phi_{nj}^2(\theta) - 1]^3 - [\phi_{nj}(\theta) - 1]^3| = |\phi_{nj}(\theta) - 1|^3 \cdot \{[\phi_{nj}(\theta) + 1]^3 - 1\} < 26 \cdot |\phi_{nj}(\theta) - 1|^3$ on the set $A_n^c(\theta)$. Then $\sum_{j=1}^n|[\phi_{nj}^2(\theta) - 1]^3$

$- [\phi_{nj}(\theta) - 1]^3| < 26 \cdot \sum_{j=1}^n |\phi_{nj}(\theta) - 1|^3 \to 0$ in $P_\theta$-probability, as $n \to \infty$, and therefore $\sum_{j=1}^n \beta_{nj}[\phi_{nj}^2(\theta) - 1]^3 \to 0$ in $P_\theta$-probability, as $n \to \infty$, and this completes the proof of (ii).

Now we are going to show a convergence used in the proof of the previous lemma. Namely

LEMMA 3.1.2. (i) $\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2 \to \mathcal{E}_\theta|h'\phi_1(\theta)|^2$ *in $P_\theta$-probability, as* $n \to \infty$.
(ii) $\sum_{j=1}^n [\phi_{nj}^2(\theta) - 1]^2 \to \mathcal{E}_\theta|2h'\phi_1(\theta)|^2$ *in $P_\theta$-probability, as* $n \to \infty$.

PROOF. (i) We have seen that Assumption (A3)(i) implies $n^{\frac{1}{2}} \cdot [\phi_{n1}(\theta) - 1] \to h'\phi_1(\theta)$ in q.m. $[P_\theta]$, as $n \to \infty$, and hence

$$(3.1.5) \quad n \cdot [\phi_{n1}(\theta) - 1]^2 - [h'\phi_1(\theta)]^2 \to 0 \quad \text{in the first mean } [P_\theta], \quad \text{as} \quad n \to \infty.$$

Now $n^{-1} \cdot \sum_{j=1}^n [h'\phi_j(\theta)]^2 \to \mathcal{E}_\theta|h'\phi_1(\theta)|^2$ a.s. $[P_\theta]$, as $n \to \infty$, by the ergodic theorem. But $\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2 - n^{-1} \cdot \sum_{j=1}^n [h'\phi_j(\theta)]^2 = n^{-1} \cdot \sum_{j=1}^n \{n \cdot [\phi_{nj}(\theta) - 1]^2 - [h'\phi_j(\theta)]^2\}$, and $P_\theta[n^{-1}|\sum_{j=1}^n n \cdot [\phi_{nj}(\theta) - 1]^2 - [h'\phi_j(\theta)]^2\}| > \epsilon] \leqq n^{-1}\epsilon^{-1} \cdot n \cdot \mathcal{E}_\theta|n \cdot [\phi_{n1}(\theta) - 1]^2 - [h'\phi_1(\theta)]^2| \to 0$, as $n \to \infty$ by (3.1.5). Then $\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2 \to \mathcal{E}_\theta|h'\phi_1(\theta)|^2$ in $P_\theta$-probability, as $n \to \infty$ which is (i).

(ii) From (i) $\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2 \to \mathcal{E}_\theta|h'\phi_1(\theta)|^2$ in $P_\theta$-probability, as $n \to \infty$, and hence $4\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2 \to \mathcal{E}_\theta|2h'\phi_1(\theta)|^2$ in $P_\theta$-probability, as $n \to \infty$. So it suffices to show

$$(3.1.6) \quad \sum_{j=1}^n [\phi_{nj}^2(\theta) - 1]^2 - 4\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^2 \to 0 \quad \text{in } P_\theta\text{-probability}, \quad \text{as} \quad n \to \infty.$$

By observing that $[\phi_{nj}^2(\theta) - 1]^2 = [\phi_{nj}(\theta) - 1]^4 + 4\phi_{nj}(\theta) \cdot [\phi_{nj}(\theta) - 1]^2$, (3.1.6) becomes $\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^4 + 4\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^3 \to 0$ in $P_\theta$-probability, as $n \to \infty$. On the set $A_n^c(\theta)$ on which we are working we have $|\phi_{nj}(\theta) - 1|^4 < |\phi_{nj}(\theta) - 1|^3$, and since $\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^3 \to 0$ in $P_\theta$-probability, as $n \to \infty$, we also have $\sum_{j=1}^n [\phi_{nj}(\theta) - 1]^4 \to 0$ in $P_\theta$-probability, as $n \to \infty$. Hence (3.1.6) is true and so is (ii).

We are now going to establish a further result which will be used on various occasions in the sequel, and then we will make some comments on the results obtained so far and those that are still to be established, in order to complete the proof of Theorem 3.1.1.

We have seen that $n^{\frac{1}{2}} \cdot [\phi_{nj}(\theta) - 1] \to h'\phi_j(\theta)$ in q.m. $[P_\theta]$, as $n \to \infty$. Also the following is true:

LEMMA 3.1.3. $n^{\frac{1}{2}} \cdot [\phi_{nj}^2(\theta) - 1] \to 2h'\phi_j(\theta)$ *in the first mean* $[P_\theta]$, *as* $n \to \infty$, *for all* $j$.

PROOF. It is easily seen that $n^{\frac{1}{2}} \cdot [\phi_{nj}^2(\theta) - 1] - 2h'\phi_j(\theta) = \phi_{nj}(\theta) \cdot \{n^{\frac{1}{2}} \cdot [\phi_{nj}(\theta) - 1] - h'\phi_j(\theta)\} + h'\phi_j(\theta) \cdot [\phi_{nj}(\theta) - 1] + n^{\frac{1}{2}} \cdot [\phi_{nj}(\theta) - 1] - h'\phi_j(\theta)$. Therefore $\mathcal{E}_\theta |n^{\frac{1}{2}} \cdot [\phi_{nj}^2(\theta) - 1] - 2h'\phi_j(\theta)| \leqq \mathcal{E}_\theta |\phi_{nj}(\theta)\{n^{\frac{1}{2}} \cdot [\phi_{nj}(\theta) - 1] - h'\phi_j(\theta)\}| + \mathcal{E}_\theta |h'\phi_j(\theta) \cdot [\phi_{nj}(\theta) - 1]| + \mathcal{E}_\theta |n^{\frac{1}{2}}[\phi_{nj}(\theta) - 1] - h'\phi_j(\theta)| \leqq \mathcal{E}_\theta^{\frac{1}{2}} |n^{\frac{1}{2}} \cdot [\phi_{nj}(\theta) - 1] - h'\phi_j(\theta)|^2 + \mathcal{E}_\theta^{\frac{1}{2}} |h'\phi_j(\theta)|^2 \cdot \mathcal{E}_\theta^{\frac{1}{2}} |\phi_{nj}(\theta) - 1|^2 + \mathcal{E}_\theta |n^{\frac{1}{2}} \cdot [\phi_{nj}(\theta) - 1] - h'\phi_j(\theta)|$ by Hölder's inequality and the fact that $\mathcal{E}_\theta\phi_{nj}^2(\theta) = 1$. Since $\mathcal{E}_\theta |n^{\frac{1}{2}} \cdot [\phi_{nj}(\theta) - 1] - h'\phi_j(\theta)| \leqq \mathcal{E}_\theta^{\frac{1}{2}} |n^{\frac{1}{2}} \cdot [\phi_{nj}(\theta) - 1] - h'\phi_j(\theta)|^2 \to 0$, as

$n \to \infty$, $\mathcal{E}_\theta |\phi_{nj}(\theta) - 1|^2 \to 0$, as $n \to \infty$, and since $\mathcal{E}_\theta |h'\dot{\phi}_1(\theta)|^2 < \infty$, we get $\mathcal{E}_\theta |n^{\frac{1}{2}} \cdot [\phi_{nj}^2(\theta) - 1] - 2h'\dot{\phi}_j(\theta)| \to 0$, as $n \to \infty$, as was to be shown. By Lemma 3.1.1(i) and Lemma 3.1.2(i) Theorem 3.1.1 becomes

$$\text{(3.1.7)}\quad \begin{aligned} &\textstyle\sum_{j=1}^n [\phi_{nj}(\theta) - 1] - n^{-\frac{1}{2}} \cdot \sum_{j=1}^n h'\dot{\phi}_j(\theta) \\ &\qquad \to -\tfrac{1}{2}\mathcal{E}_\theta |h'\dot{\phi}_1(\theta)|^2 \quad \text{in } P_\theta\text{-probability}, \quad \text{as} \quad n \to \infty. \end{aligned}$$

The proof of (3.1.7) is based on the following facts: We set, as usual, $\mathfrak{A}_j = \mathfrak{B}(X_0, X_1, \cdots, X_j)$, $j \geqq 0$. Then we will see that $\mathcal{E}_\theta\{\dot{\phi}_j(\theta) | \mathfrak{A}_{j-1}\} = 0$ a.s. $[P_\theta]$, $j \geqq 1$. This property of $\dot{\phi}_j(\theta)$ will also be used heavily in the next subsection. Furthermore, while $\sum_{j=1}^n [\phi_{nj}(\theta) - 1]$ may not converge anywhere and in any sense, it will be shown that, upon conditioning the random variables $\phi_{nj}(\theta)$ by the $\sigma$-fields $\mathfrak{A}_{j-1}$, we will be able to make a convergence statement about the resulting conditioned sum. In addition to that it will also be shown that if we center the random variables $[\phi_{nj}(\theta) - 1] - n^{-\frac{1}{2}} \cdot h'\dot{\phi}_j(\theta)$ at their conditional expectations, given $\mathfrak{A}_{j-1}$, the sum of them over $j$ converges, as $n \to \infty$. Formally we have

LEMMA 3.1.4. (i) *For every* $j \geqq 1$, $\mathcal{E}_\theta\{\dot{\phi}_j(\theta) | \mathfrak{A}_{j-1}\} = 0$ (= *the zero* $k \times 1$ *column vector*), a.s. $[P_\theta]$. *We set* $\psi_{nj}(\theta) = \mathcal{E}_\theta\{\phi_{nj}(\theta) | \mathfrak{A}_{j-1}\}$, $1 \leqq j \leqq n$. *Then*

(ii) $\sum_{j=1}^n [\psi_{nj}(\theta) - 1] \to -\frac{1}{2}\mathcal{E}_\theta |h'\dot{\phi}_1(\theta)|^2$ *in* $P_\theta$*-probability, as* $n \to \infty$.

(iii) $\sum_{j=1}^n [\{[\phi_{nj}(\theta) - 1] - n^{-\frac{1}{2}}h'\dot{\phi}_j(\theta)\} - \mathcal{E}_\theta\{[\phi_{nj}(\theta) - 1] - n^{-\frac{1}{2}}h'\dot{\phi}_j(\theta)\} | \mathfrak{A}_{j-1}]$ $= \sum_{j=1}^n \{[\phi_{nj}(\theta) - 1] - n^{-\frac{1}{2}}h'\dot{\phi}_j(\theta)\} - \sum_{j=1}^n [\psi_{nj}(\theta) - 1] \to 0$ *in* $P_\theta$*-probability, as* $n \to \infty$.

PROOF. (i) It is easily seen, by means of the Markov property, that $\mathcal{E}_\theta\{\phi_{nj}^2(\theta) | \mathfrak{A}_{j-1}\} = 1$ a.s. $[P_\theta]$. On the other hand, from Lemma 3.1.3 we get $n^{\frac{1}{2}} \cdot [\phi_{nj}^2(\theta) - 1] \to 2h'\dot{\phi}_j(\theta)$ in the first mean $[P_\theta]$, as $n \to \infty$, and hence $\mathcal{E}_\theta\{n^{\frac{1}{2}} \cdot [\phi_{nj}^2(\theta) - 1] | \mathfrak{A}_{j-1}\} \to \mathcal{E}_\theta\{2h'\dot{\phi}_j(\theta) | \mathfrak{A}_{j-1}\}$ in the first mean $[P_\theta]$, as $n \to \infty$, by a well-known property of conditional expectations. Since $\mathcal{E}_\theta\{n^{\frac{1}{2}} \cdot [\phi_{nj}^2(\theta) - 1] | \mathfrak{A}_{j-1}\} = 0$ a.s. $[P_\theta]$, we get $\mathcal{E}_\theta\{h'\dot{\phi}_j(\theta) | \mathfrak{A}_{j-1}\} = 0$ a.s. $[P_\theta]$, for any $h \in \mathcal{E}_k$ which gives (i).

(ii) From the ergodic theorem it follows that

$$\text{(3.1.8)}\quad n^{-1} \cdot \textstyle\sum_{j=1}^n \mathcal{E}_\theta\{|h'\dot{\phi}_j(\theta)|^2 \mid \mathfrak{A}_{j-1}\} \to \mathcal{E}_\theta |h'\dot{\phi}_1(\theta)|^2 \quad \text{a.s. } P[\theta] \quad \text{as} \quad n \to \infty.$$

But

$$\text{(3.1.9)}\quad \begin{aligned} &\textstyle\sum_{j=1}^n \mathcal{E}_\theta\{[\phi_{nj}(\theta) - 1]^2 \mid \mathfrak{A}_{j-1}\} - n^{-1} \cdot \sum_{j=1}^n \mathcal{E}_\theta\{|h'\dot{\phi}_j(\theta)|^2 \mid \mathfrak{A}_{j-1}\} \to 0 \\ &\qquad \text{in } P_\theta\text{-probability}, \quad \text{as} \quad n \to \infty, \end{aligned}$$

and this because $P_\theta[|\sum_{j=1}^n \mathcal{E}_\theta\{[\phi_{nj}(\theta) - 1]^2 \mid \mathfrak{A}_{j-1}\} - n^{-1} \cdot \sum_{j=1}^n \mathcal{E}_\theta\{|h'\dot{\phi}_j(\theta)|^2 \cdot \mid \mathfrak{A}_{j-1}\}| > \epsilon] \leqq \epsilon^{-1} \cdot \mathcal{E}_\theta |n \cdot [\phi_{n1}(\theta) - 1]^2 - |h'\dot{\phi}_1(\theta)|^2| \to 0$, as $n \to \infty$, by (3.1.2). Then (3.1.8) and (3.1.9) taken together give

$$\text{(3.1.10)}\quad \textstyle\sum_{j=1}^n \mathcal{E}_\theta\{[\phi_{nj}(\theta) - 1]^2 \mid \mathfrak{A}_{j-1}\} \to \mathcal{E}_\theta |h'\dot{\phi}_1(\theta)|^2 \quad \text{a.s. } [P_\theta], \quad \text{as} \quad n \to \infty.$$

We now write $[\phi_{nj}^2(\theta) - 1] = [\phi_{nj}(\theta) - 1]^2 + 2[\phi_{nj}(\theta) - 1]$. Conditioning both sides by $\mathfrak{A}_{j-1}$ and taking into account that $\mathcal{E}_\theta\{[\phi_{nj}^2(\theta) - 1] | \mathfrak{A}_{j-1}\} = 0$ a.s. $[P_\theta]$, we get $0 = \mathcal{E}_\theta\{[\phi_{nj}(\theta) - 1]^2 \mid \mathfrak{A}_{j-1}\} + 2[\psi_{nj}(\theta) - 1]$ a.s. $[P_\theta]$.

From this it follows

$$2\cdot\sum_{j=1}^{n}[\psi_{nj}(\theta)-1]+\sum_{j=1}^{n}\mathcal{E}_\theta\{[\phi_{nj}(\theta)-1]^2\mid\mathcal{A}_{j-1}\}=0\quad\text{a.s. }[P_\theta].$$

By taking the limits, as $n\to\infty$, and on account of (3.1.10) we get $2\cdot\sum_{j=1}^{n}[\psi_{nj}(\theta)-1]\to-\mathcal{E}_\theta\,|h'\dot\phi_1(\theta)|^2$ in $P_\theta$-probability, as $n\to\infty$, which is (ii).

(iii) We have to show that

$$\sum_{j=1}^{n}\{[\phi_{nj}(\theta)-1]-n^{-\frac{1}{2}}\cdot h'\dot\phi_j(\theta)-[\psi_{nj}(\theta)-1]\to0$$

$$\text{in } P_\theta\text{-probability},\quad\text{as}\quad n\to\infty.$$

This is an immediate consequence of the extended Kolmogorov inequality ([6], p. 386). In fact, from the definition of $\psi_{nj}(\theta)$ and Part (i) of Lemma 3.1.4, we have that the random variables $[\phi_{nj}(\theta)-1]-n^{-\frac{1}{2}}\cdot h'\dot\phi_j(\theta)-[\psi_{nj}(\theta)-1]$ are centered at conditional expectations given the predecessors. Therefore, $P_\theta[|\sum_{j=1}^{n}\{[\phi_{nj}(\theta)-1]-n^{-\frac{1}{2}}\cdot h'\dot\phi_j(\theta)-[\psi_{nj}(\theta)-1]\}|>\epsilon]\leqq\epsilon^{-2}\cdot\{\mathcal{E}_\theta\,|n^{\frac{1}{2}}\cdot[\phi_{n1}(\theta)-1]-h'\dot\phi_1(\theta)|^2-\mathcal{E}_\theta\,|n^{\frac{1}{2}}\cdot[\psi_{n1}(\theta)-1]|^2\}\to0$, as $n\to\infty$, because $n^{\frac{1}{2}}\cdot[\phi_{n1}(\theta)-1]\to h'\dot\phi_1(\theta)$ in q.m. $[P_\theta]$, as $n\to\infty$, and hence $n^{\frac{1}{2}}\cdot[\psi_{n1}(\theta)-1]\to0$ in q.m. $[P_\theta]$, as $n\to\infty$. Now adding up (ii) and (iii) of the present lemma we get (3.1.7). Since we can replace $P_\theta$ by $P_{n,\theta}$ everywhere, the proof of Theorem 3.1.1 is complete.

3.2. The meaning of Theorem 3.1.1 is that for $n$ sufficiently large and high $P_\theta$-probability, the log-likelihood ratio $\Lambda[P_{n,\theta+h_n\cdot n^{-\frac{1}{2}}};P_{n,\theta}]$ is equal to a certain random variable $h'\Delta_n(\theta)$ plus a nonrandom quantity $A(h,\theta)$. In this subsection we will show that the random variables $\Delta_n(\theta)$ are asymptotically normal, and this fact will force the log-likelihood ratios to be asymptotically normal themselves. The proof of the relevant theorem will be based upon the central limit theorem for martingales, whose proof can be found in [2], pp. 788–792. For easy reference we state it below as

Lemma 3.2.1. *Let $(\mathfrak{X},\mathcal{A},P)$ be a probability space and $\{u_n\}$, $n\geqq1$, be a stationary, ergodic stochastic process defined on $(\mathfrak{X},\mathcal{A},P)$ into $(R,\mathcal{B})$ such that $\mathcal{E}(u_1)=0$, $\mathcal{E}\,|u_1|^2<\infty$, and $\mathcal{E}\{u_j\mid u_1,\cdots,u_{j-1}\}=0$ a.s. $[P]$.*

Then $\mathcal{L}[n^{\frac{1}{2}}\sum_{j=1}^{n}u_j\mid P]\to N(0,\mathcal{E}\,|u_1|^2)$, as $n\to\infty$. Now we give a formal expression of what we vaguely described above. Namely,

Theorem 3.2.1. *$\mathcal{L}[\Delta_n(\theta)\mid P_{n,\theta}]\to N(0,\Gamma(\theta))$, as $n\to\infty$, where $\Gamma(\theta)$ is given in Theorem 3.1.1, and hence $\mathcal{L}[\Lambda[P_{n,\theta+h_n\cdot n^{-\frac{1}{2}}};P_{n,\theta}]\mid P_{n,\theta}]\to N(-\frac{1}{2}h'\Gamma(\theta)h,h'\Gamma(\theta)h)$, as $n\to\infty$.*

Proof. Assume we have shown

$$\text{(3.2.1)}\qquad\mathcal{L}[\Delta_n(\theta)\mid P_{n,\theta}]\to N(0,\Gamma(\theta)),\quad\text{as}\quad n\to\infty.$$

Then $\mathcal{L}[h'\Delta_n(\theta)\mid P_{n,\theta}]\to N(0,h'\Gamma(\theta)h)$, and in view of Theorem 3.1.1 our second assertion in Theorem 3.2.1 is an immediate consequence of the first assertion. So we have only to establish (3.2.1). For any $h\,\varepsilon\,\mathcal{E}_k$ we set $u_j=2h'\dot\phi_j(\theta)$, $j\geqq1$. Then, in order for (3.2.1) to be true, it suffices to show that

$$\text{(3.2.2)}\qquad\mathcal{L}[n^{-\frac{1}{2}}\cdot\sum_{j=1}^{n}u_j\mid P_{n,\theta}]\to N(0,h'\Gamma(\theta)h),\quad\text{as}\quad n\to\infty.$$

We will demonstrate (3.2.2) by showing that the stochastic process $\{u_j\}, j \geqq 1$, defined as above satisfies the conditions of Lemma 3.2.1. That this is so follows easily because $\mathcal{E}_\theta |u_1|^2 = \mathcal{E}_\theta |2h'\dot{\phi}_1(\theta)|^2 < \infty$, $\mathcal{E}_\theta(u_1) = \mathcal{E}_\theta[2h'\dot{\phi}_1(\theta)] = 0$ by Lemma 3.1.4(i), and finally, $\mathcal{E}_\theta\{u_j \mid u_1, \cdots, u_{j-1}\} = \mathcal{E}_\theta\{\mathcal{E}_\theta\{u_j \mid \mathcal{A}_{j-1}\} \mid u_1, \cdots, u_{j-1}\}$, since $\mathcal{A}_{j-1} \supseteq \mathcal{B}(u_1, \cdots, u_{j-1})$, $= \mathcal{E}_\theta\{\mathcal{E}_\theta\{2h'\dot{\phi}_j(\theta) \mid \mathcal{A}_{j-1}\} \mid u_1, \cdots, u_{j-1}\} = 0$ a.s. $[P_\theta]$, since $\mathcal{E}_\theta\{h'\dot{\phi}_j(\theta) \mid \mathcal{A}_{j-1}\} =$ a.s. $[P_\theta]$ by Lemma 3.1.4(i). Hence (3.2.2) is true and so is the theorem.

Concluding this subsection it is worth noticing that in establishing Theorems 3.1.1 and 3.2.1 we have not used the whole force of our assumptions. Those parts we made use of so far are (A1), (A2)(i), (A3)(i), and only the continuity in $P_\theta$-probability of $f_1(\theta, \theta')$, as $\theta' \to \theta$, from (A4), in order to show the continuity in $P_\theta$-probability, as $\theta' \to \theta$, of $q(X_0; \theta, \theta')$.

3.3. In this subsection we will work on the continuity of the covariance $\Gamma(\theta)$. In the process of establishing this continuity we will derive another result to be used later. Also some further properties of $\Delta_n(\theta)$ will be shown.

THEOREM 3.3.1. *Under either one of the assumptions* (A3)(iii), (A3)(iii)$'$, *the function* $\Gamma(\theta)$ *is continuous in* $\theta$.

PROOF. The function $\Gamma(\theta)$ has been defined to be $\Gamma(\theta) = 4\mathcal{E}_\theta\{\dot{\phi}_1(\theta) \cdot \dot{\phi}_1'(\theta)\}$, or equivalently, $h'\Gamma(\theta)h = \mathcal{E}_\theta |2h'\dot{\phi}_1(\theta)|^2$, for any $h \varepsilon \mathcal{E}_k$. So under (A3)(iii), the continuity of $\Gamma(\theta)$ is immediate. It remains for us to show that (A3)(iii)$'$ implies the continuity of $\Gamma(\theta)$, and we will do that by establishing the relation

$$\text{(3.3.1)} \qquad \text{(A3)(iii)}' \text{ implies (A3)(iii).}$$

First we prove an easy result that we formulate as a lemma for later reference, i.e.,

LEMMA 3.3.1. *Under the Assumptions* (A2)(i) and (A4), $\|P_{1,\theta'} - P_{1,\theta}\| \to 0$, *as* $\theta' \to \theta$, *and hence* $P_{1,\theta'} \to P_{1,\theta}$, *as* $\theta' \to \theta$.

PROOF. In Lemma 3.1.3 we have shown that $n^{\frac{1}{2}} \cdot [\phi_{n1}^2(\theta) - 1] \to 2h'\dot{\phi}_1(\theta)$ in the first mean $[P_\theta]$, as $n \to \infty$, as a consequence of differentiability in q.m. $[P_\theta]$ of the random function $\phi_1(\theta, \theta')$ with respect to $\theta'$ at $(\theta, \theta)$. Thus $\phi_{n1}^2(\theta) \to 1$ in the first mean $[P_\theta]$, as $n \to \infty$. It is clear that we will have the same relation by allowing $\lambda \to 0$ in an arbitrary way, and not necessarily in the same rate as the $n^{-\frac{1}{2}}$; i.e., we will have $\phi_1^2(\theta, \theta + \lambda h) \to 1$ in the first mean $[P_\theta]$, as $\lambda \to 0$, uniformly on bounded sets of $h \varepsilon \mathcal{E}_k$. This implies that $\phi_1^2(\theta, \theta') \to 1$ in the first mean $[P_\theta]$, as $\theta' \to \theta$. Since all arguments go through exactly the same way by using $f_1$ instead of $\phi_1$, we have $f_1^2(\theta, \theta') \to 1$ in the first mean $[P_\theta]$, as $\theta' \to \theta$. By noticing that $1 = dP_{1,\theta}/dP_{1,\theta}$ this last relation is the same as $\|P_{1,\theta'} - P_{1,\theta}\| \to 0$, as $\theta' \to \theta$. Finally, that $\|P_{1,\theta'} - P_{1,\theta}\| \to 0$, as $\theta' \to \theta$, implies $P_{1,\theta'} \to P_{1,\theta}$, as $\theta' \to \theta$, is well known and trivial. In proving the implication (3.3.1) we will also need

LEMMA 3.2.2. *Under Assumption* (A3)(ii) *and Lemma* 3.3.1, $\mathcal{L}[h'\dot{\phi}_1(\theta') \mid P_{1,\theta'}] \to \mathcal{L}[h'\dot{\phi}_1(\theta) \mid P_{1,\theta}]$, *as* $\theta' \to \theta$, *for any* $h \varepsilon \mathcal{E}_k$.

PROOF. We have to show that $\int f(x)\, d\mathcal{L}[h'\dot{\phi}_1(\theta') \mid P_{1,\theta'}] \to \int f(x)\, d\mathcal{L} \cdot [h'\dot{\phi}_1(\theta) \mid P_{1,\theta}]$, as $\theta' \to \theta$, for any numerical function $f$, bounded and continuous on $R$; or

$$(3.2.2)\qquad \int f(h'\phi_1(\theta'))\, dP_{1,\theta'} \to \int f(h'\phi_1(\theta))\, dP_{1,\theta}, \quad \text{as} \quad \theta' \to \theta.$$

We write $\int f(h'\phi_1(\theta'))\, dP_{1,\theta'} - \int f(h'\phi_1(\theta))\, dP_{1,\theta} = \int [f(h'\phi_1(\theta')) - f(h'\phi_1\cdot(\theta))]\, dP_{1,\theta} + \int f(h'\phi_1(\theta'))\, d(P_{1,\theta'} - P_{1,\theta})$. By (A3)(ii) $h'\phi_1(\theta') \to h'\phi_1(\theta)$ in $P_\theta$-probability, as $\theta' \to \theta$, and hence $f(h'\phi_1(\theta')) - f(h'\phi_1(\theta)) \to 0$ in $P_\theta$-probability, as $\theta' \to \theta$. Since $|f(h'\phi_1(\theta')) - f(h'\phi_1(\theta))| \leqq 2M(<\infty) = P_{1,\theta}$-integrable (where $M$ is a bound for $f$), we get

$$(3.3.3)\qquad \int [f(h'\phi_1(\theta')) - f(h'\phi_1(\theta))]\, dP_{1,\theta} \to 0, \quad \text{as} \quad \theta' \to \theta$$

by the dominated convergence theorem.

Also $|\int f(h'\phi_1(\theta'))\, d(P_{1,\theta'} - P_{1,\theta})| \leqq M\cdot\|P_{1,\theta'} - P_{1,\theta}\| \to 0$, as $\theta' \to \theta$, by Lemma 3.3.1, and this together with (3.3.3) imply (3.3.2), hence the lemma.

Now we are in a position to prove (3.3.1). We have

$$\mathcal{E}_{\theta'}\, |h'\phi_1(\theta')|^2 = \int x^2\, dF_{1,\theta'},$$

$$\mathcal{E}_{\theta}\, |h'\phi_1(\theta)|^2 = \int x^2\, dF_{1,\theta},$$

where $F_{1,\theta'}$ is the distribution function of the random variable $h'\phi_1(\theta')$ under $P_{1,\theta'}$, and $F_{1,\theta}$ is the distribution function of the random variable $h'\phi_1(\theta)$ under $P_{1,\theta}$. By (A3)(iii)$'$ $\mathcal{E}_{\theta'}\, |h'\phi_1(\theta')|^{2+\delta} = \int |x|^{2+\delta}\, dF_{1,\theta'} \leqq C(\theta, h)(<\infty)$, for all $\theta'$ in a neighborhood $N_\theta$ of $\theta$, for some $0 < \delta(<1)$, and this together with $F_{1,\theta'} \to F_{1,\theta}$, as $\theta' \to \theta$, of Lemma 3.3.2 imply $\int x^2\, dF_{1,\theta'} \to \int x^2\, dF_{1,\theta}$, as $\theta' \to \theta$, or $\mathcal{E}_{\theta'}\, |h'\phi_1(\theta')|^2 \to \mathcal{E}_\theta\, |h'\phi_1(\theta)|^2$, as $\theta' \to \theta$, as was to be seen.

We conclude this subsection with the following:

THEOREM 3.3.2. (i) *For each $n$ and $A \in \mathcal{A}_n$ the function $\theta \to P_{n,\theta}(A)$ is Borel measurable in $\theta$.*

(ii) *For every $\theta \in \Theta$ and $v \in \mathcal{E}_k$ we have $\Delta_n(\theta + v\cdot n^{-\frac{1}{2}}) - \Delta_n(\theta) \to -\Gamma(\theta)\cdot v$ in $P_{n,\theta}$-probability, as $n \to \infty$.*

(iii) *If $\mathcal{C}$ denotes the $\sigma$-field of Borel sets of $\Theta$, the function $\Delta_n(\theta)$ is $\mathcal{A}_n \times \mathcal{C}$-measurable.*

PROOF. (i) A straightforward generalization of Lemma 3.3.1 gives $\|P_{n,\theta'} - P_{n,\theta}\| \to 0$, as $\theta' \to \theta$, for any fixed $n$. In fact, $[dP_{n,\theta'}/dP_{n,\theta}] = q(X_0\,;\, \theta, \theta')\cdot\prod_{j=1}^n \phi_j^2(X_{j-1}, X_j\,;\, \theta, \theta') \to 1 = [dP_{n,\theta}/dP_{n,\theta}]$ in $P_{n,\theta}$-probability, as $\theta' \to \theta$. Since both sides integrate to 1, with respect to $P_{n,\theta}$, this last convergence implies convergence in the first mean, hence $\|P_{n,\theta'} - P_{n,\theta}\| \to 0$, as $\theta' \to \theta$, and this furnishes more than is asserted in (i); namely, that the function $\theta \to P_{n,\theta}(A)$ is continuous in $\theta$, uniformly in $A \in \mathcal{A}_n$.

(ii) This is true, according to the theory developed in [5], for almost all (Lebesgue) $\theta \in \Theta$. However, it would be necessary to check that this is true for all $\theta \in \Theta$ for this particular choice of $\Delta_n(\theta)$.

Let $\{\epsilon_i, i = 1, 2, \cdots, k\}$ be a base for $\mathcal{E}_k$, and define the vector $\Delta_n^*(\theta)$ by the equality $h'\Delta_n^*(\theta) = \sum_{i=1}^k h_i \Lambda[P_{n,\theta+\epsilon_i n^{-\frac{1}{2}}}\,;\, P_{n,\theta}]$, where $h = \sum_{i=1}^k h_i\epsilon_i$. Then it is known (see [5], pp. 56–57) that

$$(3.3.4)\qquad \begin{aligned} &\Delta_n^*(\theta + v\cdot n^{-\frac{1}{2}}) - \Delta_n^*(\theta) \to -\Gamma(\theta)v \quad \text{in } P_{n,\theta}\text{-probability,}\\ &\qquad\qquad \text{as} \quad n \to \infty, \quad \text{for every} \quad \theta \in \Theta \quad \text{and} \quad v \in \mathcal{E}_k. \end{aligned}$$

We have

$$\begin{aligned}(3.3.5)\quad &\Lambda[P_{n,\theta+\epsilon_i\cdot n^{-\frac{1}{2}}+v\cdot n^{-\frac{1}{2}}}; P_{n,\theta+v\cdot n^{-\frac{1}{2}}}] - \Lambda[P_{n,\theta+\epsilon_i\cdot n^{-\frac{1}{2}}}; P_{n,\theta}] \\ &= [\log q(X_0; \theta + v\cdot n^{-\frac{1}{2}}; \theta + \epsilon_i\cdot n^{-\frac{1}{2}} + v\cdot n^{-\frac{1}{2}}) \\ &\quad - \log q(X_0; \theta, \theta + \epsilon_i\cdot n^{-\frac{1}{2}}) + \textstyle\sum_{j=1}^n \{\log \phi_j^2(\theta + v\cdot n^{-\frac{1}{2}}; \theta \\ &\quad + \epsilon_i\cdot n^{-\frac{1}{2}} + v\cdot n^{-\frac{1}{2}}) - \log \phi_j^2(\theta, \theta + \epsilon_i\cdot n^{-\frac{1}{2}})\}.\end{aligned}$$

From (A3)(i) and (A4)(i) the first term in the above sum tends to zero in $P_{n,\theta}$-probability, as $n \to \infty$. Multiplying both sides of (3.3.5) by $h_i$ and summing over $i$, we have that $h'\Delta_n^*(\theta + v\cdot n^{-\frac{1}{2}}) - h'\Delta_n^*(\theta)$ differs from $\sum_{j=1}^n \{\sum_{i=1}^k h_i \log \phi_j^2\cdot(\theta + v\cdot n^{-\frac{1}{2}}; \theta + \epsilon_i\cdot n^{-\frac{1}{2}} + v\cdot n^{-\frac{1}{2}}) - \sum_{i=1}^k h_i \log \phi_j^2(\theta, \theta + \epsilon_i\cdot n^{-\frac{1}{2}})\}$ by a quantity which tends to zero in $P_{n,\theta}$-probability, as $n \to \infty$. Next $\sum_{j=1}^n \{\sum_{i=1}^k h_i \log \phi_j^2(\theta, \theta + \epsilon_i n^{-\frac{1}{2}})\}$ differs from $h'\Delta_n(\theta) - \sum_{i=1}^k h_i\{\sum_{j=1}^n [\phi_j(\theta, \theta + \epsilon_i\cdot n^{-\frac{1}{2}}) - 1]^2\}$ by a quantity which goes to zero in $P_{n,\theta}$-probability, as $n \to \infty$, from Lemma 3.1.1 and (A3)(i).

We have denoted by $\dot\phi_1(\theta)$ the derivative of $\phi_1(\theta, \theta')$ with respect to $\theta'$ at $(\theta, \theta)$. It is easy to see that the derivative of $\phi_1(\theta', \theta)$ with respect to $\theta'$ at $(\theta, \theta)$ is $-\dot\phi_1(\theta)$. This fact, together with Assumption (A3)(i), and continuity in probability of $\dot\phi_1(\theta)$ permits us to write $n^{\frac{1}{2}}\cdot[\phi_j(\theta + v\cdot n^{-\frac{1}{2}}; \theta + v\cdot n^{-\frac{1}{2}} + \epsilon_i\cdot n^{-\frac{1}{2}}) - 1] + \epsilon_i'\dot\phi_j(\theta + v\cdot n^{-\frac{1}{2}}) + R_{nj}(\theta, \epsilon_i, v)$, where $\mathcal{E}_\theta |R_{nj}(\theta, \epsilon_i, v|^2 \to 0$, as $n \to \infty$.

Since $\sum_{j=1}^n \log \phi_j^2(\theta + v\cdot n^{-\frac{1}{2}}; \theta + v\cdot n^{-\frac{1}{2}} + \epsilon_i\cdot n^{-\frac{1}{2}})$ differs from $2\sum_{j=1}^n [\phi_j\cdot(\theta + v\cdot n^{-\frac{1}{2}}; \theta + v\cdot n^{-\frac{1}{2}} + \epsilon_i\cdot n^{-\frac{1}{2}}) - 1] - \sum_{j=1}^n [\phi_j(\theta + v\cdot n^{-\frac{1}{2}}; \theta + v\cdot n^{-\frac{1}{2}} + \epsilon_i\cdot n^{-\frac{1}{2}}) - 1]^2$ by a quantity which goes in probability to zero, a result analogous to one obtained already, we have that $\sum_{j=1}^n \{\sum_{i=1}^k h_i \log \phi_j^2(\theta + v\cdot n^{-\frac{1}{2}}; \theta + v\cdot n^{-\frac{1}{2}} + \epsilon_i\cdot n^{-\frac{1}{2}})\}$ differs from $h'\Delta_n^*(\theta + v\cdot n^{-\frac{1}{2}}) - \sum_{i=1}^k h_i\{\sum_{j=1}^n [\phi_j(\theta + v\cdot n^{-\frac{1}{2}}; \theta + v\cdot n^{-\frac{1}{2}} + \epsilon_i\cdot n^{-\frac{1}{2}}) - 1]^2\}$ by a quantity which tends to zero in probability. Therefore $h'\Delta_n^*(\theta + v\cdot n^{-\frac{1}{2}}) - h'\Delta_n^*(\theta)$ differs from $h'\Delta_n(\theta + v\cdot n^{-\frac{1}{2}}) - h'\Delta_n(\theta) - \sum_{i=1}^k h_i\{\sum_{j=1}^n [\phi_j(\theta + v\cdot n^{-\frac{1}{2}}; \theta + v\cdot n^{-\frac{1}{2}} + \epsilon_i\cdot n^{-\frac{1}{2}}) - 1]^2\}$ by a quantity which goes to zero in probability. But $\sum_{j=1}^n [\phi_j(\theta, \theta + \epsilon_i\cdot n^{-\frac{1}{2}}) - 1]^2 \to \frac{1}{4}\epsilon_i'\Gamma(\theta)\epsilon_i$ in $P_{n,\theta}$-probability, as $n \to \infty$, as it has been seen already, and in a similar fashion, $\sum_{j=1}^n [\phi_j(\theta + v\cdot n^{-\frac{1}{2}}; \theta + v\cdot n^{-\frac{1}{2}} + \epsilon_i\cdot n^{-\frac{1}{2}}) - 1]^2 \to \frac{1}{4}\epsilon_i'\Gamma(\theta)\epsilon_i$ in $P_{n,\theta}$-probability, as $n \to \infty$. Therefore $h'\Delta_n(\theta + v\cdot n^{-\frac{1}{2}}) - h'\Delta_n(\theta) \to -h'\Gamma(\theta)v$ in $P_{n,\theta}$-probability, as $n \to \infty$, for every $\theta \in \Theta$ and $v \in \mathcal{E}_k$, as was to be seen.

REMARK. In the proof, we used, tacitly, the property that $\{P_{m,\theta+h_m\cdot m^{-\frac{1}{2}}}\}$ and $\{P_{m,\theta}\}$ are contiguous, where $\{h_m\}$ is a bounded sequence of elements of $\mathcal{E}_k$. However, this is an immediate consequence of Theorem 3.2.1. More explicitly, if $\{h_n\} \subset \{h_m\}$ with $h_n \to h$, as $n \to \infty$, $h \in \mathcal{E}_k$, then $\mathcal{L}[\Lambda[P_{n,\theta+h_n\cdot n^{-\frac{1}{2}}}; P_{n,\theta}] | P_{n,\theta}] \to N(-\frac{1}{2}h'\Gamma(\theta)h, h'\Gamma(\theta)h)$, as $n \to \infty$. If now $\mathcal{L}[\chi] = N(-\frac{1}{2}h'\Gamma(\theta)h, h'\Gamma(\theta)h)$, then $\mathcal{E}\{\exp \chi\} = \int \exp \chi \, d\mathcal{L}[\chi] = (2\pi)^{-\frac{1}{2}}\cdot\sigma^{-1} \int \exp[-(x - \sigma^2/2)^2/2\sigma^2]\, dx = 1$, where we have set $\sigma^2 = h'\Gamma(\theta)h$. So the defining property (5), [5], p. 40, is satisfied, and this completes the proof.

(iii) This is an immediate consequence of (A3)(ii).

3.4. In this subsection we are dealing with the construction of a consistent estimate of the parameter $\theta$ with the further property of converging to $\theta$ in a certain prescribed rate. The problem will be dealt with in two steps. Firstly, we will exhibit the construction of a consistent estimate, and secondly, we will indicate a way of obtaining a consistent estimate attaining the required rate of convergence.

LEMMA 3.4.1. *Let Assumptions* (A1), (A2) *and* (A4) *be satisfied. There exists a sequence* $\{\theta_n^*\}$ *of measurable functions taking values in* $\Theta$ *such that*

$$\theta_n^* \to \theta \quad \text{in } P_{n,\theta}\text{-probability}, \quad \text{as} \quad n \to \infty, \quad \text{for every} \quad \theta \in \Theta.$$

PROOF. The proof of this lemma is basically that of Lemma 4 of [4], pp. 136–137. It will be necessary, however, to replace the sample distribution functions being used there by some other functions of the observations, having similar convergence properties, and, in general, to secure the conditions for the application of that lemma.

Given $X_0, X_1, \cdots, X_n$ from the Markov process $\{X_n, n \geqq 0\}$ in question, we consider the (random) points $(X_{j-1}, X_j)$, $j = 1, 2, \cdots, n$, in the plane and define $\mu_n(\{(X_{j-1}, X_j)\}) = n^{-1}$, $j = 1, 2, \cdots, n$. Obviously $\{(X_{j-1}, X_j)\} \in \mathcal{B}^2$ $(= \mathcal{B} \times \mathcal{B})$ and $\mu_n$ is a random probability measure on $\mathcal{B}^2$; i.e., for a fixed $x \in \mathfrak{X}$ it is a probability measure on $\mathcal{B}^2$, and it is $\mathcal{A}$-measurable as a function of $x \in \mathfrak{X}$. What we are aiming at is to show that

$$\mu_n \to P_{1,\theta} \quad \text{a.s.} \quad [P_\theta], \quad \text{as} \quad n \to \infty. \tag{3.4.1}$$

That is, if $\mathcal{U} = \{u \mid u: R \times R \to R, \text{ bounded and continuous}\}$, then $\int u\, d\mu_n \to \int u\, dP_{1,\theta}$ a.s. $[P_\theta]$, as $n \to \infty$, for every $u \in \mathcal{U}$. But $\int u\, d\mu_n = n^{-1} \cdot \sum_{j=1}^n u(X_{j-1}, X_j)$, while $\int u\, dP_{1,\theta} = \mathcal{E}_\theta u(X_0, X_1)$ with $\mathcal{E}_\theta |u(X_0, X_1)| < \infty$. Since now $n^{-1} \cdot \sum_{j=1}^n u(X_{j-1}, X_j) \to \mathcal{E}_\theta u(X_0, X_1)$ a.s. $[P_\theta]$, as $n \to \infty$, we get $\int u\, d\mu_n \to \int u\, dP_{1,\theta}$ a.s. $[P_\theta]$, as $n \to \infty$, which is (3.4.1).

The random measure $\mu_n$, as defined above, with the property that $\mu_n \to P_{1,\theta}$ a.s. $[P_\theta]$, as $n \to \theta$, are the ones that play the role of the sample distribution functions in the lemma quoted above.

Next we intend to establish a function on $\Theta$ into the space of bounded signed measures on $\mathcal{B}^2$ which we will require to be 1 : 1 and continuous with respect to the topology $\mathfrak{J}_c$ determined by weak convergence. For this purpose we let $\mathfrak{M}$ stand for the set of bounded signed measures on $\mathcal{B}^2$ and we make $\mathfrak{M}$ into a normed space by defining $\|\mu\| = \sup \cdot \{\mu(B); B \in \mathcal{B}^2\} = \mu^+(R \times R) + \mu^-(R \times R)$, for $\mu \in \mathfrak{M}$. In fact, $(\mathfrak{M}, \|\cdot\|)$ is a linear metric Banach space, and so is $(\mathcal{U}, \|\cdot\|)$ if $\|u\| = \sup \cdot \{u(z); z \in R \times R\}$. Now we denote by $\mathfrak{M}_1$ the subset of $\mathfrak{M}$ consisting of probability measures, and we use (A2)(ii) as well as Lemma 3.3.1. These, together with (3.4.1), provide the tools for the application of Lemma 4 of [4], which gives the estimates $\theta_n^*$ of the present lemma.

In the following we will need another result that we are going to establish now.

Given the random variables $X_0, X_1, \cdots, X_n$ from the Markov process $\{X_n, n \geqq 0\}$, we set, for convenience, $Y_j = (X_{j-1}, X_j)$, $j = 1, \cdots, n$. Then, if $g$ is any numerical, measurable function on $R \times R$ which is bounded, the ergodic theorem gives $n^{-1}\cdot\sum_{j=1}^{n} g(Y_j) \to \mathcal{E}g(Y_1)$ a.s. $[P]$, as $n \to \infty$, where we let $P$ denote any one of the measures $P_\theta$. Here we will show a stronger result which states, in effect, that $n^{\frac{1}{2}}\cdot[n^{-1}\cdot\sum_{j=1}^{n} g(Y_j) - \mathcal{E}g(Y_j)]$ does not grow too big, as $n \to \infty$, provided $g$ is of a special form. More precisely

LEMMA 3.4.2. *We fix an arbitrary $\theta \varepsilon \Theta$ and set $P$ for the probability measure $P_\theta$. Let $B \varepsilon \mathcal{B}$ and set $A_1 = B \times B$. Then if $g = I_{A_1}$, the indicator function of $A_1$, we have: For every $\epsilon > 0$ there exists $b(\epsilon) > 0$ such that*

$$P[|n^{\frac{1}{2}}[n^{-1}\cdot\textstyle\sum_{j=1}^{n} g(Y_j) - \mathcal{E}g(Y_j)]| > b(\epsilon)] < \epsilon, \qquad \text{all} \quad n.$$

PROOF. We have $g(Y_j) = I_{[Y_j \varepsilon A_1]}$, and we set $Z_j = g(Y_j) - \mathcal{E}g(Y_j)$. Then for any $c > 0$ we get

$$\begin{aligned} &P[|n^{\frac{1}{2}}\cdot[n^{-1}\textstyle\sum_{j=1}^{n} g(Y_j) - \mathcal{E}g(Y_j)]| > c] \\ &\quad = P[|\textstyle\sum_{j=1}^{n} Z_j| > c\cdot n^{\frac{1}{2}}] \leqq c^{-2}\cdot n^{-1}\mathcal{E}\,|\textstyle\sum_{j=1}^{n} Z_j|^2 \\ &\quad = c^{-2}\cdot n^{-1}\cdot\{n\cdot\mathcal{E}\,|Z_1|^2 + 2\textstyle\sum_{j=1}^{n-1}\sum_{i=1}^{n-j}\mathcal{E}(Z_jZ_{j+i})\}. \end{aligned} \tag{3.4.2}$$

It is seen that $2\sum_{j=1}^{n-1}\sum_{i=1}^{n-j}\mathcal{E}(Z_jZ_{j+1}) \leqq M^*\mathcal{E}\,|Z_1|^2$, $(0 < M^* < \infty)$, and hence the right side of (3.4.2) is bounded by $c^{-2}(M^* + 1)\mathcal{E}\,|Z_1|^2$, for all $n$. Thus, if for $\epsilon > 0$ we take $b(\epsilon) = b^{\frac{1}{2}}\cdot\epsilon^{-\frac{1}{2}}$, where $b = (M^* + 1)\mathcal{E}\,|Z_1|^2$, we get the result claimed in the lemma.

If, in the last lemma, we use $P_\theta$ and let $\theta$ vary over $\Theta$, we have that $b(\epsilon)$ is a function of $\theta$; i.e., we have $b(\epsilon, \theta)$ rather than $b(\epsilon)$, and consequently

$$P_\theta[|n^{\frac{1}{2}}\cdot[n^{-1}\textstyle\sum_{j=1}^{n} g(Y_j) - \mathcal{E}_\theta g(Y_j)]| > b(\epsilon, \theta)] < \epsilon, \qquad \text{all} \quad n.$$

From the definition of $b(\epsilon, \theta)$ it follows that it is a continuous function of $\theta$. So by restricting attention to a compact subset $K$ of $\Theta$, we have that $b(\epsilon, \theta)$ is bounded by $b_K^*(\epsilon)$, say. Thus we have

COROLLARY 3.4.1. *Let $g$ be as in Lemma 3.4.2. Then for every $\epsilon > 0$ and any compact subset $K$ of $\Theta$ there exists $b_K^*(\epsilon) > 0$ such that*

$$P_\theta[|n^{\frac{1}{2}}\cdot[n^{-1}\cdot\textstyle\sum_{j=1}^{n} g(Y_j) - \mathcal{E}_\theta g(Y_j)]| > b_K^*(\epsilon)] < \epsilon, \qquad \text{all} \quad n, \quad \theta \varepsilon K.$$

We let again $g$ be a numerical, measurable function on $R \times R$ that is bounded, and set $\beta(\theta) = \mathcal{E}_\theta g(Y_1)$. Now we will see that $\beta(\theta)$ is differentiable and we will give an explicit expression for its derivative $\dot\beta(\theta)$. This fact is based upon the Assumption (A4)(i). To demonstrate it we write

$$\begin{aligned} &|\lambda^{-1}\cdot\{\beta(\theta + \lambda h) - \beta(\theta)\} - \textstyle\int 2gh'\dot f_1(\theta)\, dP_\theta| \\ &\qquad = |\textstyle\int [\lambda^{-1}\cdot\{f_1(\theta, \theta + \lambda h) - 1\} - 2h'\dot f_1(\theta)]g\, dP_\theta| \\ &\qquad \leqq M\cdot\textstyle\int |\lambda^{-1}\cdot\{f_1(\theta, \theta + \lambda h) - 1\} - 2h'\dot f_1(\theta)|\, dP_\theta \to 0, \quad \text{as} \quad \lambda \to 0, \end{aligned}$$

uniformly on bounded sets of $h \varepsilon \mathcal{E}_k$. ($M$ is a bound for $g$.) This shows that $\beta(\theta)$ is differentiable and its derivative is given by

$$\dot\beta(\theta) = \int 2g\dot f_1(\theta)\, dP_\theta. \tag{3.4.3}$$

Now let $y = (x_1, x_2)$ represent the points of $R \times R$ and we take $g$ to be of the following form $g = I_{[(-\infty, x_1] \times (-\infty, x_2]]}$ (i.e., the indicator of the set $(-\infty, x_1] \times (-\infty, x_2]$). Then $\mathcal{E}_\theta g(Y_1) = F(y; \theta)$, the distribution function of $Y_1$ under $P_\theta$, evaluated at $y$, and $\dot{F}(y; \theta) = \int 2g\dot{f}_1(\theta)\, dP_\theta$. From (A4)(ii) we have that $|\dot{f}_1(\theta)| \neq 0$ with $P_\theta$-probability $>0$, for every $\theta \in \Theta$. This property of $\dot{f}_1(\theta)$ together with the fact that $|\mathcal{E}_\theta \dot{f}_1(\theta)| = 0$, which is a consequence of (A4)(i), imply that for every $\theta \in \Theta$ there exist $y_\theta$, $d(\theta) > 0$ and a (spherical) neighborhood $S(\theta, \delta(\theta))$ such that

$$(3.4.4) \quad |F(y_\theta; \theta'') - F(y_\theta; \theta')| \geqq |\theta'' - \theta'| \cdot d(\theta), \quad \text{for all } \theta'', \ \theta' \in S(\theta, \delta(\theta)).$$

Now we are in a position to state

THEOREM 3.4.1. *Let Assumptions* (A1) *to* (A4) *be satisfied. There exists a sequence* $\{\hat{\theta}_n\}$ *of measurable functions taking values in* $\Theta$ *such that for every* $\epsilon > 0$ *there exists* $b(\epsilon) > 0$ *so that*

$$P_\theta[|n^{\frac{1}{2}} \cdot (\hat{\theta}_n - \theta)| > b(\epsilon)] < \epsilon, \qquad \textit{for all } n.$$

PROOF. Since $\Theta$ is an open subset of $\mathcal{E}_k$ we can write $\Theta = \sum_{j=1}^{\infty} K_j$, where $\{K_j\}$ is an increasing sequence of compact subsets of $\Theta$. For $n$ sufficiently large and $P_\theta$-probability $\geqq 1 - \epsilon$, the consistent estimate $\theta_n^*$ of Lemma 3.4.1 will lie in a neighborhood of the true parameter $\theta$. This neighborhood, in turn, will be a subset of $K_j$ for some $j \geqq 1$, call it $K$, and therefore it suffices to carry through the arguments for $\theta \in K$.

We consider the collection $\{S(\theta, \delta(\theta)/3), \theta \in K\}$ of neighborhoods provided by (3.4.4). Then there exists a finite number of them covering $K$, say $S(\theta_j, \delta(\theta_j)/3)$, $j = 1, 2, \cdots, m$, and let $y_j = y_{\theta_j}$ be the corresponding $y$'s. We consider the distribution functions $F(y_j; \theta)$ and let $F_n(y_j)$ be the corresponding empirical distribution functions. If $\theta_n^* \in S(\theta_i, \delta(\theta_i)/3)$ we consider the sphere $S(\theta_i, \delta(\theta_i)$ and let $S(\theta_{j_\ell}, \delta(\theta_{j_\ell}))$, $\ell = 1, \cdots, r$ be those spheres out of $\{S(\theta_j, \delta(\theta_j))\}$, $j = 1, \cdots, m$, which intersect $S(\theta_i, \delta(\theta_i))$. Then we define $\hat{\theta}_n$ by the relation

$$(3.4.5) \quad |F_n(y_i) - F(y_i; \hat{\theta}_n)| \leqq \inf \sup |F_n(y_{j_\ell}) - F(y_{j_\ell}; \theta)| + n^{-1},$$

where the sup is taken over $\ell = 1, \cdots, r$ and the inf over $\theta \in K$. It is possible to give a specific rule for the choice of $\hat{\theta}_n$, and show that the resulting estimate is a measurable function. In the sequel it will be assumed that this is the case.

From Corollary 3.4.1 it follows that

$$P_\theta[|n^{\frac{1}{2}} \cdot [F_n(y_{j_\ell}) - F(y_{j_\ell}; \theta)]| > b_K^*(\epsilon, \ell)] < \epsilon, \qquad \text{all } n, \quad \theta \in K.$$

By taking $b_K^*(\epsilon) = \max \cdot \{b_K^*(\epsilon, \ell); 1 \leqq \ell \leqq r\} + 1$, we have

$$(3.4.6) \qquad P_\theta[|n^{\frac{1}{2}}[F_n(y_{j_\ell}) - F(y_{j_\ell}; \theta)]| > b_K^*(\epsilon)] < \epsilon, \qquad \text{all } n, \quad \theta \in K.$$

From (3.4.5) it follows then that

$$(3.4.7) \qquad P_\theta[|n^{\frac{1}{2}} \cdot [F_n(y_i) - F(y_i; \hat{\theta}_n)]| > b_K^*(\epsilon)] < \epsilon, \qquad \text{all } n, \quad \theta \in K.$$

Now (3.4.6) and (3.4.7) imply

(3.4.8) $$P_\theta[|n^{\frac{1}{2}}\cdot[F(y_i\,;\hat{\theta}_n) - F(y_i\,;\theta)]| > 2b_K{}^*(\epsilon)] < \epsilon, \quad \text{all } n, \quad \theta \varepsilon K.$$

For $n$ sufficiently large and $P_\theta$-probability $\geqq 1 - \epsilon$, $\hat{\theta}_n$ lies in $S(\theta_i\,, \delta(\theta_i))$, and thus (3.4.4) gives

(3.4.9) $$n^{\frac{1}{2}}\cdot|F(y_i\,;\hat{\theta}_n) - F(y_i\,;\theta)| \geqq n^{\frac{1}{2}}\cdot|\hat{\theta}_n - \theta|\cdot d(\theta_i).$$

Therefore (3.4.8) and (3.4.9) imply

(3.4.10) $$P_\theta[|n^{\frac{1}{2}}\cdot(\hat{\theta}_n - \theta)| \geqq 2b_K{}^*(\epsilon)/d(\theta_i)] < \epsilon, \quad \text{for } n \text{ sufficiently large.}$$

By increasing $b_K{}^*(\epsilon)$ so that to take care of the finitely many exceptional $n$'s which may not satisfy (3.4.10), and setting $b(\epsilon)$ for $2b_K{}^*(\epsilon)/d(\theta_i)$, we get $P_\theta[|n^{\frac{1}{2}}\cdot(\hat{\theta}_n - \theta)| \geqq b(\epsilon)] < \epsilon$, for all $n$, as was to be established.

**4. Acknowledgment.** I am deeply indebted to Professor Lucien LeCam who suggested the problem to me, held many helpful discussions, and generously imparted his ideas. However, for any omissions or errors which may occur in this work, the responsibility lies with this author alone.

## APPENDIX

In this appendix we give the definition of differentially asymptotically normal families of distributions (DAN families of distributions, for short), and briefly explain its meaning.

It is well known that in cases where we are dealing with a family of probability measures which are exponential—in particular, normal—we can use this family in a satisfactory way for problems of statistical inferences. However, this is not so if the family of probability measures in question is not of this special form. In order to be able to handle the same problems in more general situations, Professor L. LeCam has introduced the notion of DAN families of distributions, the meaning of it being that, roughly speaking, DAN families of distributions can be, for certain problems, approximately treated, in some neighborhood of the underlying parameter, as if they were normal. This is made precise by the defining properties of DAN families of distributions to be given below.

Let $(\mathfrak{X}, \mathfrak{A})$ be a measurable space. Then

DEFINITION 1. Let $\mu$, $\nu$ and $\lambda$ be any three finite positive measures on $\mathfrak{A}$ such that $\mu \ll \lambda$, $\nu \ll \lambda$ (e.g., $\lambda = \mu + \nu$). Then we denote by $f = d\mu/d\lambda$, $g = d\nu/d\lambda$ and we define

$$\begin{aligned} \Lambda[\mu;\nu] &= \log\,[f/g], &&\text{if } f\cdot g > 0, \\ &= \text{arbitrary}, &&\text{if } f\cdot g = 0. \end{aligned}$$

Next let $\mathfrak{A}_n$ be a sequence of sub-$\sigma$-fields of $\mathfrak{A}$ and $P_n$ be a sequence of probability measures on $\mathfrak{A}_n$. If $T_n : (\mathfrak{X}, \mathfrak{A}_n) \to (\mathcal{E}_k\,, \mathfrak{B}^k)$, $\mathfrak{A}_n$-measurable, where $\mathfrak{B}^k = \prod_{i=1}^k \mathfrak{B}$, we denote by $\mathfrak{L}[T_n \mid P_n]$ the probability distribution of $T_n$ under $P_n$. Then, if $P$ is a probability measure on $\mathfrak{B}^k$, we give the

DEFINITION 2. $\mathfrak{L}[T_n \mid P_n] \to P$, as $n \to \infty$, if for every numerical bounded and continuous function $f$ on $\mathcal{E}_k$, $\int f\, d\mathfrak{L}[T_n \mid P_n] \to \int f\, dP$, as $n \to \infty$.

DEFINITION 3. Let $T_n$ be as in Definition 2. We say that the sequence $\{\mathcal{L}[T_n \mid P_n]\}$ is relatively compact if for every subsequence $\{n'\} \subset \{n\}$, there is a further subsequence $\{m\} \subset \{n'\}$ such that $\mathcal{L}[T_m \mid P_m]$ converges to a probability distribution.

Equivalently, for every $\epsilon > 0$ there exists $b(\epsilon) > 0$ such that $P_n[|T_n| > b(\epsilon)] < \epsilon$ for all $n$. Let now $\Theta$ be an open subset of $\mathcal{E}_k$ and for each $\theta \varepsilon \Theta$ let $P_{n,\theta}$ be probability measures on $\mathcal{A}_n$. Then we give

DEFINITION 4. (DAN families of distributions.) The sequence of families $\{P_{n,\theta}, \theta \varepsilon \Theta\}$, is called DAN if the following (DN1) to (DN7) conditions are satisfied.

(DN1) For each $\theta \varepsilon \Theta$ there exists a $k \times 1$ column vector $\mu(\theta)$, a covariance $\Gamma(\theta)$, and a sequence $\{\Delta_n(\theta)\}$ of $\mathcal{A}_n$-measurable functions taking values in $\mathcal{E}_k$ such that if $h_n \to h$, as $n \to \infty$, $h_n, h \varepsilon \mathcal{E}_k$, then $\Lambda[P_{n,\theta+h_n \cdot n^{-\frac{1}{2}}}; P_{n,\theta}] - h'\Delta_n(\theta) \to -A(h, \theta)$ in $P_{n,\theta}$-probability, as $n \to \infty$, where $A(h, \theta) = h'\mu(\theta) + \frac{1}{2}h'\Gamma(\theta)h$.

(DN2) Both $\mu$ and $\Gamma$ are continuous functions of $\theta \varepsilon \Theta$.

(DN3) $\mathcal{L}[\Delta_n(\theta) \mid P_{n,\theta}] \to N(\mu(\theta), \Gamma(\theta))$, as $n \to \infty$.

(DN4) For every $\theta \varepsilon \Theta$ and $v \varepsilon \mathcal{E}_k$ we have $\Delta_n(\theta + v \cdot n^{-\frac{1}{2}}) - \Delta_n(\theta) \to -\Gamma(\theta)v$ in $P_{n,\theta}$-probability, as $n \to \infty$.

(DN5) For each $n$ and $A \varepsilon \mathcal{A}_n$ the function $\theta \to P_{n,\theta}(A)$ is Borel measurable in $\theta$.

REMARK. (DN5) is slightly different from the corresponding definition in [5]. We give it in this form, in order to avoid the introduction of further definitions which are not needed in this paper.

(DN6) If $\mathcal{C}$ denotes the $\sigma$-field of Borel sets of $\Theta$, the function $\Delta_n(\theta)$ is $\mathcal{A}_n \times \mathcal{C}$-measurable.

(DN7) There is a sequence $\{\hat{\theta}_n\}$ of $k$-dimensional, $\mathcal{A}_n$-measurable functions such that for each $\theta \varepsilon \Theta$ the distributions $\{\mathcal{L}[n^{\frac{1}{2}} \cdot (\hat{\theta}_n - \theta) \mid P_{n,\theta}]\}$ form a relatively compact sequence.

DEFINITION 5. Let $\mu$, $\nu$ and $\lambda$ be as in Definition 1. We define the $L_1$-norm of $\mu - \nu$, denoted by $\|\mu - \nu\|$, as follows: $\|\mu - \nu\| = \int |f - g| \, d\lambda$. If $\mu, \nu$ are probability measures we also have $\|\mu - \nu\| = 2 \sup \{|\mu(A) - \nu(A)|; A \varepsilon \mathcal{A}\}$.

## REFERENCES

[1] BILLINGSLEY, P. (1961). *Statistical Inferences for Markov Processes.* Univ. of Chicago Press.

[2] BILLINGSLEY, P. (1961). The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.* **12** 788–792.

[3] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[4] LECAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 129–156.

[5] LECAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.* **3** 37–98.

[6] LOÈVE, M. (1960). *Probability Theory,* 2nd ed. Van Nostrand, New York.