Ann. Inst. Statist. Math.
Vol. 53, No. 4, 799–809 (2001)
©2001 The Institute of Statistical Mathematics

# UNIFORM ASYMPTOTIC EXPANSION OF LIKELIHOOD RATIO FOR MARKOV DEPENDENT OBSERVATIONS

ANDREI NOVIKOV

*Departamento de Matemáticas, UAM-Iztapalapa, av. Michoacán y la Purísima s/n, col. Vicentina, 09340 México D. F., Mexico,* e-mail: an@xanum.uam.mx

(Received February 15, 1999; revised July 24, 2000)

**Abstract.** An asymptotic expansion of the logarithm of the likelihood ratio for Markov dependent observation is obtained. A functional limit theorem for the likelihood ratio is proved, which gives a way to study limiting distributions of the likelihood ratio based on stopping times, in particular, that of sequential probability ratio test.

*Key words and phrases*: Likelihood ratio, Markov process, asymptotic expansion, functional limit theorem, sequential probability ratio test.

## 1. Introduction

In the paper Akritas and Roussas (1979), an asymptotic expansion of the logarithm of the likelihood ratio is obtained for a random number of Markov observations. This is done on the basis of the expansion for non-random sample size of quantities in Roussas (1972), using additional conditions on the family of distributions. In this paper, we obtain a more general result, namely, a uniform (over sample size number) asymptotic expansion of the likelihood ratio (Theorems 1 and 2), from which the result of Akritas and Roussas (1979) follows easily (see Theorem 4). More than that, our results provide us with tools of asymptotic analysis of the likelihood ratio based on stopping times of more complicated nature, e.g., stopping time of the sequential probability ratio test (SPRT). In this case, the result of Akritas and Roussas (1979) does not apply, as is suggested by the form of the limiting distribution of the likelihood ratio at the stopping time, which is a two-point distribution, in contrast to the limiting normal distribution of the stopped log-likelihood in Akritas and Roussas (1979). We prove a functional limit theorem for the likelihood ratio (Theorem 3), which is a base for such an investigation.

We make use of conditions even less restrictive than those in Roussas (1972), and do not suppose that the distributions are absolutely continuous with respect to each other, similar to Ibragimov and Has′minskiĭ (1981).

It should be mentioned in passing that the concept of contiguity employed in this paper was introduced and exploited by Lucien Le Cam in his fundamental paper (Le Cam (1960)). See also Chapter 6 in Le Cam (1986), as well as Le Cam and Yang (1990).

## 2. Assumptions and notation

Let $\Theta \subset \mathbb{R}^k$ be an open set of parameters, and let $x_0, x_1, x_2, \dots, x_n, \dots$ be a random sequence on a probability space $(\Omega, \mathfrak{A}, P_\theta)$, which is a stationary Markov process given any $\theta \in \Theta$, taking values in a complete separable space $\mathcal{X}$ supplied with the $\sigma$-algebra of

Borel subsets $\mathfrak{X}$. Without loss of generality, let us suppose that $\Omega = \mathcal{X}^{\mathbb{N}} = \mathcal{X} \times \mathcal{X} \times \cdots$, and that the $\sigma$-algebra $\mathfrak{A} = \mathfrak{X} \otimes \mathfrak{X} \otimes \cdots$ is generated by cylinder sets.

Let the distribution of the sequence $x_0, x_1, \dots$ be defined by the transition density functions $f(x_1|x_0;\theta)$ and the initial distribution with the density function $f(x_0;\theta)$ with respect to some measure $\mu$ on $(\mathcal{X}, \mathfrak{X})$. Let us suppose that the function $f(\cdot|\cdot;\theta)$ is $\mathfrak{X}\otimes\mathfrak{X}$-measurable, and $f(\cdot;\theta)$ is $\mathfrak{X}$-measurable. Let us denote by $P_\theta^n$ the restriction of the measure $P_\theta$ to the $\sigma$-algebra $\mathfrak{A}_n = \mathfrak{X} \otimes \mathfrak{X} \otimes \cdots \mathfrak{X} = \mathfrak{X}^{(n+1)}$.

For the points $\theta \in \Theta$ and $\theta + h \in \Theta$, let us define the likelihood ratio corresponding to a fixed number $n$ of observations as the Radon-Nikodym derivative of the absolutely continuous part $P_{\theta+h,a}^n$ of the distribution $P_{\theta+h}^n$ with respect to $P_\theta^n$:

$$Z_\theta(n,h) = \frac{dP_{\theta+h,a}^n}{dP_\theta^n}(x_0, x_1, \dots, x_n) = \exp\{\Lambda_\theta(n,h)\},$$

where

$$\Lambda_\theta(n,h) = \ln \frac{f(x_0;\theta+h)}{f(x_0;\theta)} + \sum_{i=1}^{n} \ln \frac{f(x_i \mid x_{i-1};\theta+h)}{f(x_i \mid x_{i-1};\theta)},$$

if $f(x_0;\theta)\prod_{i=1}^n f(x_i \mid x_{i-1};\theta) > 0$, and $\Lambda_\theta(n,h) = -\infty$, otherwise. Under the hypothesis that all the distributions of the family $\{P_\theta^n, \theta \in \Theta\}$ are absolutely continuous with respect to each other, in Roussas (1972) an asymptotic expansion of the random variable $\Lambda_\theta(n, h_n/\sqrt{n})$ is obtained for each sequence $h_n \to h$ as $n \to \infty$. The main aim of this paper is to obtain a uniform over $1 \le k \le n$ asymptotic expansion of the random variable $\Lambda_\theta(k, h_n/\sqrt{n})$ and to prove corresponding functional limit theorems.

Let us introduce the following notation. Let $h'$ be the transposed vector of $h$, $(h,g) = g'h$ be the scalar product of the vectors $h$ and $g$, $|h| = |(h,h)|^{1/2}$ be the length of the vector $h$ , and let us assume that integration without indicating limits corresponds to the integral over the entire space.

The regularity conditions below are close to those of Roussas (1972), but do not require that the distributions be mutually absolutely continuous.

A1. The sequence $x_0, x_1, \dots, x_n \dots$ is ergodic with respect to any probability distribution $P_\theta$.

A2. For each $\theta \in \Theta$, there exists a vector-function $\psi(\theta) = \psi(x_0, x_1;\theta)$ of dimension $k$ such that

$$\int f(x_0;\theta)[f^{1/2}(x_1 \mid x_0;\theta+h) - f^{1/2}(x_1 \mid x_0;\theta) - h'\psi(\theta)]^2 d\mu^2(x_0,x_1)$$
$$= o(|h|^2), \qquad h \to 0$$

(here and throughout $\mu^2 = \mu \times \mu$ denotes the product-measure of $\mu$ with itself on the product $\sigma$-algebra $\mathfrak{X} \times \mathfrak{X}$).

A3. For each $\theta \in \Theta$, the Fisher information matrix

$$I(\theta) = 4\int f(x_0;\theta)\psi(x_0,x_1;\theta)\psi'(x_0,x_1;\theta)d\mu^2(x_0,x_1)$$

is positive definite.

A4. For any $\theta \in \Theta$, $f(x;\theta_n) \to f(x,\theta)$ in $\mu$-measure as $\theta_n \to \theta$, $n \to \infty$.

Assumption A2 is essentially the quadratic mean differentiability of the square root of the transition density, and is a slight generalization of the regularity condition commonly used for independent observations (see, for example, Ibragimov and Has'minskiĭ (1981)).

## 3. Uniform asymptotic expansion of the likelihood ratio

Let $\varphi_i(\theta)$ be defined as $\psi(x_{i-1}, x_i; \theta)/\sqrt{f(x_i \mid x_{i-1}; \theta)}$ if the denominator is positive, and as $\infty$, otherwise, $i = 1, 2, \ldots$, and let $\Delta_k(\theta) = 2\sum_{i=1}^k \varphi_i(\theta)$.

THEOREM 1. *For any sequence $h_n \to h$, $n \to \infty$ and for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} P_\theta^n \left( \max_{1\le k\le n} \left| \Lambda_\theta(k, h_n/\sqrt{n}) - \frac{1}{\sqrt{n}} h'\Delta_k(\theta) + \frac{k}{2n} h' I(\theta) h \right| > \varepsilon \right) = 0. \tag{3.1}$$

The proof of Theorem 1 follows that of Theorem 4.1 in Roussas (1972) using "uniform" generalizations of corresponding lemmas in Roussas (1972) (see Lemmas 1, 4, 5, 6).

Let $h_n$, $n = 1, 2, \ldots$ be a sequence of vectors, $h_n \to h$, $n \to \infty$. Let us denote by $\theta_n = \theta + h_n/\sqrt{n}$, $n = 1, 2, \ldots$. Let

$$r_i(\theta_n, \theta) = \left[ \frac{f(x_i \mid x_{i-1}; \theta_n)}{f(x_i \mid x_{i-1}; \theta)} \right]^{1/2},$$

if the denominator is positive, and $r_i(\theta_n, \theta) = \infty$, otherwise, $i = 1, 2, \ldots$.

LEMMA 1. *For any $\varepsilon > 0$ and any $h_n \to h \in \mathbb{R}^k$,*

$$\lim_{n\to\infty} P_\theta \left( \max_{1\le k\le n} \left| \sum_{i=1}^k (r_i(\theta_n, \theta) - 1)^2 - \frac{k}{n} E_\theta (h'\varphi_1(\theta))^2 \right| > \varepsilon \right) = 0.$$

PROOF. By Condition A2 for any $i = 1, 2, \ldots$,

$$E_\theta(\sqrt{n}(r_i(\theta_n, \theta) - 1) - h'\varphi_i(\theta))^2 \to 0,$$

as $n \to \infty$. Therefore,

$$n(r_1(\theta_n, \theta) - 1)^2 \to (h'\varphi_1(\theta))^2, \tag{3.2}$$

as $n \to \infty$, in the first mean with respect to the measure $P_\theta$ by Vitali's theorem (Theorem 2.1 A in Roussas (1972)). Let $\varepsilon > 0$ be fixed. Then

$$\begin{aligned}
&P_\theta \left( \max_{1\le k\le n} \left| \sum_{i=1}^k (r_i(\theta_n, \theta) - 1)^2 - \frac{1}{n} \sum_{i=1}^k (h'\varphi_i(\theta))^2 \right| > \varepsilon \right) \\
&\le P_\theta \left( \sum_{i=1}^n \left| (r_i(\theta_n, \theta) - 1)^2 - \frac{1}{n} (h'\varphi_i(\theta))^2 \right| > \varepsilon \right) \\
&\le \frac{1}{\varepsilon} E_\theta | n(r_1(\theta_n, \theta) - 1)^2 - (h'\varphi_1(\theta))^2 | \to 0,
\end{aligned}$$

as $n \to \infty$, as a consequence of (3.2). To prove Lemma 1, it remains to show that

$$\frac{1}{n} \max_{1 \le k \le n} \left| \sum_{i=1}^{k} (h'\varphi_i(\theta))^2 - kE_\theta(h'\varphi_i(\theta))^2 \right| \to 0, \tag{3.3}$$

as $n \to \infty$, in probability $P_\theta$.

This is a consequence of the following simple lemma.

LEMMA 2. *If the sequence $\{S_n\}$ of real numbers is such that $\frac{1}{n}S_n \to 0$, as $n \to \infty$, then*

$$\frac{1}{n}S_n^* = \frac{1}{n} \max_{1 \le k \le n} |S_k| \to 0, \qquad n \to \infty.$$

PROOF. Let $\varepsilon$ be any positive number. Let us choose the number $m$ so that

$$\sup_{k \ge m} \frac{1}{k}|S_k| \le \varepsilon.$$

Then for $n > m$, we have

$$\frac{1}{n}S_n^* = \frac{1}{n} \max_{1 \le k \le m} |S_k| + \max_{m \le k \le n} \frac{1}{n}|S_k| \le \frac{1}{n} \max_{1 \le k \le m} |S_k| + \max_{m \le k \le n} \frac{1}{k}|S_k| \le \frac{1}{n} \max_{1 \le k \le m} |S_k| + \varepsilon,$$

hence, for any $\varepsilon > 0$, $\limsup_{n \to \infty} \frac{1}{n}S_n^* < \varepsilon$, which implies the statement of Lemma 2.

To prove (3.3), it remains to apply Lemma 2 to the sequence $S_n = \sum_{i=1}^n \xi_i$ with $\xi_i = (h'\varphi_i(\theta))^2 - E_\theta(h'\varphi_i(\theta))^2$, for which $\frac{1}{n}S_n \to 0$ with probability 1 by the ergodic theorem (Loève (1960), Chapter 9, §33).

LEMMA 3. *For any $h \in \mathbb{R}^k$, $E_\theta(h'\varphi_1(\theta))^2 = \frac{1}{4}h'I(\theta)h$.*

PROOF. It is sufficient to show that for any $i$, $j \in \{1, 2, \dots, k\}$,

$$E_\theta \varphi_1^i(\theta)\varphi_1^j(\theta) = \int f(x_0; \theta)\psi^i(\theta)\psi^j(\theta)d\mu^2(x_0, x_1) \tag{3.4}$$

(superscript $i$ denotes the number of the coordinate $a^i$ of the vector $a = (a^1, a^2, \dots, a^k)$). To do this, let us show that on the set $A = \{(x_0, x_1) : f(x_1 \mid x_0; \theta) = 0\}$, $\psi^i(x_0, x_1; \theta) = 0$ almost everywhere with respect to the measure $P_\theta^0 \times \mu$ on $\mathfrak{X} \times \mathfrak{X}$. Let $e(h) = (0, 0, \dots, 0, h, 0, \dots, 0)'$, where the $i$-th coordinate is non-zero. By Condition A2,

$$\int_A \left( \frac{1}{h} f^{1/2}(x_1 \mid x_0; \theta + e(h)) - \psi^i(x_0, x_1; \theta) \right)^2 dP_\theta^0 \times \mu(x_0, x_1) \to 0,$$

as $h \to 0$, which implies that on the set $A$,

$$\frac{1}{h} f^{1/2}(x_1 \mid x_0; \theta + e(h)) \to \psi^i(x_0, x_1; \theta) \tag{3.5}$$

in the measure $P_\theta^0 \times \mu$ as $h \to 0$.

Supposing that there exists a subset $B \subset A$ of a positive $P_\theta^0 \times \mu$-measure, where, say, $\psi^i(x_0, x_1; \theta) > 0$, we easily get to a contradiction to (3.5) observing that in such a case, for any $(x_0, x_1) \in B$,

$$\psi^i(x_0, x_1; \theta) = \lim_{x \to 0, h<0} \frac{1}{h} f^{1/2}(x_1 \mid x_0; \theta + e(h)) \le 0.$$

In the same way, it is proved that the set $\{(x_0, x_1) : \psi^i(x_0, x_1; \theta) < 0\} \cap A$ has $P_\theta^0 \times \mu$-measure 0.

Hence, $\psi^i(x_0, x_0; \theta) = 0$ on $A$ $P_\theta^0 \times \mu$-almost everywhere, which proves (3.4).

LEMMA 4. *Let* $\bar{r}_i(\theta_n, \theta) = E_\theta\{r_i(\theta_n, \theta) \mid \mathfrak{A}_{i-1}\}$, $i = 1, 2, \ldots$. *Then for any* $\varepsilon > 0$,

$$\lim_{n\to\infty} P_\theta \left( \max_{1\le k\le n} \left| \sum_{i=1}^{k} (\bar{r}_i(\theta_n, \theta) - 1) + kE_\theta(h'\varphi_1(\theta))^2/2n \right| > \varepsilon \right) = 0.$$

PROOF. From the ergodic theorem and Lemma 2, it follows that

$$\lim_{n\to\infty} \frac{1}{n} \max_{1\le k\le n} \left| \sum_{i=1}^{k} E_\theta\{(h'\varphi_i(\theta))^2 \mid \mathfrak{A}_{i-1}\} - kE_\theta(h'\varphi_1(\theta))^2 \right| \to 0 \tag{3.6}$$

in probability $P_\theta$, as $n \to \infty$. In addition, for any fixed $\varepsilon > 0$,

$$\begin{aligned} &P_\theta \left( \max_{1\le k\le n} \left| \sum_{i=1}^{k} \left[ E_\theta\{(r_i(\theta_n, \theta) - 1)^2 | \mathfrak{A}_{i-1}\} - \frac{1}{n} E_\theta\{(h'\varphi_i(\theta))^2 \mid \mathfrak{A}_{i-1}\} \right] \right| > \varepsilon \right) \\ &\quad \le P_\theta \left( \sum_{i=1}^{n} E_\theta \left\{ \left| (r_i(\theta_n, \theta) - 1)^2 - \frac{1}{n} (h'\varphi_i(\theta))^2 \right| \mathfrak{A}_{i-1} \right\} > \varepsilon \right) \\ &\quad \le \frac{1}{\varepsilon} E_\theta |n(r_1(\theta_n, \theta) - 1)^2 - (h'\varphi_1(\theta))^2| \to 0, \end{aligned} \tag{3.7}$$

as $n \to \infty$, where the last convergence is due to (3.2).

Combining (3.6) and (3.7), we get

$$P_\theta \left( \max_{1\le k\le n} \left| \sum_{i=1}^{k} E_\theta\{(r_i(\theta_n \theta) - 1)^2 \mid \mathfrak{A}_{i-1}\} - kE_\theta(h'\varphi_1(\theta))^2/n \right| > \varepsilon \right) \to 0,$$

as $n \to \infty$. To prove Lemma 4, it remains to show that

$$\max_{1\le k\le n} \left| \sum_{i=1}^{k} (E_\theta\{(r_i(\theta_n \theta))^2 \mid \mathfrak{A}_{i-1}\} - 1) \right| \to 0 \tag{3.8}$$

in $P_\theta$-probability, as $n \to \infty$.

Let $\varepsilon > 0$ be any number. Then

$$P_\theta \left( \max_{1\le k\le n} \left| \sum_{i=1}^{k} (E_\theta\{(r_i(\theta_n, \theta))^2 \mid \mathfrak{A}_{i-1}\} - 1) \right| > \varepsilon \right) \tag{3.9}$$

$$
\begin{aligned}
&\leq P_\theta\left(\sum_{i=1}^{n}\left|(E_\theta\{(r_i(\theta_n,\theta))^2 \mid \mathfrak{A}_{i-1}\}-1)\right|>\varepsilon\right)\\
&\leq \frac{n}{\varepsilon}E_\theta|E_\theta\{(r_1(\theta_n,\theta))^2 \mid \mathfrak{A}_0\}-1|\\
&=\frac{n}{\varepsilon}E_\theta\left|\int_{\{f(x_1|x_0;\theta)\neq 0\}} f(x_1\mid x_0;\theta_n)d\mu(x_1)-1\right|\\
&=\frac{n}{\varepsilon}E_\theta\int_{\{f(x_1|x_0;\theta)=0\}} f(x_1\mid x_0;\theta_n)d\mu(x_1),
\end{aligned}
$$

where the second inequality follows from Chebyshev's inequality.

From Condition A2, it follows, in particular, that for the set $A=\{(x_0,x_1): f(x_1 \mid x_0;\theta)=0\}$,

$$
\int_A\left(f^{1/2}(x_1\mid x_0;\theta_n)-\frac{1}{\sqrt{n}}h'\psi(\theta)\right)^2 dP_\theta^0\times\mu(x_0,x_1)=o\left(\frac{1}{n}\right),\qquad n\to\infty.
$$

As was shown in the proof of Lemma 3, $\psi(\theta)=0$ on $A$ $P_\theta^0\times\mu$-almost everywhere, so

$$
n\int_A f(x_1\mid x_0;\theta_n)dP_\theta^0\times\mu(x_0,x_1)=o(1),\qquad n\to\infty,
$$

so the right-hand side of (3.9) tends to 0, as $n\to\infty$, which proves (3.8), and hence the lemma.

LEMMA 5. *For any* $\varepsilon>0$,

$$
\lim_{n\to\infty}P_\theta\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k}\left[(r_i(\theta_n,\theta)-1)-\frac{1}{\sqrt{n}}h'\varphi_i(\theta)\right]-\sum_{i=1}^{k}(\bar{r}_i(\theta_n,\theta)-1)\right|>\varepsilon\right)=0.
$$

PROOF. Put

$$
Y_i=(r_i(\theta_n,\theta)-1)-\frac{1}{\sqrt{n}}h'\varphi_i(\theta)-(\bar{r}_i(\theta_n,\theta)-1),\qquad i=1,2,\ldots.
$$

We have then

$$
E_\theta\{Y_i\mid\mathfrak{A}_{i-1}\}=-\frac{1}{\sqrt{n}}E_\theta\{h'\varphi_i(\theta)\mid\mathfrak{A}_{i-1}\}. \tag{3.10}
$$

Let us show that the right-hand side of (3.10) is equal to 0 $P_\theta$-almost everywhere. To this end, note first that

$$
E_\theta|\sqrt{n}(r_1^2(\theta_n,\theta)-1)-2h'\varphi_1(\theta)|\to 0, \tag{3.11}
$$

as $n\to\infty$. Convergence (3.11) is proven in the same way as Lemma 5.4 in Roussas (1972), with the change that in our case $E_\theta r_1^2(\theta_n,\theta)\leq 1$. Further, by (3.11),

$$
\sqrt{n}E_\theta\{(r_1^2(\theta_n,\theta)-1)\mid\mathfrak{A}_0\}\to 2E_\theta\{h'\varphi_1(\theta)\mid\mathfrak{A}_0\}, \tag{3.12}
$$

as $n \to \infty$, in the first mean with respect to $P_\theta$. On the other hand, $P_\theta^0$-almost sure

$$E_\theta\{(r_1^2(\theta_n,\theta)-1) \mid \mathfrak{A}_0\} = \int_{\{f(x_1|x_0;\theta)=0\}} f(x_1 \mid x_0; \theta + h_n/\sqrt{n}) d\mu(x_1). \tag{3.13}$$

As shown above (see (3.9) and below), the integral on the right-hand side of (3.13), multiplied by $n$, tends to zero in the first mean with respect to $P_\theta$, as $n \to \infty$. Because the left-hand side of (3.12) is, by virtue of (3.13), the same integral, multiplied by $\sqrt{n}$, it tends to 0 as well. Hence from (3.12), it follows that $P_\theta^0$-almost sure, $E_\theta\{h'\varphi_1(\theta) \mid \mathfrak{A}_0\} = 0$, and, by (3.10), $E_\theta\{Y_i \mid \mathfrak{A}_{i-1}\} = 0$ $P_\theta$-almost sure for any $i = 1, 2, \ldots$. Consequently, the sequence of sums $S_n = \sum_{i=1}^n Y_i$, $i = 1, 2\ldots$ is a martingale with respect to the system of $\sigma$-algebras $\{\mathfrak{A}_n, n = 0, 1, 2, \ldots\}$ and probability $P_\theta$. Thus, the sequence $S_n^2, n = 1, 2, \ldots$ is a submartingale. By Doob's inequality ( Loève (1960), §29.3), for any $\varepsilon > 0$,

$$P_\theta\left(\max_{1\le k\le n} |S_k| > \varepsilon\right) = P_\theta\left(\max_{1\le k\le n} S_k^2 > \varepsilon^2\right) \le \varepsilon^{-2} E_\theta S_n^2 = n\varepsilon^{-2} E_\theta Y_1^2. \tag{3.14}$$

The right-hand side of (3.14) does not exceed

$$\begin{aligned}&\varepsilon^{-2} E_\theta[(r_1(\theta_n,\theta)-1)\sqrt{n} - h'\varphi_1(\theta) - (\bar{r}_1(\theta_n,\theta)-1)\sqrt{n}]^2 \\ &\quad \le 2\varepsilon^{-2}\left[E_\theta[(r_1(\theta_n,\theta)-1)\sqrt{n} - h'\varphi_1(\theta)]^2 + E_\theta[(\bar{r}_1(\theta_n,\theta)-1)\sqrt{n}]^2\right].\end{aligned} \tag{3.15}$$

The first term on the right-hand side of (3.15) tends to zero, as $n \to \infty$, by Condition A2, and the second term, by Jenssen's inequality, does not exceed the first one, because $P_\theta$-almost sure,

$$(\bar{r}_1(\theta_n,\theta)-1)\sqrt{n} = E_\theta\{(r_1(\theta_n,\theta)-1)\sqrt{n} - h'\varphi_1(\theta) \mid \mathfrak{A}_0\}.$$

LEMMA 6. *For any* $\varepsilon > 0$,

$$\lim_{n\to\infty} P_\theta\left(\max_{1\le k\le n}\left|\Lambda_\theta(k, h_n/\sqrt{n}) - 2\sum_{i=1}^k (r_i(\theta_n,\theta)-1) + \sum_{i=1}^k (r_i(\theta_n,\theta)-1)^2\right| > \varepsilon\right) = 0.$$

PROOF. By Lemma 5.2 in Roussas (1972), for any $\varepsilon > 0$, the probability of the event $A_n = \{\max_{1\le k\le n} |r_k(\theta_n,\theta) - 1| > \varepsilon\}$, for $n$ large enough, does not exceed $\varepsilon$: $P_\theta(A_n) < \varepsilon$. For $\ln x = (x-1) - \frac{1}{2}(x-1)^2 + c(x-1)^3$, where $|c| < 3$ as $\frac{1}{2} \le x \le \frac{3}{2}$, then

$$\begin{aligned}&P_\theta\left(\max_{1\le k\le n}\left|\Lambda_\theta(k, h_n/\sqrt{n}) - 2\sum_{i=1}^k (r_i(\theta_n,\theta)-1) + \sum_{i=1}^k (r_i(\theta_n,\theta)-1)^2\right| > 2\varepsilon\right) \\ &\quad \le P_\theta(A_n) + P_\theta\left(\max_{1\le k\le n}\sum_{i-1}^k |r_i(\theta_n,\theta)-1|^3 > \varepsilon/3\right) \\ &\quad \le \varepsilon + P_\theta\left(\max_{1\le k\le n}|r_k(\theta_n,\theta)-1|\sum_{i=1}^n [r_i(\theta_n,\theta)-1]^2 > \varepsilon/3\right).\end{aligned} \tag{3.16}$$

Because, by Lemma 1, the sums $\sum_{i=1}^n [r_i(\theta_n,\theta)-1]^2$ are bounded in probability and $\max_{1\le k\le n}|r_k(\theta_n,\theta)-1| \to 0$, as $n \to \infty$, in $P_\theta$-probability, then the second term on the right-hand side of (3.16) tends to 0, as $n \to \infty$. Because $\varepsilon$ is arbitrarily small, this proves the Lemma.

PROOF OF THEOREM 1. First, from Lemmas 6 and 1, it follows that

$$\max_{1\le k\le n}\left|\Lambda_\theta(k, h_n/\sqrt{n}) - 2\sum_{i=1}^{k}(r_i(\theta_n,\theta)-1) + \frac{k}{n}E_\theta(h'\varphi_1(\theta))^2\right| \to 0, \tag{3.17}$$

as $n \to \infty$, in $P_\theta$-probability . From Lemmas 4 and 5, it follows that

$$\max_{1\le k\le n}\left|\sum_{i=1}^{n}(r_i(\theta_n,\theta)-1) - \frac{1}{\sqrt{n}}\sum_{i=1}^{k}h'\varphi_i(\theta) + \frac{k}{2n}E_\theta(h'\varphi_1(\theta))^2\right| \to 0, \tag{3.18}$$

as $n \to \infty$, in $P_\theta$-probability. From (3.17) and (3.18), we get

$$\max_{1\le k\le n}\left|\Lambda_\theta(k, h_n/\sqrt{n}) - \frac{2}{\sqrt{n}}\sum_{i=1}^{k}h'\varphi_i(\theta) + \frac{2k}{n}E_\theta(h'\varphi_1(\theta))^2\right| \to 0. \tag{3.19}$$

As $2\sum_{i=1}^{k}h'\varphi_i(\theta) = h'\Delta_k(\theta)$, and, by Lemma 3, $E_\theta(h'\varphi_1(\theta))^2 = \frac{1}{4}h'I(\theta)h$, then expression (3.19) is equivalent to the statement of Theorem 1.

It is natural to investigate asymptotic behaviour of the likelihood ratio under the alternative hypothesis that $x_0, x_1, \dots, x_n$ follow the distribution $P^n_{\theta_n}$. The following Theorem establishes this behaviour.

THEOREM 2. *For any sequence $h_n \to h$, as $n \to \infty$, and for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} P_{\theta_n}\left(\max_{1\le k\le n}\left|\Lambda_\theta(k, h_n/\sqrt{n}) - \frac{1}{\sqrt{n}}h'\Delta_k(\theta) + \frac{k}{2n}h'I(\theta)h\right| > \varepsilon\right) = 0. \tag{3.20}$$

PROOF. Denote by $\mathfrak{L}(\xi \mid P)$ the distribution law of the random variable $\xi$ on some space with probability $P$.

By Theorem 1, the sequences

$$\{\mathfrak{L}(\Lambda_\theta(n, h_n/\sqrt{n}) \mid P^n_\theta)\} \quad \text{and} \quad \left\{\mathfrak{L}\left(\frac{1}{\sqrt{n}}h'\Delta_n(\theta) - \frac{1}{2}h'I(\theta)h \mid P^n_\theta\right)\right\}$$

have the same weak limit. Because $E_\theta\{h'\varphi_i(\theta) \mid \mathfrak{A}_{i-1}\} = 0$ (see the proof of Lemma 5) and $E_\theta(h'\varphi_i(\theta))^2 = h'I(\theta)h/4$ (Lemma 3), the sequence $h'\Delta_n = 2\sum_{i=1}^{n}h'\varphi_i(\theta)$, $n = 1, 2, \dots$, is a martingale. By the central limiting theorem for martingales (Roussas (1972), Theorem 2.2)

$$\mathfrak{L}\left(\frac{1}{\sqrt{n}}h'\Delta_n(\theta) \mid P^n_\theta\right) \to N(0,\sigma^2), \qquad n \to \infty,$$

where $\sigma^2 = h'I(\theta)h$ and $N(m,\sigma^2)$ is the normal distribution with mean $m$ and variance $\sigma^2$, and the arrow denotes weak convergence of distributions. Then

$$\mathfrak{L}(\Lambda_\theta(n, h_n/\sqrt{n}) \mid P^n_\theta) \to N\left(-\frac{1}{2}\sigma^2, \sigma^2\right), \qquad n \to \infty,$$

where, denoting the limiting distribution by $\mathfrak{L}$, we have $\int e^x d\mathfrak{L}(x) = 1$. From Theorem 6 in Roussas (1972), it follows from this that sequences $\{P_\theta^n\}$ and $\{P_{\theta_n}^n\}$ are contiguous, and hence (3.20) is a consequence of Theorem 1 and the definition of contiguity in Roussas (1972).

## 4. Functional limit theorem for likelihood ratio

In this section, we consider random functions, naturally connected with the likelihood ratio, and weak convergence of their distributions. First, let us introduce some notation. Let $D[0,1]$ be Skorokhod's space with the Skorokhod's metric $\rho(x,y), x,y \in D[0,1]$. For a random element $\xi \in D[0,1]$ on a space with a measure $P$, let us denote by $\mathfrak{L}_D(\xi \mid P)$ its distibution on the $\sigma$-algebra of Borelian subsets $D[0,1]$.

Let

$$\Lambda_{\theta,n}^*(t; h_n/\sqrt{n}) = \Lambda_\theta([nt], h_n/\sqrt{n}), \qquad 0 \le t \le 1,$$

(here $[a]$ denotes the integer part of the number $a$).

THEOREM 3. *Under the conditions of Theorem* 1, *and as* $n \to \infty$,

$$\mathfrak{L}_D(\Lambda_{\theta,n}^*(\cdot; h_n/\sqrt{n}) \mid P_\theta^n) \to \mathfrak{L}_D(h'w - v), \tag{4.1}$$

$$\mathfrak{L}_D(\Lambda_{\theta,n}^*(\cdot; h_n/\sqrt{n}) \mid P_{\theta_n}^n) \to \mathfrak{L}_D(h'w + v), \tag{4.2}$$

*where* $w = w(h)$ *is the k-dimensional Wiener process with* $Ew(t) = 0$ *and covariance matrix* $Ew'(t)w(t) = tI(\theta)$, $v = v(t) = th'I(\theta)h/2$.

PROOF. Let us denote by $\Delta_n^*(t) = \frac{1}{\sqrt{n}}\Delta_{[nt]}(\theta)$, $n = 1, 2, \ldots$. By virtue of Theorem 4.1 in Billingsley (1968), it follows from Theorems 1 and 2 that the limiting distributions of the processes $\Lambda_{\theta,n}^*(t; h_n/\sqrt{n})$ and $h'\Delta_n^*(t) - [tn]h'I(\theta)h/2n$, $0 \le t \le 1$ are the same. Because the sequence of the function $v_n(t) = [tn]h'I(\theta)h/2n$ obviously converges to $v(t)$ uniformly, as $n \to \infty$, it remains to prove the convergence

$$\mathfrak{L}_D(h'\Delta_n^* \mid P_\theta^n) \to \mathfrak{L}_D(h'w), \tag{4.3}$$

$$\mathfrak{L}_D(h'\Delta_n^* \mid P_{\theta_n}^n) \to \mathfrak{L}_D(h'w + 2v), \tag{4.4}$$

as $n \to \infty$.

Convergence of finite-dimensional distributions, corresponding to (4.3) and (4.4) follows from Theorems 4.2 and 4.6 in Roussas (1972), respectively. Let us prove tightness of corresponding distributions families in $D[0,1]$.

Because, as shown above, the sequence $\{h'\Delta_n(\theta)\}$, $n = 1, 2, \ldots$, forms a martingale with respect to $(\mathfrak{A}_n, P_\theta^n)$, then tightness (and hence, convergence) of the sequence of the distributions $\mathfrak{L}_D(h'\Delta_n^* \mid P_\theta^n)$ follows from Theorem 23.1 Billingsley (1968). Tightness of $\mathfrak{L}_D(h'\Delta_n^* \mid P_{\theta_n}^n)$ follows from that of $\mathfrak{L}_D(h'\Delta_n^*) \mid P_\theta^n)$ because of contiguity of $P_\theta^n$ and $P_{\theta_n}^n$.

## 5. Asymptotic expansion of the likelihood ratio based on random number of observations

In this section, we show how the results of Sections 1–3 easily imply those of Akritas and Roussas (1979); as suggested by the authors, the auxiliary condition A5 in Akritas and Roussas (1979) turned out to be unnecessary.

THEOREM 4. *Let the conditions of Theorem 1 be fulfilled, and let $\nu_n$, $n = 1, 2, \ldots$, be a sequence of Markov moments with respect to the system of $\sigma$-algebras $\mathfrak{A}_n$, $n = 0, 1, 2, \ldots$, such that $\nu_n/n \to 1$ in $P_\theta$-probability. Then, as $n \to \infty$,*

$$\Lambda_\theta(\nu_n, h_n/\sqrt{n}) - \frac{1}{\sqrt{n}} h' \Delta_{\nu_n}(\theta) \to -\frac{1}{2} h' I(\theta) h, \tag{5.1}$$

*both in $P_\theta$-probability and in $P_{\theta_n}$-probability, whereas*

$$\mathcal{L}\left(\frac{1}{\sqrt{n}} h' \Delta_{\nu_n}(\theta) \mid P_\theta^n\right) \to N\left(-\frac{1}{2}\sigma^2, \sigma^2\right), \tag{5.2}$$

$$\mathcal{L}\left(\frac{1}{\sqrt{n}} h' \Delta_{\nu_n}(\theta) \mid P_{\theta_n}^n\right) \to N\left(\frac{1}{2}\sigma^2, \sigma^2\right), \tag{5.3}$$

*where $\sigma^2 = h' I(\theta) h$.*

PROOF. Let $\delta > 0$ be any number. By the condition $\nu_n/n \to 1$ in $P_\theta$-probability, it follows that $\lim_{n\to\infty} P_\theta(\nu_n > 2n) = 0$. Hence,

$$\begin{aligned}
&\limsup_{n\to\infty} P_\theta\left(\left|\Lambda_\theta(\nu_n, h_n/\sqrt{n}) - \frac{1}{\sqrt{n}} h' \Delta_{\nu_n}(\theta) + \frac{\nu_n}{2n} h' I(\theta) h\right| > \delta\right) \\
&\le \lim_{n\to\infty} \Bigg[P_\theta(\nu_n > 2n) + P_\theta\Bigg(\max_{1\le k\le 2n} \Big|\Lambda_\theta(k, h_n/\sqrt{n}) \\
&\qquad - \frac{1}{\sqrt{n}} h' \Delta_k(\theta) + \frac{k}{2n} h' I(\theta) h\Big| > \delta\Bigg)\Bigg] = 0
\end{aligned}$$

according to Theorem 1. As $\nu_n/n \to 1$ in $P_\theta$-probability, this implies convergence (5.1) in $P_\theta$-probability.

Convergence in $P_{\theta_n}$-probability follows from Theorem 2 in a similar way. To show that $\lim_{n\to\infty} P_{\theta_n}(\nu_n > 2n) = 0$, the contiguity of $P_{\theta_n}^n$ and $P_\theta^n$ is used.

Because, by virtue of Theorem 3, convergence (5.2) holds with $n$ instead of $\nu_n$, (5.2) will be established, if we show that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P_\theta\left(\frac{1}{\sqrt{n}} \mid h' \Delta_{\nu_n}(\theta) - h' \Delta_n(\theta) \mid > \varepsilon\right) = 0. \tag{5.4}$$

Let us estimate the probability in (5.4). Let $\delta > 0$ be fixed. Then

$$\begin{aligned}
&\limsup_{n\to\infty} P_\theta\left(\frac{1}{\sqrt{n}} |h' \Delta_{\nu_n}(\theta) - h' \Delta_n(\theta)| > \varepsilon\right) \le \lim_{n\to\infty} P_\theta(|\nu_n/n - 1| > \delta) \\
&\quad + \limsup_{n\to\infty} P_\theta\left(\frac{1}{\sqrt{n}} \sup_{n(1-\delta)\le k\le n(1+\delta)} |h' \Delta_k(\theta) - h' \Delta_n(\theta)| > \varepsilon\right) \\
&\le 2 \limsup_{n\to\infty} P_\theta\left(\frac{1}{\sqrt{n}} \sup_{1\le k\le n\delta} |h' \Delta_k(\theta)| > \varepsilon\right) \\
&= 2 \lim_{n\to\infty} P_\theta\left(\sup_{0\le t\le \delta} |h' \Delta_n^*(t)| > \varepsilon\right) = 2P\left(\sup_{0\le t\le \delta} |h' w(t)| > \varepsilon\right).
\end{aligned} \tag{5.5}$$

The last equality follows from Theorem 3 here and Theorem 5.1 in Billingsley (1968), because the functional $\omega$ on $D[0,1]$, defined for any $x \in D[0,1]$ as $\omega(x) = \sup_{0\le t\le\delta} |x(t)|$, is obviously continuous on the set $C[0, 1]$ of continuous functions on $[0, 1]$, to which almost all paths of the limiting process $h'w$ belong. Letting $\delta$ on the right-hand side of (5.5) tend to zero, we get (5.4), and hence (5.2). The statement (5.3) of Theorem 4 is proven analogously.

## Acknowledgements

The author wishes to thank the anonymous referees for their valuable suggestions, which contributed to the improvement of the final form of this paper.

## References

Akritas, M. G. and Roussas, G. G. (1979). Asymptotic expansion of the log-likelihood function based on stopping times defined on a Markov process, *Ann. Inst. Statist. Math.*, **31**(1), 21–38.

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

Ibragimov, I. A. and Has′minskiĭ, R. Z. (1981). *Statistical Estimation. Asymptotic Theory*, Springer, New York, Berlin.

Le Cam, L. (1960). Locally asymptotically normal families of distributions, *Univ. of California Publ. in Statistics*, **3**, 37–98

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer Ser. Statist., Springer, New York, Berlin.

Le Cam, L. and Yang, G. L. (1990). *Asymptotics in Statistics. Some Basic Concepts*, Springer Ser. Statist., Springer, New York.

Loève, M. (1960). *Probability Theory*, 2nd ed., Van Nostrand, Princeton, N. J.-Toronto, Ont.-London.

Roussas, G. G. (1972). *Contiguity of Probability Measures. Some Applications in Statistics*, Cambridge University Press, Massachusetts.