JOURNAL OF MULTIVARIATE ANALYSIS **56**, 120–152 (1996)
ARTICLE NO. 0007

On the Rate of Approximations for Maximum Likelihood Tests in Change-Point Models

EDIT GOMBAY

University of Alberta

AND

LAJOS HORVÁTH

University of Utah

We study the asymptotics of maximum-likelihood ratio-type statistics for testing a sequence of observations for no change in parameters against a possible change while some nuisance parameters remain constant over time. We obtain extreme value as well as Gaussian-type approximations for the likelihood ratio. We get necessary and sufficient conditions for the weak convergence of supremum and L_p-functionals of the likelihood ration process. We also approximate the maximum likelihood ratio with Ornstein–Uhlenbeck processes and obtain bounds for the rate of approximation. We show that the Ornstein–Uhlenbeck approach is superior to the extreme value limit in case of moderate sample sizes. © 1996 Academic Press, Inc.

1. APPROXIMATIONS FOR THE LIKELIHOOD RATIO PROCESS

Let $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ be independent random vectors in R^m with distribution functions $F(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\eta}_1), ..., F(\mathbf{x}; \boldsymbol{\theta}_n, \boldsymbol{\eta}_n)$, where $\boldsymbol{\theta}_i \in \Theta^{(1)} \subseteq R^d$ and $\boldsymbol{\eta}_i \in \Theta^{(2)} \subseteq R^p$, $1 \leqslant i \leqslant n$. We want to test

$$H_o: \boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_n, \qquad \boldsymbol{\eta}_1 = \cdots = \boldsymbol{\eta}_n$$

against the change-point alternative

H_A: there is k^*, $1 \leqslant k^* < n$ such that

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \cdots = \boldsymbol{\theta}_{k^*} \neq \boldsymbol{\theta}_{k^*+1} = \cdots = \boldsymbol{\theta}_n, \qquad \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 = \cdots = \boldsymbol{\eta}_n.$$

Received April 18, 1994; revised September 1995.

AMS 1991 subject classifications: primary 62A10; secondary 62F03.

Key words and phrases: likelihood ratio processes, maximum likelihood estimators, weighted approximations, extreme value, Brownian bridge.

0047-259X/96 $12.00

Copyright © 1996 by Academic Press, Inc.
All rights of reproduction in any form reserved.

The alternative means that a given set of unknown parameters change after an unknown time k^* and the rest of the parameters remain constant under the alternative. The parameters are unknown under the null, as well as the alternative, hypothesis. We use the likelihood ratio test to check H_o against H_A. We assume that the observations have probability densities $f(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\eta}_1), ..., f(\mathbf{x}; \boldsymbol{\theta}_n, \boldsymbol{\eta}_n)$ with respect to ν, where ν is a σ-finite measure. If we know that the change occurs at time $k^* = k$, then we should reject H_o for small values of

$$\Lambda = \frac{\sup_{(\boldsymbol{\theta}, \boldsymbol{\eta}) \in \Theta^{(1)} \times \Theta^{(2)}} \prod_{1 \leqslant i \leqslant n} f(\mathbf{X}_i; \boldsymbol{\theta}, \boldsymbol{\eta})}{\sup_{(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\eta}) \in \Theta^{(1)} \times \Theta^{(1)} \times \Theta^{(2)}} \prod_{1 \leqslant i \leqslant k} f(\mathbf{X}_i; \boldsymbol{\theta}, \boldsymbol{\eta}) \prod_{k \leqslant i \leqslant n} f(\mathbf{X}_i; \boldsymbol{\tau}, \boldsymbol{\eta})}. \tag{1.1}$$

We consider the case when the densities are smooth functions of the parameters and the parameters can be estimated consistently. Hence following Lehmann (1991, p. 409), we assume

C.1. $F(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\eta})$ generates distinct measures if $(\boldsymbol{\theta}, \boldsymbol{\eta}) \in \Theta^{(1)} \times \Theta^{(2)}$.

Let

$$g(\mathbf{x}; \mathbf{y}) = \log f(\mathbf{x}; \mathbf{y}), \qquad \mathbf{x} \in R^m, \quad \mathbf{y} \in \Theta^{(1)} \times \Theta^{(2)}, \tag{1.2}$$

and

$$g_i(\mathbf{x}; \mathbf{y}) = \frac{\partial}{\partial y_i} g(\mathbf{x}; \mathbf{y}), \qquad \mathbf{y} = (y_i, ..., y_{d+p}). \tag{1.3}$$

We also assume that we have unique maximum likelihood estimators under the null hypothesis and also under the union of the null and the alternative hypothesises:

C.2. For each $k = 1, 2, ..., n$ we can find unique $\hat{\boldsymbol{\theta}}_k, \boldsymbol{\theta}_k^*$, and $\hat{\boldsymbol{\eta}}_k$ such that

$$\sum_{1 \leqslant j \leqslant k} g_i(\mathbf{X}_j; \hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_k) = 0, \qquad 1 \leqslant i \leqslant d, \tag{1.4}$$

$$\sum_{k < j \leqslant n} g_i(\mathbf{X}_j; \boldsymbol{\theta}_k^*, \hat{\boldsymbol{\eta}}_k) = 0, \qquad 1 \leqslant i \leqslant d, \tag{1.5}$$

and

$$\sum_{1 \leqslant j \leqslant k} g_i(\mathbf{X}_j; \hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_k) + \sum_{k < j \leqslant n} g_i(\mathbf{X}_j; \boldsymbol{\theta}_j^*, \hat{\boldsymbol{\eta}}_k) = 0, \qquad d < i \leqslant d + p. \tag{1.6}$$

(If $k = n$, then we must solve only (1.4), (1.6) and $\boldsymbol{\theta}_n^*$ is undefined.) Hence the log likelihood ratio can be written as

$$-2 \log \Lambda_k = 2\{L_k(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_k) + L_k^*(\boldsymbol{\theta}_k^*, \hat{\boldsymbol{\eta}}_k) - L_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\eta}}_n)\}, \tag{1.7}$$

where

$$L_k(\mathbf{y}) = \sum_{1 \leqslant j \leqslant k} g(\mathbf{X}_j; \mathbf{y}) \tag{1.8}$$

and

$$L_k^*(\mathbf{y}) = \sum_{k < j \leqslant n} g(\mathbf{X}_j; \mathbf{y}). \tag{1.9}$$

Since k^* is unknown, we reject H_o if

$$Z_n = \max_{1 \leqslant k < n} (-2 \log \Lambda_k) \tag{1.10}$$

is large.

Assuming that the observations are univariate normals with constant variance, Sen and Srivastava (1975a, 1975b), Hawkis (1977), Siegmund (1985), Yao and Davis (1986), and James, James, and Siegmund (1987) studied the behavior of the likelihood ratio. Worsley (1986a, 1986b), Haccou, Meelis, and Van de Geer (1988), and Gombay and Horváth (1990) considered tests to detect changes in the mean of exponential observations. Worsley (1983) and Horváth (1989) obtained similar results in case of binomial r.v.'s. Yao and Davis (1986), Csörgő and Horváth (1988), and Gombay and Horváth (1990) obtained the double exponential limit distribution for $Z_n^{1/2}$ in case of univariate normal observations with constant variance and possible change in the mean. Horváth (1993) proved similar results when the mean and the variance can change at an unknown time. Gombay and Horváth (1994) generalized his result to the general case when all parameters must change at the same time under the alternative hypothesis. Srivastava and Worsley (1986) and James *et al.* (1992) considered tests for change in the mean of multivariate normal observations. Our results cover the general case when there is no assumption on the form of $f(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\eta})$ and $\boldsymbol{\eta}$ remains constant but $\boldsymbol{\theta}$ must change under the alternative.

Before we discuss the properties of Z_n under the null hypothesis we must introduce further regularity conditions on the underlying class of distributions. The true values of the unknown parameters are $\boldsymbol{\theta}_o$ and $\boldsymbol{\eta}_o$ under H_o. Let $\Theta = \Theta^{(1)} \times \Theta^{(2)}$ and

$$g_{i_1, \ldots, i_r}(\mathbf{x}; \mathbf{y}) = \frac{\partial^r g(\mathbf{x}; \mathbf{y})}{\partial y_{i_1} \cdots \partial y_{i_r}}, \qquad \mathbf{y} = (y_1, \ldots, y_{d+p}).$$

We assume the following regularity conditions:

C.3. There is an open interval $\Theta_o \subseteq \Theta \subseteq R^{d+p}$ containing $(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)$ such that $g_i(\mathbf{x}; \mathbf{y})$, $g_{i,j}(\mathbf{x}; \mathbf{y})$ and $g_{i,j,k}(\mathbf{x}; \mathbf{y})$, $1 \leqslant i, j, k \leqslant d+p$ exist and are continuous in $\mathbf{y}$ for all $\mathbf{x} \in R^m$ and $\mathbf{y} \in \Theta_o$.

C.4. There are functions $M_1(\mathbf{x})$ and $M_2(\mathbf{x})$ such that $|g_i(\mathbf{x}; \mathbf{y})| \leqslant M_1(\mathbf{x})$, $|g_{i,j}(\mathbf{x}; \mathbf{y})| \leqslant M_2(\mathbf{x})$ and $|g_{i,j,k}(\mathbf{x}; \mathbf{y})| \leqslant M_2(\mathbf{x})$ for all $\mathbf{x} \in R^m$, $\mathbf{y} \in \Theta_o$, $1 \leqslant i, j, k \leqslant d+p$, and

$$\int_{R^m} M_1(\mathbf{x})\, \nu(d\mathbf{x}) < \infty, \qquad E_{(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)} M_2(\mathbf{X}_1) < \infty.$$

C.5. $E_{\mathbf{y}} g_i(\mathbf{X}_1; \mathbf{y}) = 0$ for all $1 \leqslant i \leqslant d+p$ and $\mathbf{y} \in \Theta_o$.

C.6. $J_{ij}(\mathbf{y}) = E_{\mathbf{y}} g_i(\mathbf{X}_1; \mathbf{y})\, g_j(\mathbf{X}_1; \mathbf{y}) = -E_{\mathbf{y}} g_{i,j}(\mathbf{X}_1; \mathbf{y})$, $1 \leqslant i, j \leqslant d+p$, and $J^{-1}(\mathbf{y})$ exist and they are continuous for all $\mathbf{y} \in \Theta_o$, where $J(\mathbf{y}) = \{J_{ij}(\mathbf{y}),\ 1 \leqslant i, j \leqslant d+p\}$ is the information matrix.

C.7. $\operatorname{var}_{(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)} g_{i,j}(\mathbf{X}_1; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o) < \infty$ for all $1 \leqslant i, j \leqslant d+p$, and

C.8. $E_{(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)} |g_i(\mathbf{X}_1; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o)|^{\mu} < \infty$ for all $1 \leqslant i \leqslant d+p$ with some $\mu > 2$.

The main goal of our paper is to get approximations for the distribution of Z_n and to provide some information about the rate of convergence of these approximations. The proofs are based on the observation that $-2 \log \Lambda_k$ can be approximated with quadratic forms of sums of independent random vectors. Let

$$D_{11} = \{J_{ij}(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o), \quad 1 \leqslant i, j \leqslant d\} \tag{1.11}$$

and

$$\mathbf{V}_{k,1} = \left(\sum_{1 \leqslant j \leqslant k} g_1(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o), \ldots, \sum_{1 \leqslant j \leqslant k} g_d(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o) \right), \tag{1.12}$$

$$\mathbf{V}_{k,2} = \left(\sum_{k < j \leqslant n} g_1(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o), \ldots, \sum_{k < j \leqslant n} g_d(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o) \right). \tag{1.13}$$

It follows from C.6 that D_{11}^{-1} exists. Let

$$R_k = \frac{1}{k} \mathbf{V}_{k,1} D_{11}^{-1} \mathbf{V}_{k,1}^{\mathrm{T}} + \frac{1}{n-k} \mathbf{V}_{k,2} D_{11}^{-1} \mathbf{V}_{k,2}^{\mathrm{T}}$$
$$- \frac{1}{n} (\mathbf{V}_{k,1} + \mathbf{V}_{k,2})\, D_{11}^{-1} (\mathbf{V}_{k,1} + \mathbf{V}_{k,2})^{\mathrm{T}}, \tag{1.14}$$

where $\mathbf{x}^{\mathrm{T}}$ denotes the transpose of $\mathbf{x}$.

THEOREM 1.1. *If H_o and* C.1–C.8 *hold, then for all* $0 \leqslant \alpha < \frac{1}{2}$ *we have*

$$n^{\alpha} \max_{1 \leqslant k \leqslant n} \left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\alpha} |(-2 \log \Lambda_k) - R_k| = O_P(1) \tag{1.15}$$

and

$$\max_{1 \leqslant k < n} \frac{k}{n}\left(1-\frac{k}{n}\right) |(-2 \log \Lambda_k) - R_k| = O_P(n^{-1/2}(\log \log n)^{3/2}). \tag{1.16}$$

Theorem 1.1 can be used to get Gaussian approximations for $-2 \log \Lambda_k$. Following Vostrikova (1983) we define the likelihood ratio process

$$V_n(t) = -2 \log \Lambda_{[(n+1)t]}, \qquad 1/(n+1) \leqslant t < n/(n+1) \tag{1.17}$$

and $V_n(t) = 0$, if $0 \leqslant t < 1/(n+1)$, and $V_n(t) = 0$, if $n/(n+t) < t \leqslant 1$. The Gaussian approximation for $V_n(t)$ will imply necessary and sufficient condition for the weak convergence of weighted functionals of $V_n(t)$. Let $\{B_i(t),\ 0 \leqslant t \leqslant 1\}$, $1 \leqslant i \leqslant d$, be independent Brownian bridges and define

$$B^{(d)}(t) = \sum_{1 \leqslant i \leqslant d} B_i^2(t). \tag{1.18}$$

THEOREM 1.2. *If H_o and* C.1–C.8 *hold, then we can find a sequence of stochastic processes* $\{B_n^{(d)}(t),\ 0 \leqslant t \leqslant 1\}$ *such that*

$$\{B_n^{(d)}(t), 0 \leqslant t \leqslant 1\} \overset{\mathscr{D}}{=} \{B^{(d)}(t), 0 \leqslant t \leqslant 1\} \tag{1.19}$$

for each n,

$$n^{\alpha} \sup_{\lambda/n \leqslant t \leqslant 1-\lambda/n} (t(1-t))^{\alpha}\, |V_n(t) - B_n^{(d)}(t)/(t(1-t))| = O_P(1) \tag{1.20}$$

for all $\lambda > 0$ *and* $0 \leqslant \alpha < 1/2 - 1/\mu$ *and*

$$\sup_{0 \leqslant t \leqslant 1} |t(1-t)\, V_n(t) - B_n^{(d)}(t)| = o_P(n^{1/\mu - 1/2}). \tag{1.21}$$

Theorem 1.2 implies immediately the weak convergence of $t(1-t)\, V_n(t)$ and the convergence of weighted functionals of $V_n(t)$ in distribution. Let

$$I_{0,1}(q, c) = \int_0^1 \frac{1}{t(1-t)} \exp(-cq^2(t)/(t(1-t)))\, dt. \tag{1.22}$$

The integral in (1.22) is the Chibisov–O'Reilly version of the Kolmogorov–Erdős–Feller–Petrovsky integral test (cf., for example, Csörgő and Horváth (1993, Chap. 4)).

THEOREM 1.3. *We assume that* H_o, C.1–C.8 *hold and* q *is positive on* $(0, 1)$ *increases in a neighborhood of zero and decreases in a neighborhood of one.*

(i) *We can define a sequence of stochastic processes* $\{B_n^{(d)}(t), 0 \leqslant t \leqslant 1\}$ *such that* (1.19) *holds and*

$$\sup_{0 \leqslant t \leqslant 1} |t(1-t)\, V_n(t) - B_n^{(d)}(t)|/q^2(t) = o_P(1) \tag{1.23}$$

if and only if $I_{0,1}(q, c) < \infty$ *for all* $c > 0$.

(ii) $I_{0,1}(q, c) < \infty$ *for some* $c > 0$ *if and only if*

$$\sup_{0 \leqslant t \leqslant 1} |t(1-t)\, V_n(t)|/q^2(t) \xrightarrow{\mathscr{D}} \sup_{0 \leqslant t \leqslant 1} B^{(d)}(t)/q^2(t). \tag{1.24}$$

THEOREM 1.4. *We assume that* H_o, C.1–C.8 *hold,* q *is positive on* $(0, 1)$, *and* $0 < \alpha < \infty$. *Then*

$$\int_0^1 V_n^\alpha(t)/q(t)\, dt \xrightarrow{\mathscr{D}} \int_0^1 (B^{(d)}(t))^\alpha/((t(1-t))^\alpha\, q(t))\, dt \tag{1.25}$$

if and only if

$$\int_0^1 \frac{1}{q(t)}\, dt < \infty. \tag{1.26}$$

It follows immediately from Theorems 1.3 and 1.4 that

$$\sup_{0 \leqslant t \leqslant 1} t(1-t)\, V_n(t) \xrightarrow{\mathscr{D}} \sup_{0 \leqslant t \leqslant 1} B^{(d)}(t), \tag{1.27}$$

$$\int_0^1 t(1-t)\, V_n(t)\, dt \xrightarrow{\mathscr{D}} \int_0^1 B^{(d)}(t)\, dt, \tag{1.28}$$

and

$$\int_0^1 V_n(t)\, dt \xrightarrow{\mathscr{D}} \int_0^1 \frac{B^{(d)}(t)}{t(1-t)}\, dt. \tag{1.29}$$

Kiefer (1959a, 1959b) obtained formulas for the distribution functions of the limiting random variables in (1.27) and (1.28). Scholz and Stephens (1987) tabulated the distribution of the limit in (1.29).

Since $I_{0,1}((t(1-t))^{1/2}, c) = \infty$ for all $c > 0$, the limit distribution of Z_n cannot follow from (1.23) or (1.24). The following section contains the limit distribution of $Z_n^{1/2}$ and we also discuss Gaussian-type approximation for

Z_n. In Section 3 we provide methods to get critical values for $Z_n^{1/2}$ and Monte Carlo simulations show that out asymptotic critical values are acceptable in case of moderate sample sizes. The proofs of our results are given in Sections 4–7.

2. Approximations for $Z_n^{1/2}$

We start with the limit distribution of $Z_n^{1/2}$. Let $a(t) = (2 \log t)^{1/2}$ and $b_d(t) = 2 \log t + (d/2) \log \log t - \log \Gamma(d/2)$, where

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y}\, dy, \qquad 1 \leqslant t < \infty.$$

Theorem 2.1. *If H_o and* C.1–C.8 *hold, then we have*

$$\lim_{n \to \infty} P\{a(\log n)\, Z_n^{1/2} \leqslant t + b_d(\log n)\} = \exp(-2e^{-t}) \tag{2.1}$$

for all t.

The rate of convergence to extreme value distributions is usually slow and therefore we may need very large sample sizes if we want to use (2.1). However, we see in the following section that (2.1) gives conservative rejection regions in case of small and moderate sample sizes. Since (2.1) works only for large sample sizes it is important to get further approximations for $Z_n^{1/2}$ and get bounds for the rate of convergence. It is clear that (1.20) yields

$$|Z_n^{1/2} - \sup_{1/n \leqslant t \leqslant 1-1/n} (B_n^{(d)}(t)/(t(1-t)))^{1/2}| = O_P(1), \tag{2.2}$$

which gives no convincing evidence that Gaussian approximation is better than (2.1). The nex two theorems show that (2.2) can be improved.

Theorem 2.2. *If H_o and* C.1–C.8 *hold, then we have*

$$|Z_n^{1/2} - \sup_{1/n \leqslant t \leqslant 1-1/n} (B_n^{(d)}(t)/(t(1-t)))^{1/2}| = O_P(\exp(-\log n)^{1-\varepsilon})) \tag{2.3}$$

for all $0 < \varepsilon < 1$, where $\{B_n^{(d)}(t),\ 0 \leqslant t \leqslant 1\}$ are defined in Theorem 1.2. *Also, if $h(n) \geqslant 1/n$, $l(n) \geqslant 1/n$, and*

$$\limsup_{n \to \infty} n(h(n) + l(n)) \exp(-(\log n)^{1-\varepsilon^*}) < \infty \tag{2.4}$$

for some $0 < \varepsilon^* \leqslant 1$, *then we have*

$$|Z_n^{1/2} - \sup_{h(n) \leqslant t \leqslant 1 - l(n)} (B_n^{(d)}(t)/(t(1-t)))^{1/2}| = O_P(\exp(-(\log n)^{1-\varepsilon}) \quad (2.5)$$

for all $0 < \varepsilon < \varepsilon^*$.

Monte Carlo simulations demonstrate that the approximation in (2.5) works well even for moderate sample sizes if we choose $l(n) = h(n) = (\log n)^{3/2}/n$.

3. Asymptotic Critical Values for $Z_n^{1/2}$

Let $0 < \alpha < 1$ and define

$$z_n = z_n(1-\alpha) = \sup(x\colon P\{Z_n^{1/2} \leqslant x\} \leqslant 1-\alpha) \quad (3.1)$$

and

$$\begin{aligned} r(h, l) &= r(h, l; 1-\alpha) \\ &= \sup(x\colon P\{\sup_{h \leqslant t \leqslant 1-t} \{B^{(d)}(t)/(t(1-t))\}^{1/2} \leqslant x\} = 1-\alpha), \end{aligned} \quad (3.2)$$

where $\{B^{(d)}(t), 0 \leqslant t \leqslant 1\}$ is defined in (1.18). First we show that $r(h, l)$ is an asymptotically correct critical value of size α.

Theorem 3.1. *We assume that* H_o *and* C.1–C.8 *hold. If* $h(n) \geqslant 1/n$, $l(n) \geqslant 1/n$, *and* (2.4) *is satisfied with some* $0 < \varepsilon^* \leqslant 1$, *then we have*

$$\lim_{n \to \infty} P\{Z_n^{1/2} > r(h(n), l(n))\} = \alpha \quad (3.3)$$

and

$$|z_n - r(h(n), l(n))| = o((\log\log n)^{-1/2}). \quad (3.4)$$

It follows from Theorems 2.1 and 2.2 that

$$\lim_{n \to \infty} \frac{z_n}{(2\log\log n)^{1/2}} = 1 \quad (3.5)$$

and

$$\lim_{n \to \infty} \frac{r(h(n), l(n))}{(2\log\log n)^{1/2}} = 1, \quad (3.6)$$

and therefore it is not immediate that (3.3) implies (3.4).

TABLE I

Critical Values for $Z_n^{1/2}$ in Case of Exponential Observations

Sample size	$1-\alpha$	Worsely's critical values	Asymptotic critical values	$r\left(\frac{(\log n)^{3/2}}{n}, \frac{(\log n)^{3/2}}{n}; 1-\alpha\right)$	$z_n(1-\alpha)$
20	0.90	2.63	3.11 (0.976)	2.52 (0.866)	2.64
	0.95	2.90	3.60 (0.993)	2.81 (0.939)	2.87
	0.99	3.43	4.70 (1.000)	3.38 (0.985)	3.47
50	0.90	2.78	3.18 (0.968)	2.69 (0.876)	2.74
	0.95	3.04	3.62 (0.993)	2.97 (0.944)	2.99
	0.99	3.56	4.69 (1.000)	3.52 (0.991)	3.52
100	0.90	2.86	3.23 (0.968)	2.79 (0.885)	2.86
	0.95	3.12	3.64 (0.991)	3.06 (0.940)	3.10
	0.99	3.63	4.57 (1.000)	3.59 (0.990)	3.60
500	0.90	NA	3.31 (0.961)	2.95 (0.875)	3.03
	0.95	NA	3.69 (0.988)	3.20 (0.941)	3.25
	0.99	NA	4.54 (1.000)	3.71 (0.988)	3.76

The computation of the distribution of $\sup(B^{(d)}(t)/(t(1-t)))^{1/2}$ is based on a representation using Ornstein–Uhlenbeck processes. Let $V_1, ..., V_d$ be independent identically distributed Ornstein–Uhlenbeck processes. This means that $V_i(t)$ is a Gaussian process with $EV_i(t)=0$ and $EV_i(t)\ V_i(s)=\exp(-\frac{1}{2}|t-s|)$. Next we define

$$\Delta(t)=\left(\sum_{1\leqslant i\leqslant d} V_i^2(t)\right)^{1/2}.$$

TABLE II

Critical Values for $Z_n^{1/2}$ in Case of Poisson Observations

Sample size	$1-\alpha$	Asymptotic critical values	$r\left(\frac{(\log n)^{3/2}}{n}, \frac{(\log n)^{3/2}}{n}; 1-\alpha\right)$	$z_n(1-\alpha)$
20	0.90	3.11	2.52	2.57
		(0.977)	(0.833)	
	0.95	3.60	2.81	2.82
		(0.996)	(0.948)	
	0.99	4.70	3.38	3.44
		(1.000)	(0.988)	
50	0.90	3.18	2.69	2.71
		(0.976)	(0.896)	
	0.95	3.62	2.97	2.95
		(0.997)	(0.952)	
	0.99	4.60	3.52	3.38
		(1.000)	(0.994)	
100	0.90	3.23	2.79	2.83
		(0.965)	(0.891)	
	0.95	3.64	3.06	3.06
		(0.991)	(0.950)	
	0.99	4.57	3.59	3.62
		(1.000)	(0.989)	
500	0.90	3.31	2.95	2.98
		(0.964)	(0.893)	
	0.95	3.69	3.20	3.18
		(0.992)	(0.953)	
	0.99	4.54	3.71	3.67
		(1.000)	(0.993)	

It is easy to see that for all $0 < h < 1 - l < 1$ we have

$$\sup_{h \leqslant t \leqslant 1-l} \left(\frac{B^{(d)}(t)}{t(1-t)}\right)^{1/2} \stackrel{\mathscr{D}}{=} \sup_{0 \leqslant t \leqslant \log(1-h)(1-l)/hl} \Delta(t). \tag{3.7}$$

There is no known simple formula for the distribution function of $\sup_{0 \leqslant t \leqslant T} \Delta(t)$. However, it is relatively easy to get its Laplace transform and, inverting it numerically, we get selected values of the distribution function of $\sup_{0 \leqslant t \leqslant T} \Delta(t)$ (cf. Keilson and Ross, 1975, and DeLong, 1981). Vostrikova (1981) showed that for all $T > 0$

$$P\{\sup_{0 \leqslant t \leqslant T} \Delta(t) > x\} = \frac{x^d \exp(-x^2/2)}{2^{d/2}\Gamma(d/2)} \left\{T - \frac{d}{x^2} T + \frac{4}{x^2} + O\left(\frac{1}{x^4}\right)\right\} \tag{3.8}$$

TABLE III

Critical Values for $Z_n^{1/2}$ in Case of Normal Observations
(Change in the Mean, the Variance Is Constant and Known)

Sample size	$1-\alpha$	Yao and Davis critical values	Asymptotic critical values	$r\left(\frac{(\log n)^{3/2}}{n}, \frac{(\log n)^{3/2}}{n}; 1-\alpha\right)$	$z_n(1-\alpha)$
20	0.90[a]	2.53	3.11	2.52	2.54
			(0.977)	(0.897)	
	0.95	2.82	3.60	2.81	2.81
			(0.997)	(0.950)	
	0.99	3.38	4.70	3.38	3.37
			(1.000)	(0.991)	
50	0.90[b]	2.75	3.18	2.69	2.71
			(0.976)	(0.899)	
	0.95[c]	3.02	3.62	2.97	2.94
			(0.997)	(0.954)	
	0.99	3.56	4.60	3.52	3.50
			(1.000)	(0.991)	
100	0.90[d]	2.85	3.23	2.79	2.82
			(0.971)	(0.894)	
	0.95	3.12	3.64	3.06	3.06
			(0.992)	(0.950)	
	0.99	3.58	4.57	3.59	3.58
			(1.000)	(0.991)	
500	0.90[e]	3.02	3.31	2.95	2.93
			(0.963)	(0.907)	
	0.95	3.28	3.69	3.20	3.21
			(0.995)	(0.949)	
	0.99	3.80	4.54	3.71	3.69
			(1.000)	(0.991)	

[a] Siegmund's (1985) critical value is 2.55; Hawkins' (1977) bound is 2.79.
[b] Hawkins' (1977) critical value is 2.73.
[c] Siegmund's (1985) critical value is 2.97.
[d] Hawkins' (1977) critical value is 2.85.
[e] Hawkins' (1977) critical value is 3.03.

as $x \to \infty$. It turns out that the tail approximation in (3.8) works very well even in case of moderate x.

We used Theorem 2.1, as well as Theorem 3.1, to get critical values for $Z_n^{1/2}$. We chose $h(n) = l(n) = (\log n)^{3/2}/n$ in (3.2) and compared $r(h, l)$ to the asymptotic critical values r^* based on Theorem 2.1 and values obtained by Monte-Carlo simulations. We used 5000 repetitions in the simulations and the results are reported in Tables I–VI.

The probabilities $P\{Z_n^{1/2} \leqslant r^*\}$ and $P\{Z_n^{1/2} \leqslant r\}$ are also given in parentheses beneath the values of r and r^*. It turned out that the critical value obtained from the extreme value limit distribution overestimates the true one. The only exception is the case when the mean and the variance of normal observation can change under H_A. In this case the asymptotic distribution gives better critical values for small sample sizes, because the possible change in the variance increases Z_n. In Table I we compare Worsley's (1986a) critical values to ours when the observations follow exponential distribution. Worsley (1986a) overestimates the true critical values and, because of the recursive nature of his calculations, this method works only if the sample size is small. Table II gives the results in case of

TABLE IV

Critical Values for $Z_n^{1/2}$ in Case of Normal Observations
(Change in the Mean, the Variance Is Constant but Unknown)

Sample size	$1-\alpha$	Yao and Davis critical values	Asymptotic critical values	$r\left(\frac{(\log n)^{3/2}}{n}, \frac{(\log n)^{3/2}}{n}; 1-\alpha\right)$	$z_n(1-\alpha)$
20	0.90	2.53	3.11	2.52	2.76
			(0.959)	(0.820)	
	0.95	2.82	3.60	2.81	3.01
			(0.993)	(0.913)	
	0.99	3.38	4.70	3.38	3.55
			(1.000)	(0.980)	
50	0.90	2.75	3.18	2.69	2.82
			(0.968)	(0.865)	
	0.95	3.02	3.62	2.97	3.05
			(0.994)	(0.936)	
	0.99	3.56	4.60	3.52	3.54
			(1.000)	(0.989)	
100	0.90	2.85	3.23	2.79	2.84
			(0.964)	(0.882)	
	0.95	3.12	3.64	3.06	3.11
			(0.988)	(0.945)	
	0.99	3.58	4.57	3.59	3.68
			(1.000)	(0.986)	
500	0.90	3.02	3.31	2.95	2.94
			(0.958)	(0.905)	
	0.95	3.28	3.69	3.20	3.23
			(0.992)	(0.943)	
	0.99	3.80	4.54	3.71	3.66
			(1.000)	(0.992)	

Poisson observations. The simulations were run under the assumption that we have Poisson random variables with parameter 10. (We note that in the exponential and the normal cases the distribution of $Z_n^{1/2}$ does not depend on the values of the parameters under H_o.) Tables III–V cover the case of univariate normal observations. Our results are compared to the critical value (bounds) reported by Hawkins (1977), Siegmund (1985), and Yao and Davis (1986). Table VI contains the critical values when we test for change in the mean vector of bivariate normal random observations with a known covariance matrix. We note that James *et al.* (1992) underestimates the critical values when $n = 20$ and it is very difficult to compute them in case of larger sample sizes.

TABLE V

Critical Values for $Z_n^{1/2}$ in Case of Normal Observations
(Change in the Mean and the Variance)

Sample size	$1-\alpha$	Asymptotic critical values	$r\left(\frac{(\log n)^{3/2}}{n}, \frac{(\log n)^{3/2}}{n}; 1-\alpha\right)$	$z_n(1-\alpha)$
20	0.90	3.53	3.02	3.66
		(0.869)	(0.697)	
	0.95	4.02	3.29	3.99
		(0.954)	(0.804)	
	0.99	5.12	3.83	4.59
		(0.999)	(0.933)	
50	0.90	3.62	3.18	3.47
		(0.936)	(0.813)	
	0.95	4.06	3.44	3.72
		(0.979)	(0.891)	
	0.99	5.04	3.96	4.27
		(0.999)	(0.975)	
100	0.90	3.67	3.27	3.49
		(0.932)	(0.821)	
	0.95	4.09	3.53	3.77
		(0.976)	(0.907)	
	0.99	5.02	4.03	4.21
		(0.999)	(0.977)	
500	0.90	3.77	3.43	3.58
		(0.940)	(0.856)	
	0.95	4.14	3.67	3.83
		(0.983)	(0.919)	
	0.99	5.00	4.15	4.41
		(1.000)	(0.983)	

TABLE VI

Critical Values for $Z_n^{1/2}$ in Case of Bivariate Normal Observations (Change in the Mean, the Covariance Matrix Is Constant and Known)

Sample size	$1-\alpha$	Asymptotic critical values	$r\left(\frac{(\log n)^{3/2}}{n}, \frac{(\log n)^{3/2}}{n}; 1-\alpha\right)$	$z_n(1-\alpha)$
20[a]	0.90	3.53	3.02	3.02
		(0.981)	(0.900)	
	0.95	4.02	3.29	3.28
		(0.998)	(0.951)	
	0.99	5.12	3.83	3.70
		(1.000)	(0.994)	
50	0.90	3.62	3.18	3.26
		(0.968)	(0.879)	
	0.95	4.06	3.44	3.48
		(0.991)	(0.944)	
	0.99	5.04	3.96	4.04
		(1.000)	(0.987)	
100	0.90	3.67	3.27	3.30
		(0.963)	(0.889)	
	0.95	4.09	3.53	3.55
		(0.991)	(0.946)	
	0.99	5.02	4.03	4.05
		(1.000)	(0.989)	
500	0.90	3.77	3.43	3.41
		(0.968)	(0.904)	
	0.95	4.14	3.67	3.665
		(0.990)	(0.952)	
	0.99	5.00	4.15	4.19
		(1.000)	(0.990)	

[a] The critical values in James *et al.* (1992) are 2.90, 3.08, and 3.40.

It is clear from Tables I–VI that $r((\log n)^{3/2}/n, (\log n)\, t^{3/2}/n; 1-\alpha)$ gives very good critical values and they are the same or better than critical values obtained earlier by using different methods. By (3.8) it is easy to compute $r(h, l; 1-\alpha)$ which is also distribution-free (it does not depend on $f(\mathbf{x}, \boldsymbol{\theta})$). We can use the methods in Yao and Davis (1986), Siegmund (1985), and James *et al.* (1992) only in case of normal observations. Worsley (1986a) is more general, but it is based on recursive integrals, is very difficult to compute in the case of large sample sizes, and is not distribution-free.

4. Preliminary Lemmas

Let $\mathbf{x}^{\mathrm{T}}$ and $|\mathbf{x}|$ denote the transpose and the maximum norm of vectors and matrices.

Lemma 4.1. *If H_o and* C.1–C.3 *hold, then for all $\varepsilon > 0$ and $\delta > 0$ we can find $T_1 = T_1(\varepsilon, \delta)$ and $n_1 = n_1(\varepsilon, \delta)$ such that*

$$P\left\{\max_{t \leqslant k \leqslant n-T} |(\hat{\boldsymbol{\theta}}_k, \boldsymbol{\theta}_k^*, \hat{\boldsymbol{\eta}}_k) - (\boldsymbol{\theta}_o, \boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| > \varepsilon\right\} \leqslant \delta, \tag{4.1}$$

if $T \geqslant T_1$ and $n \geqslant n_1$.

Proof. We follow the proof of Theorem 6.2.2 of Lehmann (1991). Let γ be so small that

$$\Gamma = \{(\boldsymbol{\theta}, \boldsymbol{\eta}) \colon |(\boldsymbol{\theta}, \boldsymbol{\eta}) - (\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| \leqslant \gamma\} \subseteq \Theta_o. \tag{4.2}$$

Define

$$u(\boldsymbol{\theta}, \boldsymbol{\eta}) = E_{(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)} \{g(\mathbf{X}_1; \boldsymbol{\theta}, \boldsymbol{\eta}) - g(\mathbf{X}_1; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o)\}. \tag{4.3}$$

It follows from Jensen's inequality and C.1 that

$$u(\boldsymbol{\theta}, \boldsymbol{\eta}) < 0 \qquad \text{if} \quad (\boldsymbol{\theta}, \boldsymbol{\eta}) \neq (\boldsymbol{\theta}_o, \boldsymbol{\eta}_o). \tag{4.4}$$

By C.3 we have that $g(\mathbf{X}_i; \boldsymbol{\theta}, \boldsymbol{\eta})$ is differentiable with respect to $(\boldsymbol{\theta}, \boldsymbol{\eta})$ and, therefore, by the strong law of large numbers, for each $\varepsilon > 0$ and $\delta > 0$ we can find $T_1 = T_1(\varepsilon, \delta)$ such that

$$P\left\{\max_{T_1 \leqslant k < \infty} \sup_{\mathbf{y} \in \Gamma} \frac{1}{k} |L_k(\mathbf{y}) - L_k(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o) - ku(\mathbf{y})| > \varepsilon\right\} \leqslant \delta. \tag{4.5}$$

Let

$$u^* = \sup_{\mathbf{y} \in \partial\Gamma} u(\mathbf{y}), \tag{4.6}$$

where $\partial\Gamma$ is the boundary of Γ. The continuity of $u(\mathbf{y})$ and (4.4) give that

$$u^* < 0. \tag{4.7}$$

If we choose any $0 < \varepsilon \leqslant |u^*|/2$ in (4.5) we get

$$P\left\{\max_{T \leqslant k < \infty} \frac{1}{k} \sup_{\mathbf{y} \in \partial\Gamma} (L_k(\mathbf{y}) - L_k(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)) < u^*/2\right\} \geqslant 1 - \delta, \tag{4.8}$$

if $T \geqslant T_1$. Similar arguments give

$$P\left\{\max_{1 \leqslant k \leqslant n-T} \sup_{\mathbf{y} \in \partial\Gamma} \frac{1}{n-k}\left(L_k^*(\mathbf{y}) - (L_n(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o) - L_k(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o))\right) < u^*/2\right\} \geqslant 1-\delta, \tag{4.9}$$

if $T \geqslant T_1$ and $n \geqslant n_1 = n_1(\varepsilon, \delta)$. Putting together (4.8) and (4.9) we obtain

$$P\left\{\max_{T \leqslant k \leqslant n-T} \sup_{(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\eta}) \in \partial\Gamma^*} (L_k(\boldsymbol{\theta}, \boldsymbol{\eta}) + L_k^*(\boldsymbol{\tau}, \boldsymbol{\eta}) - L_n(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)) < Tu^*\right\} \geqslant 1-2\delta, \tag{4.10}$$

if $T \geqslant T_1$ and $n \geqslant n_1$, where $\partial\Gamma^* = \{(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\eta})\colon |(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\eta}) - (\boldsymbol{\theta}_o, \boldsymbol{\tau}_o, \boldsymbol{\eta}_o)| = \gamma\}$. Now C.2 and (4.10) yield

$$P\left\{\max_{T \leqslant k \leqslant n-T} |(\hat{\boldsymbol{\theta}}_k, \boldsymbol{\theta}_k^*, \hat{\boldsymbol{\eta}}_k) - (\boldsymbol{\theta}_o, \boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| \leqslant \gamma\right\} \geqslant 1-2\delta. \tag{4.11}$$

Since γ in (4.11) can be as small as we want, the proof of Lemma 4.1 is complete.

LEMMA 4.2. *If H_o and* C.1–C.8 *hold, then for all $\delta > 0$ we can find $C_1 = C_1(\delta)$, $T_2 = T_2(\delta)$, and $n_2 = n_2(\delta)$ such that*

$$P\left\{\max_{T \leqslant k \leqslant n-T} (k/\log\log k)^{1/2}\, |\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_o| > C_1\right\} \leqslant \delta, \tag{4.12}$$

$$P\left\{\max_{T \leqslant k \leqslant n-T} ((n-k)/\log\log(n-k))^{1/2}\, |\boldsymbol{\theta}_k^* - \boldsymbol{\theta}_o| > C_1\right\} \leqslant \delta, \tag{4.13}$$

$$P\left\{n^{-1/2} \max_{T \leqslant k \leqslant n-T} k\, |\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_o| > C_1\right\} \leqslant \delta, \tag{4.14}$$

$$P\left\{n^{-1/2} \max_{T \leqslant k \leqslant n-T} (n-k)\, |\boldsymbol{\theta}_k^* - \boldsymbol{\theta}_o| > C_1\right\} \leqslant \delta \tag{4.15}$$

and

$$P\left\{n^{1/2} \max_{T \leqslant k \leqslant n-T} |\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_o| > C_1\right\} \leqslant \delta, \tag{4.16}$$

if $T \geqslant T_2$ and $n \geqslant n_2$.

Proof. By Lemma 4.1 we can assume that

$$(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_n) \in \Theta_o, \qquad (\boldsymbol{\theta}_k^*, \hat{\boldsymbol{\eta}}_n) \in \Theta_o \qquad \text{for all} \quad T \leqslant k \leqslant n-T \tag{4.17}$$

and

$$\max_{T \leqslant k \leqslant n-T} |(\hat{\boldsymbol{\theta}}_k, \boldsymbol{\theta}_k^*, \hat{\boldsymbol{\eta}}_k) - (\boldsymbol{\theta}_o, \boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| \leqslant \varepsilon. \tag{4.18}$$

Let

$$\hat{A}_{il,k} = \sum_{1 \leqslant j \leqslant k} g_{i,l}(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o), \tag{4.19}$$

$$\tilde{A}_{il,k} = \sum_{k < j \leqslant n} g_{i,l}(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o), \tag{4.20}$$

and

$$\hat{A}_k = \{\hat{A}_{il,k}, 1 \leqslant i, l \leqslant d+p\}, \qquad \tilde{A}_k = \{\tilde{A}_{il,k}, 1 \leqslant i, l \leqslant d+p\}. \tag{4.21}$$

By the law of iterated logarithm for partial sums we have

$$\max_{1 < k \leqslant n} (k \log\log k)^{-1/2} |\hat{A}_k + kJ| = O_P(1), \tag{4.22}$$

$$\max_{1 \leqslant k < n} ((n-k) \log\log(n-k))^{-1/2} |\tilde{A}_k + (n-k) J| = O_P(1), \tag{4.23}$$

and the weak convergence of partial sums yields

$$n^{1/2} \max_{1 \leqslant k \leqslant n} \left| \frac{1}{k} \hat{A}_k + J \right| = O_P(1), \tag{4.24}$$

$$n^{1/2} \max_{1 \leqslant k < n} \left| \frac{1}{n-k} \tilde{A}_k + J \right| = O_P(1), \tag{4.25}$$

where $J = J(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)$. Similarly to (1.11) we define $D_{12} = \{J_{il}(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o), 1 \leqslant i \leqslant d, d < l \leqslant d+p\}$, $D_{21} = D_{12}^{\mathrm{T}}$, and $D_{22} = \{J_{il}(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o), d < i, l \leqslant d+p\}$. Let $\hat{\boldsymbol{\theta}}_k = (\theta_{k,1}, ..., \theta_{k,d})$, $\hat{\boldsymbol{\eta}}_k = (\hat{\eta}_{k,1}, ..., \hat{\eta}_{k,p})$, and $\boldsymbol{\theta}_o = (\theta_{o,1}, ..., \theta_{o,d})$, $\boldsymbol{\eta}_o = (\boldsymbol{\eta}_{o,1}, ..., \boldsymbol{\eta}_{o,p})$.

For all $1 \leqslant i \leqslant d$, a two-term Taylor expansion and (1.4) give

$$\begin{aligned} -\sum_{1 \leqslant j \leqslant k} g_i(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o) = & \sum_{1 \leqslant l \leqslant d} (\hat{\theta}_{k,l} - \theta_{o,l}) \hat{A}_{il,k} \\ & + \sum_{d < l \leqslant d+p} (\hat{\eta}_{k,l-d} - \eta_{o,l-d}) \hat{A}_{il,k} \\ & + \tfrac{1}{2} \sum_{1 \leqslant l \leqslant d} (\hat{\theta}_{k,l} - \theta_{o,l}) \hat{R}_{il,k} \\ & + \tfrac{1}{2} \sum_{d < l \leqslant d+p} (\hat{\eta}_{k,l-d} - \eta_{o,l-d}) \hat{R}_{il,k} \end{aligned} \tag{4.26}$$

and by C.4 we also have

$$|\hat{R}_{il,k}| \leqslant |(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_k) - (\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| \sum_{1 \leqslant j \leqslant k} M_2(\mathbf{X}_j). \tag{4.27}$$

By (4.26) and (4.27) we can find matrices $F_{11,k}$ and $F_{12,k}$ such that

$$-\frac{1}{k} \mathbf{V}_{k,1} = (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_o) F_{11,k} + (\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_o) F_{21,k} \tag{4.28}$$

and

$$\begin{aligned} &\max \{ |F_{11,k} + D_{11}|, |F_{21,k} + D_{21}| \} \\ &\quad \leqslant (p+d)^2 \left\{ \left| \frac{1}{k} \hat{A}_k + J \right| + |(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_k) - (\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| \frac{1}{k} \sum_{1 \leqslant j \leqslant k} M_2(\mathbf{X}_j) \right\}. \end{aligned} \tag{4.29}$$

Similar arguments give

$$-\frac{1}{n-k} \mathbf{V}_{k,2} = (\boldsymbol{\theta}_k^* - \boldsymbol{\theta}_o) \tilde{F}_{11,k} + (\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_o) \tilde{F}_{21,k} \tag{4.30}$$

and

$$\begin{aligned} &\max \{ |\tilde{F}_{11,k} + D_{11}|, |\tilde{F}_{21,k} + D_{21}| \} \\ &\quad \leqslant (p+d)^2 \left\{ \left| \frac{1}{n-k} \tilde{A}_k + J \right| + |(\boldsymbol{\theta}_k^*, \hat{\boldsymbol{\eta}}_k) - (\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| \right. \\ &\qquad \left. \times \frac{1}{n-k} \sum_{k < j \leqslant n} M_2(\mathbf{X}_j) \right\}. \end{aligned} \tag{4.31}$$

Let

$$\mathbf{V}_{n,3} = \left(\sum_{1 \leqslant j \leqslant n} g_{d+1}(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o), \ldots, \sum_{1 \leqslant j \leqslant n} g_{d+p}(\mathbf{X}_j; \boldsymbol{\theta}_o, \boldsymbol{\eta}_o) \right). \tag{4.32}$$

Now we use (1.6) and get

$$-\frac{1}{n} \mathbf{V}_{n,3} = \frac{k}{n} (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_o) F_{12,k} + \frac{n-k}{n} (\boldsymbol{\theta}_k^* - \boldsymbol{\theta}_o) \tilde{F}_{12,k} + (\boldsymbol{\eta}_k - \boldsymbol{\eta}_o) F_{22,k} \tag{4.33}$$

and

$$|F_{12,k}+D_{12}| \leqslant (p+d)^2 \left\{ \left| \frac{1}{k} \hat{A}_k + J \right| \right.$$
$$\left. + |(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_k) - (\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| \frac{1}{k} \sum_{1 \leqslant j \leqslant k} M_2(\mathbf{X}_j) \right\}, \tag{4.34}$$

$$|\tilde{F}_{12,k}+D_{12}| \leqslant (p+d)^2 \left\{ \left| \frac{1}{n-k} \tilde{A}_k + J \right| \right.$$
$$\left. + |(\boldsymbol{\theta}_k^*, \hat{\boldsymbol{\eta}}_k) - (\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)| \frac{1}{n-k} \sum_{k < j \leqslant n} M_2(\mathbf{X}_j) \right\}, \tag{4.35}$$

$$|F_{22,k}+D_{22}| \leqslant (p+d)^2 \left\{ \left| \frac{1}{n} \hat{A}_n + J \right| + \frac{k}{n} |\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_o| \frac{1}{k} \sum_{1 \leqslant j \leqslant k} M_2(\mathbf{X}_j) \right.$$
$$+ \frac{n-k}{n} |\hat{\boldsymbol{\theta}}_k^* - \boldsymbol{\theta}_o| \frac{1}{n-k} \sum_{k < j \leqslant n} M_2(\mathbf{X}_j)$$
$$\left. + |\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_o| \frac{1}{n} \sum_{1 \leqslant j \leqslant n} M_2(\mathbf{X}_j) \right\}. \tag{4.36}$$

Condition C.6, (4.18), (4.24), (4.25), (4.29), (4.31), and (4.36) imply that $F_{11,k}^{-1}$, $\tilde{F}_{11,k}^{-1}$, and $F_{22,k}^{-1}$ exist. Thus we can define

$$U_k = \left(F_{22,k} - \frac{k}{n} F_{21,k} F_{11,k}^{-1} F_{12,k} - \frac{n-k}{n} \tilde{F}_{21,k} \tilde{F}_{11,k} \tilde{F}_{12,k} \right)^{-1}, \tag{4.37}$$

and we can solve (4.28), (4.30), and (4.31). We get

$$\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_o = \frac{1}{n} (\mathbf{V}_{k,1} F_{11,k}^{-1} F_{12,k} + \mathbf{V}_{k,2} \tilde{F}_{11,k}^{-1} \tilde{F}_{12,k} - \mathbf{V}_{n,3})\, U_k. \tag{4.38}$$

It is clear that

$$\max_{T \leqslant k \leqslant n-T} |U_k| = O_P(1) \tag{4.39}$$

and, therefore, the weak convergence of partial sums of i.i.d.r. vectors and (4.38) yield (4.16). Combining (4.28) with (4.16) and the law of iterated logarithm we get (4.12). Using the weak convergence of partial sums with (4.28) and (4.16) we obtain immediately (4.14). Similar arguments give (4.13) and (4.15). Now the proof of Lemma 4.2 is complete.

Let

$$U = -(D_{22} - D_{21} D_{11}^{-1} D_{12})^{-1} \tag{4.40}$$

and

$$\mathbf{R}_{k,1} = \frac{1}{k}\mathbf{V}_{k,1}D_{11}^{-1} - \frac{1}{n}(\mathbf{V}_{n,3} - \mathbf{V}_{n,1}D_{11}^{-1}D_{12})\, UD_{21}D_{11}^{-1}, \tag{4.41}$$

$$\mathbf{R}_{k,2} = \frac{1}{n-k}\mathbf{V}_{k,2}D_{11}^{-1} - \frac{1}{n}(\mathbf{V}_{n,3} - \mathbf{V}_{n,1}D_{11}^{-1}D_{12})\, UD_{21}D_{11}^{-1}, \tag{4.42}$$

$$\mathbf{R}_{k,3} = \frac{1}{n}(\mathbf{V}_{n,3} - \mathbf{V}_{n,1}D_{11}^{-1}D_{12})\, U. \tag{4.43}$$

LEMMA 4.3. *If H_o and* C.1–C.8 *hold, then for all $\delta > 0$ we can find $C_2 = C_2(\delta)$, $T_3 = T_3(\delta)$, and $n_3 = n_3(\delta)$ such that*

$$P\{\max_{T \leqslant k \leqslant n-T} (k/\log\log k)\, |\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_o - \mathbf{R}_{k,1}| > C_2\} \leqslant \delta, \tag{4.44}$$

$$P\{\max_{T \leqslant k \leqslant n-T} ((n-k)/\log\log(n-k))\, |\boldsymbol{\theta}_k^* - \boldsymbol{\theta}_o - \mathbf{R}_{k,3}| > C_2\} \leqslant \delta, \tag{4.45}$$

and

$$P\{\max_{T \leqslant k \leqslant n-T} (n/\log\log n)\, |\boldsymbol{\eta}_k - \boldsymbol{\eta}_o - \mathbf{R}_{k,3}| > C_2\} \leqslant \delta, \tag{4.46}$$

if $T \geqslant T_3$ and $n \geqslant n_3$.

Proof. It follows from the law of iterated logarithm, Lemma 4.2, and (4.22)–(4.25), (4.29), (4.31), (4.34)–(4.38) that

$$P\left\{\max_{T \leqslant k \leqslant n-T} (n/\log\log n) \left| \frac{1}{n}(\mathbf{V}_{k,1}F_{11,k}^{-1}F_{12,k} + \mathbf{V}_{k,2}\tilde{F}_{11,k}^{-1}\tilde{F}_{12,k} - \mathbf{V}_{n,3})\, U_k - \mathbf{R}_{k,3}\right| > C_2\right\} \leqslant \delta. \tag{4.47}$$

Putting together (4.28), (4.29), Lemma 4.2, and (4.46) with the law of iterated logarithm we get immediately (4.44). Similar arguments yield (4.45).

LEMMA 4.4. *If H_o and* C.1–C.8 *hold, then for all $\delta > 0$ we can find $C_3 = C_3(\delta)$, $T_4 = T_4(\delta)$, and $n_4 = n_4(\delta)$ such that*

$$P\left\{\max_{T \leqslant k \leqslant n-T} k^{1/2}(\log\log k)^{-3/2}\, |L_k(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_k) - L_k(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o) + \sum_{d < i \leqslant d+p} (\hat{\eta}_{k,i-d} - \eta_{o,i-d}) \sum_{1 \leqslant j \leqslant k} g_i(\mathbf{X}_j; \hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\eta}}_k) - \frac{k}{2}(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_o, \hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_o)\, J(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_o, \hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_o)^{\mathrm{T}}| > C_3\right\} \leqslant \delta \tag{4.48}$$

and

$$P\left\{\max_{T\leqslant k\leqslant n-T}(n-k)^{1/2}(\log\log(n-k))^{-3/2}\,|L_k^*(\theta_k^*,\hat{\boldsymbol{\eta}}_k)-L_k^*(\boldsymbol{\theta}_o,\boldsymbol{\eta}_o)\right.$$
$$+\sum_{d<i\leqslant d+p}(\hat{\eta}_{k,i-d}-\eta_{o,i-d})\sum_{k<j\leqslant n}g_i(\mathbf{X}_j;\boldsymbol{\theta}_k^*,\hat{\boldsymbol{\eta}}_k)$$
$$\left.-\frac{n-k}{2}(\boldsymbol{\theta}_k^*-\boldsymbol{\theta}_o,\hat{\boldsymbol{\eta}}_k-\boldsymbol{\eta}_o)\,J(\boldsymbol{\theta}_o,\boldsymbol{\eta}_o)(\boldsymbol{\theta}_k^*-\boldsymbol{\theta}_o,\hat{\boldsymbol{\eta}}_k-\boldsymbol{\eta}_o)^{\mathrm{T}}|>C_3\right\}\leqslant\delta, \tag{4.49}$$

if $T\geqslant T_4$ *and* $n\geqslant n_4$.

Proof. We prove only (4.48) because the proof of (4.49) is similar. First we apply a three-term, then we apply a two-term Taylor expansion, and we get

$$g(\mathbf{X}_i;\boldsymbol{\theta}_o,\boldsymbol{\eta}_o)-g(\mathbf{X}_i;\hat{\boldsymbol{\theta}}_k,\hat{\boldsymbol{\eta}}_k)-\sum_{d<j\leqslant d+p}(\hat{\eta}_{k,j-d}-\eta_{o,j-d})\,g_j(\mathbf{X}_i;\hat{\boldsymbol{\theta}}_k,\hat{\boldsymbol{\eta}}_k)$$
$$=\sum_{1\leqslant j\leqslant d}g_j(\mathbf{X}_i;\hat{\boldsymbol{\theta}}_k,\hat{\boldsymbol{\eta}}_k)(\theta_{o,j}-\hat{\theta}_{k,j})$$
$$+\tfrac{1}{2}\sum_{1\leqslant j\leqslant d}\sum_{1\leqslant l\leqslant d}g_{j,l}(\mathbf{X}_i;\boldsymbol{\theta}_o,\boldsymbol{\eta}_o)(\theta_{o,j}-\hat{\theta}_{k,j})(\theta_{o,l}-\hat{\theta}_{k,l})$$
$$+\tfrac{1}{2}\sum_{1\leqslant j\leqslant d}\sum_{d<l\leqslant d+p}g_{j,l}(\mathbf{X}_i;\boldsymbol{\theta}_o,\boldsymbol{\eta}_o)(\theta_{o,j}-\hat{\theta}_{k,j})(\eta_{o,l-d}-\hat{\eta}_{k,l-d})$$
$$+\tfrac{1}{2}\sum_{d<j\leqslant d+p}\sum_{d<l\leqslant d+p}g_{j,l}(\mathbf{X}_i;\boldsymbol{\theta}_o,\boldsymbol{\eta}_o)$$
$$\times(\eta_{o,j-d}-\hat{\eta}_{k,j-d})(\eta_{o,l-d}-\hat{\eta}_{k,l-d})+U_i,$$

and by C.4 and Lemma 4.2 we have

$$P\left\{\max_{T\leqslant k\leqslant n-T}k^{1/2}(\log\log k)^{-3/2}\left|\sum_{1\leqslant i\leqslant k}U_i\right|>C_4\right\}\leqslant\delta \tag{4.50}$$

with some $C_4=C_4(\delta)$ if $T\geqslant T_4$, $n\geqslant n_4$. Using (1.4), (4.22), and Lemma 4.2 we get immediately (4.48) from (4.50) and (4.51).

LEMMA 4.5. *If* H_o *and* C.1–C.8 *hold, then we have*

$$n^{1/2}(\log\log n)^{-3/2}\left|L_n(\hat{\boldsymbol{\theta}}_n,\hat{\boldsymbol{\eta}}_n)-L_n(\boldsymbol{\theta}_o,\boldsymbol{\eta}_o)\right.$$
$$\left.-\frac{n}{2}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_o,\hat{\boldsymbol{\eta}}_n-\boldsymbol{\eta}_o)\,J(\boldsymbol{\theta}_o,\boldsymbol{\eta}_o)(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_o,\hat{\boldsymbol{\eta}}_n-\boldsymbol{\eta}_o)^{\mathrm{T}}\right|=O_P(1) \tag{4.51}$$

and

$$n^{1/2}(\log\log n)^{-3/2}\left|L_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\eta}}_n) - L_n(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)\right.$$
$$\left. -\frac{1}{2n}(\mathbf{V}_{n,1}, \mathbf{V}_{n,3})\, J^{-1}(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)(\mathbf{V}_{n,1}, \mathbf{V}_{n,3})^{\mathrm{T}}\right| = O_P(1). \tag{4.52}$$

Proof. Since $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\eta}}_n$ satisfy

$$\sum_{1 \leqslant j \leqslant n} g_i(\mathbf{X}_j, \hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\eta}}_n) = 0 \tag{4.53}$$

for all $1 \leqslant i \leqslant d+p$, (4.51) and (4.52) follow immediately from Ibragimov and Hašminskii (1973) (cf. also Lemmas 2.2 and 2.3 in Gombay and Horváth, 1994).

The following lemma is well known in linear algebra.

LEMMA 4.6. *If* C.6 *holds, then we have*

$$J^{-1}(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)$$
$$= \begin{bmatrix} D_{11}^{-1} + D_{11}^{-1} D_{12}(D_{22} - D_{21} D_{11}^{-1} D_{12})^{-1} D_{21} D_{11}^{-1}, & -D_{11}^{-1} D_{12}(D_{22} - D_{21} D_{11}^{-1} D_{12})^{-1} \\ -(D_{22} - D_{21} D_{11}^{-1} D_{12})^{-1} D_{21} D_{11}^{-1}, & (D_{22} - D_{21} D_{11}^{-1} D_{12})^{-1} \end{bmatrix}. \tag{4.54}$$

Horváth (1993) obtained the following result.

LEMMA 4.7. *Let* $\{\boldsymbol{\zeta}_i = (\zeta_{i,1}, ..., \zeta_{i,d}),\ 1 \leqslant i < \infty\}$ *be a sequence of i.i.d.r. vectors satisfying* $E\zeta_{1,j} = 0$, $E\zeta_{1,j}^2 = 1$, $E\zeta_{1,j}\zeta_{1,k} = 0$, $1 \leqslant j,\ k \leqslant d$, $j \neq k$, *and* $\max_{1 \leqslant i \leqslant d} E\,|\zeta_{1j}|^{\mu} < \infty$ *for some* $\mu > 2$. *Then we have*

$$\lim_{n \to \infty} P\left\{a(\log n) \max_{1 \leqslant k \leqslant n}\left(\frac{1}{k}\sum_{1 \leqslant j \leqslant d}\left(\sum_{1 \leqslant i \leqslant k} \zeta_{i,j}\right)^2\right)^{1/2} - b_d(\log n) \leqslant t\right\}$$
$$= \exp(-e^{-t})$$

for all t.

5. PROOFS OF THE RESULTS IN SECTION 1

Proof of Theorem 1.1. It is easy to see that for each fixed k we have that

$$|\Lambda_k| = O_P(1) \qquad \text{as} \quad n \to \infty \tag{5.1}$$

and by symmetry

$$|\Lambda_{n-k}| = O_P(1) \qquad \text{as} \quad n \to \infty. \tag{5.2}$$

Thus we get

$$\sup_{1 \leqslant k \leqslant T} |-2 \log \Lambda_k| + \sup_{n-T \leqslant k < n} |-2 \log \Lambda_k| = O_P(1) \qquad \text{as} \quad n \to \infty. \tag{5.3}$$

Using Lemma 4.6 one can verify that

$$\begin{aligned} R_k = {} & k(\mathbf{R}_{k,1}, \mathbf{R}_{k,3})\, J(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)(\mathbf{R}_{k,1}, \mathbf{R}_{k,3})^{\mathrm{T}} \\ & + (n-k)(\mathbf{R}_{k,2}, \mathbf{R}_{k,3})\, J(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)(\mathbf{R}_{k,2}, \mathbf{R}_{k,3})^{\mathrm{T}} \\ & - \frac{1}{n} (\mathbf{V}_{n,1}, \mathbf{V}_{n,3})\, J^{-1}(\boldsymbol{\theta}_o, \boldsymbol{\eta}_o)(\mathbf{V}_{n,1}, \mathbf{V}_{n,3})^{\mathrm{T}}. \end{aligned} \tag{5.4}$$

Next we apply (1.6) and Lemmas 4.2–4.5 and we obtain Theorem 1.1.

Proof of Theorem 1.2. We follow the proof of the theorem of Csörgő and Horváth (1986) (cf. also the second proof of Theorem 4.2.1 in Csörgő and Horváth, 1993). First we note that we can find a sequence of i.i.d.r. vectors $\{\boldsymbol{\xi}_i = (\xi_{i,1}, ..., \xi_{i,d}),\ 1 \leqslant i \leqslant \infty\}$ satisfying $E\boldsymbol{\xi}_1 = \mathbf{0}$, $E\xi_{1,i}^2 = 1$, $E\xi_{1,i}\xi_{1,j} = 0$ $(i \neq j)$ such that

$$R_k = n \sum_{1 \leqslant i \leqslant d} \left(S_i(k) - \frac{k}{n} S_i(n) \right)^2 / (k(n-k)), \tag{5.5}$$

where

$$S_i(k) = \sum_{1 \leqslant j \leqslant k} \xi_{j,i}. \tag{5.6}$$

By Einmahl (1987, 1989) we can define $2d$ independent Wiener processes $W_{1,1}, ..., W_{d,1}, W_{1,2}, ..., W_{d,2}$ such that for all $0 < \lambda < \infty$

$$\max_{1 \leqslant i \leqslant d} \sup_{\lambda \leqslant x \leqslant n/2} |S_i(x) - W_{i,1}(x)| / x^{1/\mu} = O_P(1) \tag{5.7}$$

and

$$\max_{1 \leqslant i \leqslant d} \sup_{n/2 \leqslant x \leqslant n-\lambda} |S_i(n) - S_i(x) - W_{i,2}(n-x)| / (n-x)^{1/\mu} = O_P(1). \tag{5.8}$$

Putting together (5.7) and (5.8) we get

$$\begin{aligned} n^{\beta} \sup_{\lambda/n \leqslant t \leqslant 1/2} & n^{-1/2} |S_i(nt) - t(S_i(n/2) + S_i(n) - S_i(n/2)) \\ & - (W_{i,1}(nt) - t(W_{i,1}(n/2) + W_{i,2}(n/2)))|/(t(1-t))^{1/2-\beta} \\ & = O_P(1) \sup_{\lambda \leqslant x \leqslant n/2} x^{1/\mu+\beta-1/2} = O_P(1), \end{aligned} \tag{5.9}$$

if $0 \leqslant \beta \leqslant 1/2 - 1/\mu$. Similarly,

$$\begin{aligned} n^{\beta} \sup_{1/2 \leqslant t \leqslant 1-\lambda/n} & n^{-1/2} |S_i(nt) - S_i(n) + (1-t)(S_i(n/2) + S_i(n) - S_i(n/2)) \\ & - (-W_{i,2}(n-nt) + (1-t)(W_{i,2}(n/2) + W_{i,1}(n/2)))|/(t(1-t))^{1/2-\beta} \\ & = O_P(1), \end{aligned} \tag{5.10}$$

if $0 \leqslant \beta \leqslant 1/2 - 1/\mu$. It is easy to see that

$$B_{n,i}(t) = \begin{cases} n^{-1/2}(W_{i,1}(nt) - t(W_{i,1}(n/2) + W_{i,2}(n/2))), \\ \qquad 0 \leqslant t \leqslant \frac{1}{2}, \\ n^{-1/2}(-W_{i,2}(n-nt) + (1-t)(W_{i,1}(n/2) + W_{i,2}(n/2))), \\ \qquad \frac{1}{2} \leqslant t \leqslant 1, \end{cases} \tag{5.11}$$

is a Brownian bridge and $B_{n,1}(t), ..., B_{n,d}(t)$ are independent processes for each n. By (5.9) and (5.10) we get for all $\lambda > 0$ and $0 \leqslant \beta \leqslant 1/2 - 1/\mu$ that

$$n^{\beta} \sup_{\lambda/n \leqslant t \leqslant 1-\lambda/n} |n^{-1/2}(S_i(nt) - tS_i(n)) - B_{n,i}(t)|/(t(1-t))^{1/2-\beta} = O_P(1) \tag{5.12}$$

for all $1 \leqslant i \leqslant d$. Next we write

$$\begin{aligned} n^{\alpha} \sup_{\lambda/n \leqslant t \leqslant 1-\lambda/n} & |(n^{-1/2}(S_i(nt) - tS_i(n)))^2 - B_{n,i}^2(t)|/(t(1-t))^{1-\alpha} \\ & \leqslant (n^{\alpha/2} \sup_{\lambda/n \leqslant t \leqslant 1-\lambda/n} |n^{-1/2}(S_i(nt) - tS_i(n)) - B_{n,i}(t)|/(t(1-t))^{1/2-\alpha/2})^2 \\ & \quad + n^{\alpha} \sup_{\lambda/n \leqslant t \leqslant 1-\lambda/n} |(n^{-1/2}(S_i(nt) - tS_i(n)) - B_{n,i}(t)) \\ & \quad \times B_{n,i}(t)|/(t(1-t))^{1-\alpha} \\ & = U_{n,1} + U_{n,2}. \end{aligned} \tag{5.13}$$

By (5.12) we have

$$U_{n,1} = O_P(1). \tag{5.14}$$

It is easy to see that for all $\varepsilon > 0$

$$n^{-\varepsilon} \sup_{\lambda/n \leqslant t \leqslant 1-\lambda/n} |B_{n,i}(t)|/(t(1-t))^{1/2+\varepsilon}$$
$$\xrightarrow{\mathscr{D}} \max(\sup_{\lambda \leqslant t < \infty} |W_1(t)|/t^{1/2+\varepsilon}, \sup_{\lambda \leqslant t < \infty} |W_2(t)|/t^{1/2+\varepsilon}), \quad (5.15)$$

where W_1 and W_2 are independent Wiener processes. Let $0 < \varepsilon < 1/2 - 1/\mu - \alpha$. Now (5.12) and (5.15) give

$$\begin{aligned} U_{n,2} &\leqslant n^{\alpha+\varepsilon} \sup_{\lambda/n \leqslant t \leqslant 1-\lambda/n} |n^{-1/2}(S_i(nt) - tS_i(n)) - B_{n,i}(t)|/(t(1-t))^{1/2-(\alpha+\varepsilon)} \\ &\quad \times n^{-\varepsilon} \sup_{\lambda/n \leqslant t \leqslant 1-\lambda/n} |B_{n,i}(t)|/(t(1-t))^{1/2+\varepsilon} \\ &= O_P(1). \end{aligned} \quad (5.16)$$

Putting together (1.15), (5.5), (5.13), (5.14), and (5.16) we get immediately (1.20).

According to Einmahl (1987, 1989), the weighted approximations in (5.7) and (5.8) can be replaced by

$$\max_{1 \leqslant i \leqslant d} \sup_{0 \leqslant x \leqslant n/2} |S_i(x) - W_{i,1}(x)| = o_P(n^{1/\mu}) \quad (5.17)$$

and

$$\max_{1 \leqslant i \leqslant d} \sup_{n/2 \leqslant x \leqslant n} |S_i(n) - S_i(x) - W_{i,2}(n-x)| = o_P(n^{1/\mu}). \quad (5.18)$$

Hence we get, similarly to (5.9), (5.10), and (5.12), that

$$\sup_{0 \leqslant t \leqslant 1} |n^{-1/2}(S_i(nt) - tS_i(n)) - B_{n,i}(t)| = o_P(n^{1/\mu - 1/2}), \quad (5.19)$$

and therefore (1.16) and (3.5) imply (1.21).

Proof of Theorem 1.3. If $I_{0,1}(q, c) < \infty$ for some $c > 0$, then we have

$$\lim_{t \to 0} q(t)/t^{1/2} = \infty, \qquad \lim_{t \to 1} q(t)/(1-t)^{1/2} = \infty \quad (5.20)$$

(cf. Lemma 4.1.3 in Csörgő and Horváth, 1993). Let $0 < \varepsilon < 1$. Then Theorem 1.2 yields

$$\sup_{\varepsilon \leqslant t \leqslant 1-\varepsilon} |t(1-t)\, V_n(t) - B_n^{(d)}(t)|/q^2(t) = o_P(1). \quad (5.21)$$

Choosing $\lambda = \frac{1}{2}$ in Theorem 1.2 we can write

$$U_{n,3} = \sup_{1/(2n) \leqslant t \leqslant \varepsilon} |t(1-t)\, V_n(t) - B_n^{(d)}(t)|/q^2(t)$$

$$\leqslant \sup_{1/(2n) \leqslant t \leqslant \varepsilon} \frac{t(1-t)}{q^2(t)} \sup_{1/(2n) \leqslant t \leqslant \varepsilon} |V_n(t) - B_n^{(d)}(t)/(t(1-t))|. \quad (5.22)$$

Using (5.20) and Theorem 1.2 with $\alpha = 0$ we get for all $\delta > 0$ that

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} P\{U_{n,3} > \delta\} = 0. \quad (5.23)$$

Similarly,

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} P\{U_{n,4} > \delta\} = 0, \quad (5.24)$$

where

$$U_{n,4} = \sup_{1-\varepsilon \leqslant t \leqslant 1-1/(2n)} |t(1-t)\, V_n(t) - B_n^{(d)}(t)|/q^2(t).$$

It follows from the definition of $V_n(t)$ that

$$\sup_{0 \leqslant t/(2n)} |V_n(t)| = \sup_{1-1/(2n) \leqslant t \leqslant 1} |V_n(t)| = 0. \quad (5.25)$$

If $I_{0,1}(q, c) < \infty$ for some $c > 0$, then q is an upper class function for Brownian bridges (cf. Theorem 4.1.1 in Csörgő and Horváth, 1993), and therefore by (5.21)–(5.25) we have

$$\sup_{0 \leqslant t \leqslant 1} |t(1-t)\, V_n(t)|/q^2(t) = \sup_{1/(2n) \leqslant t \leqslant 1-1/(2n)} |t(1-t)\, V_n(t)|/q^2(t)$$

$$= \sup_{1/(2n) \leqslant t \leqslant 1-1/(2n)} |B_n^{(d)}(t)|/q^2(t) + o_P(1), \quad (5.26)$$

which immediately implies (1.24).

If $I_{0,1}(c, q) < \infty$ for all $c > 0$, then (cf. Corollary 4.1.2 in Csörgő and Horváth, 1993)

$$\sup_{0 \leqslant t \leqslant 1/(2n)} B_n^{(d)}(t)/q^2(t) = o_P(1) \quad (5.27)$$

and

$$\sup_{1-1/(2n) \leqslant t \leqslant 1} B_n^{(d)}(t)/q^2(t) = o_P(1). \quad (5.28)$$

Hence (1.23) follows from (5.21)–(5.25).

Next we show that if (1.23) holds with a sequence of stochastic procesess satisfying (1.19), then $I_{0,1}(q, c) < \infty$ for all $c > 0$. Since we have (3.25), (1.23) can hold if and only if (5.27) and (5.28) are satisfied. The distribution of $B_n^{(d)}(t)$ does not depend on n, and therefore we have (5.27) and (5.28) if and only if

$$\lim_{t \to 0} B^{(d)}(t)/q^2(t) = 0 \qquad \text{a.s.} \tag{5.29}$$

and

$$\lim_{t \to 1} B^{(d)}(t)/q^2(t) = 0 \qquad \text{a.s.} \tag{5.30}$$

Hence $I_{0,1}(q, c) < \infty$ for all $c > 0$.

The limit distribution in (1.24) is almost surely finite if and only if

$$\limsup_{t \to 0} B^{(d)}(t)/q^2(t) < \infty \qquad \text{a.s.} \tag{5.31}$$

$$\limsup_{t \to 1} B^{(d)}(t)/q^2(t) < \infty \qquad \text{a.s.} \tag{5.32}$$

According to the definition of $B^{(d)}(t)$ and Corollary 4.1.1 in Csörgő and Horváth (1993), (5.31) and (5.32) hold if and only if $I_{0,1}(q, c) < \infty$ for some $c > 0$.

Proof of Theorem 1.4. Let $0 < \varepsilon < \frac{1}{2}$. It is easy to see that

$$\int_0^\varepsilon V_n^\alpha(t)/q(t)\, dt \leqslant 2^\alpha \int_0^\varepsilon |V_n(t) - B_n^{(d)}(t)/(t(1-t))|^\alpha/q(t)\, dt$$
$$+ 2^\alpha \int_0^\varepsilon (B_n^{(d)}(t))^\alpha \frac{(t(1-t))^\alpha}{q(t)}\, dt. \tag{5.33}$$

Now (1.20) with $\alpha = 0$ yields

$$\int_0^\varepsilon |V_n(t) - B_n^{(d)}(t)/(t(1-t))|^\alpha/q(t)\, dt = O_P(1) \int_0^\varepsilon \frac{1}{q(t)}\, dt \tag{5.34}$$

and (1.21) gives

$$\int_\varepsilon^{1-\varepsilon} \left| V_n^\alpha(t) - \left(\frac{B_n^{(d)}(t)}{t(1-t)}\right)^\alpha \right| \Big/ q(t)\, dt = o_P(1). \tag{5.35}$$

Similarly to (5.33) and (5.34) we have

$$\int_{1-\varepsilon}^{1} V_n^{\alpha}(t)/q(t)\,dt = O_P(1)\int_{1-\varepsilon}^{1}\frac{1}{q(t)}\,dt + 2^{\alpha}\int_{1-\varepsilon}^{1}\left(\frac{B_n^{(d)}(t)}{t(1-t)}\right)^{\alpha}\Big/ q(t)\,dt. \tag{5.36}$$

Rajput (1972) and Csörgő, Horváth, and Shao (1993) showed that (1.26) implies

$$\lim_{\varepsilon\to 0}\int_0^{\varepsilon}\left(\frac{B^{(d)}(t)}{t(1-t)}\right)^{\alpha}\Big/ q(t)\,dt = 0 \qquad \text{a.s.} \tag{5.37}$$

and

$$\lim_{\varepsilon\to 0}\int_{1-\varepsilon}^{1}\left(\frac{B^{(d)}(t)}{t(1-t)}\right)^{\alpha}\Big/ q(t)\,dt = 0 \qquad \text{a.s.} \tag{5.38}$$

Hence (1.25) follows from (5.33)–(5.38).

By Rajput (1972) and Csörgő, Horváth, and Shao (1993) the limiting r.v. in (1.25) exists almost surely, if and only if (1.26) holds. Thus (1.25) implies (1.26).

6. Proofs of the Results in Section 2

Proof of Theorem 2.1. Let $c_1(n) = \log n$ and $c_2(n) = n/\log n$. We write

$$Z_n = \max(T_{n,1}, \ldots, T_{n,6}), \tag{6.1}$$

where

$$T_{n,1} = \max_{1\leqslant k\leqslant c_1(n)} (-2\log \Lambda_k),$$

$$T_{n,2} = \max_{c_1(n)\leqslant k\leqslant c_2(n)} (-2\log \Lambda_k),$$

$$T_{n,3} = \max_{c_2(n)\, k\leqslant n/2} (-2\log \Lambda_k),$$

$$T_{n,4} = \max_{n/2\leqslant k\leqslant n-c_2(n)} (-2\log \Lambda_k),$$

$$T_{n,5} = \max_{n-c_2(n)\leqslant k\leqslant n-c_1(n)} (-2\log \Lambda_k)$$

and

$$T_{n,6} = \max_{n-c_1(n) \leqslant k < n} (-2 \log \Lambda_k).$$

Darling and Erdős (1956) and Theorem 1.2 with $\alpha = 0$ imply

$$T_{n,1} = O_P((\log \log \log n)^{1/2}) \tag{6.2}$$

and, therefore, for all t we have

$$a^2(\log n)\, T_{n,1} - (t + b_d(\log n))^2 \xrightarrow{P} -\infty. \tag{6.3}$$

Using Theorem 1.1 with any $0 < \alpha < \frac{1}{2}$ we get

$$|T_{n,2} - \max_{c_1(n) \leqslant k \leqslant c_2(n)} R_k| = o_P(1/\log \log n) \tag{6.4}$$

and

$$|T_{n,3} - \max_{c_2(n) \leqslant k \leqslant n/2} R_k| = o_P(1/\log \log n). \tag{6.5}$$

Then central limit theorem yields

$$\max_{c_1(n) \leqslant k \leqslant c_2(n)} \left| R_k - \frac{1}{k} \sum_{1 \leqslant j \leqslant d} S_j^2(k) \right| = O_P(1/\log n), \tag{6.6}$$

and by Darling and Erdős (1956) and (6.5) we have

$$a^2(\log n)\, T_{n,3} - (t + b_d(\log n))^2 \xrightarrow{P} -\infty. \tag{6.7}$$

Similar arguments give

$$a^2(\log n)\, T_{n,4} - (t + b_d(\log n))^2 \xrightarrow{P} -\infty, \tag{6.8}$$

$$a^2(\log n)\, T_{n,6} - (t + b_d(\log n))^2 \xrightarrow{P} -\infty, \tag{6.9}$$

and

$$\left| T_{n,5} - \max_{n-c_2(n) \leqslant k \leqslant n-c_1(n)} \frac{1}{n-k} \sum_{1 \leqslant i \leqslant d} (S_i(n) - S_i(k))^2 \right| = o_P(1/\log \log n). \tag{6.10}$$

Using again Darling and Erdős (1956) (cf. also Lemma 4.7) one gets immediately that

$$a^2(\log n) \max_{1 \leqslant k \leqslant c_1(n)} \frac{1}{k} \sum_{1 \leqslant i \leqslant d} S_i^2(k) - (t + b_d(\log n))^2 \xrightarrow{P} -\infty \tag{6.11}$$

and, similarly,

$$a^2(\log n) \max_{n - c_1(n) \leqslant k \leqslant n} \frac{1}{n-k} \sum_{1 \leqslant i \leqslant d} (S_i(n) - S_i(k))^2 - (t + b_d(\log n))^2 \xrightarrow{P} -\infty. \tag{6.12}$$

Thus we showed that

$$\begin{aligned} &\lim_{n \to \infty} P\left\{a^2(\log n)\, Z_n \leqslant (t + b_d(\log n))^2\right\} \\ &\quad = \lim_{n \to \infty} P\left\{a^2(\log n) \max\left(\max_{1 \leqslant k \leqslant n/2} \frac{1}{k} \sum_{1 \leqslant i \leqslant d} S_i^2(k),\right.\right. \\ &\qquad \left.\left. \max_{n/2 < k \leqslant n} \frac{1}{n-k} \sum_{1 \leqslant i \leqslant d} (S_i(n) - S_i(k))^2\right) \leqslant (t + b_d(\log n))^2\right\}, \end{aligned} \tag{6.13}$$

and therefore (2.1) follows immediately from Lemma 4.7.

Proof of Theorem 2.2. Since (2.5) contains (2.3), it is enough to prove the second half of Theorem 2.2. Let $\tau(n)$ and $\tau^*(n)$ be defined by

$$\max_{1 \leqslant k < n} (-2 \log \Lambda_k) = -2 \log \Lambda_{\tau(n)} \tag{6.14}$$

and

$$\sup_{h(n) \leqslant t \leqslant 1 - l(n)} B_n^{(d)}(t)/(t(1-t)) = B_n^{(d)}(\tau^*(n))/(\tau^*(n)(1 - \tau^*(n))), \tag{6.15}$$

where $B_n^{(d)}$ is defined in Theorem 1.2. Theorem 2.2 is proven if we can show

$$\left|(-2 \log \Lambda_{\tau(n)}) - \frac{B_n^{(d)}(\tau^*(n))}{\tau^*(n)(1 - \tau^*(n))}\right| = O_P(\exp(-(\log n)^{1-\varepsilon})). \tag{6.16}$$

For any $0 < \varepsilon' < \varepsilon^*$ we define

$$d(n) = \exp((\log n)^{1-\varepsilon'}). \tag{6.17}$$

It is clear that $[d(n), n-d(n)] \subseteq [nh(n), n(1-l(n))]$, if n is large. Lemma 4.7 yields

$$\frac{1}{2\log\log n} \sup_{1/n \leqslant t \leqslant 1-1/n} B_n^{(d)}(t)/(t(1-t)) \xrightarrow{P} 1, \tag{6.18}$$

$$\frac{1}{2(1-\varepsilon')\log\log n} \sup_{1/n \leqslant t \leqslant d(n)/n} B_n^{(d)}(t)/(t(1-t)) \xrightarrow{P} 1, \tag{6.19}$$

and

$$\frac{1}{2(1-\varepsilon')\log\log n} \sup_{1-d(n)/n \leqslant t \leqslant 1-1/n} B_n^{(d)}(t)/(t(1-t)) \xrightarrow{P} 1. \tag{6.20}$$

Putting together (6.18)–(6.20) we get immediately

$$\lim_{n\to\infty} P\{d(n) \leqslant n\tau^*(n) \leqslant n-d(n)\} = 1. \tag{6.21}$$

Applying Theorem 1.2 with $\alpha = 0$ and (6.18)–(6.20) we obtain

$$\frac{1}{2\log\log n} \max_{1 \leqslant k < n} (-2\log \Lambda_k) \xrightarrow{P} 1, \tag{6.22}$$

$$\frac{1}{2(1-\varepsilon')\log\log n} \max_{1 \leqslant k \leqslant d(n)} (-2\log \Lambda_k) \xrightarrow{P} 1, \tag{6.23}$$

and

$$\frac{1}{2(1-\varepsilon')\log\log n} \max_{n-d(n) \leqslant k < n} (-2\log \Lambda_k) \xrightarrow{P} 1, \tag{6.24}$$

which yield

$$\lim_{n\to\infty} P\{d(n) \leqslant \tau(n) \leqslant n-d(n)\} = 1. \tag{6.25}$$

Hence (2.5) follows from Theorem 1.2, (6.19), (6.22), (6.23), and (6.25).

7. Proofs of the Results in Section 3

Proof of Theorem 3.1. Let

$$s(n) = \exp(-(\log n)^{1-\varepsilon})$$

with some $0<\varepsilon<\varepsilon^*$. According to Theorem 2.2 for each $\delta>0$ we can find an integer $n_o=n_o(\delta)$ such that

$$P\left\{|Z_n^{1/2}-\sup_{h(n)\leqslant t\leqslant 1-l(n)}(B_n^{(d)}(t)/(t(1-t)))^{1/2}|>s(n)\right\}\leqslant\delta$$

if $n\geqslant n_o$. Thus we get

$$\begin{aligned}-\delta+P\Big\{&\sup_{h(n)\leqslant t\leqslant 1-l(n)}(B^{(d)}(t)/(t(1-t)))^{1/2}\leqslant r(h,l)-s(n)\Big\}\\ &\leqslant P\left\{Z_n^{1/2}\leqslant r(h,l)\right\}\\ &\leqslant\delta+P\Big\{\sup_{h(n)\leqslant t\leqslant 1-l(n)}(B^{(d)}(t)/(t(1-t)))^{1/2}\leqslant r(h,l)+s(n)\Big\}.\end{aligned}\tag{7.1}$$

Using Theorem 2.1 and (3.2) we get

$$\lim_{n\to\infty}(r(h,l)\,a(\log n)-b_d(\log n))=-\log\log\frac{1}{(1-\alpha)^{1/2}}.\tag{7.2}$$

Since

$$\lim_{n\to\infty}s(n)\log\log n=0,$$

Theorem 3.1 follows from (7.1) and (7.2).

References

Csörgő, M., and Horváth, L. (1986). Approximations of weighted empirical and quantile processes. *Statist. Probab. Lett.* **4** 275–280.

Csörgő, M., and Horváth, L. (1988). Nonparametric methods for changepoint problems. In *Handbook of Statistics*, Vol. 7, pp. 275–280. North Holland, Amsterdam.

Csörgő, M., and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics.* Wiley, New York.

Csörgő, M., Horváth, L., and Shao, Q. M. (1993). Convergence of integrals of uniform empirical and quantile processes. *Stochastic Proc. Appl.* **45** 283–294.

Darling, D. A., and Erdős, P. (1956). A limit theorem for maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155.

DeLong, D. M. (1981). Crossing probabilities for a square root boundary by a Bessell process. *Comm. Statist. Theory Methods A* **10** (21), 2197–2213.

Einmahl, U. (1987). Strong invariance principles for partial sums of independent random vectors. *Ann. Probab.* **15** 1419–1440.

Einmahl, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68.

Gombay, E., and Horváth, L. (1990). Asymptotic distributions of maxium likelihood tests for change in the mean. *Biometrika* **77** 411–414.

Gombay, E., and Horváth, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Proc. Appl.* **50** 161–171.

HACCOU, P., MEELIS, E., AND VAN DE GEER, S. (1988). The likelihood ratio test for the change point problem for exponentially distributed random variables. *Stochastic Proc. Appl.* **27** 121–139.

HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.* **27** 180–186.

HORVÁTH, L. (1989). The limit distributions of the likelihood ratio and cumulative sum tests for a change in binomial probability. *J. Multivariate Anal.* **31** 148–159.

HORVÁTH, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann. Statist.* **21** 671–680.

IBRAGIMOV, I. A., AND HAŠMINSKII, R. Z. (1973). On the approximation of statistical estimators by sums of independent variables. *Dokl. Akad. Nauk SSSR* **210** 883–887.

JAMES, B., JAMES, K. L., AND SIEGMUND, D. (1987). Tests for a change-point. *Biometrika* **74** 71–83.

JAMES, B., JAMES, K. L., AND SIEGMUND, D. (1992). Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. *Statist. Sinica* **2** 69–90.

KEILSON, J., AND ROSS, H. F. (1975). Passage time distributions for Gaussian Markov (Ornstein–Uhlenbeck) statistical processes. In *Selected Tables in Mathematical Statistics*, Vol. III, pp. 233–327, Amer. Math. Soc., Providence, RI.

KIEFER, J. (1959a). A functional equation technique for obtaining Wiener process probabilities associated with theorems of Kolmogorov–Smirnov type. *Math Proc. Cambridge Philos. Soc.* **55** 328–332.

KIEFER, J. (1959b). *K*-sample analogues of the Kolmogorov-Smirnov and Cramér–von Mises tests. *Ann. Math. Statist.* **30** 420–447.

LEHMANN, E. L. (1991). *Theory of Point Estimation*. Pacific Grove, Wadsworth & Brooks/Cole, CA.

RAJPUT, B. (1972). Gaussian measures on L_p spaces, $1 \leqslant p < \infty$. *J. Multivariate Anal.* **2** 382–403.

SCHOLZ, F. W., AND STEPHENS, M. A. (1987). *K*-sample Anderson–Darling tests. *J. Amer. Statist. Assoc.* **82** 918–924.

SEN, A., AND SRIVASTAVA, M. S. (1975a). On tests for detecting change in the mean. *Ann. Statist.* **3** 98–108.

SEN, A., AND SRIVASTAVA, M. S. (1975b). Some one-sided tests for change in level. *Technometrics* **17** 61–64.

SIEGMUND, D. (1985). *Sequential Analysis*. Springer-Verlag, New York.

SRIVASTAVA, M. S., AND WORSLEY, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Statist. Assoc.* **81** 199–204.

VOSTRIKOVA, L. J. (1981). Detection of a "disorder" in a Wiener process. *Theory Probab. Appl.* **26** 356–362.

VOSTRIKOVA, L. J. (1983). Functional limit theorems for the "disorder" problem. *Stochastics* **9** 103–124.

WORSLEY, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika* **70** 455–464.

WORSLEY, K. J. (1986a). Confidence regions and tests for a change-point in a sequence of exponential random variables. *Biometrika* **73** 91–104.

WORSLEY, K. J. (1986b). Confidence regions and tests for a change. In *Pacific Statistical Congress*, pp. 266–272. Elsevier Science, New York.

YAO, Y. C., AND DAVIS, R. A. (1986). The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhyā Ser. A* **48** 339–353.