Ann. Inst. Statist. Math.
Vol. 50, No. 1, 71 86 (1998)

THE BAHADUR-KIEFER REPRESENTATION OF THE TWO DIMENSIONAL SPATIAL MEDIANS

MIGUEL A. ARCONES

Department of Mathematics, University of Texas, Austin, TX 78712-1082, U.S.A.

(Received January 8, 1996; revised February 26, 1997)

Abstract. We consider the asymptotic behavior, both in distribution and almost sure, of the Bahadur-Kiefer representation of the two dimensional spatial medians. The rates appearing in this expansion are non-standard. The rate in the almost sure expansion is $n(2\log n)^{-1/2}(2\log\log n)^{-1}$. The set of clusters points in the almost sure representation is obtained. The distribution of the Bahadur-Kiefer representation of the two dimensional spatial medians converges with rate $n(2\log n)^{-1/2}$ to a limit that is determined precisely.

Key words and phrases: Bahadur-Kiefer representations, spatial medians, empirical processes.

1. The Bahadur-Kiefer representation of the two dimensional spatial medians

In this note, we investigate the Bahadur-Kiefer representation of the two dimensional spatial medians. Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in $\mathbb{R}^d$. Let X be a copy of X_1. Haldane (1948) extended the definition of median to the multivariate case, defining the spatial median $\hat{\theta}_n$ as a statistic such that

$$(1.1) \qquad \sum_{j=1}^{n} |X_j - \hat{\theta}_n| = \inf_{\theta\in\mathbb{R}^d} \sum_{j=1}^{n} |X_j - \theta|,$$

where $|\cdot|$ is the Euclidean distance. The main advantage of this estimator over the sample mean is that it is more robust. It is less sensible to the presence of outliers. Another interesting property of the spatial median is that it has a breakdown point of 1/2. It also has nice invariance properties. The spatial median is invariant by translations, rotations (the choice and the location of the coordinates axes) and dilations (choice of units). Of course, this estimator $\hat{\theta}_n$ is estimating a parameter θ_0 characterized by

$$(1.2) \qquad E[|X-\theta| - |X-\theta_0|] > 0, \qquad \text{for} \quad \theta \neq \theta_0.$$

If the distribution of X is symmetric about x_0, then (1.2) holds for $\theta_0 = x_0$. It is known that if X_1 is not concentrated in a subspace of dimension one, then there is a unique θ_0 satisfying (1.2), i.e. the spatial median is well (and uniquely) defined when the dimension is bigger than one (see for example Milasevic and Ducharme (1987)).

Given an estimator $\hat{\theta}_n$ of a parameter θ_0, under regularity conditions, there is a function ϕ such that

$$(1.3) \qquad n^{1/2}\left(\hat{\theta}_n - \theta_0 - n^{-1}\sum_{j=1}^{n}(\phi(X_j) - E[\phi(X_j)])\right) \overset{\Pr}{\to} 0.$$

If $E[|\phi(X)|^2] < \infty$ and (1.3) holds, then the statistic $\hat{\theta}_n$ is asymptotically normal. Sometimes, it is interesting to know the exact rate of convergence of the term in (1.3). The study of the rate of convergence of terms as the one in (1.3) started with Bahadur (1966) and Kiefer (1967), which considered the p-th quantile, i.e. $\hat{\theta}_n := \inf\{t : F_n(t) \geq p\}$, where F_n be the empirical distribution function. Let F be the distribution function of X_1. Kiefer (1967) showed that if θ_0 is the value such that $F(\theta_0) = p$, F is second differentiable at θ_0 and $F'(\theta_0) > 0$, then

$$(1.4) \qquad \limsup_{n\to\infty} \pm(n/2\log\log n)^{3/4}(\hat{\theta}_n - \theta_0 + (F'(\theta_0))^{-1}(F_n(\theta_0) - F(\theta_0))) = 2^{1/2}3^{-3/4}p^{1/4}(1-p)^{1/4} \quad \text{a.s.}$$

He also proved that

$$(1.5) \qquad n^{3/4}(\hat{\theta}_n - \theta_0 + (F'(\theta_0))^{-1}(F_n(\theta_0) - F(\theta_0))) \overset{d}{\to} p^{1/4}(1-p)^{1/4}|g_1|^{1/2}g_2,$$

where g_1 and g_2 are two independent standard normal r.v.'s. Finding the rate of convergence of (1.3), we grasp a very good insight into the effect of the influence curve in the asymptotics of the statistic. It is a way to measure the differentiability of the statistical functional. Bahadur-Kiefer representations are needed in the construction of sequential fixed width confidence intervals for a parameter (see Chow and Robbins (1965); and Carroll (1978)). Second order representations can be very useful in gaining statistical insights (see for example the four problems in Pfanzagl (1985), pages 4 and 5).

The Bahadur-Kiefer representation of the spatial median was considered by Arcones and Mason (1992). Under the condition $E[|X - \theta_0|^{-2}] < \infty$ and $d \geq 2$, they obtained that

$$(1.6) \qquad \limsup_{n\to\infty} n(2\log\log n)^{-1}|H'(\theta_0)(\hat{\theta}_n - \theta_0) + H_n(\theta_0)| = c \quad \text{a.s.}$$

where c is a constant, $H'(\theta_0)$ is the $d \times d$ matrix determined by

$$(1.7) \qquad H'(\theta_0)t = E[|X - \theta_0|^{-3}((X - \theta_0)'t)(X - \theta_0) - |X - \theta_0|^{-1}t],$$

$$H_n(\theta_0) = n^{-1}\sum_{j=1}^{n} h(X_j - \theta_0),$$

and

(1.8) $$h(x) = |x|^{-1}x, \quad \text{if} \quad x \neq 0 \text{ and } h(0) = 0.$$

Observe that there is no abuse of notation, under the condition $E[|X-\theta_0|^{-1}] < \infty$, $H'(\theta_0)$ is the derivative of $H(\theta) := E[h(X-\theta)]$ (see the remark after Lemma 2.1 below). The derivative of $E[|X-\theta|]$ at θ_0, which is $E[h(X-\theta_0)]$, is zero. They also showed that

(1.9) $$\{n(H'(\theta_0)(\hat{\theta}_n - \theta_0) + H_n(\theta_0))\}, \qquad n \geq 1,$$

converges in distribution. If $d = 1$, the spatial median is the usual median (or the p-th quantile for $p = 1/2$). Thus, the asymptotics of the terms in (1.6) and (1.9) when $d = 1$, are given in (1.4) and (1.5) taking $p = 1/2$ (with different rates).

It is easy to see that the condition $E[|X-\theta_0|^{-2}] < \infty$ is not satisfied naturally if X has values in $\mathbb{R}^2$. For example, it does not hold if X has a positive continuous density in a neighborhood of θ_0. $E[|X-\theta_0|^{-2}] < \infty$ holds in some situations. This condition holds for some discrete distributions. It also holds when X has a first differentiable density in a neighborhood of θ_0 which is zero at θ_0. Niemiro (1992) and Koltchinskii (1994*a*, 1994*b*) mentioned that another rate appears in (1.9) in this two dimensional situation. Here, we obtain the exact rate in the almost sure and distributional behavior of the Bahadur-Kiefer representation of two dimensional spatial medians under mild conditions. We use the approach in Arcones (1994*a*, 1994*b*). We will obtain that the rate of the Bahadur-Kiefer representation of the two dimensional spatial median is that of the one dimensional $L_{3/2}$ estimators, which were considered in these last references. One reason to determine the asymptotic limit behavior of the Bahadur-Kiefer representation of two dimensional spatial medians is to find out whether the limit depends or not on the underlying distribution. We will obtain that there is distributional dependence at difference with the one dimensional case. Our main result is the following:

THEOREM 1.1. *Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in $\mathbb{R}^2$. Suppose that:*

(i) *There exists $\theta_0 \in \mathbb{R}^2$ such that*

$$E[|X-\theta| - |X-\theta_0|] > 0,$$

for each $\theta \neq \theta_0$.

(ii) *$H(\theta) = E[h(X-\theta)]$ (h is as in (1.8)) has a second order expansion at θ_0 of the form:*

$$H(\theta) = H(\theta_0) + H'(\theta_0)\cdot(\theta-\theta_0) + B(\theta-\theta_0, \theta-\theta_0) + o(|\theta-\theta_0|^2),$$

as $\theta \to \theta_0$, where $B : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ is a positive definite bilinear form.

(iii) *X has a density $f(x)$ in a neighborhood of θ_0 which is continuous at θ_0.*

Then,

(a) *With probability one,*

$$\{n(2\log n)^{-1/2}(2\log\log n)^{-1}(H'(\theta_0)(\hat{\theta}_n - \theta_0) + H_n(\theta_0))\},$$

is relatively compact in $\mathbb{R}^2$ *and its limit set is*

$$\Big\{\pi^{1/2}(f(\theta_0))^{1/2}2^{-3/2}\begin{pmatrix} x_1+2^{1/2}x_2 & x_3 \\ x_3 & x_1+2^{1/2}x_4\end{pmatrix} \cdot (H'(\theta_0))^{-1}\begin{pmatrix} c_{1,1}x_5+c_{1,2}x_6 \\ c_{2,1}x_5+c_{2,2}x_6\end{pmatrix} : \sum_{j=1}^{4} x_j^2 + c_{1,1}x_5^2 + 2c_{1,2}x_5x_6 + c_{2,2}x_6^2 \le 1\Big\}$$

where $c_{i,j} = \mathrm{Cov}(|X-\theta_0|^{-1}(X^{(i)}-\theta_0^{(i)}), |X-\theta_0|^{-1}(X^{(j)}-\theta_0^{(j)}))$, *for* $1 \le i,j \le 2$; $X = (X^{(1)}, X^{(2)})'$ *and* $\theta = (\theta_0^{(1)}, \theta_0^{(2)})'$. *In particular,*

$$\text{(1.10)}\quad \limsup_{n\to\infty} n(2\log n)^{-1/2}(2\log\log n)^{-1}|(H'(\theta_0))^{-1}(\hat\theta_n-\theta_0)+H_n(\theta_0)| = \sup\Big\{\pi^{1/2}(f(\theta_0))^{1/2}2^{-3/2}\Big|\begin{pmatrix} x_1+2^{1/2}x_2 & x_3 \\ x_3 & x_1+2^{1/2}x_4\end{pmatrix} \cdot (H'(\theta_0))^{-1}\begin{pmatrix} c_{1,1}x_5+c_{1,2}x_6 \\ c_{2,1}x_5+c_{2,2}x_6\end{pmatrix}\Big| : \sum_{j=1}^{4} x_j^2 + c_{1,1}x_5^2 + 2c_{1,2}x_5x_6 + c_{2,2}x_6^2 \le 1\Big\} \quad a.s.,$$

which is positive if $f(\theta_0) > 0$.

(b) $\{n(2\log n)^{-1/2}(H'(\theta_0)(\hat\theta_n-\theta_0)+H_n(\theta_0))\}$ *converges in distribution to*

$$\pi^{1/2}(f(\theta_0))^{1/2}2^{-3/2}\begin{pmatrix} g_1+2^{1/2}g_2 & g_3 \\ g_3 & g_1+2^{1/2}g_4\end{pmatrix}(H'(\theta_0))^{-1}W,$$

where g_1, g_2, g_3, g_4 *are independent standard normal r.v.'s, and* W *is a* $\mathbb{R}^2$*-valued Gaussian r.v., independent of* g_1, g_2, g_3, g_4, *with mean zero and covariance given by*

$$E[WW'] = E[h(X-\theta_0)(h(X-\theta_0))'] - E[h(X-\theta_0)]E[(h(X-\theta_0))'].$$

In some situations, to know either the distributional of almost sure behavior of the Bahadur-Kiefer representation is not enough. For example, Duttweiler (1973) needed to determine the mean-square error of the Bahadur's expansion.

2. Proof of Theorem 1.1

Given a $d \times m$ matrix A, we define the following norm

$$\|A\| := \sup_{\substack{b \in \mathbb{R}^m \\ |b| \le 1}} |b' A|.$$

Observe that if $m = 1$, last norm is just the Euclidean norm.

We use the approach in Arcones (1994*a*, 1994*b*), which extends the one in Arcones and Mason (1992). In order to prove part (a) in Theorem 1.1, we use the following:

THEOREM 2.1. (Arcones, 1994*a*, Theorem 9) *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in a measurable space $(S, \mathcal{S})$, let Θ be a subset of $\mathbb{R}^d$, let $h : S \times \Theta \to \mathbb{R}^d$ be a jointly measurable function, let $H(\theta) = E[h(X_1, \theta)]$, let $H_n(\theta) = n^{-1} \sum_{i=1}^n h(X_i, \theta)$, let $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ be a sequence of estimators and let $\{b_n\}$ be a sequence of real numbers tending to infinity. Suppose that the following conditions hold:*

(i) *There exists a θ_0 in the interior of Θ such that $H(\theta_0) = 0$.*

(ii) *$H(\theta)$ has a second order expansion at θ_0, meaning that $H(\theta)$ is first differentiable at θ_0 and there exists a bilinear form $B : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ such that*

$$H(\theta) = H(\theta_0) + H'(\theta_0) \cdot (\theta - \theta_0) + B(\theta - \theta_0, \theta - \theta_0) + o(|\theta - \theta_0|^2).$$

(iii) $\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{\theta : |\theta - \theta_0| \le \delta} a_n |H_n(\theta) - H_n(\theta_0) - H(\theta) + H(\theta_0)| = 0$ *a.s., where*

$$a_n = (n/2 \log\log n)^{1/2}. \tag{2.1}$$

(iv) *$\hat{\theta}_n \to \theta_0$ a.s. and $b_n H_n(\hat{\theta}_n) \to 0$ a.s.*

(v) *There exists a real number M, $M > \|(H'(\theta_0))^{-1}\| (E[|h(X, \theta_0)|^2])^{1/2}$, and a compact set $K \subset l_\infty(T_M^*)$, where $T_M^* = \{t \in \mathbb{R}^d : |t| \le M\} \cup \{\infty\}$, such that, with probability one, $\{Z_n(t) : t \in T_M^*\}$, $n \ge 1$, is almost surely relatively compact in $l_\infty(T_M^*)$ and its limit set is K, where*

$$Z_n(t) := b_n(H_n(\theta_0 + t a_n^{-1}) - H_n(\theta_0) - H(\theta_0 + t a_n^{-1}) + H(\theta_0)), \qquad \text{if} \quad |t| \le M,$$

and

$$Z_n(\infty) := a_n(H_n(\theta_0) - H(\theta_0)).$$

We make two cases:

Case a. *If $b_n(2 \log\log n / n) \to 0$, then, with probability one,*

$$\{b_n(H_n(\theta_0) + H'(\theta_0) \cdot (\hat{\theta}_n - \theta_0))\}_{n=1}^{\infty}$$

is relatively compact in $\mathbb{R}^d$ and its limit set is

$$L(K) := \{-\alpha(-(H'(\theta_0))^{-1} \cdot \alpha(\infty)) : \alpha \in K\}.$$

Consequently, $\limsup_{n\to\infty} b_n |H_n(\theta_0) + H'(\theta_0)\cdot(\hat\theta_n - \theta_0)| = \sup\{|x| : x \in L(K)\}$ *a.s.*

Case b. *If* $b_n(2\log\log n/n) \to 1$, *then, with probability one,*

$$\{b_n(H_n(\theta_0) + H'(\theta_0)\cdot(\hat\theta_n - \theta_0))\}_{n=1}^{\infty}$$

is relatively compact in $\mathbb{R}^d$ *and its limit set is*

$$\begin{aligned} L(K) := \{-\alpha(-(H'(\theta_0))^{-1}\cdot\alpha(\infty)) - B((H'(\theta_0))^{-1}\cdot\alpha(\infty),\\ (H'(\theta_0))^{-1}\cdot\alpha(\infty)) : \alpha \in K\}. \end{aligned}$$

Consequently, $\limsup_{n\to\infty} b_n |H_n(\theta_0) + H'(\theta_0)\cdot(\hat\theta_n - \theta_0)| = \sup\{|x| : x \in L(K)\}$ *a.s.*

We apply Theorem 2.1, Case b, with $\Theta = S = \mathbb{R}^2$, $b_n = n(2\log\log n)^{-1}\cdot(2\log n)^{-1/2}$ and $h(x,\theta) = h(x-\theta)$. First, we present the following estimations on the variation of the function h.

LEMMA 2.1. *For each* $x, y \in \mathbb{R}^d$,

$$|h(y) - h(x)| \le 2 \wedge (2|x-y|/|x|) \tag{2.2}$$

and

$$\begin{aligned} |h(y) - h(x) + x'(y-x)|x|^{-3}x - |x|^{-1}(y-x)| \\ \le (4|y-x|/|x|) \wedge (4|y-x|^2/|x|^2). \end{aligned} \tag{2.3}$$

PROOF. Since $|h(x)| \le 1$, $|h(x) - h(y)| \le 2$. We also have that

$$h(x) - h(y) = \frac{(|y| - |x|)y + |y|(x-y)}{|x||y|}. \tag{2.4}$$

So, (2.2) follows. (2.4) implies that

$$|h(y) - h(x) + x'(y-x)|x|^{-3}x - |x|^{-1}(y-x)| \le 4|y-x|/|x|.$$

By an elementary computation

$$\begin{aligned} h(y) - h(x) + x'(y-x)|x|^{-3}x - |x|^{-1}(y-x) \\ = \frac{(|y|-|x|)(x'(y-x))y}{|x|^2|y|(|x|+|y|)} + \frac{(|y|-|x|)(x'(y-x))y}{|x|^3|y|} \\ - \frac{|y-x|^2 y}{|x||y|(|x|+|y|)} - \frac{(x'(y-x))(y-x)}{|x|^3}, \end{aligned}$$

and (2.3) follows. □

It follows from Lemma 2.1 that if $E[|X-\theta_0|^{-1}]<\infty$, then $H(\theta)=E[h(X-\theta)]$ is differentiable at θ_0 and $H'(\theta_0)$ is given by (1.7). Condition (iii) in Theorem 1.1 implies that $E[|X-\theta_0|^{-1}]<\infty$. So, condition (ii) in Theorem 2.1 is satisfied. Next, we consider condition (iii).

LEMMA 2.2. *Under the conditions in Theorem* 1.1,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\theta:|\theta-\theta_0|\le\delta} a_n|H_n(\theta)-H_n(\theta_0)-H(\theta)+H(\theta_0)|=0 \qquad a.s.$$

PROOF. We have that the class of functions $\{h(x,\theta)-h(x,\theta_0):\theta\in\mathbb{R}^2\}$ is a VC subgraph class of functions (see Lemma 22 in Nolan and Pollard (1987)) and $E[|h(X,\theta)-h(X,\theta_0)|^2]\to 0$, as $\theta\to\theta_0$. So, the claim follows from the law of the iterated logarithm for empirical processes indexed by VC subgraph classes. □

LEMMA 2.3. *Let $\hat\theta_n$ be a sequence of r.v.'s such that $\sum_{j=1}^n|X_j-\hat\theta_n|=\inf_{\theta\in\mathbb{R}^2}\sum_{j=1}^n|X_j-\theta|$. Then, under the conditions in Theorem* 1.1,

$$\hat\theta_n\to\theta_0 \quad a.s. \qquad and \qquad n(\log n)^{-1/2}H_n(\hat\theta_n)\to 0 \quad a.s.$$

PROOF. It is well known that in this situation $\hat\theta_n\to\theta_0$ a.s. (see for example Proposition 2 in Arcones (1995)). For each $v\in\mathbb{R}^2$, $v\neq 0$,

$$\begin{aligned}(2.5)\qquad 0&\le\lim_{t\to 0+}t^{-1}\sum_{j=1}^n(|X_j-\hat\theta_n+tv|-|X_j-\hat\theta_n|)\\&=\sum_{j=1}^n v'h(X_j-\hat\theta_n)I_{X_j\neq\hat\theta_n}+|v|\sum_{j=1}^n I_{X_j=\hat\theta_n}.\end{aligned}$$

Let V be a neighborhood of θ_0, in which X has a density. If $i\neq j$ and $X_i,X_j\in V$, then $X_i\neq X_j$. So, eventually $\sum_{j=1}^n I_{X_j=\hat\theta_n}\le 1$. Taking $v=-\sum_{j=1}^n h(X_j-\hat\theta_n)I_{X_j\neq\hat\theta_n}$ in (2.5), we get that

$$\left|\sum_{j=1}^n h(X_j-\hat\theta_n)I_{X_j\neq\hat\theta_n}\right|\le\sum_{j=1}^n I_{X_j=\hat\theta_n}\le 1,$$

for n large enough. So, the claim follows. □

The funny rate of convergence in the Bahadur-Kiefer representation of the two dimensional spatial medians comes from the following lemma.

LEMMA 2.4. *Under the conditions in Theorem* 1.1,

$$\begin{aligned}\lim_{\lambda\to\infty}(2\log\lambda)^{-1}\lambda^2 E[(h(X-\theta_0-t\lambda^{-1})-h(X-\theta_0))\\ \cdot(h(X-\theta_0-s\lambda^{-1})-h(X-\theta_0))']\\ =\pi f(\theta_0)2^{-3}\begin{pmatrix}3t_1s_1+t_2s_2 & t_1s_2+t_2s_1\\ t_1s_2+t_2s_1 & t_1s_1+3t_2s_2\end{pmatrix}\end{aligned}$$

for each $s, t \in \mathbb{R}^2$.

PROOF. By Lemma 2.1,

$$(2.6)\qquad \begin{aligned}&(\log\lambda)^{-1}\lambda^2 E[|h(X-\theta_0-t\lambda^{-1})-h(X-\theta_0)|\\ &\qquad\cdot|h(X-\theta_0-s\lambda^{-1})-h(X-\theta_0)|I_{|X-\theta_0|\le M\lambda^{-1}}]\\ &\le(\log\lambda)^{-1}\lambda^2 4\Pr\{|X-\theta_0|\le M\lambda^{-1}\}\to 0.\end{aligned}$$

By the change of variables $x=\theta_0+\lambda^{-1}y$

$$\begin{aligned}&(2\log\lambda)^{-1}\lambda^2 E[(h(X-\theta_0-t\lambda^{-1})-h(X-\theta_0))\\ &\qquad\times(h(X-\theta_0-s\lambda^{-1})-h(X-\theta_0))' I_{M\lambda^{-1}\le|X-\theta_0|<(\log\lambda)^{-1/2}}]\\ &=(2\log\lambda)^{-1}\lambda^2\int_{M\lambda^{-1}\le|x-\theta_0|<(\log\lambda)^{-1/2}}(h(x-\theta_0-t\lambda^{-1})-h(x-\theta_0))\\ &\quad\times(h(x-\theta_0-s\lambda^{-1})-h(x-\theta_0))'f(x)dx\\ &=(2\log\lambda)^{-1}\int_{M\le|y|<\lambda(\log\lambda)^{-1/2}}(h(y-t)-h(y))\\ &\quad\times(h(y-s)-h(y))'f(\theta_0+\lambda^{-1}y)dy,\end{aligned}$$

which, by Lemma 2.1, is

$$\begin{aligned}&=o(1)+f(\theta_0)(2\log\lambda)^{-1}\\ &\cdot\int_{M\le|y|<\lambda(\log\lambda)^{-1/2}}(|y|^{-3}(y't)y-|y|^{-1}t)(|y|^{-3}(y's)y-|y|^{-1}s)'dy.\end{aligned}$$

Changing to polar coordinates, i.e. $y=(r\cos\theta, r\sin\theta)'$, we get that the last equation is

$$o(1)+f(\theta_0)(2\log\lambda)^{-1}\int_M^{\lambda(\log\lambda)^{-1/2}}\int_0^{2\pi}((v_\theta' t)v_\theta-t)((v_\theta' s)v_\theta-s)'r^{-1}d\theta dr,$$

where $v_\theta=(\cos\theta,\sin\theta)'$, which converges, as $\lambda\to\infty$, to

$$2^{-1}f(\theta_0)\int_0^{2\pi}((v_\theta' t)v_\theta-t)((v_\theta' s)v_\theta-s)'d\theta.$$

Using that

$$\begin{aligned}&\int_0^{2\pi}\sin^2\theta d\theta=\int_0^{2\pi}\cos^2\theta d\theta=\pi,\qquad\int_0^{2\pi}\sin^2\theta\cos^2\theta d\theta=\pi/4,\\ &\int_0^{2\pi}\sin^4\theta d\theta=\int_0^{2\pi}\cos^4\theta d\theta=3\pi/4,\end{aligned}$$

and

$$\int_0^{2\pi} \sin^i \theta \cos^j \theta d\theta = 0$$

if either i or j (or both) are odd; we have that

$$\begin{aligned}
&\int_0^{2\pi} ((v_\theta' t)v_\theta - t)((v_\theta' s)v_\theta - s)' d\theta \\
&= \int_0^{2\pi} (v_\theta' t)(v_\theta' s)v_\theta v_\theta' - (v_\theta' t)v_\theta s' - (v_\theta' s)t v_\theta' + ts' d\theta \\
&= \int_0^{2\pi} (t_1 \cos\theta + t_2 \sin\theta)(s_1 \cos\theta + s_2 \sin\theta) \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \cos^2\theta \end{pmatrix} d\theta \\
&\quad - \int_0^{2\pi} (t_1 \cos\theta + t_2 \sin\theta) \begin{pmatrix} s_1\cos\theta & s_2\cos\theta \\ s_1\sin\theta & s_2\sin\theta \end{pmatrix} d\theta \\
&\quad - \int_0^{2\pi} (s_1 \cos\theta + s_2 \sin\theta) \begin{pmatrix} t_1\cos\theta & t_1\sin\theta \\ t_2\cos\theta & t_2\sin\theta \end{pmatrix} d\theta \\
&\quad + \int_0^{2\pi} \begin{pmatrix} t_1 s_1 & t_1 s_2 \\ t_2 s_1 & t_2 s_2 \end{pmatrix} d\theta \\
&= 2^{-2}\pi \begin{pmatrix} 3t_1 s_1 + t_2 s_2 & t_1 s_2 + t_2 s_1 \\ t_1 s_2 + t_2 s_1 & t_1 s_1 + 3t_2 s_2 \end{pmatrix}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
(2.7)\qquad &(2\log\lambda)^{-1}\lambda^2 E[(h(X-\theta_0 - t\lambda^{-1}) - h(X-\theta_0)) \\
&\qquad \times (h(X-\theta_0 - s\lambda^{-1}) - h(X-\theta_0))' \\
&\qquad \times I_{M\lambda^{-1} \le |X-\theta_0| < (\log\lambda)^{1/2}}] \\
&\to \pi f(\theta_0) 2^{-3} \begin{pmatrix} 3t_1 s_1 + t_2 s_2 & t_1 s_2 + t_2 s_1 \\ t_1 s_2 + t_2 s_1 & t_1 s_1 + 3t_2 s_2 \end{pmatrix}.
\end{aligned}$$

By Lemma 2.1,

$$\begin{aligned}
(2.8)\qquad &(\log\lambda)^{-1}\lambda^2 E[|h(X-\theta_0 - t\lambda^{-1}) - h(X-\theta_0)| \\
&\qquad \cdot |h(X-\theta_0 - s\lambda^{-1}) - h(X-\theta_0)| I_{|X-\theta_0| > (\log\lambda)^{1/2}}] \\
&\le (\log\lambda)^{-1} 4|t||s| E[|X-\theta_0|^{-2} I_{|X-\theta_0| > (\log\lambda)^{-1/2}}] \\
&\le (\log\lambda)^{-1/2} 4|t||s| E[|X-\theta_0|^{-1}] \to 0.
\end{aligned}$$

The claim follows from (2.6)–(2.8). □

In order to determine the limit set K in condition (v) in Theorem 2.1, we need to use the concept of the unit ball of the reproducing kernel Hilbert space of a mean-zero Gaussian process $\{Z(t) : t \in T\}$. Let $\mathcal{L}$ be the linear subspace of L_2, generated by $\{Z(t) : t \in T\}$. Then, the reproducing kernel Hilbert space (r.k.h.s.) of the Gaussian process $\{Z(t) : t \in T\}$ is the following class of functions on T

$$(2.9)\qquad \{(E[Z(t)\xi])_{t\in T} : \xi \in \mathcal{L}\}.$$

This space is endowed of the inner product

$$\langle f_1, f_2 \rangle := E[\xi_1 \xi_2], \tag{2.10}$$

where $f_i(t) = E[Z(t)\xi_i]$ for each $t \in T$ and each $i = 1, 2$. The unit ball of this r.k.h.s. is

$$K := \{(E[Z(t)\xi])_{t \in T} : \xi \in \mathcal{L} \text{ and } E[\xi^2] < 1\}. \tag{2.11}$$

LEMMA 2.5. *For each $M < \infty$, with probability one,*

$$\{Z_n(t) := a_n^2 (2 \log n)^{-1/2} \cdot (H_n(\theta_0 + t a_n^{-1}) - H_n(\theta_0) - H(\theta_0 + t a_n^{-1}) + H_n(\theta_0)) : t \in T_M\}$$

is relatively compact in $l_\infty(T_M)$, where a_n was defined in (2.1), $T_M = \{t \in \mathbb{R}^2 : |t| \le M\}$, *and its limit set is the unit ball of the $\mathbb{R}^2$-valued mean-zero Gaussian process $\{Z(t) : t \in T_M\}$ with covariance given by*

$$E[Z(t)(Z(s))'] = \pi f(\theta_0) 2^{-3} \begin{pmatrix} 3t_1 s_1 + t_2 s_2 & t_1 s_2 + t_2 s_1 \\ t_1 s_2 + t_2 s_1 & t_1 s_1 + 3 t_2 s_2 \end{pmatrix}$$

for $|s|, |t| \le M$.

PROOF. Since the class of functions $\{h(x - \theta) - h(x - \theta_0) : \theta \in \mathbb{R}^2\}$ is a VC subgraph class, by Corollary 4 in Arcones (1994*a*) and Lemma 2.4, it suffices to show that

$$\limsup_{n \to \infty} a_n^2 (\log n)^{-1} (\log \log n)^{-1} \cdot E\left[\sup_{|t| \le M} |h(X - \theta_0 - t a_n^{-1}) - h(X - \theta_0)|^2\right] = 0 \tag{2.12}$$

and

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{\substack{|s - t| \le \delta \\ |s|, |t| \le M}} (\log n)^{-1} a_n^2 \cdot E[|h(X - \theta_0 - t a_n^{-1}) - h(X - \theta_0 - s a_n^{-1})|^2] = 0. \tag{2.13}$$

By Lemma 2.1,

$$\begin{aligned} & a_n^2 (\log n)^{-1} (\log \log n)^{-1} E\left[\sup_{|t| \le M} |h(X - \theta_0 - t a_n^{-1}) - h(X - \theta_0)|^2\right] \\ &\le a_n^2 (\log n)^{-1} (\log \log n)^{-1} E[4 \wedge (4 M^2 a_n^{-2} |X - \theta_0|^{-2})] \\ &= a_n^2 (\log n)^{-1} (\log \log n)^{-1} 4 \Pr\{|X - \theta_0| \le M a_n^{-1}\} \\ &\quad + 4 M^2 (\log n)^{-1} (\log \log n)^{-1} E[|X - \theta_0|^{-2} I_{|X - \theta_0| > M a_n^{-1}}]. \end{aligned}$$

Obviously,

$$a_n^2(\log n)^{-1}(\log\log n)^{-1}\Pr\{|X-\theta_0|\le Ma_n^{-1}\}\to 0.$$

Taking $\epsilon>0$, such that X has a density in $\{x\in\mathbb{R}^2:|x-\theta_0|\le\epsilon\}$, we have that

$$\begin{aligned}(2.14)\qquad &(\log n)^{-1}(\log\log n)^{-1}E[|X-\theta_0|^{-2}I_{|X-\theta_0|>Ma_n^{-1}}]\\ &\le(\log n)^{-1}(\log\log n)^{-1}\int_{\epsilon>|x-\theta_0|>Ma_n^{-1}}|x-\theta_0|^{-2}f(x)dx\\ &\quad+(\log n)^{-1}(\log\log n)^{-1}\epsilon^{-2}.\end{aligned}$$

Changing to polar coordinates, i.e. $x=\theta_0+(r\cos\phi,r\sin\phi)'$,

$$\begin{aligned}(2.15)\qquad &(\log n)^{-1}(\log\log n)^{-1}\int_{\epsilon>|x-\theta_0|>Ma_n^{-1}}|x-\theta_0|^{-2}f(x)dx\\ &\le\sup_{\epsilon>|x-\theta_0|>Ma_n^{-1}}f(x)(\log n)^{-1}(\log\log n)^{-1}\\ &\quad\cdot\int_{Ma_n^{-1}}^{\epsilon}\int_0^{2\pi}r^{-1}d\phi dr\to 0.\end{aligned}$$

So, (2.12) follows.

By Lemma 2.1 and the computation in (2.14) and (2.15), for $|s-t|\le\delta$, $|s|,|t|\le M$,

$$\begin{aligned}&(\log n)^{-1}a_n^2E[|h(X-\theta_0-ta_n^{-1})-h(X-\theta_0-sa_n^{-1})|^2]\\ &\quad=(\log n)^{-1}a_n^2E[|h(X-\theta_0-ta_n^{-1})-h(X-\theta_0-sa_n^{-1})|^2I_{|X-\theta_0|\le 2Ma_n^{-1}}]\\ &\qquad+(\log n)^{-1}a_n^2E[|h(X-\theta_0-ta_n^{-1})-h(X-\theta_0-sa_n^{-1})|^2I_{|X-\theta_0|>2Ma_n^{-1}}]\\ &\quad\le(\log n)^{-1}a_n^2 4\Pr\{|X-\theta_0|\le 2Ma_n^{-1}\}\\ &\qquad+4\delta^2(\log n)^{-1}E[|X-\theta_0-ta_n^{-1}|^{-2}I_{|X-\theta_0|>2Ma_n^{-1}}]\\ &\quad\le o(1)+16\delta^2(\log n)^{-1}E[|X-\theta_0|^{-2}I_{|X-\theta_0|>2Ma_n^{-1}}]=o(1)+\delta^2O(1).\end{aligned}$$

So, (2.13) follows. □

PROOF OF PART (A) OF THEOREM 1.1. We have already checked conditions (i)–(iv) in Theorem 2.1. Let

$$Z_n(t):=a_n^2(2\log n)^{-1/2}(H_n(\theta_0+ta_n^{-1})-H_n(\theta_0)-H(\theta_0+ta_n^{-1})+H(\theta_0)),\quad\text{if}\quad |t|\le M,$$

and let

$$Z_n(\infty):=a_n(H_n(\theta_0)-H(\theta_0)).$$

Let g_1, g_2, g_3, g_4 be independent standard normal r.v.'s. Let

$$Z(t)=\pi^{1/2}(f(\theta_0))^{1/2}2^{-3/2}\begin{pmatrix}(g_1+2^{1/2}g_2)t_1+g_3t_2\\ g_3t_1+(g_1+2^{1/2}g_4)t_2\end{pmatrix},$$

for $|t| \le M$. Observe that

$$E[Z(t)(Z(s))'] = \pi f(\theta_0) 2^{-3} \begin{pmatrix} 3t_1 s_1 + t_2 s_2 & t_1 s_2 + t_2 s_1 \\ t_1 s_2 + t_2 s_1 & t_1 s_1 + 3t_2 s_2 \end{pmatrix}$$

for $|s|, |t| \le M$. Let $Z(\infty)$ be a $\mathbb{R}^2$-valued Gaussian random vector, independent of g_1, g_2, g_3, g_4, with mean zero and covariance

$$E[Z(\infty)(Z(\infty))'] = E[h(X - \theta_0)(h(X - \theta_0))'] - E[h(X - \theta_0)]E[(h(X - \theta_0))'].$$

By Lemma 2.5 and Remark 12 in Arcones (1994a), we have that with probability one, $\{Z_n(t) : t \in T_M^*\}$ is relatively compact in $l_\infty(T_M^*)$ and its limit set is the unit ball of the $\mathbb{R}^2$-valued Gaussian process $\{Z(t) : t \in T_M^*\}$. Observe that we require $E[Z(t)(Z(\infty))'] = 0$ because, by Lemma 2.1,

$$a_n (2 \log n)^{-1/2} \operatorname{Cov}(h(X - \theta_0 - a_n^{-1} t) - h(X - \theta_0), h(X - \theta_0)) \to 0.$$

Thus condition (v) in Theorem 2.1 holds. Therefore, with probability one,

$$\{a_n^2 (2 \log n)^{-1/2} (H_n(\theta_0) + H'(\theta_0) \cdot (\hat{\theta}_n - \theta_0))\}_{n=1}^{\infty}$$

is relatively compact in $\mathbb{R}^2$ and its limit set is

$$L(K) := \{-\alpha(-(H'(\theta_0))^{-1} \cdot \alpha(\infty)) : \alpha \in K\},$$

where K is the unit ball of the r.k.h.s. of the Gaussian process $\{Z(t) : t \in T_M^*\}$. The linear space in L_2 generated by the coordinates of this process is

$$\left\{ \sum_{j=1}^{4} x_j g_j + x_5 W_1 + x_6 W_2 : x_i \in \mathbb{R}, 1 \le i \le 6 \right\},$$

where $W' = (W_1, W_2)$. So,

$$K = \left\{ \alpha : \text{there exist } x_i \text{ with } E\left[\left(\sum_{j=1}^{4} x_j g_j + x_5 W_1 + x_6 W_2 \right)^2 \right] \le 1, \right.$$
$$\left. \text{such that } \alpha(t) = E\left[Z(t) \sum_{j=1}^{6} x_j g_j \right], \text{ for each } t \in T_M^* \right\}.$$

If $|t| \le M$,

$$\alpha(t) = E\left[Z(t) \left(\sum_{j=1}^{4} x_j g_j + x_5 W_1 + x_6 W_2 \right) \right]$$
$$= \pi^{1/2} (f(\theta_0))^{1/2} 2^{-3/2} \begin{pmatrix} x_1 + 2^{1/2} x_2 & x_3 \\ x_3 & x_1 + 2^{1/2} x_4 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

and

$$\alpha(\infty) = E\left[Z(\infty)\left(\sum_{j=1}^{4} x_j g_j + x_5 W_1 + x_6 W_2\right)\right] = \begin{pmatrix} c_{1,1}x_5 + c_{1,2}x_6 \\ c_{2,1}x_5 + c_{2,2}x_6 \end{pmatrix}.$$

Thus,

$$\begin{aligned} &-\alpha(-(H'(\theta_0))^{-1}\cdot\alpha(\infty)) \\ &\quad= \pi^{1/2}(f(\theta_0))^{1/2}2^{-3/2}\begin{pmatrix} x_1 + 2^{1/2}x_2 & x_3 \\ x_3 & x_1 + 2^{1/2}x_4 \end{pmatrix} \\ &\quad\cdot (H'(\theta_0))^{-1}\begin{pmatrix} c_{1,1}x_5 + c_{1,2}x_6 \\ c_{2,1}x_5 + c_{2,2}x_6 \end{pmatrix} \end{aligned}$$

and the set in (1.10) is as claimed.

Next, we show that

$$\sup\left\{\left|\begin{pmatrix} x_1 + 2^{1/2}x_2 & x_3 \\ x_3 & x_1 + 2^{1/2}x_4 \end{pmatrix}(H'(\theta_0))^{-1}\begin{pmatrix} c_{1,1}x_5 + c_{1,2}x_6 \\ c_{2,1}x_5 + c_{2,2}x_6 \end{pmatrix}\right| \right.$$
$$\left. : \sum_{j=1}^{4} x_j^2 + c_{1,1}x_5^2 + c_{1,2}x_5x_6 + c_{2,1}x_5x_6 + c_{2,2}x_6^2 \le 1\right\}$$

is positive. If $x_1 = 1/2$ and $x_2 = x_3 = x_4 = 0$, then

$$\begin{aligned} &\begin{pmatrix} x_1 + 2^{1/2}x_2 & x_3 \\ x_3 & x_1 + 2^{1/2}x_4 \end{pmatrix}(H'(\theta_0))^{-1}\begin{pmatrix} c_{1,1}x_5 + c_{1,2}x_6 \\ c_{2,1}x_5 + c_{2,2}x_6 \end{pmatrix} \\ &\quad= 2^{-1}(H'(\theta_0))^{-1}\begin{pmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{pmatrix}\begin{pmatrix} x_5 \\ x_6 \end{pmatrix}. \end{aligned}$$

Since $X^{(1)} - E[X^{(1)}]$ and $X^{(2)} - E[X^{(2)}]$ are linearly independent vectors of L_2, $\begin{pmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{pmatrix}$ is positive definite, and $(H'(\theta_0))^{-1}\begin{pmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{pmatrix}\begin{pmatrix} x_5 \\ x_6 \end{pmatrix} \neq 0$, for $x_5^2 + x_6^2 \neq 0$. So, we can take x_5 and x_6, such that $2^{-2} + c_{1,1}x_5^2 + 2c_{1,2}x_5x_6 + c_{2,2}x_6^2 \le 1$ and

$$(H'(\theta_0))^{-1}\begin{pmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{pmatrix}\begin{pmatrix} x_5 \\ x_6 \end{pmatrix} \neq 0. \qquad \square$$

In the next theorem, we need to use a definition of weak convergence of stochastic processes with arbitrary index set. We use the definition in Hoffmann-Jørgensen (1991). This definition is as follows. Let T be a parameter set. Let $\{Z_n(t) : t \in T\}$, $n \ge 1$, be a sequence of stochastic processes, and let $\{Z(t) : t \in T\}$ be another stochastic process. We say that the sequence of stochastic processes $\{Z_n(t) : t \in T\}$, $n \ge 1$, converges weakly to $\{Z(t) : t \in T\}$ in $l_\infty(T)$ if:

(i) $\sup_{t\in T}|Z_n(t)| < \infty$ a.s. for each n large enough.

(ii) There exists a separable set A of $l_\infty(T)$ such that $\Pr^*\{Z \in A\} = 1$, where $\Pr^*$ means outer probability.

(iii) $E^*[H(Z_n)] \to E[H(Z)]$ for each bounded, continuous function H in $l_\infty(T)$, where E^* means outer expectation.

Now we consider the proof of part (b) of Theorem 1.1. We will need the following:

THEOREM 2.2. (Arcones (1994*b*), Theorem 9) *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in a measurable space $(S,\mathcal{S})$, let Θ be a subset of $\mathbb{R}^d$, let $h : S\times\Theta \to \mathbb{R}^d$ be a jointly measurable function, let $H(\theta) = E[h(X_1,\theta)]$, let $H_n(\theta) = n^{-1}\sum_{i=1}^n h(X_i,\theta)$, let $\hat\theta_n = \hat\theta_n(X_1,\dots,X_n)$ be a sequence of estimators and let $\{b_n\}$ be a sequence of real numbers tending to infinity. Suppose that the following conditions hold:*

(i) *There exists a θ_0 in the interior of Θ such that $H(\theta_0)=0$.*

(ii) *$H(\theta)$ has a second order expansion at θ_0, meaning that $H(\theta)$ is first differentiable at θ_0 and there exists a bilinear form $B:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d$ such that*

$$H(\theta) = H(\theta_0) + H'(\theta_0)\cdot(\theta-\theta_0) + B(\theta-\theta_0,\theta-\theta_0) + o(|\theta-\theta_0|^2).$$

(iii) *For each $\tau>0$,*

$$\lim_{\delta\to0}\limsup_{n\to\infty}\Pr{}^*\left\{\sup_{|\theta-\theta_0|\le\delta} n^{1/2}|(P_n-P)(h(\cdot,\theta)-h(\cdot,\theta_0))|>\tau\right\}=0.$$

(iv) *$\hat\theta_n \xrightarrow{\Pr} \theta_0$ and $b_n n^{1/2}H_n(\hat\theta_n)\xrightarrow{\Pr}0$.*

(v) *There exists a $\mathbb{R}^d$-valued stochastic process $\{Z(t): t\in\mathbb{R}^d\cup\{\infty\}\}$ such that, for each $M<\infty$, $\{Z_n(t): t\in T_M^*\}$ converges weakly to $\{Z(t): t\in T_M^*\}$, where $T_M^* = \{t\in\mathbb{R}^m : |t|\le M\}\cup\{\infty\}$,*

$$Z_n(t) = b_n n^{1/2}(P_n-P)(h(\cdot,\theta_0+n^{-1/2}t)-h(\cdot,\theta_0))$$

and

$$Z_n(\infty) = n^{1/2}(P_n-P)h(\cdot,\theta_0).$$

Case a. *If $b_n n^{-1/2}\to0$, then*

$$\{b_n n^{1/2}(H_n(\theta_0)+H'(\theta_0)\cdot(\hat\theta_n-\theta_0))\}$$

converges in distribution to

$$-Z(-(H'(\theta_0))^{-1}\cdot Z(\infty)).$$

Case b. *If $b_n = n^{1/2}$, then*

$$\{n(H_n(\theta_0)+H'(\theta_0)\cdot(\hat\theta_n-\theta_0))\}$$

converges in distribution to

$$-Z(-(H'(\theta_0))^{-1}\cdot Z(\infty)) - B((H'(\theta_0))^{-1}\cdot Z(\infty),(H'(\theta_0))^{-1}\cdot Z(\infty)).$$

Using last theorem, the proof of the part (b) of Theorem 1.1 is similar to the proof of the part (a). The main difference is in the proof of (v) in Theorem 2.2. To check this condition, we use the following central limit theorem for triangular arrays indexed by VC subgraph classes of functions. It follows from Theorem 10.6 in Pollard (1990).

THEOREM 2.3. *Let $\{h(x,\theta)-h(x,\theta_0):\theta\in\mathbb{R}^m\}$ be a VC subgraph class of functions, let $M<\infty$, let $\{a_n\}$ and $\{b_n\}$ be sequence of real numbers. Suppose that:*

(i) $\lim_{n\to\infty} b_n^2\,\mathrm{Cov}(h(X,\theta_0+ta_n^{-1})-h(X,\theta_0),h(X,\theta_0+sa_n^{-1})-h(X,\theta_0))$ *exists for each* $|s|,|t|\le M$.

(ii) $b_n^2E[H^2_{Ma_n^{-1}}(X)]=O(1)$, *where* $H_r(x)=\sup_{|t|\le r}|h(x,\theta_0+t)-h(x,\theta_0)|$.

(iii) $b_n^2E[H^2_{Ma_n^{-1}}(X)I_{H_{Ma_n^{-1}}(X)\ge\tau b_n^{-1}n^{1/2}}]\to 0$, *for each* $\tau>0$.

(iv) $$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\substack{|s-t|\le\delta\\|s|,|t|\le M}} b_n^2E[(h(X,\theta_0+ta_n^{-1})-h(X,\theta_0+sa_n^{-1}))^2]=0.$$

Then,

$$\Big\{b_nn^{-1/2}\sum_{i=1}^n(h(X_i,\theta_0+ta_n^{-1})-h(X_i,\theta_0)-E[h(X,\theta_0+ta_n^{-1})]+E[h(X,\theta_0)]):|t|\le M\Big\}$$

converges weakly to the Gaussian process $\{Z(t):|t|\le M\}$ with mean zero and covariance given by

$$E[Z(t)Z(s)]=\lim_{n\to\infty}b_n^2\,\mathrm{Cov}(h(X,\theta_0+ta_n^{-1})-h(X,\theta_0),h(X,\theta_0+sa_n^{-1})-h(X,\theta_0)).$$

The following follows directly from Theorem 2.3 using the arguments in Lemma 2.5.

LEMMA 2.6. *Under the conditions in Theorem 1.1, for each $M<\infty$,*

$$\{n(2\log n)^{-1/2}(H_n(\theta_0+tn^{-1/2})-H_n(\theta_0)-H(\theta_0+tn^{-1/2})+H(\theta_0)):|t|\le M\}$$

converges weakly to the $\mathbb{R}^2$-valued mean zero Gaussian process $\{Z(t):|t|\le M\}$ with covariance given by

$$E[Z(t)(Z(s))']=\pi f(\theta_0)2^{-3}\begin{pmatrix}3t_1s_1+t_2s_2 & t_1s_2+t_2s_1\\ t_1s_2+t_2s_1 & t_1s_1+3t_2s_2\end{pmatrix}$$

for each $|s|, |t| \leq M$.

The rest of the conditions in Theorem 2.2 can be checked similarly to conditions checked to prove part (a) of Theorem 1.1.

References

Arcones, M. A. (1994*a*). Some strong limit theorems for M-estimators, *Stochastic Process. Appl.*, **53**, 241–268.

Arcones, M. A. (1994*b*). On the weak Bahadur-Kiefer representation for M-estimators, *Probability in Banach Spaces, 9* (eds. J. Hoffmann-Jørgensen, J. Kuelbs and M. B. Marcus), 357–372, Birkhäuser, Boston.

Arcones, M. A. (1995). Asymptotic normality of multivariate trimmed means, *Statist. Probab. Lett.*, **25**, 43–53.

Arcones, M. A. and Mason, D. M. (1992). A general approach to the Bahadur-Kiefer representations for M-estimators, *Methods of Mathematical Statistics* (to appear).

Bahadur, R. R. (1966). A note on quantiles in large samples, *Ann. Math. Statist.*, **37**, 577–580.

Carroll, R. J. (1978). On almost sure expansions for M-estimates, *Ann. Statist.*, **6**, 314–318.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed sequential confidence intervals for the mean, *Ann. Statist.*, **36**, 457–462.

Duttweiler, D. L. (1973). The mean-square error of Bahadur's order-statistic approximation, *Ann. Statist.*, **1**, 446–453.

Haldane, S. J. (1948). Note on the median of a multivariate distribution, *Biometrika*, **35**, 414–415.

Hoffmann-Jørgensen, J. (1991). *Stochastic Processes on Polish Spaces*, *Aarhus Universitet Matematisk Institut Various Publications Series*, No. 39, Aarhus, Denmark.

Kiefer, J. (1967). On Bahadur's representation of sample quantiles, *Ann. Math. Stat.*, **38**, 1323–1342.

Koltchinskii, V. (1994*a*). Bahadur-Kiefer approximation for spatial quantiles, *Probability in Banach Spaces, 9* (eds. J. Hoffmann-Jørgensen, J. Kuelbs and M. B. Marcus), 401–415, Birkhäuser, Boston.

Koltchinskii, V. (1994*b*). Nonlinear transformations of empirical processes: functional inverses and Bahadur-Kiefer representations, *Probability Theory and Mathematical Statistics, Proceedings of the Sixth Vilnius Conference (1993)*, 423–445, VSP BV, Zeist, The Netherlands, and TEV Ltd., Vilnius, Lithuania.

Milasevic, P. and Ducharme, G. R. (1987). Uniqueness of the spatial median, *Ann. Statist.*, **15**, 1332–1333.

Niemiro, W. (1992). Asymptotics for M-estimators defined by convex minimization, *Ann. Statist.*, **20**, 1514–1533.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence, *Ann. Statist.*, **15**, 780–799.

Pfanzagl, J. with the assistance of W. Wefelmeyer (1985). Asymptotic expansions for general statistical models, *Lecture Notes in Statist.*, **31**, Springer, New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*, NSF CBMS Regional Conference Series in Probab. and Statist., Vol. 2, Institute of Mathematical Statistics, Hayward, California.