Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Multivariate Analysis 92 (2005) 205–225
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

# A semiparametric density estimator based on elliptical distributions

Eckhard Liebscher*

*Institute of Mathematics, Technical University Ilmenau, D-98684 Ilmenau/Thür, Germany*

Received 3 December 2002

**Abstract**

In the paper we study a semiparametric density estimation method based on the model of an elliptical distribution. The method considered here shows a way to overcome problems arising from the curse of dimensionality. The optimal rate of the uniform strong convergence of the estimator under consideration coincides with the optimal rate for the usual one-dimensional kernel density estimator except in a neighbourhood of the mean. Therefore the optimal rate does not depend on the dimension. Moreover, asymptotic normality of the estimator is proved.


*AMS 2000 subject classifications:* 62G07; 62H12

*Keywords:* Elliptical distributions; Kernel density estimator

## 1. Introduction

It is well known that in high dimensions nonparametric kernel density estimators have a poor performance for small samples and a very slow optimal convergence rate (cf. [14,15]). This is one phenomenon of the so-called "curse of dimensionality". Thus there is a need for new methods of density estimation in order to overcome this problem. In this paper we choose a semiparametric approach which is based on elliptical densities. Our approach goes partially back to Stute and Werner [17], and to Cui and He [1]. The new idea of the estimator introduced in this paper is to transform the data before applying the nonparametric estimator in order to avoid

*Fax: +49-3677-693270.
*E-mail address:* lieb@mathematik.tu-ilmenau.de.

0047-259X/$ - see front matter 
doi:10.1016/j.jmva.2003.09.007

convergence problems in boundary regions. Moreover, in the two papers mentioned before, it is supposed that at least a part of the parameters of the elliptical distribution is known. By contrast all parameters are assumed to be unknown in this paper. The estimator established in this paper offers interesting applications in areas where density estimators are needed for high-dimensional data [6]. The discriminant analysis is one such potential field. It should be pointed out that fitting an elliptical distribution is a useful alternative to the multivariate normal distribution which is frequently used. Accounts of the parametric estimation theory of elliptical distributions may be found in [1]; [3, p. 206]; [8].

The density $f$ of an elliptical distribution on $\mathbb{R}^d$ is given by

$$f(x) = \det(\Sigma)^{-1/2} g((x-\mu)^T \Sigma^{-1}(x-\mu)) \quad (x \in \mathbb{R}^d), \tag{1.1}$$

where $\mu \in \mathbb{R}^d$ is the mean and $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a measurable function with

$$\int_{\mathbb{R}^d} g(x^T x)\, dx = 1,$$

$\mathbb{R}^+ = [0, +\infty)$. We restrict ourselves to the case $d \geqslant 2$. Assume that $g$ is chosen such that $\Sigma$ is the covariance matrix of the distribution determined by (1.1). This additional condition on $g$ ensures the identifiability of the parameters in the distribution model (cf. [8, Theorems 2.6.2 and 2.6.5]). The theory of elliptical distributions is presented in the monograph by Fang and Zhang [8] where further references are given (see also [7]). We combine the components of $\mu$ and $\Sigma = (\sigma_{ij})_{i,j=1\dots d}$ in a parameter vector $\theta = (\mu_1, \dots, \mu_d, \sigma_{11}, \sigma_{12}, \dots, \sigma_{dd})^T \in \mathbb{R}^{d(d+3)/2}$ without repeating identical quantities ($\Sigma$ is symmetric). The main idea for estimating $f$ is to use nonparametric methods as well as parametric estimators. The estimation procedure works as follows: first an estimator for $\theta$ is computed, then $g$ is estimated in a nonparametric way and finally, the estimators for $\mu$ and $\Sigma$ are plugged in. It turns out that the optimal rate of uniform strong convergence of the estimator for $f$ coincides with the optimal rate for the usual *one-dimensional* kernel density estimator. Therefore the optimal rate does not depend on the dimension. A further advantage of our estimator is that the methods of bandwidth selection known from one-dimensional density estimation theory apply (cf. [9,10]).

The paper is organized as follows: The estimator is developed in Section 2. In Section 3 we provide a theorem about the asymptotic normality of the density estimator. Moreover, we give the rate of uniform strong convergence. The proofs are deferred to Section 4.

## 2. Estimators

Let us consider a random vector $X$ having the density given by (1.1). It is well known that $Z = \Sigma^{-1/2}(X - \mu)$ has a spherical distribution. Moreover, $Z \stackrel{d}{=} Ru^{(d)}$, where $u^{(d)}$ is uniformly distributed on the unit sphere of $\mathbb{R}^d$, $R$ is a random variable taking values in $\mathbb{R}^+$, and $R, u^{(d)}$ are independent. $Y_1 \stackrel{d}{=} Y_2$ means that $Y_1$ and $Y_2$ have the same distribution. Let $Y = (X-\mu)^T \Sigma^{-1}(X-\mu)$. Now $Y \stackrel{d}{=} R^2$, which has the density

$$f_Y(y) = s_d y^{d/2-1} g(y) \quad (y \in \mathbb{R}^+), \quad s_d = \frac{\pi^{d/2}}{\Gamma(d/2)} \tag{2.1}$$

(cf. [8, Theorem 2.5.5 and Corollaries 1, 2, p. 65]). To estimate $f$, we need some estimator for $g$. At first glance one could have two ideas:

- Idea 1: If $\hat{f}_Y$ is an estimator for $f_Y$, then $\hat{g}(y) = s_d^{-1} y^{-d/2+1} \hat{f}_Y(y)$ is an estimator for $g$. But then $\hat{g}(y) \to \infty$ as $y \to 0$ if $\hat{f}_Y(y)$ is bounded away from 0 in a neighbourhood of 0.
- Idea 2: We consider $\tilde{Y} = Y^{d/2}$ instead of $Y$. Let $\hat{f}_{\tilde{Y}}(y)$ be an estimator for the density $f_{\tilde{Y}}(y)\,\tilde{Y}$. Then

$$f_{\tilde{Y}}(y) = \frac{2}{d} s_d g(y^{2/d}) \quad \text{and} \quad \check{g}(y) = \frac{d}{2} s_d^{-1} \hat{f}_{\tilde{Y}}(y^{d/2}).$$

  This estimator $\check{g}$ for $g$ behaves well near zero. But the estimator $\check{g}$ has the disadvantage that it becomes wiggly for large values of $y$ since the data points are stretched by the power with exponent $d/2$ if $d>2$.

Now the final idea is to combine the advantages of the two estimators above and to introduce a rather general type of estimators. For this purpose, let $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ be a function having a derivative $\psi'$ with $\psi'(y) > 0$ for $y \geqslant 0$, and the property $\psi(0) = 0$. Then the density $h$ of $\tilde{Y} = \psi(Y)$ is given by

$$h(t) = \Psi'(t) f_Y(\Psi(t)) = s_d \Psi'(t) \Psi(t)^{d/2-1} g(\Psi(t)),$$

$\Psi$ is the inverse function of $\psi$. Further

$$g(x) = s_d^{-1} x^{-d/2+1} \psi'(x)\, h(\psi(x)). \tag{2.2}$$

This formula shows how to compute $g$ from $h$. We will see that $h$ can be estimated nonparametrically. Then we obtain an estimator for $g$ by applying (2.2). The function $\psi$ should be chosen such that the disadvantages described above are avoided. If $\lim_{x\to 0+0} x^{-d/2+1}\psi'(x)$ is a positive constant and $\psi'$ is bounded, then we can expect a good behaviour of an estimator for $g$ in a neighbourhood of 0. $\lim_{x\to\infty} \psi(x)/x = \text{const}$ ensures good properties of $\hat{f}_n$ for large values of the argument. The precise conditions on $\psi$ are given in the next section.

Now we turn to establish the specific estimator for $f$. Let $X_1, \dots, X_n$ be a sample of $\mathbb{R}^d$-valued random vectors having an elliptical distribution according to (1.1). Suppose that $g$ is bounded. Let $\hat{\mu}_n$ and $\hat{\Sigma}_n$ be estimators for $\mu$ and $\Sigma$, respectively. Then $\hat{\theta}_n = (\hat{\mu}_1, \dots, \hat{\mu}_d, \hat{\sigma}_{11}, \hat{\sigma}_{12}, \dots, \hat{\sigma}_{dd})^T \in \mathbb{R}^{d(d+3)/2}$ is an estimator for $\theta$. Suppose that $\hat{\theta}_n$ fulfils the following property:

$$\limsup_{n\to\infty} \sqrt{\frac{n}{\ln\ln(n)}}\,||\hat{\theta}_n - \theta|| = C_0 \quad \text{a.s.} \tag{2.3}$$

with a finite nonrandom constant $C_0 > 0$. For example, $\hat{\theta}_n$ arising from sample mean and sample covariance matrix satisfies this condition in view of Strassen's law of the iterated logarithm. Another appropriate choice for $\hat{\mu}_n$ and $\hat{\Sigma}_n$ could be the robust M-estimators (cf. [11]). The problem of efficient semiparametric estimation is examined in the monograph by Bickel et al. [2]. This monograph also provides a comprehensive overview of literature on semiparametric estimation.

The next step is to establish a nonparametric estimator for the density $h$ of $\tilde{Y}$. Let $Y_{ni} = \psi((X_i - \hat{\mu}_n)^T \hat{\Sigma}_n^{-1}(X_i - \hat{\mu}_n)),\quad i = 1, \dots, n.$ Using the transformed sample $Y_{n1}, \dots, Y_{nn}$, we define the kernel estimate for $h$:

$$\hat{h}_n(y) = \frac{1}{nB(n)} \sum_{i=1}^{n} (K((y - Y_{in})B(n)^{-1}) + K((y + Y_{in})B(n)^{-1})) \quad (y \in \mathbb{R}^+) \tag{2.4}$$

with a random bandwidth $B(n)$ and a kernel function $K$. The additional term $K((y + Y_{in})B(n)^{-1})$ is inserted in order to avoid boundary effects in the neighbourhood of zero and according to the idea of reflection methods. The reader interested in reflection methods is referred to [4,18]. Using the estimator $\hat{h}_n$ from (2.4), we get the estimator for $f$ as follows:

$$\hat{g}_n(z) = s_d^{-1} z^{-d/2+1} \psi'(z)\, \hat{h}_n(\psi(z)) \quad (z \in \mathbb{R}^+),$$
$$\hat{f}_n(x) = \det(\hat{\Sigma}_n)^{-1/2} \hat{g}_n((x - \hat{\mu}_n)^T \hat{\Sigma}_n^{-1}(x - \hat{\mu}_n)) \quad (x \in \mathbb{R}^d). \tag{2.5}$$

The asymptotic properties of $\hat{f}_n(x)$ are studied in the next sections. An other transformation-based estimator for a density is considered in El Barmi and Simonoff (2000).

Figs. 1 and 2 below show an example of estimators for $g$. The data were taken from the UCI Machine Learning Repository (Dataset "breast cancer"—new diagnostic database—variables 3,8,16,29). Obviously, there is a significant difference between the estimated function $g$ and the function $g$ arising from multivariate normal distribution.

## 3. Asymptotic properties of the density estimators

Prior to formulating the main results of the paper, we provide the assumptions on $K$ and $\psi$ of the estimator $\hat{f}_n(x)$ defined in Section 2 by (2.4) and (2.5). Here $p$ is some positive even integer.

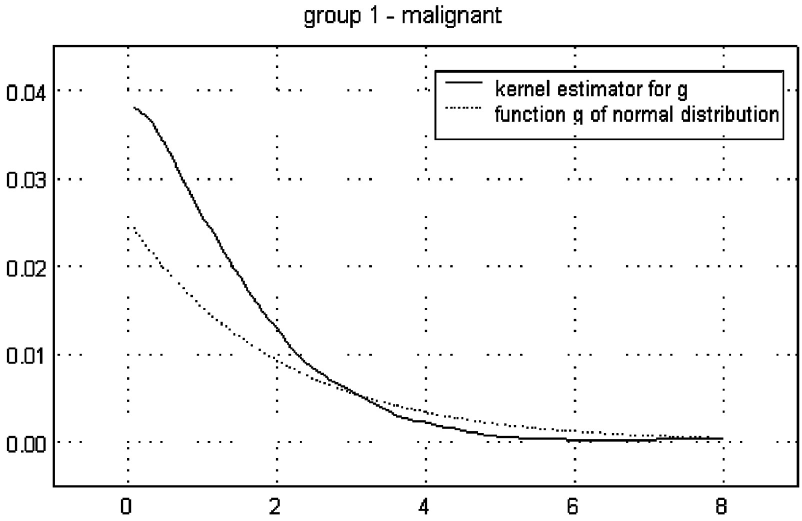


Fig. 1. Kernel estimator $\hat{g}_n$—group 1.

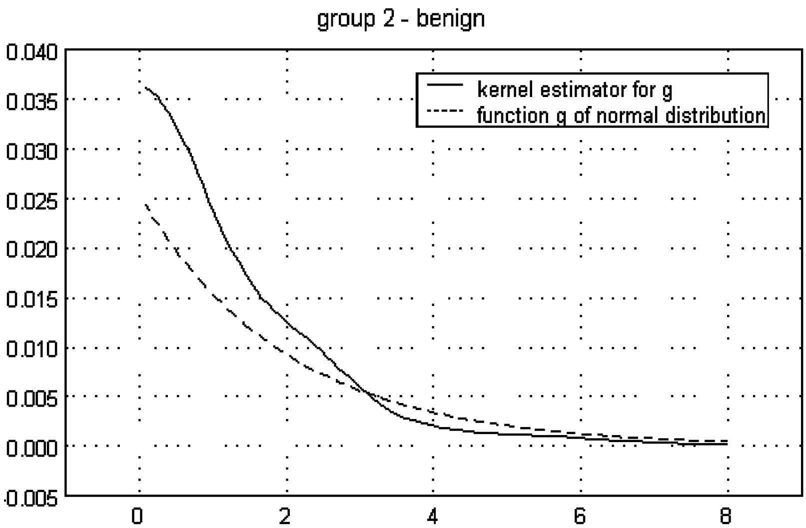


Fig. 2. Kernel estimator $\hat{g}_n$—group 2.

**Condition** $\mathscr{K}(p)$**.** The kernel function $K : \mathbb{R} \to \mathbb{R}$ has a Lipschitz continuous derivative on $\mathbb{R}$ and vanishes outside the interval $[-1, 1]$. Moreover,

$$\int_{-1}^{1} K(t)\,dt = 1, \quad \int_{-1}^{1} t^k K(t)\,dt = 0 \quad \text{for } k = 1, \ldots, p-1.$$

**Condition $\mathscr{T}(p)$.** The $(p+1)$th order derivative of $\Psi$ exists and is continuous on $(0,\infty)$, $\Psi$ is the inverse function of $\psi$, $\psi'$ is positive and bounded on $(0,+\infty)$, and $\psi''$ is bounded on $(0,+\infty)$. The function $x\mapsto x^{d/2-1}\psi'(x)^{-1}$ has a bounded derivative on $[0,M_1]$ with some $M_1>0$. Moreover,

$$\lim_{x\downarrow 0} x^{-d/2+1}\psi'(x) = C_1 > 0. \tag{3.1}$$

There are constants $\alpha\in(0,1]$, $C_2, M_2>0$ such that

$$|\Psi(t)| \leqslant C_2|t|^{\alpha} \quad \text{for } t\in[0,M_2].$$

**Example** (For $\psi$).

$$\psi(x) = -a + (a^{d/2} + x^{d/2})^{2/d} \tag{3.2}$$

with a constant $a>0$. Then

$$\lim_{x\downarrow 0} x^{-d/2+1}\psi'(x) = a^{1-d/2}, \quad \lim_{x\to\infty}\frac{\psi(x)}{x} = 1 \quad \text{and}$$

$$\Psi(t) = ((t+a)^{d/2} - a^{d/2})^{2/d} = a^{1-2/d}\left(\tfrac{d}{2}t\right)^{2/d} + o(t) \quad \text{as } t\downarrow 0.$$

Hence Condition $\mathscr{T}(p)$ is satisfied with $\alpha = 2/d$.

The random bandwidth $B(n)$ is assumed to fulfil the conditions

$$C_3 b(n) \leqslant B(n) \leqslant C_4 b(n), \tag{3.3}$$

$$\lim_{n\to\infty} b(n)\ln\ln n = 0 \quad \text{and } b(n) \geqslant C_5 n^{-1/5}. \tag{3.4}$$

where $C_3, C_4, C_5 > 0$ are constants and $\{b(n)\}_{n=1,2\ldots}$ is a sequence of positive real numbers. Now the theorem about rates of uniform strong convergence of $\hat{f}_n$ defined in (2.5) reads as follows:

**Theorem 3.1.** *Suppose that the pth order derivative $g^{(p)}$ of $g$ exists and is bounded on $\mathbb{R}^+$ for some even integer $p\geqslant 2$. Let conditions $\mathscr{K}(p)$, $\mathscr{T}(p)$, (2.3), (3.3), (3.4) and* $\mathbb{E}||Z||^{\tau} < +\infty$ be satisfied for some $\tau>4$. Then, for any compact set $D$ with $\mu\notin D$,

$$\sup_{x\in D} |\hat{f}_n(x) - f(x)| = O\left(\sqrt{\ln n}(nb(n))^{-1/2} + b^p(n)\right) \quad a.s. \tag{3.5}$$

*For any compact set $D$ with $\mu\in D$, we still have*

$$\sup_{x\in D} |\hat{f}_n(x) - f(x)| = O\left(\sqrt{\ln n}(nb(n))^{-1/2} + b^{\gamma}(n)\right) \quad a.s.$$

*with $\gamma = \min\{\alpha, \alpha+1-\alpha d/2\}$, $\alpha$ from Condition $\mathscr{T}(p)$.*

This theorem improves the rates given in [5]. For the function $\psi$ determined by (3.2), $\gamma$ is equal to $\frac{2}{d}$. Putting $b(n) = \mathrm{const}(n/\ln(n))^{-p/(2p+1)}$, we obtain the optimized (w.r.t. the bandwidth) convergence rate $\sup_{x \in D} |\hat{f}_n(x) - f(x)| = O((n/\ln(n))^{-p/(2p+1)})$ of (3.5). This rate is the optimal one known from one-dimensional kernel density estimation. The strong convergence rate of $\hat{f}_n$ for arguments away from $\mu$ does not depend on the dimension. The reason for the slow convergence rate of $\hat{f}_n$ for arguments close to $\mu$ is the following: In many cases we have $\lim_{t \downarrow 0} h'(t) = +\infty$ or $-\infty$. This holds for example if $\lim_{x \downarrow 0} x^{-d/2+1} g'(x) = +\infty$ or $-\infty$ and the derivative of the function $\Psi' \Psi^{d/2-1}$ is bounded in a neighbourhood of zero.

The following theorem states the asymptotic normality of the estimator $\hat{f}_n$. Here $V_n = o_{\mathbb{P}}(a_n)$ means that $a_n^{-1} V_n \xrightarrow{\mathbb{P}} 0$, $\{V_n\}$ and $\{a_n\}$ are sequences of random variables and positive real numbers, respectively.

**Theorem 3.2.** *Let the assumptions of Theorem* 3.1 *be satisfied and* $b(n) = C_6 n^{-1/(2p+1)}$ *with a constant* $C_6 > 0$. *Moreover, assume that*

$$|B(n)b(n)^{-1} - 1| = o_{\mathbb{P}}\left(\sqrt{\ln \ln n}\; n^{-1/2}\right). \tag{3.6}$$

*Then for any* $x \in \mathbb{R}^d, x \neq \mu$ *such that* $g^{(p)}$ *is continuous at* $u := (x-\mu)^T \Sigma^{-1} (x - \mu)$,

$$\sqrt{nB(n)}(\hat{f}_n(x) - f(x)) \xrightarrow{\mathscr{D}} \mathscr{N}(\bar{\mu}, \bar{\sigma}^2) \quad a.s.$$

*where*

$$\bar{\sigma}^2 = \det(\Sigma)^{-1} s_d^{-1} u^{-d/2+1} \psi'(u) g(u) \int_{-1}^{1} K^2(t)\, dt$$

$$\bar{\mu} = \det(\Sigma)^{-1/2} s_d^{-1} u^{-d/2+1} \psi'(u) C_6^{(2p+1)/2} \frac{1}{p!} h^{(p)}(\psi(u)) \int_{-1}^{1} t^p K(t)\, dt.$$

A similar theorem was proved by Stute and Werner [17] in the case of known $\mu$ and $\psi(x) \equiv x$. From Theorem 3.2 one may construct confidence regions for $f$, for example. But when doing so, one needs estimators for $\bar{\mu}$ and $\bar{\sigma}^2$ which in turn requires an appropriate estimator for $h^{(p)}(\psi(u))$. Moreover, $\hat{\mu}_n$ and $\hat{\Sigma}_n$ should be used instead of $\mu$ and $\Sigma$. Bandwidth selection methods satisfying (3.6) can be found in [9].

## 4. Proofs

Assume that (2.3), (3.3) and (3.4) are satisfied, and $g'$ exists and is bounded on finite subintervals of $\mathbb{R}^+$. Thus there is some $n_0$ such that $3C_4 b(n) < 1$ for $n \geqslant n_0$. Further on let $n \geqslant n_0$. First we provide two lemmas which are used at several places below.

**Lemma 4.1.** *Under Condition $\mathscr{T}(2)$, the functions $t \mapsto \Psi'(t)\Psi(t)^{d/2-1}$ and $h$ are bounded on bounded subsets of $\mathbb{R}^+$.*

**Proof.** Condition (3.1) implies that

$$\lim_{t \downarrow 0} \Psi'(t)\Psi(t)^{d/2-1} = C_1^{-1}.$$

Since $\psi'(t) > 0$ for $t > 0$, the function $t \mapsto \Psi'(t)\ \Psi(t)^{d/2-1}$ and hence $h$ are bounded on any interval $[m_1, m_2]$, $m_1 \geqslant 0$. □

**Lemma 4.2.** *Under Condition $\mathscr{T}(2)$,*

$$\sup_{t,v \in [0,\bar{M}]} |h(t) - h(v)|\, |t - v|^{-\gamma} < +\infty$$

*for any $\bar{M} > 0$, where $\gamma = \min\{\alpha, \alpha + 1 - \alpha d/2\}$.*

**Proof.** Observe that by the Lipschitz continuity of $g$,

$$|h(t) - h(v)| \leqslant C_7 |\Psi(t) - \Psi(v)|$$

uniformly for $t, v \in [0, \bar{M}]$. By Condition $\mathscr{T}(2)$,

$$|\Psi(t) - \Psi(v)| \leqslant C_8 |t - v|^{\gamma}$$

uniformly for $t, v \in [0, \bar{M}]$. $C_7, C_8 > 0$ are constants. This completes the proof. □

Let us introduce some notations

$$K_b(y,t) = K((y-t)/b) + K((y+t)/b) \quad \text{for } y, t \geqslant 0,$$
$$\bar{Y}_{in} = (X_i - \hat{\mu}_n)^T \hat{\Sigma}_n^{-1} (X_i - \hat{\mu}_n),$$
$$\bar{Y}_i = (X_i - \mu)^T \Sigma^{-1} (X_i - \mu), \quad \tilde{Y}_i = \psi(\bar{Y}_i) \quad \text{for } i = 1, \ldots, n,$$

and

$$\tilde{h}_n(y,b) = \frac{1}{nb} \sum_{i=1}^{n} K_b(y, \tilde{Y}_i), \quad \hat{h}_n(y,b) = \frac{1}{nb} \sum_{i=1}^{n} K_b(y, Y_{in}) \quad (y \in \mathbb{R}^+).$$

Note that $Y_{in} = \psi(\bar{Y}_{ni})$ for $i = 1, \ldots, n$. Obviously, each $\tilde{Y}_i$ has the density $h$ such that $\tilde{h}_n(., b)$ is the usual density estimator for $h$ with some boundary adjustment. In the first part of this section we prove strong convergence rates for $\hat{h}_n$ and later for $\hat{f}_n$. Let $\underline{b}_n = C_3 b(n)$, $\bar{b}_n = C_4 b(n)$. The compact set $[m, M] \times [\underline{b}_n, \bar{b}_n]$ with arbitrary $m$ and $M$, $0 \leqslant m < M$ can be covered with closed rectangles $U_1, \ldots, U_{n^2}$ having sides of length $(M - m)n^{-1}, (\bar{b}_n - \underline{b}_n)n^{-1}$ and centres $(u_1, b_1), \ldots, (u_{n^2}, b_{n^2})$ such that

$\bigcup_{i=1}^{n} U_i = [m, M] \times [\underline{b}_n, \bar{b}_n]$. $m$ and $M$ will be determined later. Observe that

$$\begin{aligned}
&\sup_{y \in [m,M]} |\hat{h}_n(y) - h(y)| \\
&\quad \leqslant \sup_{y \in [m,M]} \sup_{b \in [\underline{b}_n, \bar{b}_n]} |\hat{h}_n(y, b) - h(y)| \\
&\quad \leqslant \max_{k=1,\dots,n^2} \Bigg( \sup_{(y,b) \in U_k} |\hat{h}_n(y, b) - \hat{h}_n(u_k, b_k)| + |\hat{h}_n(u_k, b_k) - \tilde{h}_n(u_k, b_k)| \\
&\qquad + |\tilde{h}_n(u_k, b_k) - h(u_k)| + \sup_{y \,:\, (y,b_k) \in U_k} |h(u_k) - h(y)| \Bigg). \qquad (4.1)
\end{aligned}$$

**Lemma 4.3.** *Assume that Condition $\mathscr{T}(2)$ is satisfied. Then*

$$\max_{k=1,\dots,n^2} \sup_{y \,:\, (y,b_k) \in U_k} |h(u_k) - h(y)| = \begin{cases} O(n^{-1}) & \text{if } m \neq 0, \\ O(n^{-\gamma}) & \text{if } m = 0, \end{cases} \quad \gamma \textit{ as above.}$$

**Proof.** By construction of the sets $U_k$, the assertion follows from Lemma 4.2. □

**Lemma 4.4.** *Assume that the pth order derivative $h^{(p)}$ of $h$ exists for some even integer $p \geqslant 2$ and is bounded on every interval $[m_1, m_2]$ with $m_1 > 0$. Moreover, let the Conditions $\mathscr{K}(p)$ and $\mathscr{T}(p)$ be fulfilled. Then*

$$\max_{k=1,\dots,n^2} |\tilde{h}_n(u_k, b_k) - h(u_k)| = O\Big(\sqrt{\ln n}(nb(n))^{-1/2} + \beta_n\Big) \quad a.s.$$

*where $\beta_n = b^p(n)$ if $m > 0$ and $\beta_n = b^\gamma(n)$ if $m = 0$, $\gamma$ as above.*

**Proof.** By standard arguments, one can show that

$$\max_{k=1,\dots,n^2} |\tilde{h}_n(u_k, b_k) - \mathbb{E}\tilde{h}_n(u_k, b_k)| = O\Big(\sqrt{\ln n}(nb(n))^{-1/2}\Big) \quad \text{a.s.} \qquad (4.2)$$

(cf. [16, Theorem 1.2]). In the case $m > 0$, we obtain

$$\begin{aligned}
&\max_{k=1,\dots,n^2} |\mathbb{E}\tilde{h}_n(u_k, b_k) - h(u_k)| \\
&\quad = \max_{k=1,\dots,n^2} \left| b_k^{-1} \int_0^\infty K((u_k - t)/b_k) h(t)\, dt - h(y) \right| = O(b^p(n)) \qquad (4.3)
\end{aligned}$$

by Taylor expansion for $n$ large enough. Eqs. (4.2) and (4.3) imply Lemma 4.4 in the case $m>0$. By Lemma 4.2,

$$\begin{aligned}
&\sup_{b\in[\underline{b}_n,\bar{b}_n]}\ \sup_{y\in[0,M]} |\mathbb{E}\tilde{h}_n(y,b)-h(y)| \\
&\quad = \sup_{b\in[\underline{b}_n,\bar{b}_n]}\ \sup_{y\in[0,M]} \left| b^{-1}\int_0^{\infty} (K((y-t)/b)+K((y+t)/b))(h(t)-h(y))\,dt \right| \\
&\quad \leqslant \sup_{b\in[\underline{b}_n,\bar{b}_n]} \left( \sup_{y\in[2b,M]} \left| \int_{-1}^{1} K(t)(h(y-tb)-h(y))\,dt \right| \right. \\
&\quad \left. + \sup_{y\in[0,2b)} \left| \int_{-1}^{y/b} (K(t)+K(-t+2y/b))(h(y-tb)-h(y))\,dt \right| \right) \\
&\quad = O(b^{\gamma}(n)).
\end{aligned}$$

Eq. (4.2) completes the proof of Lemma 4.4 in the case $m=0$. $\square$

**Lemma 4.5.** *Assume that Conditions $\mathscr{K}(1)$ and $\mathscr{T}(2)$ are satisfied and $\mathbb{E}||Z||^{\tau}<+\infty$ for some $\tau>4$. Then*

$$\max_{k=1,\dots,n^2} |\hat{h}_n(u_k,b_k)-\tilde{h}_n(u_k,b_k)| = o((nb(n))^{-1/2}) \quad a.s.$$

For the proof of this lemma, a series of further lemmas is needed. Using the Lipschitz continuity of $K'$, we obtain

$$\max_{k=1,\dots,n^2} |\hat{h}_n(u_k,b_k)-\tilde{h}_n(u_k,b_k)| \leqslant B_{1n}+O(n^{-1}b(n)^{-2})(B_{2n}+B_{3n}) \tag{4.4}$$

where $G_b(y,t)=K'((y-t)/b)-K'((y+t)/b)$,

$$\begin{aligned}
B_{1n} &= \max_{k=1,\dots,n^2} \left| n^{-1}b_k^{-2}\sum_{i=1}^{n} (\bar{Y}_{in}-\bar{Y}_i)G_{b_k}(u_k,\tilde{Y}_i)\psi'(\bar{Y}_i) \right|, \\
B_{2n} &= \max_{k=1,\dots,n^2} \sum_{i=1}^{n} ((\bar{Y}_i-\bar{Y}_{in})^2+|\bar{Y}_i-\bar{Y}_{in}|I(|\psi(\bar{Y}_{in})-\tilde{Y}_i|>b_k)),
\end{aligned}$$

$$\begin{aligned}
B_{3n} = \max_{k=1,\dots,n^2} b_k^{-1}\sum_{i=1}^{n} (\bar{Y}_i-\bar{Y}_{in})^2 \\
\times (I(|u_k-\tilde{Y}_i|\leqslant 2b_k)+I(|u_k+\tilde{Y}_i|\leqslant 2b_k))\psi'(\bar{Y}_i).
\end{aligned}$$

Let $Z_i=\Sigma^{-1/2}(X_i-\mu)$ such that $\bar{Y}_i=Z_i^TZ_i$ and

$$\bar{Y}_{in}-\bar{Y}_i = Z_i^T\Delta_nZ_i+2\tilde{\mu}_n^TZ_i+\beta_n,$$

where $\Delta_n \coloneqq \Sigma^{1/2}\hat{\Sigma}_n^{-1}\Sigma^{1/2} - I$, $\beta_n \coloneqq (\mu - \hat{\mu}_n)^T\hat{\Sigma}_n^{-1}(\mu - \hat{\mu}_n)$, $\tilde{\mu}_n^T \coloneqq (\mu - \hat{\mu}_n)^T\hat{\Sigma}_n^{-1}\Sigma^{1/2}$. By virtue of (2.3), we obtain that

$$\Delta_n = O\Big(n^{-1/2}\sqrt{\ln\ln n}\Big), \quad \tilde{\mu}_n = O\Big(n^{-1/2}\sqrt{\ln\ln n}\Big), \quad \beta_n = O\big(n^{-1}\ln\ln n\big) \quad \text{a.s.}$$

and

$$|\bar{Y}_{in} - \bar{Y}_i| \leqslant \kappa_n(||Z_i||^2 + 1), \quad \kappa_n = O\Big(n^{-1/2}\sqrt{\ln\ln n}\Big) \quad \text{a.s.} \tag{4.5}$$

$(i = 1, \ldots, n)$. $\{\kappa_n\}$ is a sequence of positive real numbers not depending on $i$. Moreover,

$$\begin{aligned} B_{1n} \leqslant\; & O\left(\frac{\sqrt{\ln\ln n}}{n^{3/2}b^2(n)}\right) \max_{k=1,\ldots,n^2} \sum_{j,l=1}^{d} \sum_{\delta=0,1} \left| \sum_{i=1}^{n} Z_{ij}^{\delta} Z_{il} G_{b_k}(u_k, \tilde{Y}_i)\psi'(\bar{Y}_i) \right| \\ & + O\left(\frac{\ln\ln n}{n^2 b^2(n)}\right) \max_{k=1,\ldots,n^2} \sum_{j,l=1}^{d} \left| \sum_{i=1}^{n} G_{b_k}(u_k, \tilde{Y}_i)\psi'(\bar{Y}_i) \right|, \end{aligned} \tag{4.6}$$

where $Z_i = (Z_{i1}, \ldots, Z_{id})^T$. In the sequel we derive the convergence rates of $B_{1n}$ to $B_{3n}$. For this purpose, we next need the following auxiliary statement.

**Lemma 4.6.** *Assume that Condition $\mathscr{T}(2)$ is satisfied. Let $K : \mathbb{R} \to \mathbb{R}$ be a bounded function with $K(t) = 0$ for $t$: $|t| > 1$. Then, for $j, l = 1, \ldots, d$ and $\delta, \kappa = 0, 1$,*

$$\max_{k=1,\ldots,n^2} \left| \sum_{i=1}^{n} (U_{nijlk} - \mathbb{E}U_{nijlk}) \right| = O\Big(\sqrt{nb(n)\ln(n)}\Big) \quad a.s.$$

*with $U_{nijlk} \coloneqq Z_{ij}^{\delta} Z_{il}^{\kappa} K((u_k - \tilde{Y}_i)/b_k)\psi'(\bar{Y}_i)$, and*

$$\max_{k=1,\ldots,n^2} \left| \sum_{i=1}^{n} (\bar{U}_{nijlk} - \mathbb{E}\bar{U}_{nijlk}) \right| = O\Big(\sqrt{nb(n)\ln(n)}\Big) \quad a.s.$$

*with $\bar{U}_{nijlk} \coloneqq Z_{ij}^{\delta} Z_{il}^{\kappa} K((u_k + \tilde{Y}_i)/b_k)\psi'(\bar{Y}_i)$.*

**Proof.** We only prove the first assertion since the proof of the second assertion proceeds similarly. Choose $\bar{M}$ such that $[0, \bar{M}] \supset \Psi([0, M + 1])$. Hence $U_{nijlk} = 0$ for $\omega$: $\bar{Y}_i(\omega) > \bar{M}$ since then $\tilde{Y}_i > M + 1$ and $u_k \leqslant M$. By Lemma 4.1,

$$|U_{nijlk}| \leqslant \bar{M}^{(\delta+\kappa)/2} \sup_{t \in [0,\bar{M}]} \psi'(t) \sup_{t \in [-1,1]} |K(t)|$$

and

$$\begin{aligned}\max_{\substack{i,k\,:\,1\leqslant i\leqslant n,\\ 1\leqslant k\leqslant n^2}} \mathbb{D}^2 U_{nijlk} &\leqslant \max_{k=1,\dots,n^2} \mathbb{E} Z_{1j}^{2\delta} Z_{1l}^{2\kappa} K^2((u_k-\tilde{Y}_i)/b_k)\,\psi'(\bar{Y}_1)^2\\ &\leqslant \bar{M}^{\delta+\kappa} \sup_{t\in[0,\bar{M}]} \psi'(t)^2 \max_{k=1,\dots,n^2} \mathbb{E}K^2((u_k-\tilde{Y}_1)/b_k)\\ &\leqslant \text{const}\cdot \max_{k=1,\dots,n^2} \int_{\max\{u_k-b_k,0\}}^{u_k+b_k} h(t)\,dt\\ &= O(b(n))\end{aligned}$$

for $j,l=1,\dots,d$. $\mathbb{D}^2$ is the symbol for the variance. Let $a_n := \sqrt{nb(n)\ln(n)}$. Applying the Bernstein inequality (cf. [13, p. 112]), we get

$$\begin{aligned}&\mathbb{P}\left\{\max_{k=1,\dots,n^2}\left|\sum_{i=1}^n (U_{nijlk}-\mathbb{E}U_{nijlk})\right|>\varepsilon a_n\right\}\\ &\leqslant \sum_{k=1}^{n^2} \mathbb{P}\left\{\left|\sum_{i=1}^n (U_{nijlk}-\mathbb{E}U_{nijlk})\right|>\varepsilon a_n\right\}\\ &\leqslant C_9 n^2 \exp\{-C_{10}\varepsilon^2 a_n^2 (nb(n)+\varepsilon a_n)^{-1}\}\\ &\leqslant C_{11}\exp\{2\ln(n)-C_{12}\varepsilon^2\ln(n)(1+\varepsilon)^{-1}\}\end{aligned}$$

for $\varepsilon>1$. $C_9$ to $C_{12}$ are positive constants not depending on $n$, $j$, $l$ or $\varepsilon$. Hence

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{\max_{k=1,\dots,n^2}\left|\sum_{i=1}^n (U_{nijlk}-\mathbb{E}U_{nijlk})\right|>\varepsilon a_n\right\}<+\infty$$

and the lemma follows by virtue of the Borel–Cantelli lemma. □

Throughout the remainder of this section, we suppose that Conditions $\mathscr{K}(1)$ and $\mathscr{T}(2)$ are satisfied.

**Lemma 4.7.** *We have*
(a)

$$\mathbb{E}Z_{1j}G_b(y,\tilde{Y}_1)\psi'(\bar{Y}_1)=0 \quad \textit{for } j=1,\dots,d, b>0,$$

(b)

$$\sup_{y\in[0,M]}\ \sup_{b\in[\underline{b}_n,\bar{b}_n]} |\mathbb{E}Z_{1j}Z_{1l}G_b(y,\tilde{Y}_1)\psi'(\bar{Y}_1)| = O(b^2(n)) \quad \textit{for } j,l=1,\dots,d$$

*and* (c)

$$\sup_{y\in[0,M]}\ \sup_{b\in[\underline{b}_n,\bar{b}_n]} |\mathbb{E}G_b(y,\tilde{Y}_1)\psi'(\bar{Y}_1)| = O(b(n)).$$

**Proof.** Obviously, $Z_1$ has a spherical distribution. Let $R_1 = ||Z_1||$. Now $R_1$ and $R_1^{-1}Z_1$ are independent random variables (cf. [8, p. 57]) and $R_1^2$ has the density given by (2.1). Moreover, $\mathbb{E}R_1^{-1}Z_1 = 0$.

(a) Hence

$$\mathbb{E}Z_{1j}G_b(y, \tilde{Y}_1)\psi'(\bar{Y}_1) = \mathbb{E}R_1^{-1}Z_{1j}\mathbb{E}R_1 G_b(y, \psi(R_1^2))\psi'(R_1^2) = 0$$

for $j = 1, \dots, n$, $y \in [0, M]$ which is the first assertion of the lemma.

(b) Let $\tilde{g}(t) = s_d t^{d/2} g(t)$. Since $|R_1^{-1}Z_{1j}| \leqslant 1$ for $j = 1, \dots, d$ and $\bar{Y}_1 = R_1^2$, we obtain the following inequalities and identities by partial integration and by Lemma 4.1:

$$\begin{aligned}
&\sup_{y\in[0,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} |\mathbb{E}Z_{1j}Z_{1l}G_b(y, \tilde{Y}_1)\psi'(\bar{Y}_1)| \\
&\quad \leqslant \sup_{y\in[0,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} |\mathbb{E}R_1^2 G_b(y, \psi(R_1^2))\psi'(R_1^2)| \\
&\quad = \sup_{y\in[0,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} \left| \int_0^\infty (K'((y-\psi(t))/b) - K'((y+\psi(t))/b))\psi'(t)\tilde{g}(t)\, dt \right| \\
&\quad = \sup_{y\in[0,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} \left| b \int_{-y/b}^1 (K'(-v) - K'(v+2yb^{-1}))\tilde{g}(\Psi(y+vb))\, dv \right| \\
&\quad = \sup_{y\in[0,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} \left| b^2 \int_{-y/b}^1 (K(-v) + K(v+2yb^{-1}))\tilde{g}'(\Psi(y+vb)) \right. \\
&\qquad \left. \times\, \Psi'(y+vb)\, dv \right| \\
&\quad \leqslant O(b^2(n)) \sup_{y\in[0,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} \int_{-y/b}^1 (|K(-v)| + |K(v+2yb^{-1})|) \\
&\qquad \times |\Psi'(y+vb)\Psi(y+vb)^{d/2-1}|\, dv \\
&\quad = O(b^2(n)).
\end{aligned}$$

Hence the proof of part (b) is complete.

(c) Analogously to part (b),

$$\begin{aligned}
&\sup_{y\in[0,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} |\mathbb{E}G_b(y, \tilde{Y}_1)\psi'(\bar{Y}_1)| \\
&\quad \leqslant O(b(n)) \\
&\sup_{y\in[0,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} \left| \int_{-y/b}^1 (K'(-v) - K'(v+2yb^{-1}))\Psi(y+vb)^{d/2-1} g(\Psi(y+vb))\, dv \right| \\
&= O(b(n))
\end{aligned}$$

which is assertion (c). $\square$

**Lemma 4.8.**

$$B_{1n} = o(n^{-1/2}b(n)^{-1/2}) \quad a.s.$$

**Proof.** An application of Lemmas 4.6 and 4.7 leads to

$$\max_{k=1,\dots,n^2}\left|\sum_{i=1}^{n} Z_{ij}^{\delta} Z_{il} G_{b_k}(u_k, \tilde{Y}_i)\psi'(\bar{Y}_i)\right| = O\left(\sqrt{nb(n)\ln(n)} + nb^2(n)\right) \quad \text{a.s.},$$

$$\max_{k=1,\dots,n^2}\left|\sum_{i=1}^{n} G_{b_k}(u_k, \tilde{Y}_i)\psi'(\bar{Y}_i)\right| = O\left(\sqrt{nb(n)\ln(n)} + nb(n)\right) \quad \text{a.s.}$$

for $\delta = 0, 1,\ j, l = 1, \dots, d$. Hence, by (4.6), we obtain the lemma. $\square$

**Lemma 4.9.** *Suppose that* $\mathbb{E}||Z||^{\tau} < +\infty$ *for some* $\tau > 4$. *Then*

$$B_{2n} = o(\sqrt{n}b(n)^{3/2}) \quad a.s.$$

**Proof.** By the law of large numbers and (4.5), we obtain

$$\sum_{i=1}^{n} (\bar{Y}_i - \bar{Y}_{in})^2 \leqslant O(\ln\ln n)\left(n^{-1}\sum_{i=1}^{n} ||Z_i||^4 + 1\right) = O(\ln\ln n) \quad \text{a.s.}$$

and by the Lipschitz continuity of $\psi$,

$$\begin{aligned}
&\max_{k=1,\dots,n^2} \sum_{i=1}^{n} |\bar{Y}_i - \bar{Y}_{in}| I(|\psi(\bar{Y}_{in}) - \psi(\bar{Y}_i)| > b_k) \\
&\leqslant O(b(n)^{-\tau/2+1}) \sum_{i=1}^{n} |\bar{Y}_i - \bar{Y}_{in}|\, |\psi(\bar{Y}_{in}) - \psi(\bar{Y}_i)|^{\tau/2-1} \\
&= O(b(n)^{-\tau/2+1}) \sum_{i=1}^{n} |\bar{Y}_i - \bar{Y}_{in}|^{\tau/2} \\
&= \sqrt{n}b(n)^{3/2} O(b(n)^{-\tau/2-1/2}(\ln\ln n)^{\tau/4} n^{-\tau/4+1/2})\left(n^{-1}\sum_{i=1}^{n} ||Z_i||^{\tau} + 1\right) \\
&= \sqrt{n}b(n)^{3/2} O((\ln\ln n)^{\tau/4} n^{-3(\tau-4)/20}) = o\left(\sqrt{n}b(n)^{3/2}\right) \quad \text{a.s.}
\end{aligned}$$

which implies the lemma. $\square$

**Lemma 4.10.**

$$B_{3n} = o(\sqrt{n}b(n)^{3/2}) \quad a.s.$$

**Proof.** By (4.5), we deduce

$$\begin{aligned} B_{3n} &\leqslant O(n^{-1}\ln\ln n) \max_{k=1,\dots,n^2} b_k^{-1} \sum_{i=1}^{n} (\bar{Y}_i^2+1) I(\bar{Y}_i \in [0,\bar{M}]) \psi'(\bar{Y}_i) \\ &\quad \times (I(|u_k - \tilde{Y}_i| \leqslant 2b_k) + I(|u_k + \tilde{Y}_i| \leqslant 2b_k)) \\ &\leqslant O(n^{-1}\ln\ln n) \\ &\quad \times \max_{k=1,\dots,n^2} b_k^{-1} \sum_{i=1}^{n} (I(|u_k - \tilde{Y}_i| \leqslant 2b_k) + I(|u_k + \tilde{Y}_i| \leqslant 2b_k)) \psi'(\bar{Y}_i) \end{aligned}$$

($\bar{M}$ as in Lemma 4.6). Observe that by Lemma 4.1,

$$\max_{k=1,\dots,n^2} b_k^{-1} \mathbb{E} I(|u_k - \tilde{Y}_i| \leqslant 2b_k) \leqslant \max_{k=1,\dots,n^2} b_k^{-1} \int_{\max\{u_k - 2b_k, 0\}}^{u_k+2b_k} h(v)\, dv \leqslant \text{const}$$

and

$$\max_{k=1,\dots,n^2} b_k^{-1} \mathbb{E} I(|u_k + \tilde{Y}_i| \leqslant 2b_k) \leqslant \text{const}.$$

Applying Lemma 4.6, we obtain

$$B_{3n} = O(n^{-1}\ln\ln n)\left(n + \sqrt{n\ln(n)}b(n)^{-1/2}\right) = O(\ln\ln n) \quad \text{a.s.}$$

which is Lemma 4.10. $\square$

**Proof of Lemma 4.5.** Combine Lemmas 4.8–4.10 and (4.4) to get Lemma 4.5. $\square$

**Lemma 4.11.**

$$\max_{k=1,\dots,n^2} \sup_{(y,b)\in U_k} |\hat{h}_n(y,b) - \hat{h}_n(u_k,b_k)| = O(n^{-1}b(n)^{-2}) \quad a.s.$$

*and*

$$\max_{k=1,\dots,n^2} \sup_{(y,b)\in U_k}, |\tilde{h}_n(y,b) - \tilde{h}_n(u_k,b_k)| = O(n^{-1}b(n)^{-2}) \quad a.s.$$

**Proof.** Observe that the Lipschitz continuity of $K$ implies

$$\begin{aligned} &|K_b(y,t) - K_b(w,t)| \\ &\quad \leqslant |K((y-\psi(t))/b) - K((w-\psi(t))/b)| + |K((y+\psi(t))/b) - K((w+\psi(t))/b)| \\ &\quad \leqslant C_{13}|y-w|b^{-1} \quad \text{for } y,w \in \mathbb{R}^+, b \in [\underline{b}_n, \bar{b}_n] \end{aligned}$$

and

$$\begin{aligned} &|K_b(y,t) - K_B(y,t)| \\ &\quad \leqslant |K((y-\psi(t))/b) - K((y-\psi(t))/B)| + |K((y+\psi(t))/b) - K((w+\psi(t))/B)| \\ &\quad \leqslant C_{13}|b-B| \max\{b^{-1}, B^{-1}\} \quad \text{for } y \in \mathbb{R}^+, b, B > 0 \end{aligned}$$

with a constant $C_{13}>0$. Hence, by construction of the sets $U_k$,

$$\begin{aligned}
&\max_{k=1,\dots,n^2} \sup_{(y,b)\in U_k} \left| \frac{1}{n}\sum_{i=1}^{n} (b^{-1}K_b(y, Y_{in}) - b_k^{-1}K_{b_k}(u_k, Y_{in})) \right| \\
&\quad \leqslant C_{13} b(n)^{-2} \max_{k=1,\dots,n^2} \sup_{(y,b)\in U_k} (|y-u_k| + |b-b_k|) \\
&\qquad + \max_{k=1,\dots,n^2} \sup_{(y,b)\in U_k} \frac{1}{n}\sum_{i=1}^{n} |b^{-1} - b_k^{-1}|\,|K_{b_k}(u_k, Y_{in})| \\
&\quad = O(n^{-1}b(n)^{-2}).
\end{aligned}$$

This also holds true if $Y_{in}$ is replaced by $\tilde{Y}_i$. Therefore the proof is complete. $\square$

**Proof of Theorem 3.1.** (i) *Case* $\mu\notin D$: Obviously,

$$\begin{aligned}
|\hat{f}_n(x) - f(x)| \leqslant{}& \det(\hat{\Sigma}_n)^{-1/2}(|\hat{g}_n(U_n(x)) - g(U_n(x))| + |g(U_n(x)) - g(u(x))|) \\
&+ |g(u(x))|\,|\det(\hat{\Sigma}_n)^{-1/2} - \det(\Sigma)^{-1/2}| \quad \text{for } x\in D,
\end{aligned}$$

where $U_n(x) := (x-\hat{\mu}_n)^T\hat{\Sigma}_n^{-1}(x-\hat{\mu}_n)$, $u(x) = (x-\mu)^T\Sigma^{-1}(x-\mu)$. Choose $\eta>0$ such that there are $M_3, M_4>0$ with $[M_3, M_4]\supset\{(x-\bar{\mu})^T\bar{\Sigma}^{-1}(x-\bar{\mu})\colon x\in D, ||\bar{\mu}-\mu||\leqslant\eta, ||\bar{\Sigma}-\Sigma||\leqslant\eta\}$. Now choose $m, M>0$ such that $\psi([M_3, M_4])\subset[m, M]$. By (2.3), $||\hat{\mu}_n-\mu||\leqslant\eta, ||\hat{\Sigma}_n-\Sigma||\leqslant\eta$ for $n\geqslant n_1(\omega)$. Then we obtain

$$\begin{aligned}
\sup_{x\in D} |\hat{f}_n(x) - f(x)| \leqslant{}& \det(\hat{\Sigma}_n)^{-1/2} \sup_{y\in[M_3,M_4]} |\hat{g}_n(y) - g(y)| \\
&+ \det(\hat{\Sigma}_n)^{-1/2} \sup_{y\in[M_3,M_4]} |g'(y)| \sup_{x\in D} |U_n(x) - u(x)| \\
&+ \sup_{y\in[M_3,M_4]} |g(y)|\,|\det(\hat{\Sigma}_n)^{-1/2} - \det(\Sigma)^{-1/2}| \qquad (4.7)
\end{aligned}$$

for $n\geqslant n_1(\omega)$. An application of Lemmas 4.3–4.5, 4.11 and (4.1) leads to

$$\begin{aligned}
\sup_{y\in[M_3,M_4]} |\hat{g}_n(y) - g(y)| &\leqslant \text{const}\cdot \sup_{x\in[m,M]} \sup_{b\in[\underline{b}_n,\bar{b}_n]} |\hat{h}_n(x) - h(x)| \\
&= O\Big(\sqrt{\ln n}(nb(n))^{-1/2} + b^q(n)\Big) \quad \text{a.s.} \qquad (4.8)
\end{aligned}$$

Using (2.3) we obtain

$$\sup_{x\in D} |U_n(x) - u(x)| = O\Big(n^{-1/2}\sqrt{\ln\ln(n)}\Big) \quad \text{a.s.}, \qquad (4.9)$$

$$|\det(\hat{\Sigma}_n)^{-1/2} - \det(\Sigma)^{-1/2}| = O\Big(n^{-1/2}\sqrt{\ln\ln(n)}\Big) \quad \text{a.s.} \qquad (4.10)$$

In case (i) Theorem 3.1 follows now from (4.7) to (4.10).

(ii) *Case* $\mu \in D$: Here the proof is similar to that of case (i). The main differences are that here $M_3 = 0$ and

$$\sup_{y \in [0, M_4]} |\hat{g}_n(y) - g(y)| = O\Big(\sqrt{\ln n}(nb(n))^{-1/2} + b^{\gamma}(n)\Big) \quad \text{a.s.} \qquad \square$$

Now we proceed with proving asymptotic normality. Suppose that Conditions $\mathscr{K}(1)$ and $\mathscr{T}(2)$ are satisfied. By (4.10),

$$\begin{aligned}
&\sqrt{nB(n)}\Big(\hat{f}_n(x) - f(x)\Big) \\
&\quad = \sqrt{nB(n)}(\det(\hat{\Sigma}_n)^{-1/2}(\hat{g}_n(U_n(x)) - g(u)) + (\det(\hat{\Sigma}_n)^{-1/2} - \det(\Sigma)^{-1/2})g(u)) \\
&\quad = \sqrt{nB(n)}\det(\hat{\Sigma}_n)^{-1/2}(\hat{g}_n(U_n(x)) - g(u)) + o(1) \quad \text{a.s.} \qquad (4.11)
\end{aligned}$$

($U_n = U_n(x)$ and $u = u(x)$ as in the previous proof). Now we have to consider the convergence of $\hat{g}_n(U_n(x)) - g(u)$ and get

$$\begin{aligned}
&\sqrt{nb(n)}(\hat{g}_n(U_n) - g(u)) \\
&\quad = \sqrt{nb(n)}s_d^{-1}U_n^{-d/2+1}\psi'(U_n)(\hat{h}_n(\psi(U_n), B(n)) - h(\psi(u))) + A_n,
\end{aligned}$$

where

$$A_n = \sqrt{nb(n)}s_d^{-1}(U_n^{-d/2+1}\psi'(U_n) - u^{-d/2+1}\psi'(u))h(\psi(u)).$$

Since the derivative of the function $t \mapsto t^{-d/2+1}\psi'(t)$ is bounded on finite intervals $[t_1, t_2]$, $t_1 > 0$, Eq. (4.9) implies

$$|A_n| \leqslant O\Big(\sqrt{nb(n)}\Big)|U_n - u| = o(1) \quad \text{a.s.}$$

Hence

$$\begin{aligned}
&\sqrt{nb(n)}(\hat{g}_n(U_n) - g(u)) \\
&\quad = \Big(\sqrt{nb(n)}s_d^{-1}u^{-d/2+1}\psi'(u) + o_{\mathbb{P}}(1)\Big)(\hat{h}_n(\psi(U_n), B(n)) - h(\psi(u))) \\
&\qquad + o_{\mathbb{P}}(1). \qquad (4.12)
\end{aligned}$$

We have $U_n(x) \in [u/2, 2u]$ for $n \geqslant n_2(\omega)$. Let $m, M > 0$ such that $\psi([u/2, 2u]) \subset [m, M]$. By Lemmas 4.5 and 4.11,

$$\sup_{y \in [m,M]} \sup_{b \in [\underline{b}_n, \bar{b}_n]} |\hat{h}_n(y, b) - \tilde{h}_n(y, b)| = o((nb(n))^{-1/2}) \quad \text{a.s.}$$

and

$$\hat{h}_n(\psi(U_n), B(n)) = \tilde{h}_n(\psi(U_n), B(n)) + o_{\mathbb{P}}((nb(n))^{-1/2}). \qquad (4.13)$$

($n \geqslant n_2(\omega)$). The next step is to prove asymptotic normality of $(\tilde{h}_n(\psi(U_n), B(n)) - h(\psi(u)))$. The following lemma is a classical result by Parzen [12] since $\tilde{h}_n(., b(n))$ is a kernel estimator for the density $h$ of $\tilde{Y}_i$.

**Lemma 4.12.** *Suppose that for some integer $p \geqslant 2$, $h^{(p)}$ exists on $(0, +\infty)$ and is continuous at $y > 0$. Assume that Condition $\mathscr{K}(p)$ is fulfilled and $b(n) = C_{14} n^{1/(2p+1)}$ with a constant $C_{14} > 0$. Then*

$$\sqrt{nb(n)}(\tilde{h}_n(y, b(n)) - h(y)) \xrightarrow{\mathscr{D}} \mathscr{N}(\mu_1, \sigma_1^2),$$

$$\mu_1 = C_{14}^{(2p+1)/2} \frac{1}{p!} h^{(p)}(y) \int_{-1}^{1} t^p K(t)\, dt, \quad \sigma_1^2 = h(y) \int_{-1}^{1} K^2(t)\, dt.$$

**Lemma 4.13.** *Assume that $h$ is Lipschitz continuous in some neighbourhood of $\psi(u)$. Then*

$$|\tilde{h}_n(\psi(U_n), b(n)) - \tilde{h}_n(\psi(u), b(n))| = o_{\mathbb{P}}((nb(n))^{-1/2}).$$

**Proof.** Let $n_3$ be such that $2b(n) \leqslant \psi(u)$ for $n \geqslant n_3$. Using the Lipschitz continuity of $K'$,

$$\begin{aligned}
&|\tilde{h}_n(\psi(U_n), b(n)) - \tilde{h}_n(\psi(u), b(n))| \\
&\quad \leqslant n^{-1} b(n)^{-2} \left| \sum_{i=1}^{n} K'((\psi(u) - \tilde{Y}_i) b(n)^{-1}) \right| |\psi(U_n) - \psi(u)| \\
&\qquad + O(n^{-1} b(n)^{-3}) \sum_{i=1}^{n} I(|\psi(u) - \tilde{Y}_i| \leqslant 2b(n)) (\psi(U_n) - \psi(u))^2 \\
&\qquad + O(b(n)^{-1}) I(|\psi(U_n) - \psi(u)| > b(n)) \\
&\quad = A_{n1} + A_{n2} + A_{n3}, \quad \text{say}
\end{aligned} \tag{4.14}$$

$(n \geqslant n_3)$. By (4.9),

$$|\psi(U_n) - \psi(u)| = O((\ln \ln n/n)^{1/2}) \quad \text{a.s.} \tag{4.15}$$

Analogously to Lemma 4.6, one proves

$$\begin{aligned}
&\left| \sum_{i=1}^{n} (K'((\psi(u) - \tilde{Y}_i) b(n)^{-1}) - \mathbb{E}K'((\psi(u) - \tilde{Y}_i) b(n)^{-1})) \right| \\
&\quad = O\left(\sqrt{nb(n) \ln(n)}\right) \quad \text{a.s.}
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathbb{E}K'((\psi(u) - \tilde{Y}_i) b(n)^{-1}) \\
&\quad = b(n) \int_{-1}^{1} K'(t)(h(\psi(u) - tb(n)) - h(\psi(u)))\, dt = O(b^2(n)).
\end{aligned}$$

Hence

$$A_{n1} = O\left(n^{-3/2}b(n)^{-2}\sqrt{\ln\ln n}\right)\left(nb^2(n) + \sqrt{nb(n)\ln(n)}\right)$$
$$= o(n^{-1/2}b(n)^{-1/2}) \quad \text{a.s.} \tag{4.16}$$

The consistency of density estimators (cf. [12]) and (4.15) lead to

$$A_{n2} = O(n^{-2}b(n)^{-3}\ln\ln n)\sum_{i=1}^{n} I(|\psi(u) - \tilde{Y}_i| \leqslant 2b(n)) \quad \text{a.s.}$$
$$= o_{\mathbb{P}}(n^{-1/2}b(n)^{-1/2}). \tag{4.17}$$

Further by (4.15), we obtain

$$A_{n3} \leqslant O(b(n)^{-5})(\psi(U_n) - \psi(u))^4 = o(n^{-1/2}) \quad \text{a.s.} \tag{4.18}$$

Therefore, the assertion of Lemma 4.13 follows from (4.14) and (4.16) to (4.18). $\square$

**Lemma 4.14.** *Assume that for some integer* $p \geqslant 2$, $h^{(p)}$ *exists on* $(0, +\infty)$ *and is continuous at* $\psi(u)$. *Let Condition* $\mathscr{K}(p)$ *and* (3.6) *be fulfilled and* $b(n) = C_{14}n^{1/(2p+1)}$ *with a constant* $C_{14} > 0$. *Then*

$$\sqrt{nb(n)}(\tilde{h}_n(\psi(U_n), B(n)) - h(\psi(u))) \xrightarrow{\mathscr{D}} \mathscr{N}(\mu_2, \sigma_2^2),$$
$$\mu_2 = C_{14}^{(2p+1)/2}\frac{1}{p!}h^{(p)}(\psi(u))\int_{-1}^{1} t^p K(t)\,dt, \quad \sigma_2^2 = h(\psi(u))\int_{-1}^{1} K^2(t)\,dt.$$

**Proof.** Observe that

$$\tilde{h}_n(\psi(U_n(x)), B(n)) - h(\psi(u))$$
$$= \frac{b(n)}{B(n)}(\breve{h}_n(\psi(U_n(x)), B(n)) - h(\psi(u))) + h(\psi(u))\left(\frac{b(n)}{B(n)} - 1\right), \tag{4.19}$$

where

$$\breve{h}_n(y, b) = n^{-1}b(n)^{-1}\sum_{k=1}^{n} K_b(y, \tilde{Y}_i).$$

Further

$$|K(v/b_1) - K(v/b_2)| \leqslant C_{15}\left|1 - \frac{b_1}{b_2}\right|\mathbf{1}_{[-b_1,b_1]}(v) \quad \text{for } v \in \mathbb{R}$$

with a constant $C_{15}>0$ provided that $b_1 \geqslant b_2 > 0$. Let $n_4(\omega)$ such that $B(n) \leqslant 2b(n)$ for $n \geqslant n_4(\omega)$. Hence, by consistency of density estimators,

$$\begin{aligned}
&\sqrt{nb(n)}|\breve{h}_n(\psi(U_n), B(n)) - \tilde{h}_n(\psi(U_n), b(n))| \\
&\quad \leqslant 2\sqrt{n}|B(n)b(n)^{-1} - 1| \max\{1, b(n)B(n)^{-1}\} \\
&\qquad \times\ n^{-1} b(n)^{-1/2} \sum_{k=1}^{n} (I(|\tilde{Y}_i - \psi(U_n)| \leqslant 2b(n)) + I(|\tilde{Y}_i + \psi(U_n)| \leqslant 2b(n))) \\
&\quad \leqslant o_{\mathbb{P}}\left(\sqrt{\ln \ln n}\, b(n)^{-1/2}\right) \\
&\qquad \times \left( n^{-1} \sum_{k=1}^{n} (I(|\tilde{Y}_i - \psi(u)| \leqslant 3b(n)) + I(|\tilde{Y}_i + \psi(u)| \leqslant 3b(n))) \right. \\
&\qquad \left. + I(|\psi(u) - \psi(U_n)| > b(n)) \right) \\
&\quad = o_{\mathbb{P}}\left(\sqrt{\ln \ln n\, b(n)}\right) = o_{\mathbb{P}}(1)
\end{aligned} \tag{4.20}$$

for $n \geqslant n_4(\omega)$ in view of (4.18). Combining Lemma 4.13, (4.19) and (4.20),

$$\sqrt{nb(n)}(\tilde{h}_n(\psi(U_n), B(n)) - h(\psi(u))) = \frac{b(n)}{B(n)} (\tilde{h}_n(\psi(u), b(n)) - h(\psi(u))) + o_{\mathbb{P}}(1).$$

Now apply Lemma 4.12 to get Lemma 4.14. □

**Proof of Theorem 3.2.** In view of (4.12), (4.13) and Lemma 4.14, we have

$$\sqrt{nb(n)}(\hat{g}_n(U_n(x)) - g(u)) \xrightarrow{\mathscr{D}} \mathscr{N}(\mu_3, \sigma_3^2),$$

where

$$\mu_3 = s_d^{-1} u^{-d/2+1} \psi'(u) \mu_2, \quad \sigma_3^2 = s_d^{-2} (u^{-d/2+1} \psi'(u))^2 \sigma_2^2.$$

By virtue of (4.11), the proof of Theorem 3.2 is complete. □

## Acknowledgments

The author is grateful to Dr. Hans Crauel for helpful hints.

## References

[1] T.W. Anderson, Nonnormal multivariate distributions: inference based on elliptically contoured distributions, in: C.R. Rao (Ed.), Multivariate analysis: future directions, North-Holland Series in Statistics and Probability, Vol. 5, North-Holland, Amsterdam, 1993, pp. 1–24.

[2] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, J.A. Wellner, Efficient and Adaptive Estimation for Semiparametric Models, Springer, Berlin, 1998.

[3] M. Bilodeau, D. Brenner, Theory of Multivariate Statistics, Springer, New York, 1999.
[4] A. Cowling, P. Hall, On pseudodata methods for removing boundary effects in kernel density estimation, J. R. Statist. Soc. Ser. B 58 (1996) 551–563.
[5] H. Cui, X. He, The consistence of semiparametric estimation of elliptic densities, Acta Math. Sin. New Ser. 11 (special issue) (1995) 44–58.
[6] H. El Barmi, J.S. Simonoff, Transformation-based density estimation for weighted distributions, J. Nonparametric Stat. 12 (2000) 861–878.
[7] K.-T. Fang, S. Kotz, K. Ng, Symmetric Multivariate and Related Distributions, Chapman & Hall, London, 1990.
[8] K.-T. Fang, Y. Zhang, Generalized Multivariate Analysis, Springer, Berlin, 1990.
[9] M.C. Jones, J.S. Marron, S.J. Sheather, Progress in data-based bandwidth selection for kernel density estimation, Comput. Statist. 11 (1996) 337–381.
[10] E. Liebscher, On a class of plug-in methods of bandwidth selection for kernel density estimators, Statist. Decisions 16 (1998) 229–243.
[11] R. Maronna, Robust M-estimators of multivariate location and scatter, Ann. Statist. 4 (1976) 51–67.
[12] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Statist. 33 (1962) 1065–1076.
[13] D. Pollard, Convergence of Stochastic Processes, Springer, Berlin, 1984.
[14] D.W. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, Wiley, New York, 1992.
[15] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, 1986.
[16] W. Stute, A law of the logarithm for kernel density estimators, Ann. Probab. 10 (1982) 414–422.
[17] W. Stute, U. Werner, Nonparametric estimation of elliptically contoured densities, in: G. Roussas (Ed.), Nonparametric Functional Estimation and Related Topics, NATO ASI Ser., Ser. C 335, Kluwer Academic, Dordrecht, 1991, pp. 173–190.
[18] S. Zhang, R.J. Karunamuni, M.C. Jones, An improved estimator of the density function at the boundary, J. Am. Stat. Assoc. 94 (1999) 1231–1241.