

# Central limit theorems for the integrated squared error of derivative estimators

Melanie Birke

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: melanie.birke@ruhr-uni-bochum.de

December 22, 2006

## Abstract

A central limit theorem for the weighted integrated squared error of kernel type estimators of the first two derivatives of a nonparametric regression function is proved by using results for martingale differences and U-statistics. The results focus on the setting of the Nadaraya-Watson estimator but can also be transferred to local polynomial estimates.

*Key words:* central limit theorem, integrated squared error, kernel estimates, local polynomial estimate, Nadaraya-Watson estimate, nonparametric regression

## 1 Introduction

The asymptotic distribution of integrated squared errors has been analysed for several kinds of kernel estimators. For example Bickel and Rosenblatt (1973) consider the integrated squared error of a Rosenblatt-Parzen estimator for the density of an i.i.d. sample  $X_1, \dots, X_n$  while Hall (1984a) studies the weighted integrated squared error of multivariate kernel density estimates. The methods used in those settings can generally be transferred to kernel estimators of regression functions. Konakov (1978) analyses the asymptotic distribution of a weighted integrated squared error for a regression estimate of the type

$$\int (\hat{m}(x) - m(x))^2 \hat{f}^2(x) w(x) dx,$$

where  $\hat{m}$  is the Nadaraya-Watson estimate and  $\hat{f}$  denotes the kernel density estimate. This is in fact the integrated squared error of the numerator of the Nadaraya-Watson estimate. Some central

limit theorems for integrated squared errors of multivariate kernel regression estimates of the type

$$\int_A (\hat{m}(x) - m(x))^2 v_n(x) dx$$

as well as

$$\int_A (\hat{m}(x) - m(x))^2 w(x) dx$$

are given in Hall (1984b) where  $v_n(x)$  is a random weight function and  $w(x)$  is deterministic. Further consideration of the limiting distribution of the integrated squared error for a random design is given in Nadaraya (1989). The case of a fixed design is mentioned in Ioannides (1992) and generally follows the structure of Hall (1984b). Recently Liero (1992) develops asymptotic theory for the weighted integrated squared errors of regression estimates with data-dependent bandwidths. In the present paper the asymptotic normality of the weighted integrated squared error of kernel estimates of derivatives of regression functions is shown which has not been considered yet. This result has for example an application in testing strict monotonicity of a regression function in a nonparametric regression model (see Birke and Dette, 2006).

## 2 Central limit theorem for the integrated squared error

Let  $A$  be an interval (also  $A = \mathbb{R}$  is possible) and assume that  $X_1, \dots, X_n$  are i.i.d. on the set  $A^\delta = \{x \in \mathbb{R} | \inf_{a \in A} |x - a| \leq \delta\}$ ,  $\delta > 0$  with density  $f$ . The regression model is given by

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i.$$

where  $m$  and  $f$  are four and three times continuously differentiable on  $A$ , respectively and  $\sigma$  is bounded on  $A^\delta$ . The random variables  $\varepsilon_i$ ,  $i = 1, \dots, n$  are i.i.d. and independent of the sample  $X_1, \dots, X_n$  and have finite moments of order four, especially

$$\int_{A^\delta} \mu_4(x) f(x) dx < \infty \tag{1}$$

with  $\mu_4(x) = E[(Y_i - m(X_i))^4 | X_i]$ . Further assume that  $K$  is a kernel of order 2 with compact support, say  $[-\lambda, \lambda]$  which is two times continuously differentiable and  $K(\pm\lambda) = K'(\pm\lambda) = 0$ . This for example holds for the biweight kernel  $K(x) = 15/16 (1 - x^2)^2 I_{[-1,1]}(x)$ . A common estimator of the  $k$ -th derivative of the regression function is given by

$$\hat{m}^{(k)}(x) = \frac{\partial^k}{\partial x^k} \hat{m}(x)$$

where

$$\hat{m}(x) = \frac{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}$$

denotes the Nadaraya-Watson estimate on the set  $A$  with a two times continuously differentiable kernel  $K$  of second order and a bandwidth  $h$ .

It is shown in the following that the expression

$$T^{(k)} = \int_A (\hat{m}^{(k)}(x) - m^{(k)}(x))^2 w(x) dx$$

is asymptotically normal for any bounded, continuous and positive weight function  $w$ . This statement is easier to see if the following representation is used. With

$$S_k(x) = \left[ \frac{1}{nh_r^{k+1}} \sum_{i=1}^n K_r^{(k)}\left(\frac{x - X_i}{h_r}\right) (Y_i - m(X_i)) + \hat{g}^{(k)}(x) \right] / \hat{f}(x)$$

the difference  $\hat{m}^{(k)}(x) - m^{(k)}(x)$  is for  $k = 0, 1, 2$  a linear combination of  $S_j(x)$ ,  $j = 0, \dots, k$ ,

$$\hat{m}^{(k)}(x) - m^{(k)}(x) = \sum_{l=0}^k a_{k,l}(x) S_l(x),$$

and the integrated squared error is given by

$$T_n^{(k)} = \sum_{l=0}^k \int_A a_{k,l}^2(x) w(x) S_l^2(x) dx + 2 \sum_{l < m} \int_A a_{k,l}(x) a_{k,m}(x) w(x) S_l(x) S_m(x) dx \quad (2)$$

with factors

$$\begin{aligned} a_{k,k}(x) &= 1, \quad k = 0, 1, 2, \\ a_{1,0}(x) &= -\hat{f}'(x)/\hat{f}(x), \\ a_{2,1}(x) &= -2\hat{f}'(x)/\hat{f}(x), \\ a_{2,0}(x) &= -(\hat{f}'(x)/\hat{f}(x))^2 + \hat{f}''(x)/\hat{f}(x) \end{aligned}$$

which converge in probability. In a first step the asymptotic normality of

$$\int_A w(x) S_k^2(x) dx$$

can be shown which is stated in the following theorem.

**Theorem 1** *If the conditions stated above are satisfied and the bandwidth fulfills  $h \rightarrow 0$ ,  $nh^{3/2+k} \rightarrow \infty$  and  $nh^{2k+5} = O(1)$  for  $n \rightarrow \infty$  we have for  $k = 0$*

$$(n^{-2}h^{-1}\alpha_{1,0} + n^{-1}h^4\alpha_{2,0})^{-1/2} \left( \int_A S_0^2(x) w(x) dx - B_{n,0}^{[1]} - B_{n,0}^{[2]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

and for  $k = 1, 2$

$$nh^{2k+1/2} \alpha_{1,k}^{-1/2} \left( \int_A S_k^2(x) w(x) dx - B_{n,k}^{[1]} - B_{n,k}^{[2]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

with

$$B_{n,k}^{[1]} = \frac{1}{nh^{2k+2}} \int_A \int_{-1}^1 \mathbb{E} \left[ K^{(k)2} \left( \frac{x - X_1}{h} \right) \sigma^2(X_1) \right] (f(x))^{-2} w(x) dx$$

$$\begin{aligned}
B_{n,k}^{[2]} &= h^4 \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (f(x))^{-2} w(x) dx \\
\alpha_{1,k} &= 2 \left( \int_A \sigma^4(x) w^2(x) f^{-2}(x) dx \right) \left( \int \left( \int K^{(k)}(x) K^{(k)}(x+y) dx \right)^2 dy \right) \\
\alpha_{2,0} &= 4 \int_A \sigma^2(x) \gamma_0^2(x) w^2(x) f^{-4}(x) dx \\
\gamma_k(x) &= \frac{1}{2} \left( \int_{-\lambda}^{\lambda} y^2 K(y) dy \right) \left( m^{(k+2)}(x) f(x) + 2m^{(1)}(x) f^{(k+1)}(x) \right. \\
&\quad \left. + \sum_{j=0}^{k-1} \binom{k}{j+1} \frac{k+2+j}{k-j} m^{(k+2-j)}(x) f^{(j)}(x) \right)
\end{aligned}$$

The below-mentioned corollary follows from Theorem 1.

**Corollary 1** *Assume that the conditions of Theorem 1 are fulfilled and that  $h = O(n^{-1/5})$ . Then there is for  $0 \leq l < k$*

$$nh^{2k+1/2} \int_0^1 S_l^2(x) w(x) dx = O_P(h_r^{2(k-l)-1/2}).$$

As a consequence of Theorem 1 and Corollary 1 the asymptotic normality of  $T^{(k)}$  is obtained from (2).

**Theorem 2** *Under the above conditions, if the bandwidth fulfills  $h \rightarrow 0$ ,  $nh \rightarrow \infty$  and  $nh^{3/2+k} \rightarrow \infty$  for  $n \rightarrow \infty$ , there is for  $k = 0$*

$$(n^{-2}h^{-1}\alpha_{1,0} + n^{-1}h^4\alpha_{2,0})^{-1/2} \left( \int_A (\hat{m}(x) - m(x))^2 w(x) dx - B_{n,0}^{[1]} - B_{n,0}^{[2]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

and for  $k \geq 1$  if  $h = O(n^{-1/5})$

$$nh^{2k+1/2} \alpha_{1,k}^{-1/2} \left( \int_A (\hat{m}^{(k)}(x) - m^{(k)}(x))^2 w(x) dx - B_{n,k}^{[1]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

If the variance function  $\sigma^2$  is not only continuous and bounded but once continuously differentiable  $B_{n,k}^{[1]}$  has the representation

$$\begin{aligned}
B_{n,k}^{[1]} &= \frac{1}{nh^{2k+2}} \int_0^1 \int_0^1 \sigma^2(y) f(y) K^{(k)2} \left( \frac{x-y}{h} \right) dy (f(x))^{-2} w(x) dx \\
&= \frac{1}{nh^{2k+1}} \int_0^1 \frac{\sigma^2(x) w(x)}{f(x)} dx \int_{-1}^1 K^{(k)2}(y) dy \\
&\quad + \frac{1}{nh^{2k}} \int_{-1}^1 \int_0^1 y K^{(k)2}(y) \frac{\sigma^{2'}(\xi_y) f(\xi_y) + \sigma^2(\xi_y) f'(\xi_y)}{(f(x))^2} w(x) dx dy \\
&= \frac{1}{nh^{2k+1}} \int_0^1 \frac{\sigma^2(x)}{f(x) m^{(6)}(x)} dx \int_{-1}^1 K^{(k)2}(y) dy + o\left(\frac{1}{nh^{2k+1/2}}\right).
\end{aligned}$$

**Corollary 2** *Under the assumptions of Theorem 2 if  $\sigma^2$  is once continuously differentiable and  $k \in \{1, 2\}$  we have*

$$nh^{2k+1/2}\alpha_{1,k}^{-1/2}\left(\int_A S_k^2(x)w(x)dx - \tilde{B}_{n,k}^{[1]}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

where

$$\tilde{B}_{n,k}^{[1]} = \frac{1}{nh^{2k+1}} \int_A \frac{\sigma^2(x)w(x)}{f(x)} dx \int_{-1}^1 K^{(k)^2}(y) dy$$

The Nadaraya-Watson estimator  $\hat{m}$  can also be replaced by other differentiable estimators of the regression function, for example local polynomial estimates. A local polynomial estimate has the representation

$$\hat{m}_p(x) = \frac{1}{nh f(x)} \sum_{i=1}^n K^*\left(\frac{x - X_i}{h}\right) Y_i (1 + o_P(1))$$

with  $K^*$  denoting the corresponding equivalent kernel (see Fan and Gijbels, 1997). Then an analogous assertion as in Theorem 1 holds for the local polynomial estimate, where  $K$  has to be replaced by its equivalent kernel  $K^*$ .

**Acknowledgements.** The author would like to thank Holger Dette and Natalie Neumeyer for helpful discussions. Her work was supported by the Sonderforschungsbereich 475, Komplexitätsreduktion in multivariaten Datenstrukturen.

## References

- M. Birke (2007). Schätz- und Testverfahren in der nichtparametrischen Regression unter qualitativen Annahmen. PhD thesis, Ruhr-Universität Bochum (in German).
- M. Birke and H. Dette (2006). Testing strict monotonicity in nonparametric regression. Preprint, <http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm>.
- J. Fan and I. Gijbels (1996). Local polynomial modelling and its applications. Chapman and Hall, London.
- P. Hall (1984a). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* 14, 1-16.
- P. Hall (1984b). Integrated square error properties of kernel estimators of regression functions. *Ann. Statist.* 12, 241-260.
- D.A. Ioannides (1992). Integrated square error of nonparametric estimators of regression function: The fixed design case. *Stat. Probab. Lett.* 15, 85-94.
- V.D. Konakov (1978). On a global measure of deviation for an estimate of the regression line. *Theory Probab. Appl.* 22, 858-868.

- H. Liero (1992). Asymptotic normality of a weighted integrated squared error of kernel regression estimates with data-dependent bandwidth. *J. Statist. Plann. Inference* 30, 307-325.
- E.A. Nadaraya (1989). Nonparametric estimation of probability densities and regression curves. Kluwer, Dordrecht.

## Appendix: Proofs

During this section let  $E_{\mathcal{X}}[Z]$  and  $\text{Var}_{\mathcal{X}}(Z)$  denote the expectation and variance conditioned on the sample  $X_1, \dots, X_n$  of the random variable  $Z$ , respectively.

### Proof of Theorem 1

The proof of this theorem is adapted to the proof of Hall (1984b), who considered central limit theorems for the integrated squared error of regression estimators. In a first step an analogous result to Theorem 1 with a stochastic instead of a deterministic weight function is proved. Therefore assume, that  $\mathcal{A}_{n,0}$  is the  $\sigma$ -algebra generated by the sample  $X_1, \dots, X_n$ . Let  $v_n$  be a stochastic function which is measurable with respect to  $\mathcal{A}_{n,0}$  such that it converges in probability to a bounded, nonnegative and deterministic function  $v$ . The statistic

$$\tilde{T}_n^{(k)} = \int_A S_k^2(x) \hat{f}^2(x) v_n(x) dx$$

has the decomposition

$$\tilde{T}_n^{(k)} = I_{n1} + 2I_{n2} + 2I_{n3} + I_{n4}$$

with

$$I_{n1} = \frac{1}{n^2 h^{2k+2}} \sum_{i=1}^n (Y_i - m(X_i))^2 \int_A K^{(k)^2} \left( \frac{x - X_i}{h} \right) v_n(x) dx \quad (3)$$

$$I_{n2} = \frac{1}{n^2 h^{2k+2}} \sum_{i < j} (Y_i - m(X_i))(Y_j - m(X_j)) \int_A K^{(k)} \left( \frac{x - X_i}{h} \right) K^{(k)} \left( \frac{x - X_j}{h} \right) v_n(x) dx \quad (4)$$

$$I_{n3} = \frac{1}{n h^{k+1}} \sum_{i=1}^n (Y_i - m(X_i)) \int_A K^{(k)} \left( \frac{x - X_i}{h} \right) \hat{g}^{(k)}(x) v_n(x) dx \quad (5)$$

$$I_{n4} = \int_A (\hat{g}^{(k)}(x))^2 v_n(x) dx \quad (6)$$

and the following theorem holds.

**Theorem 3** *Under the assumptions of Theorem 1, if  $v_n$  is a stochastic weight function which converges in probability to  $v$ , there is for  $k = 0$*

$$\int_A S_0^2(x) \hat{f}^2(x) v_n(x) dx = E_{\mathcal{X}}[I_{n1}] + I_{n4} + (n^{-2} h^{-1} \alpha_{1,0} + h^4 n^{-1} \alpha_{2,0})^{1/2} Z_n$$

and for  $k = 1$  and  $2$

$$\int_A S_k^2(x) \hat{f}^2(x) v_n(x) dx = E_{\mathcal{X}}[I_{n1}] + I_{n4} + (n^{-2} h^{-1-4k} \alpha_{1,k})^{1/2} Z_n$$

with an asymptotically standard normal random variable  $Z_n$  and with constants  $\alpha_{1,k}$  and  $\alpha_{2,0}$  defined in Theorem 1.

The proof of this theorem is the major part of this paragraph. Theorem 1 then follows in the broadest sense from Theorem 3 by choosing  $v_n(x) = w(x)/\hat{f}(x)$ .

**Proof of Theorem 3.** For  $k = 0$  the assertion of Theorem 3 is the same as of Theorem 1 in Hall (1984b) for  $p = 0$ . Therefore the proof is only stated for  $k = 1, 2$  and is organized in three parts by handling the expressions  $I_{n1}$ ,  $I_{n2}$  and  $I_{n3}$  separately.

(i) It follows by similar arguments as in the proof of Theorem 1 in Hall (1984b), if  $(K^{(k)})^2$  instead of  $K^2$  is used, that

$$n^4 h^{4k+4} E_{\mathcal{X}}[(I_{n1} - E_{\mathcal{X}}[I_{n1}])^2] = O_P(nh^2)$$

and therefore

$$E_{\mathcal{X}}[(I_{n1} - E_{\mathcal{X}}[I_{n1}])^2] = O_P\left(\frac{1}{n^3 h^{4k+2}}\right).$$

Applying the markov inequality we get

$$P\left(|I_{n1} - E_{\mathcal{X}}[I_{n1}]| > \frac{\lambda_n}{n^{3/2} h^{2k+1}} \middle| X_1, \dots, X_n\right) \leq \frac{1}{\lambda_n} O_P(1) \xrightarrow{n \rightarrow \infty} 0$$

which results in

$$I_{n1} = E_{\mathcal{X}}[I_{n1}] + o_P\left(\frac{1}{n^{3/2} h^{2k+1}}\right).$$

(ii) To show that  $I_{n2}$  has a contribution to the asymptotic distribution of  $T_n^{(k)}$  the quantities  $W_{nij}$  and  $\tilde{W}_{nij}$  are defined as

$$\begin{aligned} W_{nij} &= \int_A K^{(k)}\left(\frac{x - X_i}{h}\right) K^{(k)}\left(\frac{x - X_j}{h}\right) v_n(x) dx \\ \tilde{W}_{nij} &= \int_A K^{(k)}\left(\frac{x - X_i}{h}\right) K^{(k)}\left(\frac{x - X_j}{h}\right) v(x) dx. \end{aligned}$$

From

$$Y_{nj} = (Y_j - m(X_j)) \sum_{i=1}^{j-1} (Y_i - m(X_i)) W_{nij}$$

it follows that  $n^2 h^{2k+2} I_{n2} = \sum_{j=2}^n Y_{nj}$  and  $E[Y_{nj} | \mathcal{A}_{n,j-1}] = 0$  where  $\mathcal{A}_{n,j}$  denotes the  $\sigma$ -algebra generated by  $X_1, \dots, X_n$  and  $Y_1, \dots, Y_i$ . This characterizes the sequence  $\{(S_{ni} = \sum_{j=2}^i Y_{nj}, \mathcal{A}_{n,i}), 2 \leq i \leq n < \infty\}$  as a martingale difference array. It can be shown that the conditional variance

$$V_n^2 = \sum_{j=2}^n \sigma^2(X_j) \sum_{i=1}^{j-1} (Y_i - m(X_i))^2 \tilde{W}_{nij}^2 (1 + o_P(1))$$

converges to  $\alpha_{1,k}/4$  where the quantity  $\alpha_{1,k}$  is defined in Theorem 1. In addition, the Lindeberg condition

$$n^{-1}h^{-3} \sum_{i=2}^n \mathbb{E}_{\mathcal{X}}[Y_{ni}^2 I\{|Y_{ni}| > \varepsilon nh^{3/2}\}] \xrightarrow{P} 0$$

is fulfilled. This can be shown as in Hall (1984b) by considering that  $K^{(k)}$  need not be positive for  $k > 0$  and therefore using  $|K^{(k)}|$  in the estimation. Now a central limit theorem for martingale difference arrays yields that

$$n^{-1}h^{-3/2}S_{nn} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4}\alpha_{1,k}\right)$$

and therefore  $2I_{n2}$  has the same asymptotic distribution as  $n^{-1}h^{-2k-1/2}\alpha_{1,k}^{1/2}Z_n$ , where  $Z_n$  is an asymptotically standard normal distributed random variable. Furthermore  $I_{n2}$  is asymptotically independent of each sequence of events  $A_n \in \mathcal{A}_{n,0}$  (see Hall, 1984b).

(iii) The third part of the proof for  $k > 0$  differs from that in Hall (1984b). While it is shown there that  $I_{n3}$  contributes to the asymptotic distribution of  $T_n^{(0)}$ , it is proved in this part that for higher order derivatives of  $\hat{m}$  the quantity  $I_{n3}$  is asymptotically negligible.

The representation

$$I_{n3} = J_{n1} + J_{n2} \tag{7}$$

with

$$\begin{aligned} J_{n1} &= \frac{1}{nh^{k+1}} \sum_{i=1}^n (Y_i - m(X_i)) \int_A K^{(k)}\left(\frac{x - X_i}{h}\right) \gamma_{n,k}(x) v_n(x) dx \\ J_{n2} &= \frac{1}{nh^{k+1}} \sum_{i=1}^n (Y_i - m(X_i)) \int_A K^{(k)}\left(\frac{x - X_i}{h}\right) (\hat{g}^{(k)}(x) - \gamma_{n,k}(x)) v_n(x) dx \end{aligned}$$

and  $\gamma_{n,k}(x) = \mathbb{E}[\hat{g}^{(k)}(x)]$  simplifies the discussion of  $I_{n3}$ . Define

$$Y_{ni} = (Y_i - m(X_i)) \tilde{Z}_{ni}$$

with

$$\tilde{Z}_{ni} = \int_A K^{(k)}\left(\frac{x - X_i}{h}\right) \gamma_{n,k}(x) v_n(x) dx$$

and observe that

$$J_{n1} = \frac{1}{nh^{k+1}} \sum_{i=1}^n Y_{ni}.$$

**Lemma 1** *For  $k > 0$  there is*

$$\sum_{i=1}^n Y_{ni} = o_P(n^{1/2}h^3).$$



**Proof.** It can be shown, that  $\gamma_{n,k}(x) = h^2\gamma_k(x) + o(h^2)$ . Substituting this into  $\tilde{Z}_{ni}$  yields

$$\begin{aligned} \mathbb{E}_{\mathcal{X}} \left[ \left( \sum_{i=1}^n Y_{ni} \right)^2 \right] &= \sum_{i=1}^n \sigma^2(X_i) \tilde{Z}_{ni} \\ &= h^4 \sum_{i=1}^n \sigma^2(X_i) \left( \int_A K^{(k)} \left( \frac{x - X_i}{h} \right) \gamma_k(x) v(x) dx \right)^2 + o_P(nh^6) \\ &= h^4 M_n + o_P(nh^6). \end{aligned}$$

The expectation of  $M_n$  is given by

$$\begin{aligned} \mathbb{E}[M_n] &= nh^2 \int_A \int_{-\lambda}^{\lambda} \sigma^2(x - hu) f(x - hu) K^{(k)}(u) \\ &\quad \times \int_{-\lambda}^{\lambda} K^{(k)}(v) \gamma_k(x + h(v - u)) v(x + h(v - u)) dv du dx \\ &= o(nh^2) \end{aligned}$$

and the variance equals

$$\begin{aligned} \text{Var}(M_n) &= n \text{Var} \left( \sigma^2(X_1) \left( K^{(k)} \left( \frac{x - X_1}{h} \right) \gamma_k(x) v(x) \right)^2 \right) \\ &\leq n C_1 \int_{A^\varepsilon} \left( \int_A K^{(k)} \left( \frac{x - y}{h} \right) dx \right)^4 f(y) dy \\ &= nh^4 C \int_{A^\varepsilon} \left( \int_{-\lambda}^{\lambda} K^{(k)}(x) dx \right)^4 dy = O(nh^4) = o(n^2 h^4). \end{aligned}$$

This results in

$$\mathbb{E}_{\mathcal{X}} \left[ \left( \sum_{i=1}^n Y_{ni} \right)^2 \right] = h^4 \mathbb{E}[M_n] + o_P(nh^6) = o_P(nh^6),$$

and therefore

$$\sum_{i=1}^n Y_{ni} = o_P(n^{1/2} h^3),$$

which means

$$J_{n1} = \frac{1}{nh^{k+1}} o_P(n^{1/2} h^3) = o_P \left( \frac{1}{nh^{2k+1/2}} \right). \quad (8)$$

Similar methods as in Hall (1984b) (in principle using  $|K^{(k)}|$  instead of  $K^{(k)}$ , see Birke (2007)) provide

$$J_{n2} = o_P \left( \frac{1}{nh^{2k+1/2}} \right). \quad (9)$$

Equations (7), (8) and (9) now yield  $I_{n3} = o_P(1/nh^{2k+1/2})$  which completes the proof of Theorem 3.  $\square$

We still have to find a representation for  $\mathbb{E}_{\mathcal{X}}[I_{n1}]$  and  $I_{n4}$  so that Theorem 1 follows from Theorem 3. This is done in the following part.

**Lemma 2** *If the assumptions of Theorem 3 are fulfilled and the weight function is given by  $v_n(x) = w(x)/(\hat{f}(x))^2$ , then*

$$\mathbb{E}\mathcal{X}[I_{n1}] = \frac{1}{n^2 h^{2k+2}} \sum_{i=1}^n \int_A \mathbb{E} \left[ \sigma^2(X_i) K^{(k)^2} \left( \frac{x - X_i}{h} \right) \right] (f(x))^{-2} w(x) dx + o_P(n^{-1} h^{-2k-1/2})$$

and

$$I_{n4} = \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (f(x))^{-2} w(x) dx + o_P(n^{-1} h^{-2k-1/2}).$$

**Proof.** The proof uses the following Taylor expansions at several places,

$$(\hat{f}(x))^{-2} = (\mathbb{E}[\hat{f}(x)])^{-2} + O_P(1) |\hat{f}(x) - \mathbb{E}[\hat{f}(x)]| \quad (10)$$

$$(\hat{f}(x))^{-2} = (\mathbb{E}[\hat{f}(x)])^{-2} - 2 \frac{(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])}{(\mathbb{E}[\hat{f}(x)])^3} + O_P(1) (\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2 \quad (11)$$

$$(\mathbb{E}[\hat{f}(x)])^{-2} = (f(x))^{-2} + O_P(1) |\mathbb{E}[\hat{f}(x)] - f(x)|. \quad (12)$$

With (11) the conditional expectation of  $I_{n1}$  can be written as

$$\begin{aligned} \mathbb{E}\mathcal{X}[I_{n1}] &= \frac{1}{n^2 h^{2k+2}} \sum_{i=1}^n \sigma^2(X_i) \int_A K^{(k)^2} \left( \frac{x - X_i}{h} \right) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx \\ &\quad - \frac{2}{n^2 h^{2k+2}} \sum_{i=1}^n \sigma^2(X_i) \int_A K^{(k)^2} \left( \frac{x - X_i}{h} \right) (\mathbb{E}[\hat{f}(x)])^{-3} (\hat{f}(x) - \mathbb{E}[\hat{f}(x)]) w(x) dx \\ &\quad + O_P \left( \frac{1}{n^2 h^{2k+2}} \right) \sum_{i=1}^n \sigma^2(X_i) \int_A K^{(k)^2} \left( \frac{x - X_i}{h} \right) (\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2 w(x) dx \\ &= B_1 - 2B_2 + B_3. \end{aligned}$$

The variance of  $B_1$  is of order  $o(1/n^2 h^{4k+1})$ . Therefore a straight forward calculation using (12) gives for the dominating term  $B_1$

$$\begin{aligned} B_1 &= \mathbb{E}[B_1] + o_P(n^{-1} h^{-2k-1/2}) \\ &= \frac{1}{n^2 h^{2k+2}} \sum_{i=1}^n \int_A \mathbb{E} \left[ \sigma^2(X_i) K^{(k)^2} \left( \frac{x - X_i}{h} \right) \right] (f(x))^{-2} w(x) dx + o_P(n^{-1} h^{-2k-1/2}). \end{aligned}$$

To show that the remaining expressions are of smaller order, note that  $B_2$  can be expressed as

$$\begin{aligned} B_2 &= \frac{1}{n^3 h^{2k+3}} \sum_{i,j} \int_A \mathbb{E} \left[ \sigma^2(X_i) K^{(k)^2} \left( \frac{x - X_i}{h} \right) \right] \xi_n(X_j, x) w_n(x) dx \\ &\quad + \frac{1}{n^3 h^{2k+3}} \sum_{i=1}^n \int_A \zeta_n(X_i, x) \xi_n(X_i, x) w_n(x) dx \\ &\quad + \frac{1}{n^3 h^{2k+3}} \sum_{i \neq j} \int_A \zeta_n(X_i, x) \xi_n(X_j, x) w_n(x) dx = L_{n1} + L_{n2} + L_{n3} \end{aligned}$$

with

$$\begin{aligned}\xi_n(X_i, x) &= K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[K\left(\frac{x - X_i}{h}\right)\right] \\ \zeta_n(X_i, x) &= \sigma^2(X_i)K^{(k)^2}\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[\sigma^2(X_i)K^{(k)^2}\left(\frac{x - X_i}{h}\right)\right] \\ w_n(x) &= w(x)(\mathbb{E}[\hat{f}(x)])^{-3}.\end{aligned}$$

It can easily be seen that the expectations of  $L_{n1}$  and  $L_{n3}$  are 0 and that the variances of both terms are of order  $o(1/n^2h^{4k+1})$ . This yields  $L_{ni} = o_P(1/nh^{2k+1/2})$  for  $i = 1, 3$ . The expectation of  $L_{n2}$  is

$$\begin{aligned}\mathbb{E}[L_{n2}] &= \frac{1}{n^3h^{2k+3}} \sum_{i=1}^n \int_A \mathbb{E}[\zeta_n(X_i, x)\xi_n(X_i, x)]w_n(x)dx \\ &= \frac{1}{n^2h^{2k+2}} \int_{A^\epsilon} \int_{-\lambda}^{\lambda} \sigma^2(y)K^{(k)^2}(x)K(x)f(y)w_n(y - hx)dxdy + o\left(\frac{1}{n^2h^4}\right) \\ &= O\left(\frac{1}{n^2h^{2k+2}}\right) = o\left(\frac{1}{nh^{2k+1/2}}\right)\end{aligned}$$

and its variance is again of order  $o(1/n^2h^{4k+1})$ . This gives

$$L_{n2} = \mathbb{E}[L_{n2}] + o_P\left(\frac{1}{nh^{2k+1/2}}\right) = o_P\left(\frac{1}{nh^{2k+1/2}}\right),$$

and therefore shows that  $B_2$  is of the order  $o_P(1/nh^{2k+1/2})$ . The estimation of

$$B_3 = O_P\left(\frac{\log h^{-1}}{n^2h^{2k+2}}\right) = O_P\left(\frac{1}{nh^{2k+1/2}} \frac{\log h^{-1}}{nh^{3/2}}\right) = o_P\left(\frac{1}{nh^{2k+1/2}}\right)$$

is performed by using the uniformly almost sure convergence rate of  $\hat{f}$ . The first assertion of Lemma 2 now follows by combining the estimates of  $B_1$ ,  $B_2$  and  $B_3$ .

In a second step the stochastic expansion of  $I_{n4}$  is developed. If the Taylor expansion (10) is used we get

$$I_{n4} = \int_A (\hat{g}^{(k)}(x))^2 (\mathbb{E}[\hat{f}(x)])^{-2} dx + O_P(1) \int_A (\hat{g}^{(k)}(x))^2 |\hat{f}(x) - \mathbb{E}[\hat{f}(x)]| dx = I_{n4}^{[1]} + I_{n4}^{[2]}.$$

The variance of  $I_{n4}^{[1]}$  can be divided into

$$\text{Var}(I_{n4}^{[1]}) = \text{Var}(L_{n1} + L_{n2}) \leq \text{Var}(L_{n1}) + \text{Var}(L_{n2})$$

with

$$\begin{aligned}L_{n1} &= \frac{1}{n^2h^{2k+2}} \sum_{i=1}^n \int_A \xi_n^{[2]}(X_i, x)(\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx, \\ L_{n2} &= \frac{1}{n^2h^{2k+2}} \sum_{i \neq j} \xi_n^{[1]}(X_i, x) \xi_n^{[1]}(X_j, x) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx, \\ \xi_n^{[l]}(X_i, x) &= Z^l(x, X_i) - \mathbb{E}[Z^l(x, X_i)],\end{aligned}$$

and the random variable  $Z(x, X_i)$  is defined as

$$Z(x, X_1) = K^{(k)}\left(\frac{x - X_1}{h}\right)(m(X_l) - m(x)) - \sum_{j=0}^{k-1} h^{k-j} \binom{k}{j} m^{(k-j)}(x) K^{(j)}\left(\frac{x - X_1}{h}\right).$$

The variance of  $L_{n1}$  is of order

$$\begin{aligned} \text{Var}(L_{n1}) &= \frac{1}{n^4 h^{4k+4}} \sum_{i=1}^n \mathbb{E} \left[ \left( \int_A \xi_n^{[2]}(X_1, x) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx \right)^2 \right] \\ &= O\left(\frac{1}{n^3 h^{4k-1}}\right) = o\left(\frac{1}{n^2 h^{4k+1}}\right) \end{aligned}$$

while the variance of  $L_{n2}$  can be estimated as

$$\begin{aligned} \text{Var}(L_{n2}) &= O\left(\frac{1}{n^2 h^{4k+4}}\right) \int_A \int_A \mathbb{E}^2[Z(x, X_i) Z(y, X_i)] (\mathbb{E}[\hat{f}(x)])^{-2} w(x) w(y) dx dy \\ &= O\left(\frac{1}{n^2 h^{4k-4}}\right) = o\left(\frac{1}{n^2 h^{4k+1}}\right). \end{aligned}$$

This yields

$$I_{n4}^{[1]} = \mathbb{E}[I_{n4}^{[1]}] + o_P\left(\frac{1}{n h^{2k+1/2}}\right) = \int_A \mathbb{E}[(\hat{g}^{(k)}(x))^2] (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx + o_P\left(\frac{1}{n h^{2k+1/2}}\right).$$

A careful inspection shows, that

$$\mathbb{E}[(\hat{g}^{(k)}(x))^2] = \mathbb{E}[G_n^{[1]}(x)] + \frac{n-1}{n} (\mathbb{E}[G_n^{[2]}(x)])^2$$

with  $G_n^{[1]}(x) = Z^2(x, X_1)/n h^{2k+2}$  and  $G_n^{[2]}(x) = Z(x, X_1)/h^{k+1}$ . This gives uniformly on  $A$

$$\mathbb{E}[G_n^{[1]}(x)] = \frac{1}{n h^{2k+1}} \mathbb{E}[Z^2(x, X_1)] = O\left(\frac{1}{n h^{2k-1}}\right) = o\left(\frac{1}{n h^{2k+1/2}}\right),$$

which means that the first term is asymptotically negligible. The second term is equal to

$$\mathbb{E}[G_n^{[2]}(x)] = \frac{1}{h^{k+1}} \mathbb{E}[Z(x, X_1)] = \frac{1}{h^{k+1}} (h^{k+3} \gamma_k(x) + o(h^{k+3})) = O(h^2).$$

Therefore  $(\mathbb{E}[G_n^{[2]}(x)])^2$  is not asymptotically negligible under the assumption  $n h^{2k+5} = O(1)$  which gives

$$I_{n4}^{[1]} = \mathbb{E}[I_{n4}^{[1]}] + o_P\left(\frac{1}{n h^{2k+1/2}}\right) = \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx + o_P\left(\frac{1}{n h^{2k+1/2}}\right).$$

Now the expectation of  $\hat{f}(x)$  has still to be replaced by the true density  $f$ . This can be done by using the expansion (12) and results in

$$I_{n4}^{[1]} = \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (f(x))^{-2} w(x) dx + o_P\left(\frac{1}{n h^{2k+1/2}}\right).$$

An application of the Cauchy-Schwarz inequality determines the order of  $I_{n^4}^{[2]}$ ,

$$I_{n^4}^{[2]} \leq O_P(1) \left( \int_A (\hat{g}^{(k)}(x))^4 dx \right)^{1/2} \left( \int_A |\hat{f}(x) - E[\hat{f}(x)]|^2 dx \right)^{1/2}.$$

The squared second factor is of order  $O_P(1/nh)$  because the mean squared error of  $\hat{f}(x)$  is of order  $O(1/nh)$ , while the expectation of the squared first one is

$$\begin{aligned} \int_A E[(\hat{g}^{(k)}(x))^4] dx &= \frac{1}{n^4 h^{4k+4}} \sum_{i,j,k,l} \int_A E[Z(x, X_i) Z(x, X_j) Z(x, X_k) Z(x, X_l)] dx \\ &= O\left(\frac{1}{n^3 h^{4k-1}}\right) + O\left(\frac{1}{n^2 h^{4k-2}}\right) + O\left(\frac{1}{n h^{2k-5}}\right) + O(h^8). \end{aligned}$$

Altogether we have  $I_{n^4}^{[2]} = o_P(1/nh^{2k+1/2})$ , which completes the proof of Lemma 2.  $\square$

Theorem 1 now follows from Theorem 3 with the weight function  $v_n(x) = w(x)/\hat{f}^2(x)$  and Lemma 2.