Institut für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Pointwise confidence intervals in nonparametric regression with heteroscedastic error structure

M.H. Neumann

submitted: 22nd December 1992

Institut für Angewandte Analysis
und Stochastik
Hausvogteiplatz 5-7
D – O 1086 Berlin
Germany

Preprint No. 34
Berlin 1992

1991 Mathematics Subject Classification. Primary 62 G 15, Secondary 62 G 07, 62 E 20.
Key words and phrases. Nonparametric regression, asymptotic confidence interval, error in coverage probability, Edgeworth expansion, wild bootstrap.

Herausgegeben vom
Institut für Angewandte Analysis und Stochastik
Hausvogteiplatz 5-7
D – O 1086 Berlin

Fax: + 49 30 2004975
e-Mail (X.400): c=de;a=dbp;p=iaas-berlin;s=preprint
e-Mail (Internet): preprint@iaas-berlin.dbp.de

POINTWISE CONFIDENCE INTERVALS IN NONPARAMETRIC REGRESSION WITH HETEROSCEDASTIC ERROR STRUCTURE

MICHAEL H. NEUMANN

ABSTRACT. We assume a nonparametric model with heteroscedastic error structure and consider pointwise confidence intervals for the mean. We construct confidence intervals by using quantiles from a Cornish-Fisher expansion and from the wild bootstrap distribution, with as well as without a subsequent bias correction. It turns out that pure undersmoothing, where the full smoothness is used by the initial estimator, outperforms the method with a subsequent bias correction.

1. Introduction

In the present paper we consider a nonparametric regression model with a heteroscedastic error structure. We intend to construct pointwise confidence intervals for the regression function and compare various methods with respect to their coverage accuracy.

These procedures can be divided into two parts. The estimation part consists in the construction of an appropriate pivotal statistic, which is achieved by a normalization of a usual kernel estimator. A main problem in the construction of confidence intervals is the treatment of the bias, which is unavoidable in this context. We can take the initial statistic as it is, but we can also include at this stage a bias correction based on another kernel estimator. With an appropriate choice of the order of the kernels and the bandwidths, both approaches yield asymptotically equivalent pivotal statistics T_n. The second step for the construction of the confidence interval is the distribution recognition part, which finally yields a quantile for the interval. A simple approach is the inversion of an Edgeworth expansion of T_n. We can also try to estimate the distribution of T_n directly. Because of the heteroscedasticity we approximate it by the wild bootstrap distribution as proposed in [HM90]. This second step can also contain a bias correction, either explicitely by some bias estimator or implicitely via the wild bootstrap.

The aim of this paper is the comparison of these methods. Provided an optimal choice of the bandwidths, it is shown that the inversion of an Edgeworth series and the wild bootstrap without subsequent bias correction are equivalent, that is both methods yield the same rate of decay for the error in coverage probability. Now one may conjecture that it is possible to improve on these methods by a subsequent bias correction. Of course, if we use the same pivotal statistic T_n in both cases, such an additional correction does not harm, and if m is sufficiently smooth, we can attain an improvement of the uncorrected method. Such a result is obtained in [HHJ91]. The reason is that the bias-corrected method utilizes a greater part of the smoothness assumption on m, which pays off in a better recognition of the distribution of T_n. In the present paper we compare both approaches in the same way as [Hal91] does in the case of confidence intervals in density estimation. Using kernels of different orders, we construct both methods such that they exploit the same share of the smoothness assumption. Then it turns out that the first method, which is a pure undersmoothing technique, is superior to the second one.

2. Assumptions and definitions

We consider the nonparametric model

$$Y_i = m(x_i) + \varepsilon_i, \tag{1}$$

where the nonrandom design points x_i are spaced on the unit interval [0, 1] with

$$x_1 < x_2 < \ldots < x_n. \tag{2}$$

To cover a wide variety of practical cases we admit heteroscedastic errors ε_i with

$$E\varepsilon_i = 0 \quad , \quad E\varepsilon_i^2 = v_i \ , \tag{3}$$

where $0 < c \leq v_i \leq C$ holds for some constants c and C. As a basis we take a kernel estimator as proposed in [GM79], namely

$$\widehat{m}_h(x) = \sum_{j=1}^{n} W(x,h)_j Y_j \tag{4}$$

with the weights $W(x,h)_j = \int_{s_{j-1}}^{s_j} \frac{1}{h} w_{x,h}(\frac{z-x}{h})dz$, where $w_{x,h}$ is some usual kernel with support [-1,1] if $h \leq x \leq 1-h$, and some boundary kernel otherwise. Further, we set $s_0 = 0$, $s_j = (x_j + x_{j+1})/2$ for $j = 1, ..., n-1$ and $s_n = 1$. The bandwidth h may vary in the interval $H = [0, 1/2]$ and will be selected later. Let $x_0 \in (0,1)$ be this interior point of the unit interval, at which we intend to construct a confidence interval for m.

Throughout this paper we assume that

(A1) m is k-times continuously differentiable on the interval $(x_0 - \delta, x_0 + \delta)$ for some $\delta > 0$ and $m^{(k)}(x_0) \neq 0$. $(k \geq 2)$

For the asymptotic investigations we suppose that the design points $x_i = x_i(n)$ are regularly spaced, that is

(A2) $\int_0^{x_i} d(t)dt = (i - 1/2)/n$,

where d is a positive, continuous probability density on $[0,1]$, and assume the sample size to tend to infinity. The pivotal quantities considered here are derived from the quantity $V_n^{-1/2}(\widehat{m}_h(x_0) - m(x_0) - B_n)$, where $B_n = \sum_{j=1}^n W(x_0,h)_j m(x_j) - m(x_0)$ is the bias and $V_n = \sum_{j=1}^n W(x_0,h)_j^2 v_j$ is the variance of $\widehat{m}_h(x_0)$. To obtain an observable quantity we have to replace the unknowns B_n and V_n with appropriate estimates.

There are two methods to deal with the bias B_n . We can correct it explicitely by

$$\widehat{B}_n = \sum_{j=1}^{n} W(x_0,h)_j \widehat{m}_{h_1}(x_j) - \widehat{m}_{h_1}(x_0) \ , \tag{5}$$

where $\widehat{m}_{h_1}$ is again a kernel estimator according to (4), which is possibly based on another kernel function w_{x,h_1} . Another widely used practice is to neglect the bias at this stage, which corresponds to

$$\widehat{B}_n = 0. \tag{6}$$

To treat both cases simultaneously, we introduce the unified notation

$$\widehat{\widehat{m}}(x_0) = \widehat{m}_h(x_0) - \widehat{B}_n = \sum_{j=1}^{n} \overline{W}_{nj} Y_j \tag{7}$$

with appropriate weights $\overline{W}_{nj}$. We adopt the following assumption concerning the orders of the kernels $w_{x,h}$ and w_{x,h_1} , respectively.

For $0 < r \leq k$ we assume

(A3) (i) if (5) holds, then $w_{x,h}$ is a kernel of order r_1 and w_{x,h_1} is a kernel of order r_2 , $r = r_1 + r_2$,
(ii) if (6) holds, then $w_{x,h}$ is a kernel of order r.

In both cases, the estimator $\widehat{\widehat{m}}(x_0)$ utilizes the presence of r derivatives of m. By Lemma 6.1 from Section 6 we conclude that the remaining bias $b_n = E\widehat{\widehat{m}}(x_0) - m(x_0)$ of $\widehat{\widehat{m}}(x_0)$ as an estimator of $m(x_0)$ is of order $O(h^{r_1} h_1^{r_2} + n^{-2}h^{-1})$ in case (i), and of order $O(h^r + n^{-2}h^{-1})$ in case (ii).

The best choice for h_1, which preserves a variance of $\widehat{\widehat{m}}(x_0)$ of order $O((nh)^{-1})$, will be of the some order as h . In the sequel, we choose $h_1 \asymp h$, which leads to a bias of order $O(h^r + n^{-2}h^{-1})$ also in the first case.

In distinction to the paper of [Hal92] we admit also a heteroscedastic variance structure, and because we do not want to restrict ourselves to cases, where the variance function can be modelled by some parametric model, we estimate the variance of $\widehat{\widehat{m}}(x_0)$ nonparametrically by

$$\widehat{V}_n = \sum_{j=1}^{n} \overline{W}_{nj}^2 \, \widehat{\varepsilon}_j^2 , \tag{8}$$

where $\widehat{\varepsilon}_j = Y_j - \widehat{m}_f(x_j)$ are estimated residuals based on a kernel estimator $\widehat{m}_f$ with a kernel function $w_{x,f}$ and a bandwidth f . We assume that $\widehat{m}_f$ uses the full smoothness assumption on m , that is

(A4) $w_{x,f}$ is a kernel of order k and $f \asymp n^{-1/(2k+1)}$.

Replacing the unknowns B_n and V_n with $\widehat{B}_n$ and $\widehat{V}_n$, respectively, we obtain the pivotal quantity

$$T_n = \frac{\widehat{\widehat{m}}(x_0) - m(x_0)}{\widehat{V}_n^{1/2}} , \tag{9}$$

which is the starting point for the construction of confidence intervals.

3. Edgeworth expansion for T_n

It is clear that the methods can only be implemented if we choose all bandwidths in a reasonable way by the data. But to obtain more insight into the methods we treat first the case of nonrandom bandwidths $h = h(n), \quad g = g(n), \quad f = f(n)$. As we will see later, we have to apportion the smoothness of m between the curve estimation part due to $\widehat{\widehat{m}}(x_0)$ and the distribution recognition part for the pivotal quantity by explicit or implicit methods. To obtain a consistent procedure, the remaining bias of $\widehat{\widehat{m}}(x_0)$, possibly after a subsequent bias correction, must be of smaller order than the standard deviation of it, which determines the size of the confidence interval. In particular, without any subsequent bias correction we must not use an estimator $\widehat{\widehat{m}}(x_0)$ with an MSE-optimal bandwidth h , since it has bias and standard deviation of the same order. Hence, all the widely used bandwidth selectors fail, because they are tuned to keep the MSE or some related criterion as small as possible. Making a sensible data-dependent choice in practice seems to be difficult, but we will provide a (rough) applicable proposal at the end of this paper.

In the following we approximate the distribution of T_n by an Edgeworth series. For that we need some analogue to Cramér's condition in the i.i.d. case, namely

(A5) $\max_j \sup_{\|t\|>b} \left| E \exp\left\{ it' \begin{pmatrix} \varepsilon_j \\ \varepsilon_j^2 \end{pmatrix} \right\} \right| \leq C_b < 1$ for all $b > 0$.

To obtain consistency of the methods we assume that

(A6) all moments of the ε_j's are uniformly bounded,

which leads to an algebraic decay of the probabilities of large deviations of linear forms in the ε_j's.

To cover also the case with a subsequent bias estimation , we derive an Edgeworth expansion of the more general statistic

$$\overline{T}_n = \frac{\sum_{j=1}^n \overline{\overline{W}}_{nj} Y_j - m(x_0)}{\sqrt{\sum_{j=1}^n \overline{W}_{nj}^2 \widehat{\varepsilon}_j^2}} \tag{10}$$

where we assume

$$\sum_{j=1}^n \left| \overline{\overline{W}}_{nj} - \overline{W}_{nj} \right| = O((nh)^{-1/2}). \tag{11}$$

Now we derive the expansion of $\overline{T}_n$ in two steps. First, we infer from results of [Sko86] the validity of an expansion of arbitrary length of the random vector

$$S_n = B_n^{-1/2} \sum_{j=1}^n \alpha_j \ , \tag{12}$$

where $\alpha_j = (nh\overline{\overline{W}}_{nj}\varepsilon_j, (nh\overline{W}_{nj})^2 \, (\varepsilon_j^2 - v_j))'$ and $B_n = \sum_{j=1}^n \mathrm{Cov}(\alpha_j)$.

Having this, we can derive from results of [Sko81] the validity of the expansion of a sufficiently regular sequence of functions $f_n(S_n)$.

To get such a function, we approximate $\overline{T}_n$ by

$$(13) \qquad \tilde{\overline{T}}_n = \frac{\sum_{j=1}^n \overline{\overline{W}}_{nj} Y_j \; - \; m(x_0)}{\sqrt{\tilde{V}_n}} = \frac{\sum_{j=1}^n \overline{\overline{W}}_{nj} \varepsilon_j \; + \; b_n}{\sqrt{\sum_{j=1}^n \overline{W}^2_{nj}(\varepsilon_j^2 - v_j) + V_n}}$$

where $\tilde{V}_n = \sum_{j=1}^n \overline{W}^2_{nj} \varepsilon_j^2$ and $b_n = \sum_{j=1}^n \overline{\overline{W}}_{nj} m(x_j) \; - \; m(x_0)$.
On the basis of Theorem 3.2 of [Sko81] we obtain the following proposition.

Proposition 3.1. *Assume (A1) trough (A6) , (11) and* $h \asymp n^{-\lambda_1}$ *, where* $\lambda_1 > (2r+1)^{-1}$ *. Then*

$$(14) \qquad P\left(\tilde{\overline{T}}_n < t\right) \; = \; \Phi(t) - \frac{b_n}{V_n^{1/2}} \phi(t) + \rho_{n3} \frac{2t^2+1}{6} \phi(t)$$
$$- \frac{1}{2} \; \frac{\overline{V}_n - V_n}{V_n} t\phi(t) + \frac{1}{2} \; \frac{b_n}{V_n^{1/2}} \rho_{n3} t\phi(t) + O((nh)^{-1})$$

holds uniformly over $-\infty < t < \infty$, *where* $\rho_{n3} = V_n^{-3/2} \sum_{j=1}^n \overline{W}^3_{nj} \; E\varepsilon_j^3$ *and* $\overline{V}_n = \sum_{j=1}^n \overline{\overline{W}}^2_{nj} v_j$.

As a special case of (14) we obtain an expansion for the quantity $\tilde{T}_n = \tilde{V}_n^{-1/2} \left(\sum_{j=1}^n \overline{W}_{nj} Y_j \; - \; m(x_0)\right)$. The pivotal statistic T_n , which is of primary interest, does not admit an Edgeworth expansion immediately, because it cannot be expressed as a function of the mean of independent random vectors. To treat this and other more complicated cases we introduce the following concept.

Definition 3.1. Let $\{Y_n\}$ be a sequence of random variables and $\{\gamma_{n1}\}$, $\{\gamma_{n2}\}$ be sequences of constants. We write

$$Y_n = \tilde{O}(\gamma_{n1} \, , \, \gamma_{n2})$$

if

$$P(|Y_n| > C\gamma_{n1}) \leq C\gamma_{n2}$$

holds for some $n \geq 1$ and $C < \infty$.

Lemma 3.1. *Let* $\{X_n\}, \{Y_n\}$ *be sequences of random variables with*

$$P(X_n < t) = \Phi(t) + t_n \phi(t) + O(u_n)$$

uniformly over $-\infty < t < \infty$ *for some bounded sequence* $\{t_n\}$. *Further, we assume* $Y_n = \tilde{O}(\gamma_{n1} \, , \, \gamma_{n2})$ *. Then*

$$P(X_n + Y_n < t) = P(X_n < t) + O(\gamma_{n1} + \gamma_{n2} + u_n)$$

holds uniformly over $-\infty < t < \infty$.

This lemma can be proved analogously to Lemma 3.2 of [Neu92]. Now we return to the derivation of an expansion for T_n .

Lemma 3.2. *Under the assumptions of Proposition* 3.1 *it holds*

$$\overline{T}_n - \tilde{\overline{T}}_n = \tilde{O}((nh)^{-1}, n^{-1})$$

From Proposition 3.1 and the Lemmas 3.1 and 3.2 we conclude, observing that the fourth term on the right-hand side of (14) vanishes, the following assertion.

Proposition 3.2. *Assume* $(A1)$ *through* $(A6)$ *and* $h_n \asymp n^{-\lambda_1}, \lambda_1 > (2r+1)^{-1}$. *Then*

$$\begin{aligned} P(T_n < t) &= \Phi(t) - \frac{b_n}{V_n^{1/2}}\phi(t) + \rho_{n3}\frac{2t^2+1}{6}\phi(t) \\ &\quad + \frac{1}{2}\frac{b_n}{V_n^{1/2}}\rho_{n3}t\phi(t) + O((nh)^{-1}), \end{aligned} \tag{15}$$

where $b_n = O(h^r + n^{-2}h^{-1})$ *and* $\rho_{n3} \asymp (nh)^{-1/2}$.

4. Confidence intervals for $m(x_0)$

In this section we consider different methods for the construction of confidence intervals for $m(x_0)$ to an asymptotic level $1-\alpha$. On first sight, there seem to be two different methods to do this. First we can invert the Edgeworth expansion of T_n and estimate the unknown parameters, which leads to a Cornish-Fisher expansion. A special case of this method, which neglects the terms involving the skewness ρ_{n3} in (15), consists in the application of the normal quantile, which was also considered in [Neu92] for estimators with data-driven bandwidths.

A second approach is, to mimic the distribution of T_n by means of the wild bootstrap proposed by [HM90]. Then we do not need any explicit Edgeworth expansion, the bootstrap provides an appropriate quantile automatically. As we will see later, the above classification is not the deciding one. On the contrary, these two approaches are more or less equivalent.

In distinction to classical problems in parametric regression, there does not exist any consistent, unbiased estimator of $m(x_0)$. However we construct a confidence interval, we have always to deal with some remaining bias. There are various methods to keep the bias negligible, which lead to confidence procedures with distinct rates for the coverage errors. Firstly, we can neglect it and take it into account by the use of pivotal statistics, which use an undersmoothed estimator $\widehat{\widehat{m}}(x_0)$. Secondly, we can estimate it explicitely and correct the location of the interval correspondingly. And thirdly, we can take it into account implicitely by mimicking it by the bootstrap.

Although two-sided confidence interval are perhaps of greater practical interest, we consider throughout this paper one-sided intervals, because this provides a stronger look at the behavior of the methods, see [Hal91], Section 3.7, for a discussion of this problem.

4.1. Confidence intervals without bias correction. If we construct a confidence interval only on the basis the normal quantile $u_\alpha = \Phi^{-1}(\alpha)$, then we obtain by Proposition 3.2 for $I = [\widehat{\widehat{m}}(x_0) + \widehat{V}_n^{1/2} u_\alpha \,,\ \infty\ )$ a coverage probability of

$$P(m(x_0) \in I) = P(T_n < u_\alpha) = 1 - \alpha + O((nh)^{-1/2} + (nh)^{1/2} h^r).$$

By inverting the expansion (15) it is easy to see that

$$P(T_n < t - \rho_{n3} \frac{2t^2+1}{6}) = \Phi(t) + O((nh)^{1/2} h^r + (nh)^{-1}).$$

To perform such a skewness correction in practice we estimate ρ_{n3} simply by

$$\widehat{\rho}_{n3} = \widehat{V}_n^{-3/2} \sum_{j=1}^{n} \overline{W}_{nj}^3 \widehat{\varepsilon}_j^3 \,. \tag{16}$$

It is easy to see that $\widehat{\rho}_{n3} - \rho_{n3} = O_P((nh)^{-1})$ holds, and hence we might conjecture that

$$I_1 = [\widehat{\widehat{m}}(x_0) + \widehat{V}_n^{1/2} \widehat{t}_\alpha \,,\ \infty) \tag{17}$$

with $\widehat{t}_\alpha = u_\alpha - \widehat{\rho}_{n3} \frac{2u_\alpha^2+1}{6}$ forms a one-sided confidence interval for $m(x_0)$ to a nominal level $1-\alpha$ with a coverage error of $O((nh)^{1/2} h^r + (nh)^{-1})$. To prove this rigorously, we have to follow the proof of (4.6) in [Hal91], which yields the following theorem.

Theorem 4.1. *Under the assumptions of Proposition* 3.2 *we have*

$$P(m(x_0) \in I_1) \;=\; 1 - \alpha + O((nh)^{1/2} h^r + (nh)^{-1}).$$

Another way to find an appropriate quantile consists in the approximation of the distribution of T_n by the wild bootstrap proposed by [HM90].

Recall that $\widehat{\varepsilon}_j = Y_j - \widehat{m}_f(x_j)$. Now we draw independent random variables ε_j^* with

$$E^* \varepsilon_j^* = O \,, E^* \,{\varepsilon_j^*}^2 = \widehat{\varepsilon}_j^2 \,,\ E^* \,{\varepsilon_j^*}^3 = \widehat{\varepsilon}_j^3 \,,$$

where E^* denotes the expectation with respect to the (random) bootstrap distribution.

[HM90] proposed the following procedures to generate such a sample :

(i) a discrete bootstrap distribution
$\varepsilon_j^* \sim \gamma \delta_{a\widehat{\varepsilon}_j} \ + \ (1-\gamma)\delta_{b\widehat{\varepsilon}_j}$,
where $\gamma = \frac{\sqrt{5}}{10}$, $a = \frac{1-\sqrt{5}}{2}$, $b = \frac{1+\sqrt{5}}{2}$, δ_x being the Dirac measure at x,

(ii) a continuous bootstrap distribution
$\varepsilon_j^* \sim \widehat{\varepsilon}_j(\frac{Z_1}{\sqrt{2}} + \frac{Z_2^2-1}{2})$,
where Z_1 and Z_2 are independent Gaussian random variables with zero mean and unit variance, also independent of $\widehat{\varepsilon}_j$.

Now we define a bootstrap analogue of Y_j as

$$Y_j^* = \widehat{m}_{g^*}(x_j) + \varepsilon_j^* ,$$

where $\widehat{m}_{g^*}$ is another kernel estimator with (nonrandom) bandwidth g^*. As an analogue of the estimated residuals $\widehat{\varepsilon}_j$ we have

$$\widehat{\varepsilon}_j^* = Y_j^* - \sum_{k=1}^{n} W(x_j, f)_k Y_k^*.$$

If we neglect the bias by the bootstrap, that is if we mimic $T_{n0} = \widehat{V}_n^{\,-1/2}\left(\widehat{\widehat{m}}(x_0) - E\widehat{\widehat{m}}(x_0)\right)$ or $\widetilde{T}_{n0} = \widetilde{V}_n^{\,-1/2}\left(\widehat{\widehat{m}}(x_0) - E\widehat{\widehat{m}}(x_0)\right)$, we have the following bootstrap statistics:

$$T_{n0}^* = \sum_{j=1}^{n} \overline{W}_{nj}\varepsilon_j^* \quad /\sqrt{\widehat{V}_n^*}$$

and

$$\widetilde{T}_{n0}^* = \sum_{j=1}^{n} \overline{W}_{nj}\varepsilon_j^* \quad /\sqrt{\widetilde{V}_n^*},$$

where $\widehat{V}_n^* = \sum_{j=1}^n \overline{W}_{nj}^2\, \widehat{\varepsilon}_j^{*2}$ and $\widetilde{V}_n^* = \sum_{j=1}^n \overline{W}_{nj}^2 \varepsilon_j^{*2}$.

To draw conclusions for the distribution of T_{n0}^* and $\widetilde{T}_{n0}^*$ under the condition of the original sample $\mathcal{Y} = (Y_1, \dots, Y_n)$, respectively , we are going to apply again Edgeworth expansions. In case of the continuous bootstrap distribution the validity of these expansions follows from the same arguments given in the proof of Proposition 3.1, since the random variables ε_j^* fulfill Cramér's condition in a uniform manner.

In case of the discrete bootstrap version the situation seems to be more complicated. To make this difficulty clear, observe that the formal Edgeworth expansion of the sum of n random variables with a fixed lattice distribution is correct only to an order of $O(n^{-1/2})$, see [BR86], Chapter 5. This becomes immediately clear, if we note that the distribution function of that sum has jumps with a maximal height of order $O(n^{-1/2})$, and therefore we cannot approximate it with a better rate by the (continuous) Edgeworth series. On the other hand, the sum of the ε_j^*'s is distributed as the convolution of *different* lattice distributions, which has a distribution function with much smaller jumps than in the above case. This fact let us hope, that the formal Edgeworth expansion is valid to a sufficiently high order. It is clear that the ε_j^*'s do not fulfill Cramér's condition. To exploit the diversity of the distributions of the ε_j^*'s, we prove some version of Petrov's condition, which only requires that the product of the characteristic functions is bounded away from 1 in a certain region. On the basis of results of [Sko81] and [Sko86] we finally obtain the validity of the Edgeworth expansion , which is stated in the following proposition.

Proposition 4.1. *Under the assumptions of Proposition* 3.2 *we have*

(i)

$$P(\widetilde{T}^*_{n0} < t|\mathcal{Y}) = \Phi(t) + \rho^*_{n3}\frac{2t^2+1}{6}\phi(t) + O((nh)^{-1})$$

where $\rho^*_{n3} = E^*(\rho_{n3}|\mathcal{Y}) = \widehat{\rho}_{n3}$ *and* $\widehat{\rho}_{n3}$ *is defined by (*16*),*

(ii)

$$P(T^*_{n0} < t|\mathcal{Y}) = P(\widetilde{T}^*_{n0} < t|\mathcal{Y}) + O((nh)^{-1}),$$

both uniformly over $\mathcal{Y} \in \mathcal{C}_n$ *for some subset* $\mathcal{C}_n \subseteq I\!R^n$ *with* $P(\overline{\mathcal{C}}_n) = O(n^{-\lambda})$ *for arbitrarily large* λ .

Let $\widetilde{t}_\alpha$ and $\widetilde{\widetilde{t}}_\alpha$ be the $(1-\alpha)$ - quantiles of the bootstrap distributions of T^*_{n0} and $\widetilde{T}^*_{n0}$, respectively. Then we have by Proposition 4.1

$$\widetilde{t}_\alpha = \widetilde{\widetilde{t}}_\alpha + O((nh)^{-1}) = \widehat{t}_\alpha + O((nh)^{-1}) \tag{18}$$

uniformly over $\mathcal{Y} \in \mathcal{C}_n$. In other words, we have

$$\left.\begin{array}{l} \widetilde{t}_\alpha - \widehat{t}_\alpha \\ \widetilde{\widetilde{t}}_\alpha - \widehat{t}_\alpha \end{array}\right\} = \widetilde{O}((nh)^{-1}, n^{-\lambda}),$$

which yields by Lemma 3.1 the next assertion.

Proposition 4.2. *Under the assumptions of Proposition* 3.2 *we have*

$$\begin{aligned} P(T_n < \widetilde{t}_\alpha) &= P(T_n < \widetilde{\widetilde{t}}_\alpha) + O((nh)^{-1}) \\ &= 1 - \alpha + O((nh)^{1/2}h^r + (nh)^{-1}). \end{aligned}$$

uniformly over $\mathcal{Y} \in \mathcal{C}_n$.

Let $I_2 = [\widehat{\widehat{m}}(x_0) - \widehat{V}_n^{1/2}\widetilde{t}_\alpha,\ \infty)$ and $I_3 = [\widehat{\widehat{m}}(x_0) - \widehat{V}_n^{1/2}\widetilde{\widetilde{t}}_\alpha,\ \infty)$. Then we have the following analogue of Theorem 4.1.

Theorem 4.2. *Under the assumptions of Proposition* 3.2 *we have*

$$\left.\begin{array}{l} P(m(x_0) \in I_2) \\ P(m(x_0) \in I_3) \end{array}\right\} = 1 - \alpha + O((nh)^{1/2}h^r + (nh)^{-1}).$$

In view of Theorems 4.1 and 4.2 we can state that the inversion of the Edgeworth expansion as well as the wild bootstrap are indeed equivalent methods for obtaining confidence intervals.

We have seen that all confidence intervals without bias correction provide an error in coverage probability of order $O((nh)^{1/2}h^r + (nh)^{-1})$. This error term is minimized for

$$h_{opt} \asymp n^{-\frac{3}{2r+3}}, \tag{19}$$

which yields for $j = 1, 2, 3$

$$P(m(x_0) \in I_j) = 1 - \alpha + O(n^{-\frac{2r}{2r+3}}). \tag{20}$$

4.2. Confidence intervals based on an explicit bias correction. In the previous section we have seen that the coverage accuracy is affected by the incorrect location of the interval due to the bias b_n of $\widehat{\widehat{m}}(x_0)$ as an estimator of $m(x_0)$. This drawback is unavoidable in the context of nonparametric confidence intervals, because each consistent estimator of $m(x_0)$ is necessarily biased.

To take the bias into account we can include a subsequent bias correction. In distinction to [Hal91], who estimates only the leading term of a bias expansion, which removes the error term of order $O(h^r)$ but ignores the remainder from this expansion of order $o(h^r) + O(n^{-2}h^{-1})$, we estimate the whole bias $b_n = E\widehat{\widehat{m}}(x_0) - m(x_0)$. Let

$$\widehat{b}_n = \sum_{j=1}^{n} \overline{W}_{nj}\widehat{m}_g(x_j) - \widehat{m}_g(x_0) \tag{21}$$

be an estimator of b_n, which is based on a kernel estimator $\widehat{m}_g$ with an $(r-1)$-times continuously differentiable kernel of order s and bandwidth g. If we assume that m is $(r+s)$-times continuously differentiable, it follows by Lemma 6.1 from Section 6 that $\widehat{b}_n$ has a bias of order $O(h^r g^s + n^{-2}h^{-1})$, and by Lemma 8.8 in [Neu92] we conclude that $\widehat{b}_n$ has a standard deviation of order $O(h^r(ng^{2r+1})^{-1/2})$. These terms are balanced for $g \asymp n^{-1/(2r+2s+1)}$, but we will prove that another choice provides a better asymptotic coverage accuracy for the appropriately corrected confidence intervals.

Now we consider the following bias-corrected pivotal quantity

$$T_{nc} = \frac{\widehat{\widehat{m}}(x_0) - m(x_0) - \widehat{b}_n}{\widehat{V}_n^{1/2}} = \frac{\sum_{j=1}^n \overline{\overline{W}}_{nj}\varepsilon_j + \widetilde{b}_n}{\widehat{V}_n^{1/2}}, \tag{22}$$

where $\widetilde{b}_n = \sum_{j=1}^n \overline{\overline{W}}_{nj} m(x_j) - m(x_0)$. It follows by

$$\overline{\overline{W}}_{nj} - \overline{W}_{nj} = O(n^{-1}h^k g^{-(k+1)} + n^{-2}g^{-2})$$

that the $\overline{\overline{W}}_{nj}$'s fulfill (11) if

$$(h/g)^r = O((nh)^{-1/2}). \tag{23}$$

Under the assumptions of Proposition 3.2 and the additional assumption (23) we derive by Proposition 3.1 and Lemma 3.2 that

$$
\begin{aligned}
P(T_{nc} < t) &= \Phi(t) - \frac{\widetilde{b}_n}{V_n^{1/2}}\phi(t) + \rho_{n3}\frac{2t^2+1}{6}\phi(t) \\
&\quad + \frac{1}{2}\frac{\overline{V}_n - V_n}{V_n}t\phi(t) + \frac{1}{2}\frac{\widetilde{b}_n}{V_n^{1/2}}\rho_{n3}t\phi(t) + O((nh)^{-1}). \qquad (24)
\end{aligned}
$$

Now we can prove analogously to Theorem 4.1 that

$$
\begin{aligned}
P(T_{nc} < \widehat{t}_\alpha) &= P(T_{nc} < \widetilde{t}_\alpha) + O((nh)^{-1}) \\
&= P(T_{nc} < \widetilde{\widetilde{t}}_\alpha) + O((nh)^{-1}) \\
&= 1 - \alpha - \frac{\widetilde{b}_n}{V_n^{1/2}}\phi(u_\alpha) - \frac{1}{2}\frac{\overline{V}_n - V_n}{V_n}u_\alpha\phi(u_\alpha) \qquad (25) \\
&\quad + \frac{1}{2}\frac{\widetilde{b}_n}{V_n^{1/2}}\rho_{n3}u_\alpha\phi(u_\alpha) + O((nh)^{-1})
\end{aligned}
$$

holds. Let $I_{1c} = [\widehat{\widehat{m}}(x_0) - \widehat{b}_n - \widehat{V}_n^{1/2}\widehat{t}_\alpha,\, \infty)$, $I_{2c} = [\widehat{\widehat{m}}(x_0) - \widehat{b}_n - \widehat{V}_n^{1/2}\widetilde{t}_\alpha,\, \infty)$ and $I_{3c} = [\widehat{\widehat{m}}(x_0) - \widehat{b}_n - \widehat{V}_n^{1/2}\widetilde{\widetilde{t}}_\alpha,\, \infty)$. From

$$\widetilde{b}_n = O(h^r g^s + n^{-2}h^{-1}) \qquad (26)$$

and

$$
\begin{aligned}
\frac{\overline{V}_n - V_n}{V_n} &= O(nh)\left[2\sum_{j=1}^{n}\overline{W}_{nj}(\overline{\overline{W}}_{nj} - \overline{W}_{nj})v(x_j) + \sum_{j=1}^{n}\left(\overline{\overline{W}}_{nj} - \overline{W}_{nj}\right)^2 v(x_j)\right] \\
&= O((h/g)^{r+1} + (nh)^{-1}) \qquad (27)
\end{aligned}
$$

we obtain the following theorem.

Theorem 4.3. *Under the assumptions of Proposition* 3.2 *and* (23) *we have for* $j = 1,2,3$

$$P(m(x_0) \in I_{jc}) = 1 - \alpha + O((nh)^{1/2}h^r g^s + (h/g)^{r+1} + (nh)^{-1}). \qquad (28)$$

It is easy to see that the residual term of (28) attains the smallest possible order, if and only if the three terms are balanced. This is achieved for

$$h \asymp n^{-\frac{3+2s/(r+1)}{2(r+s)+3+2s/(r+1)}} \qquad (29)$$

and

$$g \asymp h(nh)^{1/(r+1)}, \qquad (30)$$

which yields

$$P(m(x_0) \in I_{jc}) = 1 - \alpha + O\left(n^{-\frac{2(r+s)}{2(r+s)+3+2s/(r+1)}}\right). \tag{31}$$

This may be compared with the slightly worse rate of [Hal91] who attained a coverage error of $O\left(n^{-\frac{2r(2r+2s+1)+2s}{(2r+2s+1)(2r+3)}}\right)$. The reason for this worse rate is, that the bandwidth in that paper is chosen such that the bias and the variance of the subsequent bias estimator are balanced, whereas it is shown here that another choice provides a better coverage accuracy.

4.3. Confidence intervals with an implicit bias correction via the wild bootstrap. We can also take into account the bias by mimicking it in the bootstrap world. We define

$$T_n^* = \frac{\sum_{j=1}^n \overline{W}_{nj} Y_j^* - \widehat{m}_{g^*}(x_0)}{\sqrt{\sum_{j=1}^n \overline{W}_{nj}^2 \, \widehat{\varepsilon}_j^{*\,2}}}. \tag{32}$$

Let t_α^b be the $(1-\alpha)$-quantile of the distribution of T_n^* and $I^b = \left[\widehat{\widehat{m}}(x_0) - \widehat{V}_n^{1/2} t_\alpha^b, \infty\right)$. If the quantity T_{nc} includes the bias corrector $\widehat{b}_n = b_n^* = \sum_{j=1}^n \overline{W}_{nj} \widehat{m}_{g^*}(x_j) - \widehat{m}_{g^*}(x_0)$, then

$$P\left(m(x_0) \in I^b\right) = P\left(T_n < t_\alpha^b\right) = P\left(T_{nc} < t_\alpha^b - \frac{b_n^*}{\widehat{V}_n^{1/2}}\right). \tag{33}$$

Observe that $t_\alpha^b - \frac{b_n^*}{\widehat{V}_n^{1/2}}$ is the $(1-\alpha)$-quantile of the quantity

$$T_{n0}^* + b_n^* \left(\frac{1}{\sqrt{\widehat{V}_n^*}} - \frac{1}{\sqrt{\widehat{V}_n}}\right) = T_{n0}^* + R_n.$$

Analogously to Proposition 4.1 we can expand $T_{n0}^* + R_n$ in an Edgeworth series. It is easy to see that $R_n = O_P(h^r + n^{-2}h^{-1})$, and since we have sufficient moment conditions, we infer that R_n contributes to this expansion only terms of order $O(h^r + n^{-2}h^{-1})$. Hence, we have the following proposition.

Proposition 4.3. *Under the assumptions of Proposition* 3.2 *we have*

$$P\left(T_{no}^* + R_n < t|\mathcal{Y}\right) = \Phi(t) + \widehat{\rho}_{n3}\frac{2t^2+1}{6}\phi(t) + O\left((nh)^{-1} + h^r\right).$$

uniformly over $\mathcal{Y} \in \mathcal{C}_n$.

By (33) we can prove analogously to Theorem 4.1 the following assertion.

Theorem 4.4. *Under the assumptions of Proposition* 3.2 *we have*

(34)
$$P\left(m(x_0) \in I^b\right) = 1 - \alpha + O\left((nh)^{1/2} h^r\, g^{*\,s} + (h/g^*)^{r+1} + (nh)^{-1} + h^r\right).$$

If we choose h and g^* according to (29) and (30), then the first three residual terms of (34) are balanced. The fourth residual term is of smaller or of the same order if and only if

(35)
$$r(r+1) \le 2s.$$

Hence, we obtain by I^b a coverage accuracy of the same order as by I_{jc} $(j = 1, 2, 3)$ iff $r(r+1) \le 2s$ holds.

5. Discussion

1) We distinct between two main classes of confidence procedures. The first class contains such methods that use a pure undersmoothing, that is the initial estimator $\widehat{\widehat{m}}(x_0)$ uses a suboptimal bandwidth with $h^{2r} \ll (nh)^{-1}$. The second class involves a subsequent bias correction, either by an explicit bias estimator or by an implicit correction via the wild bootstrap. In any case, the correspondingly corrected estimator has a remaining bias that is of lower order than its standard deviation. That is, all consistent methods use ultimately, more or less hidden, an undersmoothing.

2) There are some other papers, which are devoted to a comparison of the two approaches, undersmoothing and subsequent bias correction. [HHJ91] compared both methods, but they restricted themselves to second order kernels at all stages. Their bias-neglecting pivotal quantity U_n uses the presence of two derivatives of m, whereas their bias-corrected quantity T_n uses already the presence of four derivatives. The bias modelization by the wild bootstrap, which is only made for T_n, exploits then the existence of two more derivatives. Therefore it is not very surprising, that this bias correction method with optimal bandwidths outperforms the pure undersmoothing.

3) In the case of confidence intervals for densities [Hal91] compared the methods from another point of view. To make a fair comparison, he assumes that all methods under consideration exploit the same amount of smoothness of m, and he concludes that pure undersmoothing is to be preferred.

To do the same in our case, we assume that m is $(r+s)$-times continuously differentiable. According to (19) and (20), the undersmoothing method provides confidence intervals of size $O\left(n^{-U/2}\right)$ with a coverage error of order $O(n^{-U})$, where

(36)
$$U = \frac{2(r+s)}{2(r+s)+3}.$$

As may be seen by (29) and (31) as well as by the remark after Theorem 4.4, the bias correction methods can at best provide confidence intervals of size $O(n^{-V/2})$ with a coverage accuracy of $O(n^{-V})$, where

(37) $$V = \frac{2(r+s)}{2(r+s)+3+2s/(r+1)}.$$

If $s > 1$, then we have $U > V$ and therefore the undersmoothing method turns out to be superior.

4) The reason for this better behavior can easily be observed by comparing the statistics T_n with an $(r+s)$-th order kernel for $\widehat{\widehat{m}}(x_0)$ and T_{nc} with an r-th order kernel for $\widehat{\widehat{m}}(x_0)$ and an s-th order kernel for $\widehat{m}_g$. On the one hand, both statistics use the presence of $r+s$ derivatives of m and the remaining bias b_n is $O(h^{r+s}) + n^{-2}h^{-1}$ and $O(h^r g^s + n^{-2}h^{-1})$, respectively. On the other hand, if we mimic these statistics in the bootstrap world by T^*_{n0} and T^*_n, respectively, their denominators estimate the standard deviation of the respective numerators. This is also the case for the statistic T_n, whereas the denominator of T_{nc} primarily estimates the standard deviation of the numerator of the uncorrected quantity T_n. Hence, we have to choose the bandwidth g of the subsequent bias estimator such that it provides a good compromise between bias reduction and disturbation of the original variance. Such a compromise is not necessary for T_n, because T^*_{n0} is its complete analogue in the bootstrap world.

5) We have seen that the performance of the confidence procedure depends mainly on the possibility of a good estimation of the (approximate) cumulants of T_n. We note that the first approximate cumulant depends on the bias of $\widehat{\widehat{m}}(x_0)$, and can be estimated with an accuracy that depends on the smoothness of m. On the other hand, the third approximate cumulant depends only on the distribution of the pure errors ε_j, and can be estimated in any case with a mean squared error of $O((nh^{-1})$.

6) It is clear that all the methods require a choice of the bandwidths, which is fitted to the underlying smoothness of m. Since there does not exist a sensible universal rule for the bandwidths, we have to adapt them in dependence on the data. It is known that all of the usual bandwidth selectors find bandwidths of the MSE-optimal order (see [HHM88]), which need not to be optimal for the aim of an as accurate as possible coverage probability.

To give any definite rule, which is fitted to the underlying situation at least to a minimal amount, we propose to apply cross-validation or any other equivalent criterion for the bandwidth choice. As already stated, the ultimate estimator of $m(x_0)$, that is $\widehat{\widehat{m}}(x_0)$ in case of a bias-neglecting procedure or the bias-corrected quantity otherwise, must be an undersmoothed one. A method, which yields such an estimator is proposed in [Neu92]. We estimate $m(x_0)$ first by an r-th order kernel and correct this estimator by some bias estimator, which is based on an s-th order kernel. In both cases we can apply cross-validation to choose the bandwidths h and g, respectively. Then bias and standard deviation of the initial estimator are balanced, but the second estimator provides a smaller order for the remaining bias. Hence, we obtain a consistent procedure.

6. Appendix

First we state a lemma, which improves on the remainder term of order $O(n^{-1})$ of an approximation of $E\widehat{m}_h(x_0)$ in [GM79], p. 30.

Lemma 6.1. *Let (A1) be fulfilled and let $w_{x,h}$ be a kernel function of order $r \leq k$. Further, let $w_{x,h}$ and d be Lipschitz-continuous of order* 1. *Then*

$$E\widehat{m}_h(x_0) \;=\; \int_0^1 \frac{1}{h} w_{x,h}\left(\frac{x_0 - z}{h}\right) m(z)\, dz \;+\; O(n^{-2}h^{-1})$$

for $0 \leq h \leq \min\{x_0, 1 - x_0\}$, *where*

$$\int_0^1 \frac{1}{h} w_{x,h}\left(\frac{x_0 - z}{h}\right) m(z)\, dz \;-\; m(x_0) \;=\; O(h^r).$$

Proof. First, observe that under $h \leq \min\{x_0, 1 - x_0\}$ the estimator $\widehat{m}_h(x_0)$ does not include any boundary kernel. For simplicity we write w rather than $w_{x,h}$ in the following. By a Taylor expansion of m we obtain

$$\begin{aligned}
& E\widehat{m}_h(x_0) - \int_0^1 \frac{1}{h} w\left(\frac{x_0 - z}{h}\right) m(z)\, dz \\
&= \sum_{j=1}^n \int_{s_{j-1}}^{s_j} \frac{1}{h} w\left(\frac{x_0 - z}{h}\right) (m(x_j) - m(z))\, dz \\
&= \sum_{j=1}^n \int_{s_{j-1}}^{s_j} \frac{1}{h} \left(w\left(\frac{x_0 - z}{h}\right) - w\left(\frac{x_0 - x_j}{h}\right)\right) (x_j - z) m'(x_j)\, dz \\
&\quad + \sum_{j=1}^n \frac{1}{h} w\left(\frac{x_0 - x_j}{h}\right) m'(x_j) \int_0^1 (x_j - z)\, dz \\
&\quad + \sum_{j=1}^n \int_{s_{j-1}}^{s_j} \frac{1}{h} w\left(\frac{x_0 - z}{h}\right) O((z - x_j)^2)\, dz \\
&= A_1 + A_2 + A_3.
\end{aligned}$$

Because of $\left|w\left(\frac{x_0-z}{h}\right) - w\left(\frac{x_0-x_j}{h}\right)\right| |x_j - z| = O(n^{-2}h^{-1})$ for $z \in [s_{j-1}, s_j]$ we get immediately that

$$A_1 \;=\; O(n^{-2}h^{-1}).$$

From $(s_j - x_j) - (x_j - s_{j-1}) = \frac{x_{j+1}-x_j}{2} - \frac{x_j - x_{j-1}}{2} = \frac{1}{2n}\left(\frac{1}{d(\xi_j)} - \frac{1}{d(\xi_{j-1})}\right) = O(n^{-2})$ for some $\xi_{j-1} \in (x_{j-1}, x_j)$ and $\xi_j \in (x_j, x_{j+1})$ we obtain that $\int_{s_{j-1}}^{s_j} (x_j - z)\, dz = O(n^{-3})$

holds, which implies

$$A_2 = O(n^{-2}).$$

Finally, we obtain easily

$$A_3 = O(n^{-2}),$$

which proves the first assertion. The second part of the lemma is easily proved via a Taylor expansion of m at x_0. □

Proof of Proposition 3.1.
For the sake of a clear presentation we divide this proof into two parts.

1) validation of an Edgeworth expansion for S_n

In this part we prove on the basis of results of [Sko86] the validity of an Edgeworth expansion of the random vector S_n that is defined by (12).

For simplicity, we adopt the notation of the abovementioned paper as far as possible. First, we choose $s = 4$, $a_n(t) = (nh)^{1/2}$ and $\{\epsilon_n\}$ such that $\epsilon_n = O((nh)^{-1-\delta})$ for some δ, $0 < \delta \le 1/2$. Now we check the conditions of Theorem 3.4. in [Sko86]. It is obvious, that (I) is fulfilled.
Next we show that

$$nhM_1 \le B_n \le nhM_2 \tag{38}$$

holds for certain positive definite matrices M_1 and M_2. There "$\le$" denotes the usual partial order-relation of symmetric matrices. Let $\beta_j = (\varepsilon_j, \varepsilon_j^2 - v_j)'$ and $Cov_j =$ $\mathrm{Cov}(\beta_j)$. Since $\exp\{ix\} = 1 + ix + O(\|x\|^2)$ holds uniformly over $x \in (-\infty, \infty)$, we obtain by (A5) for arbitrary t, $\|t\| = 1$, that

$$1 > C_1 \ge |E\, \exp\{it'\beta_j\}| \ge 1 - t'Cov_j t,$$

which implies

$$\lambda_{\min}(Cov_j) \ge M_3 \tag{39}$$

for some $M_3 > 0$. On the other hand, since the first eight moments of the $\varepsilon_j's$ are assumed to be uniformly bounded, we have

$$\lambda_{\max}(Cov_j) \le M_4 \tag{40}$$

for some $M_4 < \infty$. Let $D_j = \mathrm{diag}\left[nh\overline{\overline{W}}_{nj}\,,\, (nh)^2\overline{W}_{nj}^2\right]$. It is easy to see that

$$\sum_{j=1}^{n} D_j^2 = \mathrm{diag}[m_{n1}, m_{n2}],$$

where $m_{n1}, m_{n2} \asymp nh$, which yields, in conjunction with (39) and (40), the relation (38).

Since the moments of the ε_j's are bounded, we obtain

$$\begin{aligned}\chi_k(t^k) &= cum_k(\langle t, S_n\rangle) \\ &= \sum_{j=1}^{n} cum_k(t' B_n^{-1/2} D_j\beta_j) \\ &= O(\| t \|^k \| B_n^{-1/2} \|^k \sum_{j=1}^{n} \| D_j \|^k) \\ &= O(\| t \|^k (nh)^{-(k-2)/2}),\end{aligned}$$

which implies by $\chi_{2,n} = cum_2(S_n) = I$ that

$$\begin{aligned}\rho_{4,n} &= \max_{k=3,4} \sup_t \left\{ (\frac{1}{k!} |\chi_{k,n}(t^k)| / \|t\|^k_{\chi_{2,n}})^{1/(k-2)} \right\} \\ &= \max_{k=3,4} \sup_t \left\{ O((\|t\|^k (nh)^{-(k-2)/2} \|t\|^{-k})^{1/(k-2)} \right\} \\ &= O((nh))^{-1/2},\end{aligned}$$

which means that condition (II) is satisfied.

Let $\delta > 0$ be arbitrary. Next we show that the characteristic function ξ_n of $S_n = B_n^{-1/2} \sum_{j=1}^{n} D_j\beta_j$ satisfies condition (III_α'') in [Sko86]. Let

$$J_n = \{j \,|\, \lambda_{\min}\{D_j\} \geq C\},$$

where we choose the constant C such that $\# J_n \asymp nh$ holds. Since for $\|t\| > \delta(nh)^{1/2}$ and $j \in J_n$ the estimate

$$\begin{aligned}\|B_n^{-1} D_j t\| &\geq \lambda_{\min}\{B_n^{-1/2}\}\lambda_{\min}\{D_j\} \|t\| \\ &\geq C_1(nh)^{-1/2} (M_2)^{-1/2} \delta(nh)^{1/2} \\ &= b = b(\delta) > 0\end{aligned}$$

holds, we have

$$|E \exp\{it' B_n^{-1/2} D_j\beta_j\}| \leq C_b \tag{41}$$

for $j \in J_n$, $\|t\| > \delta(nh)^{1/2}$, where $C_b < 1$ is defined by (A5). This yields under $\|t\| > \delta(nh)^{1/2}$ that

$$\begin{aligned}|\xi_n(t)| &\leq \prod_{j \in J_n} |E \exp\{it' B_n^{-1/2} D_j\beta_j\}| \\ &\leq (C_b)^{\# J_n} = O((nh)^{-\lambda})\end{aligned}$$

holds for arbitrary $\lambda > 0$, which implies that (III_α'') is satisfied.

Finally, we infer from Remark 5.2 in [Sko81] that (IV) is fulfilled. By Theorem 3.4 in [Sko86] we obtain that the random vector S_n admits an Edgeworth expansion that holds uniformly over any class C of Borel sets with

$$\sup_{B \in C} \int_{(\delta B)^\epsilon} \phi(u)du = O(\epsilon) \quad \text{as} \quad \epsilon \to 0, \tag{42}$$

where $(\delta B)^\epsilon = \{u | \exists t \in B : \|t-u\| \leq \epsilon\}$. The residual term of the expansion is of order $O(\epsilon_n (\log \epsilon_n^{-1})^m)$ for some $m > 0$, which is obviously also of order $o((nh)^{-1})$.

2) identification of the expansion for $\overline{\overline{T}}_n$

To conclude from part 1) the validity of such an expansion, we intend to apply results of [Sko81]. First, we approximate $\overline{\overline{T}}_n$ by

$$T'_n = \frac{\alpha_{n1} + b_n}{V_n^{1/2}} - \frac{1}{2} \frac{(\alpha_{n1} + b_n)\alpha_{n2}}{V_n^{3/2}} + \frac{3}{8} \frac{(\alpha_{n1} + b_n)\alpha_{n2}^2}{V_n^{5/2}}, \tag{43}$$

where $\alpha_{n1} = \sum_{j=1}^n \overline{\overline{W}}_{nj} \varepsilon_j$ and $\alpha_{n2} = \sum_{j=1}^n \overline{W}_{nj}^2 (\varepsilon_j^2 - v_j)$. By Lemma 8.1 in [Neu92] we have for arbitrary $\gamma > 0$ that $\alpha_{n1} = \tilde{O}((nh)^{-1/2+\gamma}, n^{-1})$ and $\alpha_{n2} = \tilde{O}((nh)^{-3/2+\gamma}, n^{-1})$ hold, and because of $h = o(n^{-1/(2r+1)})$ we obtain $b_n = o((nh)^{-1/2})$. This implies in particular $(\alpha_{n2} + V_n)^{-1} = \tilde{O}(nh, n^{-1})$.

Therefore, we have, by a simple Taylor expansion of $(\alpha_{n2} + V_n)^{-1/2}$ at the point $\alpha_{n2} = 0$, that

$$\begin{aligned} \overline{\overline{T}}_n - T'_n &= \tilde{O}\left((|\alpha_{n1}| + |b_n|)|\alpha_{n2}|^3 (nh)^{-7/2}, n^{-1}\right) \\ &= \tilde{O}\left((nh)^{-3/2+4\gamma}, n^{-1}\right). \end{aligned} \tag{44}$$

According to Lemma 3.1, the Edgeworth expansions of $\overline{\overline{T}}_n$ and T'_n coincide up to a term of order $O\left((nh)^{-3/2+4\gamma} + n^{-1}\right)$ and , hence, it suffices to state this expansion for T'_n.

Let $C_n = \operatorname{Cov}\left(\binom{\alpha_{n1}}{\alpha_{n2}}\right)$ and

$$\begin{aligned} g_n(S_n) &= T'_n - b_n V_n^{-1/2} \\ &= \frac{(C_n^{1/2})_1 S_n}{V_n^{1/2}} - \frac{1}{2} \frac{\left((C_n^{1/2})_1 S_n + b_n\right)(C_n^{1/2})_2 S_n}{V_n^{3/2}} \\ &\quad + \frac{3}{8} \frac{\left((C_n^{1/2})_1 S_n + b_n\right)\left((C_n^{1/2})_2 S_n\right)^2}{V_n^{5/2}} \end{aligned}$$

Obviously, g_n obeys (3.4) in [Sko81]. Analogously to that paper we define

$$f_n(S_n) = w_n^{-1/2} g_n(S_n),$$

where $w_n = (Dg_n(0))(Dg_n(0))' = \|V_n^{-1/2}(C_n^{1/2})_1 - \frac{1}{2}V_n^{-3/2}b_n(C_n^{1/2})_2\|^2$. From $C_n = \begin{pmatrix} O((nh)^{-1}) & O((nh)^{-2}) \\ O((nh)^{-2}) & O((nh)^{-3}) \end{pmatrix}$ we conclude $\|(C_n^{1/2})_1\| = O((nh)^{-1/2})$ and $\|(C_n^{1/2})_2\| = O((nh)^{-3/2})$, which yields $w_n = 1+o(1)$ and $\|D^j f_n(0)\| = O((nh)^{-(j-1)/2})$ for $j = 2,3$. Further, we have $D^4 f_n(t) \equiv 0$, that is, Assumption 3.1 in [Sko81] is fulfilled with $p = 4$ and $\lambda_n = O((nh)^{-1})$. By Theorem 3.2 and Remark 3.4 of that paper we obtain that

$$P(f_n(S_n) \in B) = \int_B \tilde{\eta}_n(u)du + o((nh)^{-1})$$

holds uniformly over any class C of Borel sets satisfying (42) in the one - dimensional case. Here $\tilde{\eta}_n$ is the signed measure with characteristic function.

$$\hat{\tilde{\eta}}_n(t) = \exp\left\{it\tilde{\chi}_{1,n} + \frac{(it)^2}{2!}\tilde{\chi}_{2,n}\right\}\left[1 + \frac{(it)^3}{3!}\tilde{\chi}_{3,n} + \frac{1}{2!}\left(\frac{(it)^3}{3!}\tilde{\chi}_{3,n}\right)^2 + \frac{(it)^4}{4!}\tilde{\chi}_{4,n}\right],$$

where $\tilde{\chi}_{\nu,n}$ denotes the ν-th cumulant of $f_n(S_n)$. Hence, we conclude that

$$P(T_n' \in B) = \int_B \eta_n(u)du + o((nh)^{-1}) \tag{45}$$

holds uniformly over $B \in C$, if C obeys (42), where

$$\hat{\eta}_n(t) = \exp\left\{-\frac{t^2}{2}\right\} \exp\left\{it\chi_{1,n} + \frac{(it)^2}{2!}(\chi_{2,n} - 1)\right\} \left[1 + \frac{(it)^3}{3!}\chi_{3,n} + \frac{1}{2!}\left(\frac{(it)^3}{3!}\chi_{3,n}\right)^2 + \frac{(it)^4}{4!}\chi_{4,n}\right], \tag{46}$$

and the $\chi_{\nu,n}$'s are the cumulants of T_n'.

Next, we approximate the first four cumulants $\chi_{\nu,n}$ up to a residual term of order $O((nh)^{-1})$. First, observe that the fourth moment of the third term on the right-hand side of (43) is of order $O((nh)^{-4})$. Hence, this term contributes only terms of order $O((nh)^{-1})$ to the cumulants we look for. Therefore we calculate them on the basis of the cumulants $\kappa_{\nu,n}$ of

$$T_n'' = \frac{\alpha_{n1} + b_n}{V_n^{1/2}} - \frac{1}{2}\frac{(\alpha_{n1} + b_n)\alpha_{n2}}{V_n^{3/2}}$$

which differ from those of T_n' only by a term of order $O((nh)^{-1})$.
It holds

$$\begin{aligned} \kappa_{1,n} &= b_n V_n^{-1/2} - \frac{1}{2}V_n^{-3/2}E\alpha_{n1}\alpha_{n2} + O((nh)^{-1}) \\ &= b_n V_n^{-1/2} - \frac{1}{2}\rho_{n3} + O((nh)^{-1}). \end{aligned}$$

With $\quad T_n''' = T_n'' - b_n V_n^{-1/2} \quad$ we obtain

$$\begin{aligned} E(T_n''')^2 &= \frac{E\alpha_{n1}^2}{V_n} - \frac{b_n}{V_n^{1/2}} \frac{E\alpha_{n1}\alpha_{n2}}{V_n^{3/2}} - \frac{E\alpha_{n1}^2\alpha_{n2}}{V_n^2} + O((nh)^{-1}) \\ &= \frac{\overline{V}_n}{V_n} - \frac{b_n}{V_n^{1/2}}\rho_{n3} + O((nh)^{-1}), \end{aligned}$$

which implies

$$\begin{aligned} \kappa_{2,n} &= E(T_n''')^2 - (ET_n''')^2 + O((nh)^{-1}) \\ &= \frac{\overline{V}_n}{V_n} - \frac{b_n}{V_n^{1/2}}\rho_{n3} + O((nh)^{-1}). \end{aligned}$$

Further, we have

$$\begin{aligned} E(T_n''')^3 &= \frac{E\alpha_{n1}^3}{V_n^{3/2}} - \frac{3}{2}\frac{E\alpha_{n1}^2(\alpha_{n1}+b_n)\alpha_{n2}}{V_n^{5/2}} + O((nh)^{-1}) \\ &= -\frac{7}{2}\rho_{n3} + O((nh)^{-1})\,, \end{aligned}$$

which yields

$$\begin{aligned} \kappa_{3,n} &= E(T_n''')^3 - 3E(T_n''')^2 ET_n''' + 2(ET_n''')^3 \\ &= -2\rho_{n3} + O((nh)^{-1})\,. \end{aligned}$$

Finally, we have

$$\kappa_{4,n} = O((nh)^{-1})\,.$$

Because of $|\kappa_{1,n}| + |\kappa_{2,n} - 1| + |\kappa_{3,n}| = O((nh)^{-1/2})$ and $e^x = 1 + x + O(x^2)$ as $x \to 0$ we have

$$\widehat{\eta}_n(t) = \exp\left\{-\frac{t^2}{2}\right\}\left[1 + it\kappa_{1,n} + \frac{(it)^2}{2!}(\kappa_{2,n}-1) + \frac{(it)^3}{3!}\kappa_{3,n} + O((nh)^{-1}(t^2+t^8))\right]\,,$$

which implies

$$\begin{aligned} P\left(\overline{\overline{T}}_n \le t\right) &= P(T_n' \le t) + O((nh)^{-1}) \\ &= \Phi(t) - (b_n V_n^{-1/2} - \frac{1}{2}\rho_{n3})\phi(t) - \left(\frac{\overline{V}_n - V_n}{V_n} - b_n V_n^{-1/2}\rho_{n3}\right)\frac{t}{2}\phi(t) \\ &\quad -(-2\rho_{n3})\frac{t^2-1}{6}\phi(t) + O((nh)^{-1})\,. \end{aligned}$$

This proves the assertion. $\square$

Proof of Proposition 4.1.

Analogously to the proof of Proposition 3.1, this proof is based on the validity of the Edgeworth expansion of the statistic $S_n^* = B_n^{*\,-1/2} \sum_{j=1}^n \alpha_j^*$, where $\alpha_j^* = \left(nh\overline{\overline{W}}_{nj}\, \varepsilon_j^* \,,\, (nh\overline{\overline{W}}_{nj})^2\, (\varepsilon_j^{*2} - E^*\varepsilon_j^{*2})\right)$ and $B_n^* = \sum_{j=1}^n\, Cov^*\,(\alpha_j^*)$.

First, we prove that the average moments of the α_j^*'s are bounded with a sufficiently high probability. It holds $E^*\,|\varepsilon_j^*|^p = (\gamma\, a^p + (1-\gamma)\, b^p)\; |\widehat{\varepsilon}_j|^p$, and because of

$$\widehat{\varepsilon}_j = \varepsilon_j + \widetilde{O}\,\left((nf)^{-1/2} n^{\delta}\,, n^{-\lambda}\right)$$

for arbitrarily small $\delta > 0$ and arbitrarily large λ , it suffices to consider the empiric average p-th moment of the ε_j's rather than those of the $\widehat{\varepsilon}_j$'s. By Markov's and Whittle's inequalities we have for $s \geq 2$

$$\begin{aligned} & P\left(\left|\sum_{j=1}^n |\, nh\overline{\overline{W}}_{nj}\, \varepsilon_j|^p - \sum_{j=1}^n E|\, nh\overline{\overline{W}}_{nj}\, \varepsilon_j|^p\right|^s > nh\right) \\ = \;& O\left((\sum_{j=1}^n (nh\overline{\overline{W}}_{nj})^{2p})^{s/2} (nh)^{-s}\right) \\ = \;& O\left((nh)^{-s/2}\right), \end{aligned}$$

which yields, by $\sum_{j=1}^n E|nh\overline{\overline{W}}_{nj}\, \varepsilon_j|^p = O(nh)$, that the average p-th moment of $(\alpha_j^*)_1$ is of order $O(nh)$ with exception of a set of events with probability of order $O\left((nh)^{-s/2}\right)$. In the same way we can prove an analogous bound for the average p-th moment of $(\alpha_j^*)_2$.

Now we infer that, with a probability of $1 - O\left((nh)^{-s/2}\right)$, condition (II) of [Sko86] and, by Remark 5.2 of [Sko81], also condition (IV) are satisfied. It remains to prove an analogue of condition (III_α'') for the characteristic function of S_n^* .

In case of the continuous bootstrap version, this condition follows in the same way as in the proof of Proposition 3.1. In case of the discrete bootstrap however, the situation is much more involved. We know from [BR86] that the Edgeworth expansion of a sum of n i.i.d. lattice-valued random vectors is valid only up to a residual term of order $O(n^{-1/2})$. Here we can exploit however the diversity of the distributions of the ε_j^*'s.

Let $K_n \subseteq J_n$ with $\# K_n = O(n^{\alpha})$ for some $\alpha > 0$ with $n^{\alpha} = o(nh)$. We define

$$\widehat{\varepsilon}_{j0} = Y_j - \sum_{k=1}^n W(x_j, f)_k\, m(x_k) \; - \sum_{k \notin K_n} W(x_j, f)_k\, \varepsilon_k.$$

We have on the one hand

$$\begin{aligned}\widehat{\varepsilon}_j &= \widehat{\varepsilon}_{j0} + \sum_{k\in K_n} W(x_0, f)_k\, \varepsilon_k \\ &= \widehat{\varepsilon}_{j0} + \tilde{O}\left((nh)^{-1} n^{\alpha/2}\, n^{\delta}\,,\ n^{-\lambda}\right),\end{aligned}$$

and on the other hand that the $\widehat{\varepsilon}_{j0}$'s are, conditioned on $\mathcal{Y}_n = \{Y_k\}_{k\notin K_n}$, independent. Let $\varepsilon^*_{j0} = \begin{cases} a\,\widehat{\varepsilon}_{j0} & if \quad \varepsilon^*_j = a\,\widehat{\varepsilon}_j \\ b\,\widehat{\varepsilon}_{j0} & if \quad \varepsilon^*_j = b\,\widehat{\varepsilon}_j \end{cases}$ and $\beta^*_{j0} = (\varepsilon^*_{j0}\,,\, \varepsilon^{*2}_{j0} - \widehat{\varepsilon}^2_{j0})'$. Then

$$\begin{aligned} E^*\exp\{i\,t' B_n^{-1/2} D_j \beta^*_{j0}\} &= \gamma \exp\left\{ i\,t' B_n^{-1/2} D_j \begin{pmatrix} a\,\widehat{\varepsilon}_{j0} \\ (a^2-1)\,\widehat{\varepsilon}^2_{j0} \end{pmatrix} \right\} \\ &\quad + (1-\gamma)\exp\left\{ i\,t' B_n^{-1/2} D_j \begin{pmatrix} b\,\widehat{\varepsilon}_{j0} \\ (b^2-1)\,\widehat{\varepsilon}^2_{j0} \end{pmatrix} \right\} \\ &= \exp\left\{ i\,t' B_n^{-1/2} D_j \begin{pmatrix} a\,\widehat{\varepsilon}_{j0} \\ (a^2\,\widehat{\varepsilon}^2_{j0} \end{pmatrix} \right\} \left[\gamma + (1-\gamma)\exp\{i\,\tilde{t}'\,\beta^*_{j0}\}\right], \end{aligned}$$

where $\tilde{t} = \operatorname{diag}[b-a,\ b^2-a^2]\, D_j\, B_n^{-1/2}\, t$.

Let $c_j = \widehat{\varepsilon}_{j0} - \varepsilon_j = m(x_j) - \sum_{k=1}^n W(x_j, f)_k\, m(x_k) - \sum_{k\in K} W(x_j, f)_k\, \varepsilon_k$ and $\mathcal{C}_1 = \{\mathcal{Y}_n \big|\ |c_j| \le (nf)^{-1/2} n^{\delta} \quad \forall\, j \in K_n\}$.

In the sequel let λ be an arbitrarily large constant. Then

$$P(\overline{\mathcal{C}}_1) = O\,(n^{-\lambda})\,.$$

Under fixed $\mathcal{Y}_n \in \mathcal{C}_1$, we obtain for $\|t\| > \delta(nh)^{1/2}$

$$\begin{aligned} &\left(E\left(|E^*\exp\{it'B_n^{-1/2}D_j\beta^*_{j0}\}\ |\ \big|\mathcal{Y}_n\right)\right)^2 \\ &= \left(E\left(|\gamma + (1-\gamma)\exp\{i\,\tilde{t}'\beta^*_{j0}\}|\ \big|\mathcal{Y}_n\right)\right)^2 \\ (47)\quad &\le E\left(|\gamma + (1-\gamma)\exp\{i\,\tilde{t}'\beta^*_{j0}\}|^2\ \big|\mathcal{Y}_n\right) \\ &= 1 + \gamma(1-\gamma)\left[E\left(\exp\{i\,\tilde{t}'\beta^*_{j0}\} + \exp\{-i\tilde{t}'\beta^*_{j0}\}\big|\mathcal{Y}_n\right) - 2\right] \\ &\le 1 + \gamma(1-\gamma)\left[\left|E\left(\exp\{i\,\tilde{t}'\beta^*_{j0}\}\ \big|\mathcal{Y}_n\right)\right| + \left|E\left(\exp\{i\,\tilde{t}'\beta^*_{j0}\}\ \big|\mathcal{Y}_n\right)\right| - 2\right] \\ &\le C^2 \end{aligned}$$

for some $C < 1$, which does not depend on t and $\mathcal{Y}$. Here the last inequality in (47) holds true, since from $\|t\| > \delta(nh)^{1/2}$ and $|c_j| \le (nf)^{-1/2}\, n^{\delta}$ the relation

$$\left\| \begin{pmatrix} \tilde{t}_1 + 2c_j\tilde{t}_2 \\ \tilde{t}_2 \end{pmatrix} \right\| \ \ge\ \delta'(nh)^{1/2}$$

follows, which yields by (41) from the proof of Proposition 3.1

$$\begin{aligned}
&\Big|E\Big(\exp\{i\,\tilde{t}'\beta^*_{j0}\}\,\Big|\mathcal{Y}_n\Big)\Big| \\
&= \Big|E\Big(\exp\{i\,[\,(\tilde{t}_1+2c_j\tilde{t}_2)\varepsilon_j+\tilde{t}_2\varepsilon_j^2+\tilde{t}_1c_j+\tilde{t}_2c_j^2\,]\}\,\Big|\mathcal{Y}_n\Big)\Big| \\
&= \Big|E\Big(\exp\{i\,[\,(\tilde{t}_1+2c_j\tilde{t}_2)\varepsilon_j+\tilde{t}_2\varepsilon_j^2\,]\}\,\Big|\mathcal{Y}_n\Big)\Big| \\
&\le \tilde{C} < 1
\end{aligned}$$

uniformly over $\mathcal{Y}_n \in \mathcal{C}_1$.

Let $Z_j = 1-\Big|\,E^*\exp\{i\,t'\,B_n^{-1/2}\,D_j\,\beta^*_{j0}\,\}\Big|$ and $\mathcal{C}_2 = \Big\{\,\mathcal{Y}\,\Big|\,(\#K_n)^{-1}\sum_{k\in K} Z_k \ge \frac{C}{2}\,\Big\}$.
Since the Z_j's are bounded random variables, which are, conditioned on $\mathcal{Y}_n$, independent, we obtain by Markov's and Whittle's inequalities

$$P(\overline{\mathcal{C}_2}) \le P(\overline{\mathcal{C}_1}) + P(\overline{\mathcal{C}_2}\cap\mathcal{C}_1) = O(n^{-\lambda}).$$

Let $\mathcal{C}_3 = \{\mathcal{Y}\big|\ |\hat{\varepsilon}_j - \hat{\varepsilon}_{j0}\,| \le (nh)^{-1}\,n^{\alpha/2}\,n^{\delta}$ and $|\hat{\varepsilon}_j\,| \le n^{\delta}$ for $j \in K_n\}$. By Lemma 8.1 in [Neu92] we obtain

$$P(\mathcal{C}_3) = O(n^{-\lambda}).$$

If we choose α and δ small enough such that $n^{\alpha/2}\,n^{3\delta} = O\left((nh)^{1/2}\right)$, then we have for $\beta^*_j = (nh\,\overline{\overline{W}}_{nj}\,\varepsilon^*_j\,,\,(nh\,\overline{W}_{nj}\,\varepsilon^*_j)^2\,)'$ that
$\Big|\Big|\beta^*_j - \beta^*_{j0}\Big|\Big| = O\left(|\varepsilon^*_j - \varepsilon^*_{j0}\,|(1+|\varepsilon^*_j\,|)\right) = O\left((nh\,)^{-1/2}\,n^{-\delta}\,\right)$ for $\mathcal{Y} \in \mathcal{C}_3$.
Let now $\mathcal{Y} \in \mathcal{C} = \{\mathcal{Y}|\,\mathcal{Y}_n \in \mathcal{C}_1\} \cap \mathcal{C}_2 \cap \mathcal{C}_3$. Then

$$\begin{aligned}
&(\#\,K_n)^{-1}\sum_{j\in K_n}\ |\,E^*\exp\{i\,t'\,B_n^{-1/2}\,D_j\,\beta^*_j\,\}\,| \\
&\le (\#\,K_n)^{-1}\sum_{j\in K_n}\ |\,E^*\exp\{i\,t'\,B_n^{-1/2}\,D_j\,\beta^*_{j0}\,\}\,| \\
&\quad + O\left(||\,t\,||\ ||\,B_n^{-1/2}\,||\ ||\,\beta^*_j - \beta^*_j\,||\right) \\
&\le 1-\frac{C}{2} + O\left(||\,t\,||\,(nh)^{-1}\,n^{\delta}\,\right) \\
&\le 1-\frac{C}{4}
\end{aligned}$$

for $\delta\,(nh)^{1/2} < ||\,t\,|| < \delta''\,nh\,n^{\delta}$.
In other words, we have shown that, for *fixed* t, $\delta\,(nh)^{1/2} < ||\,t\,|| < \delta''\,nh\,n^{\delta}$,

$$\begin{aligned}
(48)\qquad p_t &= P\left((\#K_n)^{-1}\sum_{j\in K_n}\ |\,E^*\,\exp\{\,i\,t'\,B_n^{-1/2}\,D_j\,\beta^*_j\,\}\,| \le 1-\frac{c}{4}\right) \\
&= O\left(n^{-\lambda}\,\right)
\end{aligned}$$

holds. To achieve uniformity in t, we state (48) on a sufficiently dense grid. Observe that

$$\left| E^* \exp\{ i\, t_1' B_n^{-1/2} D_j \beta_j^* \} - E^* \exp\{ i\, t_2' B_n^{-1/2} D_j \beta_j^* \} \right| = O(\|t_1 - t_2\| \, \|B_n^{-1/2}\| \, \|\beta_j^*\|)$$

holds. Let

$$\mathcal{C}_4 = \{\mathcal{Y} \mid |\beta_j^*| \le n^\delta \quad \text{for} \quad j \in K_n \}.$$

Then we have, again by Lemma 8.1 in [Neu92],

$$P(\overline{\mathcal{C}}_4) = O(n^{-\lambda}).$$

Now, it is enough to prove (48) on an grid $T_n = \{t_{n1}, \dots, t_{n\,c(n)}\}$ with an algebraic number $c(n)$ of points such that for all t with $\delta(nh)^{1/2} < \|t\| < \delta''(nh)n^\delta$ there exists a $\tilde{t} = \tilde{t}(t) \in T_n$ with $\|t - \tilde{t}(t)\| = O((nh)^{1/2} n^{-2\delta})$.

Having this, we conclude

$$P\left((\#K_n)^{-1} \sum_{j \in K_n} |E^* \exp\{ i\, t' B_n^{-1/2} D_j \beta_j^* \}| \le 1 - \frac{C}{8} \text{ for all } \delta(nh)^{1/2} < \|t\| < \delta'' \right)$$
$$\le \sum_{t \in T_n} p_t + P(\overline{\mathcal{C}}_4) = O(n^{-\lambda})$$

This yields, with a probability exceeding $1 - O(n^{-\lambda})$, that

$$\begin{aligned}
\left| E^* \{ i\, t' S_n^* \} \right| &= \left| \prod_{j=1}^{n} E^* \exp\{ i\, t' B_n^{-1/2} D_j \beta_j^* \} \right| \\
&\le \left((\#K_n)^{-1} \sum_{j \in K_n} |E^* \exp\{ i\, t' B_n^{-1/2} D_j \beta_j^* \}| \right)^{\#K_n} \\
&\le \left(1 - \frac{c}{8} \right)^{\#K_n} = O(n^{-\lambda})
\end{aligned}$$

for arbitrary $\lambda > 0$. Hence, (III_α'') is fulfilled with a probability exceeding $1 - O(n^{-\lambda})$, which yields the validity of the Edgeworth expansion of S_n^*. The rest of this proof goes completely analogous to part 2) of the proof of Proposition 3.1 . □

References

[BR86] R. N. Bhattacharya and R. R. Rao. *Normal Approximation and Asymptotic Expansions.* Krieger, Malabar, Florida, 1986.

[GM79] T. Gasser and H.-G. Müller. Kernel estimation of regression functions. In *Smoothing techniques for curve estimation. Lecture Notes in Math. 757*, pages 23 – 68. Springer, New York, 1979.

[Hal91] P. Hall. Edgeworth expansions for nonparametric density estimators, with applications. *Statistics*, 22:215 – 232, 1991.

[Hal92] P. Hall. On bootstrap confidence intervals in nonparametric regression. *Ann. Stat.*, 6:434 – 451, 1992.

[HHJ91] W. Härdle, S. Huet, and E. Jolivet. Better bootstrap confidence intervals for regression curve estimation. *submitted to Statistics*, 1991.

[HHM88] W. Härdle, P. Hall, and J. S. Marron. How far are automatically chosen regression smoothing parameters from their optimum? (with discussion). *J. Amer. Statist. Assoc.*, 83:86 – 99, 1988.

[HM90] W. Härdle and E. Mammen. Bootstrap methods in nonparametric regression. CORE Discussion Paper 9058, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, 1990. Ann. Statist., accepted.

[Neu92] M. H. Neumann. On completely data-driven pointwise confidence intervals in nonparametric regression. Rapport technique 92 - 02, INRA, Département de Biométrie, Jouy-en-Josas, France, 1992.

[Sko81] I. M. Skovgaard. Transformations of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Statist.*, 8:207 – 217, 1981.

[Sko86] I. M. Skovgaard. On multivariate Edgeworth expansions. *Internat. Statist. Rev.*, 54(2):169 – 186, 1986.

INSTITUT FÜR ANGEWANDTE ANALYSIS UND STOCHASTIK (IAAS), MOHRENSTRASSE 39, D – O – 1086 BERLIN

Veröffentlichungen des Instituts für Angewandte Analysis und Stochastik

Preprints 1992

1. D.A. Dawson and J. Gärtner: Multilevel large deviations.
2. H. Gajewski: On uniqueness of solutions to the drift–diffusion–model of semiconductor devices.
3. J. Fuhrmann: On the convergence of algebraically defined multigrid methods.
4. A. Bovier and J.-M. Ghez: Spectral properties of one–dimensional Schrödinger operators with potentials generated by substitutions.
5. D.A. Dawson and K. Fleischmann: A super–Brownian motion with a single point catalyst.
6. A. Bovier, V. Gayrard: The thermodynamics of the Curie–Weiss model with random couplings.
7. W. Dahmen, S. Prößdorf, R. Schneider: Wavelet approximation methods for pseudodifferential equations I: stability and convergence.
8. A. Rathsfeld: Piecewise polynomial collocation for the double layer potential equation over polyhedral boundaries. Part I: The wedge, Part II: The cube.
9. G. Schmidt: Boundary element discretization of Poincaré–Steklov operators.
10. K. Fleischmann, F. I. Kaj: Large deviation probability for some rescaled superprocesses.
11. P. Mathé: Random approximation of finite sums.
12. C.J. van Duijn, P. Knabner: Flow and reactive transport in porous media induced by well injection: similarity solution.
13. G.B. Di Masi, E. Platen, W.J. Runggaldier: Hedging of options under discrete observation on assets with stochastic volatility.
14. J. Schmeling, R. Siegmund–Schultze: The singularity spectrum of self-affine fractals with a Bernoulli measure.
15. A. Koshelev: About some coercive inequalities for elementary elliptic and parabolic operators.
16. P.E. Kloeden, E. Platen, H. Schurz: Higher order approximate Markov chain filters.

17. H.M. Dietz, Y. Kutoyants: A minimum–distance estimator for diffusion processes with ergodic properties.

18. I. Schmelzer: Quantization and measurability in gauge theory and gravity.

19. A. Bovier, V. Gayrard: Rigorous results on the thermodynamics of the dilute Hopfield model.

20. K. Gröger: Free energy estimates and asymptotic behaviour of reaction–diffusion processes.

21. E. Platen (ed.): Proceedings of the 1^{st} workshop on stochastic numerics.

22. S. Prößdorf (ed.): International Symposium "Operator Equations and Numerical Analysis" September 28 – October 2, 1992 Gosen (nearby Berlin).

23. K. Fleischmann, A. Greven: Diffusive clustering in an infinite system of hierarchically interacting diffusions.

24. P. Knabner, I. Kögel–Knabner, K.U. Totsche: The modeling of reactive solute transport with sorption to mobile and immobile sorbents.

25. S. Seifarth: The discrete spectrum of the Dirac operators on certain symmetric spaces.

26. J. Schmeling: Hölder continuity of the holonomy maps for hyperbolic basic sets II.

27. P. Mathé: On optimal random nets.

28. W. Wagner: Stochastic systems of particles with weights and approximation of the Boltzmann equation. The Markov process in the spatially homogeneous case.

29. A. Glitzky, K. Gröger, R. Hünlich: Existence and uniqueness results for equations modelling transport of dopants in semiconductors.

30. J. Elschner: The h-p–version of spline approximation methods for Mellin convolution equations.

31. R. Schlundt: Iterative Verfahren für lineare Gleichungssysteme mit schwach besetzten Koeffizientenmatrizen.

32. G. Hebermehl: Zur direkten Lösung linearer Gleichungssysteme auf Shared und Distributed Memory Systemen.

33. G.N. Milstein, E. Platen, H. Schurz: Balanced implicit methods for stiff stochastic systems: An introduction and numerical experiments.