The Annals of Statistics
1991, Vol. 19, No. 4, 2226–2231

RATE OF CONVERGENCE FOR THE WILD BOOTSTRAP IN NONPARAMETRIC REGRESSION[1]

BY R. CAO-ABAD

Universidad de Santiago de Compostela

This paper concerns the distributions used to construct confidence intervals for the regression function in a nonparametric setup. Some rates of convergence for the normal limit, its plug-in approach and the wild bootstrap are obtained conditionally on the explanatory variable X and also unconditionally. The bound found for the wild bootstrap approximation is slightly better (by a factor $n^{-1/45}$) than the bounds given by the plug-in approach or the CLT for the conditional probability. On the contrary, the unconditional bounds present a different feature: the rate obtained when approximating by the CLT improves the one given by the plug-in approach by a factor of $n^{-8/45}$, while this last one performs better than the wild bootstrap approximation and the corresponding ratio is $n^{-1/45}$. It should be mentioned that these two sequences, especially the last one, tend to zero at an extremely slow rate.

1. Motivation and background. Nonparametric regression smoothing includes many techniques to estimate the regression function without making assumptions about its shape. It is important to develop some ways of recognizing how accurate the estimation is. A way to do this is to construct confidence intervals for the unknown regression function m at each point x. We will only treat the case of univariate response variable Y and univariate explanatory variable X. The case of multivariate X may be treated in a similar way.

A classical means of constructing confidence intervals for $m(x)$ consists of using the limit distribution of the properly normalized difference between $m(x)$ and some estimator. For the kernel estimator $\hat{m}_h(x)$ studied by Nadaraya (1964) and Watson (1964), based on some estimate of the ISE or MISE bandwidth, the limit distribution of $(nh)^{1/2}(\hat{m}_h(x) - m(x))$ is $N(B, V)$ [see Härdle (1990)], where

$$B = \tfrac{1}{2}c_0^{5/2}d_K f(x)^{-1}(m''(x)f(x) + 2m'(x)f'(x)),$$

$$V = f(x)^{-1}c_K\sigma^2(x),$$

where $c_K = \int K(t)^2\,dt$, $d_K = \int t^2K(t)\,dt$, c_o is the constant such that $hn^{1/5}$ tends to c_o in probability and f is the density function of the explanatory variable X. The confidence intervals naturally lay in the limit distribution and

Received July 1989; revised November 1990.

[1]Supported by the Deutsche Forschungsgemeinschaft, SFB 303. This research was carried out during a visit of the author to the Rechts- und Staatswissenschaftliche Fakultät, Wirtschaftstheoretische Abteilung II, University of Bonn.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62G99.

Key words and phrases. Bootstrap, kernel smoothing, nonparametric regression.

their accuracy depends on how fast the theoretical distribution converges to its limit.

To use the bootstrap method is an alternative way to find confidence intervals. The so-called wild bootstrap introduced by Härdle and Mammen (1989) is available for the regression case. It has been also used by Härdle and Marron (1991) to construct simultaneous error bars.

Given the observed sample $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ of i.i.d. observations, the method proceeds as follows:

1. Construct the residuals $\hat{\varepsilon}_i = Y_i - \hat{m}_h(X_i)$, $i = 1, 2, \dots, n$, where $\hat{m}_h$ is the Nadaraya–Watson kernel estimator based on a bandwidth h of order $n^{-1/5}$ which usually appears to be some estimator of the MISE or the ISE bandwidth.
2. For each index i, draw the bootstrap residual $\hat{\varepsilon}_i^*$ from a two-point centered distribution in order that its second and third moments fit the square and the cubic power of the residual $\hat{\varepsilon}_i$. This distribution is found to be the one that gives probability $\gamma = (5 + 5^{1/2})/10$ to the point $a = \hat{\varepsilon}_i(1 - 5^{1/2})/2$ and $1 - \gamma$ to $b = \hat{\varepsilon}_i(1 + 5^{1/2})/2$.
3. Define the bootstrap observation $Y_i^* = \hat{m}_g(X_i) + \hat{\varepsilon}_i^*$ for each i. At this point, the bandwidth g has to be asymptotically larger than h as Härdle and Marron (1991) have pointed out.
4. Finally, construct the Nadaraya–Watson estimator $\hat{m}_h^*(x)$ based on the bootstrap sample $(X_1, Y_1^*), (X_2, Y_2^*), \dots, (X_n, Y_n^*)$.

The bootstrap approach consists of approximating the distribution of $(nh)^{1/2}(\hat{m}_h(x) - m(x))$ by the bootstrap distribution of $(nh)^{1/2}(\hat{m}_h^*(x) - \hat{m}_g(x))$.

The content of the paper is a study of different rates of convergence for the normal approximations as well as for those given by the wild bootstrap. Some results are mentioned in Section 2; they are not proved here because of their similarity to the proofs in Cao-Abad and González-Manteiga (1989). On the other hand, we will give the main ideas behind the proof of the main result (concerning the wild bootstrap and also stated in that section) at the end of this paper.

2. Normal and bootstrap approximations. In this section, we will denote by $\hat{f}_h(x)$ the usual kernel estimator of the density function f based on a bandwidth h [see Silverman (1986)]. We will use the letter h for the classical MISE bandwidth of order $n^{-1/5}$. The bandwidth g involved in the bootstrap resampling will be asymptotically larger than h as we will justify later. We also denote by Φ the standard normal c.d.f. and ε is the error present in the regression setup, that is to say, $Y = m(X) + \varepsilon$.

Several assumptions will be made.

ASSUMPTION 1. The functions $m(x)$ and $f(x)$ are four times continuously differentiable in their support.

ASSUMPTION 2. The kernel function is symmetric, nonnegative and satisfies $c_K < \infty$, $d_K < \infty$ and $\int K(t)^3\, dt < \infty$.

ASSUMPTION 3. $\sup_x \mathbb{E}(\varepsilon^3 | X = x) < \infty$.

ASSUMPTION 4. $\inf_x f(x) > 0$, where the inf is taken over the support of f.

ASSUMPTION 5. The functions $\mu_i(x) = \mathbb{E}(|Y|^i\, | X = x)$, $i = 1, 2, 3$ and $\sigma^2(x) = \mathrm{Var}(Y | X = x)$ are twice continuously differentiable.

Let us define the following approximations of the values B and V:

$$B_n = (nh)^{1/2} n^{-1} \hat{f}_h(x)^{-1} \sum_{i=1}^{n} K_h(x - X_i)(m(X_i) - m(x)),$$

$$V_n = n^{-1} h \hat{f}_h(x)^{-2} \sum_{i=1}^{n} \sigma^2(x) K_h(x - X_i)^2.$$

The following representation is useful in order to avoid randomness in the denominator of certain ratios:

$$\begin{aligned}\hat{m}_h(x) - m(x) &= (\hat{m}_h(x) - m(x)) \hat{f}_h(x) f(x)^{-1} \\ &\quad + (\hat{m}_h(x) - m(x))\left(1 - \hat{f}_h(x) f(x)^{-1}\right) \\ &= L_h(x) + O_P(n^{-4/5}),\end{aligned}$$

where the linearization L_h is given by

$$L_h(x) = n^{-1} f(x)^{-1} \sum_{i=1}^{n} K_h(x - X_i)(Y_i - m(x)).$$

Let us denote by $\mathbb{P}^{Y|X}$ the probability conditional on the sample $(X_1, X_2, \ldots, X_n)$. Under Assumptions 1, 2 and 3 and making use of the Berry–Esseen inequality [see Petrov (1975)], we can get a bound for the distance between the conditional probability and the distribution $N(B_n, V_n)$

$$\begin{aligned}(2.1)\quad &\sup_z \left| \mathbb{P}^{Y|X}\left\{(nh)^{1/2}(\hat{m}_h(x) - m(x)) \le z\right\} - \Phi\left(V_n^{-1/2}(z - B_n)\right)\right| \\ &\quad = O_P(n^{-2/5}).\end{aligned}$$

From this fact, and the usual variance and bias approximations, it is easy to show that

$$\begin{aligned}(2.2)\quad &\sup_z \left| \mathbb{P}^{Y|X}\left\{(nh)^{1/2}(\hat{m}_h(x) - m(x)) \le z\right\} - \Phi\left(V^{-1/2}(z - B)\right)\right| \\ &\quad = O_P(n^{-1/5}).\end{aligned}$$

Of course, expressions (2.1) and (2.2) cannot be used directly to find confidence intervals for $m(x)$. We have to estimate B and V and plug these estimators $\hat{B}$ and $\hat{V}$ in (2.2). For instance, we can estimate all the unknown terms in B and V, depending on m and f, by using the kernel technique once more. For this purpose, a common bandwidth may be used for all the terms but it makes much more sense to choose an appropriate bandwidth for each of them. In both cases, the bias and variance for each of the terms may be calculated depending on the bandwidth used. The optimal bandwidth and the order it gives can be computed by minimizing that function of the smoothing parameter. It turns out that the dominant part of the orders comes from the estimation of m''. After the mentioned calculations, these approximations may be found to be $\hat{B} - B = O_P(n^{-2/9})$ and $\hat{V} - V = O_P(n^{-2/5})$. This argument together with (2.2) leads us to the following rate for the plug-in approach:

$$\sup_z \left| \mathbb{P}^{Y|X}\left\{ (nh)^{1/2} (\hat{m}_h(x) - m(x)) \le z \right\} - \Phi\left(\hat{V}^{-1/2} (z - \hat{B}) \right) \right| = O_P(n^{-1/5}).$$

We may try to find a bound in the unconditional case for the approximations already mentioned. Under Assumptions 1, 2 and 5, the following one can be stated for the CLT approach:

$$\sup_z \left| \mathbb{P}\left\{ (nh)^{1/2} (\hat{m}_h(x) - m(x)) \le z \right\} - \Phi(V^{-1/2} (z - B)) \right| = O(n^{-2/5}).$$

It happens again that B and V have to be estimated and, then, by using the same kind of argument as above,

$$\sup_z \left| \mathbb{P}\left\{ (nh)^{1/2} (\hat{m}_h(x) - m(x)) \le z \right\} - \Phi\left(\hat{V}^{-1/2} (z - \hat{B}) \right) \right| = O_P(n^{-2/9}).$$

Let us pay attention to the bootstrap approximation and define the wild bootstrap bias and variance:

$$B_n^* = (nh)^{1/2} n^{-1} \hat{f}_h(x)^{-1} \sum_{i=1}^{n} K_h(x - X_i) \left(\hat{m}_g(X_i) - \hat{m}_g(x) \right),$$

$$V_n^* = n^{-1} h \hat{f}_h(x)^{-2} \sum_{i=1}^{n} K_h(x - X_i)^2 \hat{\varepsilon}_i^2.$$

As in the analogy between the bootstrap and nonbootstrap situations, expression (2.1) can be reformulated in terms of the bootstrap distribution

$$\sup_z \left| \mathbb{P}^*\left\{ (nh)^{1/2} (\hat{m}_h^*(x) - \hat{m}_g(x)) \le z \right\} - \Phi\left(V_n^{*\,-1/2} (z - B_n^*) \right) \right| = O_P(n^{-2/5}), \tag{2.3}$$

where $\mathbb{P}^*$ is the probability measure under the wild bootstrap resampling plan. This result can be proved using the Berry–Esseen inequality and Assumptions 1 and 2.

Making use of the approximations for bias and variance, we can state our main result.

THEOREM. *Under Assumptions* 1, 2, 3, 4 *and* 5, *the approximation of the bootstrap probability to the conditional one is given by*

$$\sup_z \Big|\mathbb{P}^{Y|X}\{(nh)^{1/2}(\hat{m}_h(x) - m(x)) \le z\} \\ - \mathbb{P}^*\Big\{(nh)^{1/2}\big(\hat{m}_h^*(x) - \hat{m}_g(x)\big) \le z\Big\}\Big| = O_P(n^{-2/9}), \tag{2.4}$$

where the bandwidth g is chosen to be of order $n^{-1/9}$.

As a consequence of this theorem, a normal distribution fitting the bootstrap in bias and variance can be used instead of the bootstrap distribution. For this approximation the same bound holds

$$\sup_z \Big|\mathbb{P}^{Y|X}\{(nh)^{1/2}(\hat{m}_h(x) - m(x)) \le z\} - \Phi\big(V_n^{*\,-1/2}(z - B_n^*)\big)\Big| \\ = O_P(n^{-2/9}).$$

The wild bootstrap procedure can approximate the unconditional probability. The following expression shows a sequence that bounds the approximation error

$$\sup_z \Big|\mathbb{P}\{(nh)^{1/2}(\hat{m}_h(x) - m(x)) \le z\} - \mathbb{P}^*\Big\{(nh)^{1/2}\big(\hat{m}_h^*(x) - \hat{m}_g(x)\big) \le z\Big\}\Big| \\ = O_P(n^{-1/5}).$$

Observe that all the rates obtained above are not proved to be precise, in the sense that the expressions on the left-hand side tend to zero no slower than the sequences on the right-hand side. This implies that we cannot directly compare the accuracy of the different methods by means of the rates already obtained. A deeper analysis, for instance, by using Edgeworth expansions, should be done to know the precise rates.

The choice of the bandwidth in the bootstrap resampling plan is worth mentioning. This pilot bandwidth makes an oversmoothed estimation which is needed to make the bootstrap work. In particular, it is seen in the next section that the optimal asymptotic order of the bootstrap bandwidth is $n^{-1/9}$. Although we know the asymptotic expression for the pilot bandwidth, it depends on the unknown curves. So some bandwidth selection methods must be extended to this setup in practice.

3. Proof of the theorem. Computing bias and variance of each one of the following quantities, it is easy to deduce that $V_n^* - V = O_P(n^{-2/5})$ and $V_n = V = O_P(n^{-2/5})$. Both facts imply that $V_n^* - V_n = O_P(n^{-2/5})$.

On the other hand, Härdle and Marron (1991) show that

$$\mathbb{E}^{Y|X}\Big((nh)^{-1}(B_n^* - B_n)^2\Big) = O_P\big(h^4(C_1 n^{-1} g^{-5} + C_2 g^4)\big),$$

where $C_1 = \frac{1}{4}c_{K''}d_K^4\sigma^2(x)f(x)^{-1}$ and

$$C_2 = \left(\tfrac{1}{2}d_K\right)^4\left((m(x)f(x))^{IV} - (m(x)f''(x))''\right)^2 f(x)^{-2}.$$

This consideration leads us to

$$B_n^* - B_n = O_P\left(\left((nh^5)\left(C_1 n^{-1}g^{-5} + C_2 g^4\right)\right)^{1/2}\right)$$
$$= O_P\left(\left(C_1 n^{-1}g^{-5} + C_2 g^4\right)^{1/2}\right)$$

and this order is minimized for

$$g = \left(5C_1(4nC_2)^{-1}\right)^{1/9}.$$

This means that the optimal bandwidth g is of order $n^{-1/9}$ and for this bandwidth $B_n^* - B_n = O_P(n^{-2/9})$. As the standard normal density function and its first derivative are bounded, then we can state

$$\sup_z\left|\Phi\left(V_n^{*-1/2}(z - B_n^*)\right) - \Phi\left(V_n^{-1/2}(z - B_n)\right)\right|$$
$$= O_P((B_n^* - B_n) + (V_n^* - V_n)) = O_P(n^{-2/9}).$$

From this result and expressions (2.1) and (2.3), follows (2.4). □

Acknowledgments. The author would like to thank Wolfgang Härdle for the suggestion of the subject and some interesting ideas and María Celia Rodríguez Campos and Wenceslao González Manteiga for discussing this work. He also greatly appreciates the helpful suggestions coming from an Associate Editor and two anonymous referees.

REFERENCES

Cao-Abad, R. and González-Manteiga, W. (1989). Bootstrap methods in regression smoothing: An alternative procedure to the wild resampling plan. Unpublished.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge Univ. Press.

Härdle, W. and Mammen, E. (1989). Comparing nonparametric versus parametric regression fits. Discussion paper A-177, Univ. Bonn.

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796.

Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **10** 186–190.

Petrov, V. (1975). *Sums of Independent Random Variables*. Springer, New York.

Silverman, B. W. (1986). *Density Estimation*. Chapman and Hall, New York.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

Departamento de Estadistica e
Investigación Operativa
Facultad de Matemáticas
Universidad de Santiago de Compostela
15771 Santiago de Compostela
Spain