Communications in Statistics—Theory and Methods, 43: 3720–3725, 2014

ISSN: 0361-0926 print / 1532-415X online
DOI: 10.1080/03610926.2012.700377

Extended Glivenko–Cantelli Theorem in Nonparametric Regression

FUXIA CHENG,[1] JIGAO YAN,[2] AND LIJIAN YANG[3,4]

[1]Department of Mathematics, Illinois State University, Normal, IL, USA
[2]School of Mathematical Sciences, Soochow University, Suzhou, P.R. China
[3]Center for Advanced Statistics and Econometrics Research, Suzhou, P.R. China
[4]Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA

In this paper, we consider the uniform strong consistency of the cumulative distribution function estimator in nonparametric regression. We obtain the extended Glivenko–Cantelli theorem for the residual-based empirical distribution function.

Keywords Empirical process; Glivenko–Cantelli theorem; Nonparametric regression; Residuals.

Mathematical Subject Classification Primary 62G08; Secondary 62G05, 60F15.

1. Introduction

We consider the nonparametric regression model defined as follows:

$$Y = m(X) + \varepsilon,$$

where X and Y are one-dimensional random variables, with X taking values in [0, 1]; the regression function m is unknown; and the error ε is a random variable with

$$E(\varepsilon|X) = 0 \quad a.s., \tag{1.1}$$

and unknown density function g and cumulative distribution function (cdf) G.

In the regression analysis we are concerned with here one observes n independent identically distributed (i.i.d.) copies $\{(X_i, Y_i);\ 1 \leq i \leq n\}$ of (X, Y), such that for some real valued function $m(x),\ x \in [0, 1]$,

$$\varepsilon_i = Y_i - m(X_i), \qquad i = 1,\ 2, \ldots, n,$$

are i.i.d. random variables with the unknown distribution G and density g.

Received February 25, 2012; Accepted May 28, 2012.
Address correspondence to Jigao Yan, School of Mathematical Sciences, Soochow University, Suzhou 215006, P. R. China; E-mail: yanjigao@suda.edu.cn

In view of (1.1), it implies that

$$m(x) = E(Y|X = x).$$

In literature, the main focus has been the estimation of the regression function $m(x)$. But it is also of interest and importance to know the nature of the error distribution after estimating a regression function nonparametricly.

In nonparametric regression, Akritas and Van Keilegom (2001) define the error distribution function estimator by using the quantile function and obtain the asymptotic distribution; Cheng (2002) considers the consistency of the histogram type error density function and the empirical error distribution function estimators while Cheng (2004) considers the weak and strong uniform consistency of a kernel error density estimation in nonparametric regression.

Akritas and Van Keilegom (2001) show that the empirical process based on the full sample residuals converges weakly to a zero-mean Gaussian process with a covariance function that depends on the underlying error density in a complicated way. In Cheng (2005), the properly standardized empirical process based on the split sample residuals is shown to be uniformly close to the similarly standardized empirical process of the errors.

The focus of this paper is to investigate the strong consistency of the empirical distribution function based on nonparametric residuals. We will show the extended Glivenko–Cantelli theorem for the residual-based empirical distribution function.

Let $m_n(x)$ denote the well-known Nadaraya–Watson (1964) kernel regression estimator:

$$m_n(x) := \sum_{i=1}^{n} \frac{Y_i\, K_{h_n}(x - X_i)}{\sum_{j=1}^{n} K_{h_n}(x - X_j)}, \quad K_h(x) := \frac{1}{h} K\left(\frac{x}{h}\right), \quad x \in [0, 1],$$

where h_n is the usual band width sequence of positive numbers tending to zero, and K is the kernel density function. Denote

$$d_n(x) := m_n(x) - m(x), \qquad x \in [0, 1]. \tag{1.2}$$

Let $\hat{\varepsilon}_i := Y_i - m_n(X_i)$, $i = 1, 2, \ldots, n$, denote the nonparametric residuals and let a_n be another sequence of positive numbers tending to zero. We define the empirical d.f. estimator based on these residuals as follows:

$$\hat{G}_n(t) : = \frac{1}{n} \sum_{i=1}^{n} I(\hat{\varepsilon}_i \leq t), \quad t \in R.$$

In the following sections, all limits are taken as the sample size n tends to ∞, unless specified otherwise.

The paper is organized as follows. The basic assumptions and main result are introduced in the next Section while Sec. 3 provides the details of the proof.

2. Basic Assumptions and Main Result

In this section, we first state some basic assumptions. Define

$$\beta_n := \sqrt{\frac{\log n}{nh_n}}. \tag{2.1}$$

The basic assumptions needed on h_n and d_n (defined in (1.2)) are the following:

$$h_n \to 0, \qquad \beta_n \to 0, \tag{2.2}$$

and there exist constants A and A' such that

$$\sup_{x\in[0,1]} |d_n(x)| \leq A\beta_n + A'h_n, \quad \text{for all large } n, \quad a.s. \tag{2.3}$$

Note 1. Under certain conditions, Härdle et al. (1988) have shown (2.3) holds. By (2.3) and choosing h_n satisfying

$$h_n \leq \left(\frac{\log n}{n}\right)^{1/3}, \tag{2.6}$$

there exists a constant $C > 0$ such that

$$\sup_{x\in[0,1]} |d_n(x)| < C\beta_n \quad \text{for all large } n, \quad a.s.$$

Define

$$B_n = \left\{\sup_{x\in[0,1]} |d_n(x)| \leq C\beta_n\right\}.$$

Then we have

$$P\left(\cap_{i=n}^{\infty} B_i\right) \to 1. \tag{2.7}$$

This property will be used to show the following extended Glivenko–Cantelli theorem, i.e., the uniform strong consistency of $\hat{G}_n$ for G with certain rate.

Theorem 2.1. *Under (2.2), (2.3), and (2.4) and the assumption that for some* $0 < \alpha < 1/2$ *and any positive constant K,*

$$(nh_n)^{\alpha} \sup_{t\in R} [G(t + K\beta_n) - G(t - K\beta_n)] = o(1), \tag{2.6}$$

we have

$$\sup_{t\in R}(nh_n)^{\alpha} \left|\hat{G}_n(t) - G(t)\right| \to 0, \quad a.s. \tag{2.7}$$

3. Proof

In this section, we shall give the detailed proof of Theorem 2.1. In order to prove the main result (the uniform strong consistency of $\hat{G}_n$ for G), we will need the following exponential type probability inequality established first by Dvoretzky et al. (1956). See also Corollary 1 of Massart (1990).

Lemma 3.1 *Let μ_n and μ be the one-dimensional empirical distribution and theoretical distribution, respectively, for a random sample of* n *i.i.d. random variables. Then, for any $\epsilon > 0$,*

$$P(\sup\{|\mu_n((a, b]) - \mu((a, b])| : a \leq b\} > \epsilon) \leq 4 \exp(-n\epsilon^2/2), \ \forall \ n \geq 1. \quad (3.1)$$

Proof of Theorem 2.1 Let $G_n(t)$ be the empirical distribution function of the errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. Recall the extended Glivenko–Cantelli lemma in Fabian and Hannan (1985, pp 80–83), that

$$\sup_{t\in R} n^\beta |G_n(t) - G(t)| \to 0 \ \ a.s., \quad \text{for any } \ 0 < \beta < 1/2. \quad (3.2)$$

Thus, if we show

$$\sup_{t\in R} (nh_n)^\alpha |\hat{G}_n(t) - G_n(t)| \to 0, \ \ a.s., \quad (3.3)$$

then, by (3.3), $\alpha < 1/2$ and the assumption $h_n \to 0$ in (2.2), we obtain the claim (2.7).

We now proceed to prove (3.3). Rewrite

$$\hat{\varepsilon}_i = \varepsilon_i - d_n(X_i)$$

and

$$G_n(t) = \frac{1}{n}\sum_{i=1}^n I(\varepsilon_i \leq t); \qquad \hat{G}_n(t) = \frac{1}{n}\sum_{i=1}^n I(\varepsilon_i \leq t + d_n(X_i)).$$

Fix any $\epsilon > 0$. By (2.5), we obtain that

$$P\left\{\cup_{i=n}^\infty \left(\sup_{x\in R} |\hat{G}_i(x) - G_i(x)| I\left(B_i^c\right) \geq \epsilon\right)\right\} \leq P\left\{\cup_{i=n}^\infty B_i^c\right\} = 1 - P\left\{\cap_{i=n}^\infty B_i\right\} \to 0.$$

Therefore,

$$\sup_{x\in R} |\hat{G}_n(x) - G_n(x)| I\left(B_n^c\right) \to 0, \quad \text{a.s.}$$

Thus, to show (3.3), we only need to prove

$$\sup_{x\in R} |\hat{G}_n(x) - G_n(x)| I(B_n) \to 0, \quad \text{a.s.}$$

Next, rewrite

$$(nh_n)^\alpha [\hat{G}_n(t) - G_n(t)] I(B_n) = Z_{1n}(t) + Z_{2n}(t),$$

where

$$Z_{1n}(t) := \frac{(nh_n)^\alpha}{n} \sum_{i=1}^{n} [I(\varepsilon_i \le t + d_n(X_i)) - G(t + d_n(X_i)) - I(\varepsilon_i \le t) + G(t)] I(B_n),$$

$$Z_{2n}(t) := \frac{(nh_n)^\alpha}{n} \sum_{i=1}^{n} [G(t + d_n(X_i)) - G(t)] I(B_n).$$

Using the monotonicity of G and the fact that

$$\sup_{x \in [0,1]} |d_n(x)| \le C\beta_n \quad \text{on } B_n, \tag{3.4}$$

we have that

$$\sup_{t \in R} |Z_{2n}(t)| \le (nh_n)^\alpha \sup_{t \in R} [G(t + C\beta_n) - G(t - C\beta_n)] \to 0, \tag{3.5}$$

by (2.6).

Hence it remains to show that

$$\sup_{t \in R} |Z_{1n}(t)| \to 0, \quad a.s. \tag{3.6}$$

For any $t \in R$, by (3.4), and the monotonicity of the functions I and G, it follows that

$$\begin{aligned} Z_{1n}(t) &\le \frac{(nh_n)^\alpha}{n} \sum_{i=1}^{n} [I(\varepsilon_i \le t + C\beta_n) - G(t - C\beta_n) - I(\varepsilon_i \le t) + G(t)] \\ &= \frac{(nh_n)^\alpha}{n} \sum_{i=1}^{n} [I(t < \varepsilon_i \le t + C\beta_n) - P(t < \varepsilon_i \le t + C\beta_n)] \\ &\quad + (nh_n)^\alpha [G(t + C\beta_n) - G(t - C\beta_n)], \end{aligned}$$

and similarly, we also get

$$\begin{aligned} Z_{1n}(t) \ge &-\frac{(nh_n)^\alpha}{n} \sum_{i=1}^{n} [I(t - C\beta_n < \varepsilon_i \le t) - P(t - C\beta_n < \varepsilon_i \le t)] \\ &-(nh_n)^\alpha [G(t + C\beta_n) - G(t - C\beta_n)]. \end{aligned}$$

Thus,

$$\begin{aligned} \sup_{t \in R} |Z_{1n}(t)| \le \frac{(nh_n)^\alpha}{n} &\sup_{t \in R} \left| \sum_{i=1}^{n} [I(t < \varepsilon_i \le t + C\beta_n) - P(t < \varepsilon_i \le t + C\beta_n)] \right| \\ &+ (nh_n)^\alpha \sup_{t \in R} [G(t + C\beta_n) - G(t - C\beta_n)] . \end{aligned} \tag{3.7}$$

By (2.6) and (3.7), to show (3.6), it is sufficient to verify that

$$\frac{(nh_n)^\alpha}{n} \sup_{t \in R} \left| \sum_{i=1}^{n} [I(t < \varepsilon_i \le t + C\beta_n) - P(t < \varepsilon_i \le t + C\beta_n)] \right| \to 0, \ a.s. \tag{3.8}$$

For any given $\eta > 0$, by Lemma 3.1, we obtain that for $\forall\ n \geq 1$:

$$P\left\{\frac{(nh_n)^\alpha}{n}\sup_{t\in R}\left|\sum_{i=1}^{n}[I(t<\varepsilon_i\leq t+C\beta_n)-P(t<\varepsilon_i\leq t+C\beta_n)]\right|>\eta\right\}$$
$$=P\left\{\sup_{t\in R}|\mu_n((t,t+C\beta_n])-\mu((t,t+C\beta_n])|\geq\frac{\eta}{(nh_n)^\alpha}\right\}$$
$$\leq 4\exp(-n^{1-2\alpha}h_n^{-2\alpha}\eta^2/2).$$

By $h_n \to 0$ and $\alpha < 1/2$, it follows that

$$\frac{n^{1-2\alpha}h_n^{-2\alpha}}{\log n}\to\infty,$$

and, therefore, we have

$$\sum\exp(-n^{1-2\alpha}h_n^{-2\alpha}\eta^2/2)<\infty. \tag{3.9}$$

Combining (3.9) with the Borel–Cantelli lemma, we have obtained that (3.8) holds. Thus we have completed the proof of Theorem 2.1. □

Note 2. If the density g is bounded, using $\|g\|_\infty$ to denote its L_∞-norm, then

$$(nh_n)^\alpha\sup_{t\in R}[G(t+K\beta_n)-G(t-K\beta_n)]$$
$$\leq 2K\|g\|_\infty(nh_n)^\alpha\beta_n=2K\|g\|_\infty(nh_n)^{\alpha-\frac{1}{2}}\sqrt{\log n}.$$

In this case, $\alpha < \frac{1}{2}$ and

$$(nh_n)^{\alpha-\frac{1}{2}}\sqrt{\log n}\to 0$$

imply the assumption (2.6).

Acknowledgment

I would like to thank the referee for the helpful comments and suggestions.

References

Akritas, M. G., Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Stat.* 28:549–567.

Cheng, F. (2002). Consistency of error density and distribution function estimation in nonparametric regression. *Stat. Prob. Lett.* 59:257–270.

Cheng, F. (2004). Weak and strong uniform consistency of a kernel error density estimate in nonparametric regression. *J. Stat. Plan. Inf.* 119:95–107.

Cheng, F. (2005). Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *J. Stat. Plan. Inf.* 128:327–349.

Dvoretzky, A., Kiefer, J., Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Sta.* 27:642–669.

Fabian, V., Hannan, J. (1985). *Introduction to Probability and Mathematical Statistics*. New York: Wiley.

Härdle, W., Janassen, P., Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Stat.* 16:1428–49.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* 18:1269–1283.

Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.* 9:141–142.