- -
DEPARTMENT OF STATISTICS
- -
University of Wisconsin
1210 West Dayton Street
Madison, WI 53706-1693

TECHNICAL REPORT NO. 869

SEPTEMBER, 1990

CONVERGENCE OF KERNEL REGRESSION ESTIMATORS

by Lajos Horváth, Brian S. Yandell and Ananda Sen

Convergence Of Kernel Regression Estimators

Lajos Horváth, Brian S. Yandell and Ananda Sen[1]

Department of Statistics, University of Wisconsin–Madison

Abstract

We show that variable kernel regression estimators converge in weighted centered L_p norm. This weighted norm, suitably standarized, converges to a standard normal, allowing us to develop a class of goodness of fit tests for a parametric regression curve against a general smooth curve. This result works well for sample sizes of 200 or more in simulations, although is less reliable for moderate samples as p increases.

1. Introduction

Consider the model

$$y_i = g(t_i) + \epsilon_i, 1 \leq i \leq n$$

with g some smooth function with two uniformly bounded derivatives and $\{t_i, 1 \leq i \leq n\}$ design points on $(0,1)$. We restrict attention to design points of the form $t_i = Q(i/(n+1))$, with $Q = F^{-1}$ and F a continuous distribution function. We wish to estimate g by a variable kernel estimator and show the asymptotic normality of its weighted centered L_p norm on $[a,b]$, with $0 < a \leq b < 1$ and $p \geq 1$. The estimator is of the form

$$g_n(t) = \sum_{i=1}^{n} W_i(t) y_i,$$

[1]Research supported by NSF grant DMS 8704341 and USDA-CSRS grant 511-100.

with $W_i(t) = ((t_{i+1} - t_i)/h_n(t_i))K((t - t_i)/h_n(t_i))$ and K a kernel satisfying

C.1 K is bounded and vanishes outside of a finite interval

C.2 K' exists and is bounded

C.3 $\int_{-\infty}^{\infty} K(t)dt = 1$

C.4 $\int_{-\infty}^{\infty} tK(t)dt = 0$

C.5 $\int_{-\infty}^{\infty} t^2 K(t)dt < \infty$ and $\int_{-\infty}^{\infty} t^2 K(t)dt \neq 0$.

We want to have a local estimator, i.e. with $W_i(t) = 0$ for large $|t - t_i|$. Thus the local bandwidth $h_n(t)$ needs to be bounded. We assume that $h_n(t)$ is essentially $\lambda_n \ell(t)$, where $\ell(t)$ is a smooth function and $\lambda_n \to 0$ as $n \to \infty$. If $\ell(t)$ is unbounded on $(0, 1)$ let $h_n(t) = \min(\delta, \lambda_n \ell(t))$, where $\delta > 0$ is small enough. It will be clear from the proof that for n large enough $g_n(t)$ does not depend on δ. We also assume

C.6 $\inf_{\varepsilon \leq t \leq 1-\varepsilon} \ell(t) > 0 \quad \sup_{\varepsilon \leq t \leq 1-\varepsilon} \ell(t) < \infty$ for all $0 < \varepsilon < \frac{1}{2}$

C.7 $\sup_{\varepsilon \leq t \leq 1-\varepsilon} |\ell'(t)| < \infty$ for all $0 < \varepsilon < \frac{1}{2}$,

where ℓ' is the derivative of ℓ. We further need F and g to be smooth functions. Let $f = F'$ and assume for all $0 < \varepsilon < \frac{1}{2}$

C.8 $\inf_{\varepsilon \leq t \leq 1-\varepsilon} f(t) > 0$

C.9 $\sup_{\varepsilon \leq t \leq 1-\varepsilon} |f'(t)| > 0$

C.10 $g', g^{(2)}$ exist and $\sup_{\varepsilon \leq t \leq 1-\varepsilon} max(|g'(t)|, |g^{(2)}(t)|) < \infty$.

We do not want to assume that the "errors" $\epsilon_i, 1 \leq i \leq n$ are independent and/or identically distributed random variables. We need only that the sum $S(x) = \sum_{1\le i\le x} \epsilon_i$ behaves nicely. Assume there is a Wiener process and a positive constant σ such that

C.11 $\quad \sup_{0\le x\le n} |S(x) - \sigma W(x)| = O_p(a(n)),$

where $a(n)$ is an increasing, regularly (or slowly) varying sequence. There is a considerable literature on approximations on partial sums. We refer to surveys by Csörgő and Révész (1981) and Philipp (1986).

Let

$$I_n(p) = \int_a^b |g_n(t) - g(t)|^p w(t)dt,$$

where $w \ge 0$ is a weight function, and suppose

C.12 $\quad \sup_{a\le t\le b} |g_{(n)}(t) - g_0(t)| \to 0 \quad \text{as} \quad n \to \infty.$

with $g_{(n)}(t) = (n\lambda_n)^{\frac{1}{2}}[Eg_n(t) - g(t)]$ and $g_0(t)$ an appropriate continuous function. Our main result is the following

THEOREM. Assume that C.1–C.12 hold, $a(n)(n\lambda_n^2)^{-\frac{1}{2}} \to 0$ as $n \to \infty$ and $1 \le p < \infty$. Then, as $n \to \infty$,

$$((n\lambda_n)^{\frac{p}{2}} I_n(p) - m_n)/(\sigma_*^2\lambda_n)^{\frac{1}{2}} \xrightarrow{D} N(0,1),$$

where $N(0,1)$ is a standard normal random variate and σ_* independent of n and m_n are as defined in Lemma 3.

The theorem is tailored for goodness-of-fit. Usually we can choose the design distribution and $\ell(t)$ in the definition of the bandwidth. The value of g is given under the null hypothesis. Thus we know everything in m_n and σ_* but σ. However, we can estimate σ by $\hat{\sigma}_n$ such that

$(\hat{\sigma}_n - \sigma)^2 = O_p(\lambda_n)$. Hence the result remains true when σ is replaced by σ_n in m_n and σ_*.

An important special case is the problem with uniform design points ($F(t) = t, \ \ 0 \leq t \leq 1$) and uniform weights ($\ell(t) \equiv 1$). Here,

$$Eg_n(t) = ((n+1)\lambda_n)^{-1} \sum_{i=1}^{n} K((t - i/(n+1))/\lambda_n) g(i/(n+1)),$$

and hence

$$\sup_{a \leq t \leq b} |Eg_n(t) - \lambda_n)^{-1} \int_0^1 K((t - x/\lambda_n) g(x) dx| = O(n^{-1}\lambda_n)^{-1}).$$

If n is large enough, then $\int_{(t-1)/\lambda n}^{t/\lambda n} K(u) g(t - u\lambda_n) du = \int_{-\infty}^{\infty} K(u) g(t - u\lambda_n) du$. A two-term Taylor expansion and C4 give

$$\sup_{a \leq t \leq b} |6 \int_{-\infty}^{\infty} K(u)(g(t - u\lambda_n) - g(t)) du - g^{(2)}(t) \lambda_n^2 \int_{-\infty}^{\infty} u^2 K(u) du| = 0(\lambda_n^2).$$

Hence

$$\sup_{a \leq t \leq b} |6 g_{(n)}(t) - (n\lambda_n^5)^{\frac{1}{2}} g^{(2)}(t) \int_{-\infty}^{\infty} u^2 K(u) du| = O((n\lambda_n)^{-\frac{1}{2}}) + o((n\lambda_n^5)^{\frac{1}{2}}).$$

C.12 holds if $(n\lambda_n^5)^{\frac{1}{2}} \to C_0 > 0$ with $6g_0(t) = g^{(2)}(t) \int_{-\infty}^{\infty} u^2 K(u) du$ or if $n\lambda_n^5 \to 0$ with $g_0(t) = 0$. The bandwidth $\lambda_n = C_0^2 n^{-\frac{1}{5}}$ has received special attention in the literature because it minimizes the mean squared error (cf. Marron and Härdle (1986)).

Assuming that $\{\epsilon_i, 1 \leq i \leq n\}$ are independent, identically distributed random variables with $E|\epsilon_i|^\nu < \infty$ for $\nu > 2$, the Komlós-Major-Tusnády construction yields C.11 with $a(n) = n^{1/\nu}$. Assuming ν is large enough, we can always have $n^{1/\nu}(n\lambda_n^2)^{-\frac{1}{2}} \to 0$. For example, in the optimal case of $\lambda_n = C_0 n^{-\frac{1}{5}}$, we need $\nu > 10/3$. The independence of $\epsilon_i, 1 \leq i \leq n$, can be relaxed; for example, we can get similar results for m-dependent random variables.

Section 2 proves the main result, with lemma proofs relegated to section 4. Section 3 contains results of simulations for modest sample sizes.

2. Proof of Theorem

Without loss of generality we can assume that $K(u) = 0$ if $|u| > 1$. It is easy to see that $W_i(t) = 0$, $t \in [a, b]$ if $t_i \in [a - \delta, b + \delta]$. Hence by C.6, $h_n(t) = \lambda_n \ell(t), t \in [a, b]$ if n is large enough. Also, in the definition of g_n we can restrict the summation to $(n+1)F(a-\delta) \leq i \leq (n+1)F(b+\delta)$. Hence we can write $g_n(t) = Eg_n(t) + \overline{g}_n(t)$, with

$$\overline{g}_n(t) = \frac{1}{\lambda_n} \int_{(n+1)F(a-\delta)}^{(n+1)F(b+\delta)} \frac{Q(\frac{x+1}{n+1}) - Q(\frac{x}{n+1})}{\ell(Q(\frac{x}{n+1}))} K(\frac{t - Q(\frac{x}{n+1})}{\lambda_n \ell(Q(\frac{x}{n+1}))}) dS(x).$$

Therefore $(n\lambda_n)^{\frac{p}{2}} I_n(p) = \int_a^b |(n\lambda_n)^{\frac{1}{2}} \overline{g}_n(t) + g_{(n)}(t)|^p w(t) dt$. First we consider $(n\lambda_n)^{\frac{1}{2}} \overline{g}_n(t)$.

<u>LEMMA</u> 1. Assume C.1, C.2, C.6, C.7 and C.11 hold. Then

$$\sup_{a \leq t \leq b} |n\lambda_n \overline{g}_n(t) - \int_{(n+1)F(a-\delta)}^{(n+1)F(b+\delta)} [f(Q(\frac{x}{n+1}))\ell(Q(\frac{x}{n+1}))]^{-1} K(\frac{t - Q(\frac{x}{n+1})}{\lambda_n \ell(Q(\frac{x}{n+1}))}) d\sigma W(x)|$$
$$= O_p(a(n)).$$

Elementary observations give

$$(n\lambda_n)^{-\frac{1}{2}} \int_{(n+1)F(a-\delta)}^{(n+1)F(b+\delta)} [f(Q(\frac{x}{n+1}))\ell(Q(\frac{x}{n+1}))]^{-1} K(\frac{t - Q(\frac{x}{n+1})}{\lambda_n \ell(Q(\frac{x}{n+1}))}) d\sigma W(x)$$
$$= (n\lambda_n)^{-\frac{1}{2}} \int_{a-\delta}^{b+\delta} [f(x)\ell(x)]^{-1} K((t-x)/(\lambda_n \ell(x))) d\sigma W((n+1)F(x))$$
$$\stackrel{D}{=} ((n+1)/n)^{\frac{1}{2}} \Gamma_n^{(1)}(t), \text{ with}$$

$$\Gamma_n^{(1)}(t) = \sigma \lambda_n^{-\frac{1}{2}} \int_{a-\delta}^{b+\delta} f^{-\frac{1}{2}}(x) \ell^{-1}(x) K((t-x)/(\lambda_n \ell(x))) dW(x),$$

We show in the next two lemmas that we can extract $f^{\frac{1}{2}}(t)\ell(t)$ from the integral, and can essentially consider only the asymptotics of

$$Z_n(a,b) = \int_a^b |\Gamma_n^{(2)}(t) + g_{(n)}(t)|^p w(t)dt, \text{ with}$$

$$\Gamma_n^{(2)}(t) = \lambda_n^{-\frac{1}{2}} \sigma f^{-\frac{1}{2}}(t)\ell^{-1}(t) \int_{a-\delta}^{b+\delta} K((t-x)/(\lambda_n \ell(x)))dW(x).$$

LEMMA 2. We assume that C.1, C.2, C.6–C.10 hold. Then

$$\sup_{a \le t \le b} |\Gamma_n^{(1)}(t) - \Gamma_n^{(2)}(t)| = o_p(\lambda_n^{-\frac{1}{2}}).$$

LEMMA 3. We assume that C.1, C.2, C.6–C.10 and C.12 hold. Then

$$\lambda_n^{-\frac{1}{2}} \sigma_*^{-2}(Z_n(a,b) - m_n) \xrightarrow{D} N(0,1),$$

with center and asymptotic variance, respectively,

$$m_n = \int_{-\infty}^{\infty} \int_a^b |\sigma[f(t)\ell(t)]^{-\frac{1}{2}} Dx + g_{(n)}^p(t)| w(t)\varphi(x)dtdx,$$

$$\sigma_*^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_a^b |\sigma^2[f(t)\ell(t)]^{-1} D^2 xy + \sigma[f(t)\ell(t)]^{-\frac{1}{2}} D(x+y) + g_0^2(t)|^p \times$$

$$\times w^2(t)\ell(t)[\Psi(\zeta(u);x,y) - \varphi(x)\varphi(y)]dxdydudt,$$

where $D^2 = \int_{-\infty}^{\infty} K^2(t)dt$, $\varphi(x) = (2\pi)^{-\frac{1}{2}} e^{-x^2/2}$, $\zeta(u) = D^{-2} \int K(y)K(y+u)dy$ and

$$\Psi(u;x,y) = [2\pi(1-u^2)]^{-1} \exp(-(x^2+y^2-2uxy)/(2(1-u^2))).$$

Now we are ready to prove the Theorem. Let $p \ge 1$. For any $h, g \in L_p(a,b)$

$$p^{-1}2^{1-p} \int_a^b ||g(t)|^p - |h(t)|^p| w(t)dt \le$$

$$\int_a^b |g(t)-h(t)|^p w(t)dt + [\int_a^b |g(t)-h(t)|^p w(t)dt]^{1/p}[\int_a^b |h(t)|^p w(t)dt]^{1-1/p}.$$

By Lemma 2, $\int_a^b ||\Gamma_n^{(1)}(t)+g_{(n)}(t)|^p - |\Gamma_n^{(2)}(t)+g_{(n)}(t)|^p|w(t)dt \le o_p(\lambda_n^{-p/2}) + o_p(\lambda_n^{-\frac{1}{2}})$. Hence by Lemma 3,

$$(\sigma_*^2\lambda_n)^{-\frac{1}{2}}\{\int_a^b |\Gamma_n^{(1)}(t)+g_{(n)}(t)|^p w(t)dt - m_n\} \xrightarrow{D} N(0,1),$$

and Lemma 1 implies the Theorem.

3. Simulations

We simulated I_n(p) using programs developed by Joan Staniswallis and used in her joint work with one of us (Stansiwallis and Yandell, 1990). We used a rescaled version of a function used by Wahba and Wold (1975),

$$g(x) = 4.26[e^{-3.25x} - 4e^{-6.5x} + 3e^{-9.75x}].$$

Independent, identically distributed N(0,σ^2) contaminating errors for sample sizes $n = 50$, 100 and 200 were generated with the random number generator `RNOR` residing in the Portable Statistical Library `CMLIB`. Noisy observations of g on $[-1,2]$ were used by the kernel smoother in order to avoid boundary modifications to the kernel. The spline fit to the presmoothed data and the global spline fit used only the region [0,1].

One hundred independent realizations of size 50, 100 and 200 respectively of the locally adaptive smoothing spline were generated to provide the $g_n(t)$ values. We used Simpson's one-third rule to numerically compute I_n(p) values for p = 1(0.5)5 . For simplicity we used uniform weights, i.e. $w(t) \equiv 1$. All simulations were performed on the Statistics Research VAX at the University of Wisconsin, Madison.

Figure 1 shows QQ plots of $I_n(\mathrm{p})$ for $p = 1, 2, 3, 4$ and $n =$ 50,100 and 200. Results for $n = 50$ were not very encouraging and discrepancy from normal distribution was quite prominent. Results for $n = 100$ were a little closer, but still not satisfactory enough. All these clearly indicated the need for using a sufficiently large sample size for achieving normality. Therefore, we ran the simulation for $n = 200$ and found out that the plots were much more satisfactory. In fact, deviation from normality was not prominent until $p = 3$. The singularly interesting feature of the simulations was the increase in discrepancy of $I_n(\mathrm{p})$ from normality with the increase in p.

These simulations reveal several important features. The rate of convergence depends on p and actually decreases as p increases; the rate of decrement slows down with an increase in n. The rate of convergence is very slow even for a moderate value of p.

4. Proof of Lemmas

PROOF OF LEMMA 1. Simple integration by parts implies that

$$n\lambda_n \sup_{a \le t \le b} \left| \frac{1}{\lambda_n} \int_{(n+1)F(a-\delta)}^{(n+1)F(b+\delta)} \frac{Q(\frac{x+1}{n+1}) - Q(\frac{x}{n+1})}{\ell(Q(\frac{X}{n+1}))} K\left(\frac{t - Q(\frac{x}{n+1})}{\lambda_n \ell(Q(\frac{x}{n+1}))}\right) d(S(x) - \sigma W(x)) \right| = O_p(a(n)).$$

Also,

$$\int_{(n+1)F(a-\delta)}^{(n+1)F(b+\delta)} \frac{Q(\frac{x+1}{n+1}) - Q(\frac{x}{n+1})}{\ell(Q(\frac{x}{n+1}))} K\left(\frac{t - Q(\frac{x}{n+1})}{\lambda_n \ell(Q(\frac{x}{n+1}))}\right) dW(x)$$

$$= \int_{a-\delta}^{b+\delta} (Q(F(u) + (n+1)^{-1}) - u)/\ell(u)) K((t-u)/(\lambda_n \ell(u))) dW((n+1)F(u)).$$

Hence integration by parts again proves Lemma 1 by noting

$$\sup_{a\le t\le b} \Big| \int_{a-\delta}^{b+\delta} [n(Q(F(u)+n^{-1})-u) - f^{-1}(Q(u))]\ell(u)^{-1}K((t-u)/(\lambda_n\ell(u)))dW((n+1)F(u))\Big|$$

$$= O_p((n\lambda_n)^{-\frac{1}{2}}).$$

PROOF OF LEMMA 2. Integration by parts gives

$$\sigma^{-1}\lambda_n f^{\frac{1}{2}}(t)\ell(t)\Gamma_n^{(2)}(t) = K((t-(b+\delta))/(\lambda_n\ell(b+\delta)))W(b+\delta)$$

$$-K((t-(a-\delta))/(\lambda_n\ell(a-\delta)))W(a-\delta) - \int_{a-\delta}^{b+\delta} W(x)dK((t-x)/(\lambda_n\ell(x))).$$

If n is large enough $K((t-(b+\delta))/(\lambda_n\ell(b+\delta))) = K((t-(a-\delta))/(\lambda_n\ell(a-\delta))) = 0$ and hence

$$\Gamma_n^{(2)}(t) = -\sigma\lambda_n^{-1}f^{-\frac{1}{2}}(t)\ell^{-1}(t)\int_{b-\delta}^{a+\delta} W(x)dK((t-x)/(\lambda_n\ell(x))).$$

A similar argument gives

$$-\sigma^{-1}\lambda_n^{\frac{1}{2}}\Gamma_n^{(1)}(t) = \int_{a-\delta}^{b+\delta} W(x)d\{f^{-\frac{1}{2}}(x)\ell^{-1}(x)K((t-x)/(\lambda_n\ell(x)))\}$$

$$= \int_{a-\delta}^{b+\delta} W(x)K((t-x)/(\lambda_n\ell(x)))d\{f^{-\frac{1}{2}}(x)\ell^{-1}(x)\}$$

$$+\int_{a-\delta}^{b+\delta} W(x)f^{-\frac{1}{2}}(x)\ell^{-1}(x)dK((t-x)/(\lambda_n\ell(x))).$$

It is easy to see $\sigma^{-1}\lambda_n^{\frac{1}{2}}[\Gamma_n^{(2)}(t) - \Gamma_n^{(1)}(t)] =$

$$\int_{(t-(b+\delta))/h(n)}^{(t-(a-\delta))/h(n)} (W(t-u\lambda_n) - W(t))[f^{-\frac{1}{2}}(t-u\lambda_n)/\ell(t-u\lambda_n) - f^{-\frac{1}{2}}(t)/\ell(t)]dK(u/\ell(t-u\lambda_n))$$

$$+\int_{(t-(b+\delta))/h(n)}^{(t-(a-\delta))/h(n)} (W(t-u\lambda_n) - W(t))K(u/\ell(t-u\lambda_n))d\{f^{-\frac{1}{2}}(t-u\lambda_n)\ell^{-1}(t-u\lambda_n)\}.$$

Using the continuity of the Wiener process and one term Taylor expension we get Lemma 2.

PROOF OF LEMMA 3. Note that $\Gamma_n^{(3)}(t) = \sigma^{-1} f^{\frac{1}{2}}(t)\ell(t)\Gamma_n^{(2)}(t)$ is a zero mean Gaussian process with covariance

$$E\Gamma_n^{(3)}(t)\Gamma_n^{(3)}(s) = \lambda_n^{-1}\int_{a-\delta}^{b+\delta} K((t-x)/(\lambda_n\ell(x)))K((s-x)/(\lambda_n\ell(x)))dx.$$

By C.1 there is a constant $C>0$ such that

$$E\Gamma_n^{(3)}(t)\Gamma_n^{(3)}(s) = 0 \text{ if } |t-s| > C\lambda_n. \tag{1}$$

Thus we find that

$$E(\Gamma_n^{(3)}(t))^2 = \ell(t)\int K^2(u)du + 0(\lambda_n),$$

$$E\Gamma_n^{(3)}(s)\Gamma_n^{(3)}(t) = \ell(t)\int K(y)K(y+(s-t)/(\lambda_n\ell(t)))dy + 0(\lambda_n), \tag{2}$$

and hence

$$EZ_n(a,b) = \int_{-\infty)}^{\infty}\int_a^b |\sigma f^{-\frac{1}{2}}(t)\ell^{-1}(t)[E(\Gamma_n^{(3)}(t))^2]^{\frac{1}{2}}x + g_{(n)}(t)|^p w(t)\varphi(x)dtdx = m_n + 0(\lambda_n) \tag{3}$$

by (2). Also, for $a \le c \le d \le b$, arguing similarly to (3),

$$|EZ_n(c,d) - \int_{-\infty}^{\infty}\int_c^d |\sigma f^{-\frac{1}{2}}(t)\ell^{-\frac{1}{2}}(t)Dx + g_{(n)}(t)|^p w(t)dt| \le C(d-c)\lambda_n. \tag{4}$$

We have by (1) and (3) that $\mathrm{var}(Z_n(c,d)) =$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int\int_{T_n} |[E(\Gamma_n^{(2)}(t))^2 E(\Gamma_n^{(2)}(s))^2]^{\frac{1}{2}}xy + [E(\Gamma_n^{(2)}(t))^2]^{\frac{1}{2}} g_{(n)}(s)x + [E(\Gamma_n^{(2)}(s))^2]^{\frac{1}{2}} y +$$

$$+ g_{(n)}(t)g_{(n)}(s)|^p w(t)w(s)(\Psi(\zeta_n(t,s)); x, y) - \psi(x)\psi(y))dtdsdxdy,$$

where $T_n = \{(t,s) : c \le t, s \le d, |s-t| \le C\lambda_n\}$ and $\zeta_n(t,s) = cor(\Gamma_n^{(3)}(t), \Gamma_n^{(3)}(s))$. By (1) and (2) we get $\mathrm{var}(Z_n(c,d)) =$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int\int_{T_n} |u(t)u(s)\ell^{\frac{1}{2}}(t)\ell^{\frac{1}{2}}(s)D^2 xy + u(t)\ell^{\frac{1}{2}}(t)Dg_{(n)}(s)x + u(s)\ell^{\frac{1}{2}}(s)Dg_{(n)}(t)y +$$

$$+g_{(n)}(t)g_n(s)|^p \times w(t)w(s)[\Psi(\zeta_{(n)}(t,s);x,y) - \psi(x)\psi(y)]dtdsdxdy + O((d-c)\lambda_n^2),$$

where $\zeta_{(n)}(t,s) = D^{-2}\int K(y)K(y+(s-t)/(\lambda_n \ell(t)))dy$. Thus we have shown that,

$$\mathrm{var}(Z_n(c,d)) = \lambda_n \sigma_*^2 + o((d-c)\lambda_n). \tag{5}$$

Let $\zeta_i = Z_n(a_i, a_{i+1}) - EZ_n(a_i, a_{i+1})$, with $a_i = a + i\lambda_n$, $0 \le i \le i_0$, with $i_0 = [(b-a)/\lambda_n]$ and $a_{i_0+1} = b$. By (1), $\{\zeta_i, 0 \le i \le i_0\}$ are m-dependent r.v.'s, where m is a large enough constant. Let $M = [i_0^\gamma], 0 \le \gamma < 1$ and define

$$\eta_i = \sum_{j=1}^{M} \zeta_{j+(i-1)(M+m)}, \gamma_i = \sum_{j=1}^{m} \zeta_{j+M+(i-1)(M+m)},$$

for $1 \le i \le k_o = [i_0/(M+m)]$, and $\gamma_{k_0+1} = \sum_{j=1+k_o(M+m)}^{i_0} \zeta_j$ Clearly the $\{\eta_i, 1 \le i \le k_0\}$ are independent r.v.'s and $\{\gamma_i, 1 \le i \le k_0\}$ are independent. By (5) and independence $E(\sum_{i=1}^{k_0}\gamma_i)^2 = 0(k_o\lambda_n^2) = 0(\lambda_n^{\gamma+1})$ and $|E\zeta_i\zeta_j| \le (E\zeta_i^2 E\eta_j^2)^{\frac{1}{2}} = 0(\lambda_n^2)$. Hence $E(\sum_{i-1}^{k_o+1}\gamma_i)^2 = 0(\lambda_n^{\gamma+1})$ and

$$E(\sum_{i=1}^{k_0+1}\gamma_i \sum_{i=1}^{k_0}\eta_i) = \sum_{i=1}^{k_0} E\gamma_i\eta_i + \sum_{i=1}^{k_0} E\eta_i\gamma_{i+1} = O(k_0\lambda_n^2) = O(\lambda_n^{1+\gamma}).$$

Further observing that $Z_n(a,b) - EZ_n(a,b) = \sum_{i=1}^{k_0}\eta_i + \sum_{i=1}^{k_0+1}\zeta_i$, we have

$$\lambda_n^{-\frac{1}{2}}(Z_n(a,b) - EZ_n(a,b)) = \lambda_n^{-\frac{1}{2}}\sum_{i=1}^{k_0}\eta_i + o_p(1), \tag{6}$$

and $\lambda_n^{-1}E(\sum_{i=1}^{k_0}\eta_i)^2 \to \sigma^2 \quad as \quad n \to \infty$.

In a similar fashion to (4) one can show that $E\zeta_i^4 \le C\lambda_n^4$, and therefore $E\eta_i^4 \le CM^2\lambda_n^4$. Now we can easily establish that $(\sum_{i-1}^{k_0} E\eta_i^h)^{\frac{1}{4}}/\lambda_n^{\frac{1}{2}} \to 0$, if $0 < \alpha < \frac{1}{2}$. Using the Liapunov central limit we obtain that $(\sigma^2\lambda_n)^{-\frac{1}{2}}\sum_{i=1}^{k_0}\eta_i \xrightarrow{D} N(0,1)$ and hence (6) implies the lemma.

References

Cook, R.D. and Weisberg, S. (1982) *Residuals and Influence in Regression.* London: Chapman and Hall.

Cox, D.D., Koh, E., Wahba, G. and Yandell, B.S. (1988) Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.,* **16**, 113-119.

Csörgő, M. and Révész, P. (1981) *Strong Approximations in Probability and Statistics.* Akadémiai Kiadó, Budapest, and Academic Press, New York.

Eubank, R.L. (1988) *Spline Smoothing and Nonparametric Regression.* New York: Marcell Dekker, Inc.

Marron, J.S. and Härdle, W. (1968) Random approximations to some measures of accuracy in nonparametric curve estimation. *J. Mult. Anal.*, 20, 91-113.

Philipp, W. (1986) Correction to "Weak and L^p-invariance principles for sums of B-valued random variables". *Ann. Probab.*, **14**, 1095-1101.

Silverman, B.W. (1985) Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B*, **47**, 1-52.

Staniswalis, J. G. and Yandell, B. S. (1990) Locally adaptive smoothing splines. TR, University of Wisconsin–Madison.

Wahba, G. and Wold, S. (1975) A completely automatic French curve: fitting spline functions by cross-validation. *Communications in Statist.*, 4, 1-17.

Figure1: QQ plots for Simulated Data
Solid:n=50 Dotted:n=100 Dashed:n=200