# PROBABILITIES OF MAXIMAL DEVIATIONS FOR NONPARAMETRIC REGRESSION FUNCTION ESTIMATES

by

Gordon J. Johnston*

Key Words and Phrases: Density estimates, nonparametric regression estimates, maximal deviations, Gaussian processes.

AMS Subject Classifications: Primary 62G05; Secondary 62J05.

## Abstract

Let (X,Y) have regression function $m(x) = E(Y|X=x)$, and let X have a marginal density $f_1(x)$. We consider two nonparametric estimates of $m(x)$: the Watson estimate when $f_1$ is known and the Yang estimate when $f_1$ is known or unknown. For both estimates the asymptotic distribution of the maximal deviation from $m(x)$ is proved, thus extending results of Bickel and Rosenblatt for the estimation of density functions.

*The work of this author was supported by the Air Force Office of Scientific Research under Grants AFOSR-75-2796 and AFOSR-80-0080.

## 1. Introduction

Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a random sample from a bivariate population with distribution function $F(x,y)$ and density function $f(x,y)$. Let $F_1,f_1$ $(F_2,f_2)$ denote the marginal distribution and density of $X(Y)$. We are interested in estimating the unknown regression function $m(x) = E(Y|X=x)$ without making assumptions about either $m$ or the distributional form of $F$. In this paper we consider two classes of estimates of $m(x)$. The first is due to Watson (1964) (see also Watson and Leadbetter (1963), Parzen (1962)); motivated by the formula

$$m(x) = \{\int yf(x,y)dy\}/f_1(x) ,$$

we define

$$(1.1) \qquad m_n(x) = \left\{(n\varepsilon_n)^{-1} \sum Y_i K((x-X_i)/\varepsilon_n)\right\}\left\{(n\varepsilon_n)^{-1} \sum K((x-X_i)/\varepsilon_n)\right\}^{-1}$$

and

$$(1.2) \qquad \bar{m}_n(x) = \{(n\varepsilon_n)^{-1} \sum Y_i K((x-X_i)/\varepsilon_n)\}/f_1(x) ,$$

the latter appropriate if $f_1$ is known. Here $\varepsilon_n \to 0$ and $K$ is a smooth density function symmetric about zero.

Analysis of $m_n(\cdot)$ is somewhat complicated by the fact that it is a ratio of two random variables. Yang (1977a) avoids this problem by defining and proving consistency of

$$(1.3) \qquad M_n(x) = (n\varepsilon_n)^{-1} \sum_{i=1}^{n} Y_i K((F_n(X_i)-F_n(x))/\varepsilon_n)$$

$$(1.4) \qquad \bar{M}_n(x) = (n\varepsilon_n)^{-1} \sum_{i=1}^{n} Y_i K((F_1(X_i) - F_1(x))/\varepsilon_n) ,$$

where $F_n$ is the empirical distribution of $X_1,\ldots,X_n$ and $\bar{M}_n(\cdot)$ is appropriate when $F_1, f_1$ are known. Briefly, Yang's estimates are motivated by consideration of statistics of the form $n^{-1} \sum J\left(\frac{i}{n+1}\right) H\left(X_{(i)}, Y_{(i:n)}\right)$, where $Y_{(i:n)}$ is the concomitant of the i-th order statistic $X_{(i)}$ (see also Yang (1977b)).

In the parametric normal linear regression model, (X,Y) has a bivariate normal distribution, m(x) is linear in x, and one can derive uniform confidence bands for m(x). In this paper, where neither F nor the form of m are known, we are able to obtain uniform confidence bands for the regression function m(x). More specifically, we extend the results of Bickel and Rosenblatt (1973) and Rosenblatt (1976) to obtain the limit distribution of the maximal deviation

$$(1.5) \qquad \sup\{|g_n(x) - m(x)| : 0 \le x \le 1\} ,$$

where $g_n$ is given by one of (1.2)-(1.4).

Obtaining the limit distribution of (1.5) when using the special estimates (1.2) and (1.4) (special because they require $f_1$ known) is a conceptually simple extension of Rosenblatt's (1976) results. However, our real interest is in the useful estimate (1.3), for which we obtain the limit distribution of (1.5) by showing that $M_n(x) - \bar{M}_n(x)$ is uniformly sufficiently close to zero. We have been unable to obtain useful results for (1.1), the major technical difficulty being its form as a ratio of two random variables.

Related Literature

Schuster (1972) and Johnston (1979) give different conditions for the pointwise asymptotic normality of (1.1) and (1.2). Schuster and Yakowitz (1978) give rates of almost sure convergence to zero for the maximal deviation (1.5) using (1.1). Priestly and Chao (1972) and Benedetti (1974) consider an estimate closely related to (1.2) for the case that X is non-stochastic. Stone (1977) and Lai (1977) give weak conditions for consistency of nearest neighbor estimates.

## 2. Assumptions and a Preliminary Result

Define $m_n^*(x) = f_1(x)\bar{m}_n(x)$ and $s(x) = E(Y^2|X=x)$. In this section we prove maximal deviation results for $m_n^*(\cdot)$, applying these results to (1.2)-(1.4) in the next section. Let $\{a_n\}$ be a sequence of constants with $a_n \to \infty$ as $n \to \infty$.

Assumptions

(A1) For all n and some $c < \infty$,

$$(\log n)\varepsilon_n^{-3} \int_{|y|\geq a_n} y^2 f_2(y)\,dy \leq c \ .$$

(A2) $a_n \varepsilon_n^{-1/2} n^{-1/6} (\log n)^2 \to 0$ as $n \to \infty$, and $(\log n)^{-1} (n\varepsilon_n^{1/2}) \to \infty$ .

(A3) $$(\log n) \sup_{0\leq x\leq 1} \int_{|y|\geq a_n} y^2 f(x,y)\,dy \to \infty \text{ as } n \to \infty \ .$$

(A4) There exists $\eta > 0$ such that $0 \le x \le 1$ and $n \ge 1$ implies

$$|g_n(x)| = |\int_{-a_n}^{a_n} y^2 f(x,y)\,dy| > \eta \ .$$

(A5) The kernel function K vanishes outside a finite interval [-A,A] and is absolutely continuous on [-A,A], A > 1.

(A6) The marginal density $f_1(x)$ is continuous and positive on an open interval containing [0,1].

(A7) For $g_n$ defined by (A4), $\{g_n^{\frac{1}{2}}\}$ have uniformly bounded and continuous first derivatives on [-A,A].

(A8) Both $f(x)s(x)$ and $E(|Y| \,|\, X=x)f(x)$ are bounded for $0 \le x \le 1$.

Note that (A1) and (A3) hold if Y is bounded and $a_n = \log\log n$, while (A1), (A3) and (A4) hold with $a_n = n^{\beta}$ ($\beta > 0$, $\beta$ near zero) if (X,Y) are jointly normally distributed.

<u>Theorem 1</u> (Johnston (1979)). Suppose (A1)-(A8) hold and $\varepsilon_n = n^{-\delta}$ $(0 < \delta < 1/3)$. Define

$$Y_n(t) = (n\varepsilon_n)^{\frac{1}{2}} (m_n^*(t) - Em_n^*(t))(s(t)f(t))^{-\frac{1}{2}} \ .$$

Then

$$\text{(2.1)} \qquad \left\{P\ (2\delta \log n)^{\frac{1}{2}} \left[\sup_{0\le t\le 1} |Y_n(t)|/(\lambda(K))^{\frac{1}{2}} - d_n\right] < x\right\} \to e^{-2e^{-x}} ,$$

where $\lambda(K) = \int K^2(u)du$ and

$$d_n = (2\delta \log n)^{\frac{1}{2}} + (2\delta \log n)^{-\frac{1}{2}} \left\{\log\left(\frac{c_1(K)}{\pi^{\frac{1}{2}}}\right) + \tfrac{1}{2}[\log \delta + \log \log n]\right\}$$

if $c_1(K) = (K^2(A)+K^2(-A))(2\lambda(K))^{-1} > 0$ and otherwise

$$d_n = (2\delta \log n)^{\frac{1}{2}} + (2\delta \log n)^{-\frac{1}{2}} \log\left(\frac{c_2(K)}{2\pi}\right) ,$$

where

$$c_2(K) = (2\lambda(K))^{-1} \int [K'(u)]^2 du .$$

The similarity of Theorem 2.1 to the main results of Bickel and Rosenblatt (1973) and Rosenblatt (1976) is obvious. The major technical difficulty in adapting their proofs for density estimates is the possible unboundedness of Y, which is the reason for the somewhat awkward form of (A1)-(A4). The proof of Theorem 2.1, which is given in Appendix A, closely follows Rosenblatt's (1976) argument.

In applications, we would want to replace $Em_n^*(t)$ (in the definition of $Y_n$) by $m(t)$; this results in the following corollary.

Corollary 2.1. Suppose in Theorem 2.1 that $1/5 < \delta < 1/3$, that $\int u^2K(u)du < \infty$ and that $m(t)f(t)$ has two bounded continuous derivatives. Then (2.1) holds for the process

$$Y_n^*(t) = (n\varepsilon_n)^{\frac{1}{2}} [m_n^*(t) - m(t)f(t)](s(t)f(t))^{-\frac{1}{2}} .$$

Remark. While all results are stated for suprema over the interval [0,1], they extend to arbitrary finite intervals [a,b] with no change except that (A4), (A6) and (A8) must hold for $a \le x \le b$, and $A > \max(|a|,|b|)$.

## 3. Applications to (1.2)-(1.4)

The limiting distribution of the maximal deviation of (1.2) is particularly simple since

$$
(3.1)\qquad
\begin{aligned}
Y_n(t) &= (n\varepsilon_n)^{\frac{1}{2}} (\bar{m}_n(t)-E\bar{m}_n(t))(f(t)/s(t))^{\frac{1}{2}} \\
Y_n^*(t) &= (n\varepsilon_n)^{\frac{1}{2}} (\bar{m}_n(t)-m(x))(f(t)/s(t))^{\frac{1}{2}} .
\end{aligned}
$$

The distribution for (1.4) is also fairly simple to derive from Theorem 2.1. One notes that if (A6) is strengthened as in Rosenblatt (1976) to

(A6)' $f_1(x)$ is continuous and positive on the smallest interval containing its support,

then $Z_i = F_1(X_i)$ is uniformly distributed, $E(Y|Z=F(u)) = m(u)$, $E(Y^2|Z=F(u)) = s(u)$ and $f_Z(u) = 1$. Thus, Theorem 2.1 and Corollary 2.1 will hold for the processes

$$
(3.2)\qquad
\begin{aligned}
Y_{n1}(t) &= (n\varepsilon_n)^{\frac{1}{2}} [\bar{M}_n(t) - E\bar{M}_n(t)]s(t)^{-\frac{1}{2}} \\
Y_{n2}(t) &= (n\varepsilon_n)^{\frac{1}{2}} [\bar{M}_n(t) - m(t)]s(t)^{-\frac{1}{2}} .
\end{aligned}
$$

Finally, we consider (1.3), which is applicable in the usual case that the marginal density $f_1(X)$ of X is unknown. Consider the following assumptions.

(B1) $E|Y| < \infty$ .

(B2) $E(Y|X=F^{-1}(u)) = g(u)$ has two bounded derivatives on $[0,1]$.

(B3) $E(|Y|\,|X=F^{-1}(u)) = h(u)$ is bounded on $[0,1]$.

(B4) There exists $a_n \to \infty$ with $a_n^2 \log n/(n\varepsilon_n^3) \to 0$ and

$$n^{\frac{1}{2}} \int_{|y|\geq a_n} |y|\ dF_2(y) \to 0 \text{ as } n \to \infty .$$

(B5) K has three continuous bounded derivatives on its support.

Theorem 3.1. Assume (A1)-(A8), (A6)', (B1)-(B5). Then if $0 < F(a) < F(b) < 1$,

$$(n\varepsilon_n \log n)^{\frac{1}{2}} \sup_{a\leq u\leq b} |M_n(u) - \overline{M}_n(u)| \xrightarrow{P} 0 ,$$

so that Theorem 2.1 and Corollary 2.1 hold for the processes defined by substituting $M_n$ for $\overline{M}_n$ in (3.2) (the proof is given in Appendix B).

Theorem 3.1 can be used to construct uniform confidence intervals for the regression function as follows.

Corollary 3.1. Assuming Theorem 3.1 holds, an approximate $(1-\alpha) \times 100\%$ confidence band over an interval $[a,b]$ is

$$M_n(u) \pm (n\varepsilon_n)^{-\frac{1}{2}} [s(u)\lambda(K)]^{\frac{1}{2}} [d_n + c(\alpha)(2\delta \log n)^{-\frac{1}{2}}] ,$$

where $c(\alpha) = \log 2 - \log|\log(1-\alpha)|$ (for practical applications, one would estimate $s(u)$).

## Acknowledgement

The author thanks R.J. Carroll for suggesting the problem and M.R. Leadbetter for many helpful conversations.

## References

Benedetti, J.K. (1974). Kernel estimation of regression functions. Ph.D. Dissertation, Department of Biostatistics, University of Washington.

Bhattacharya, P.K. (1976). An invariance principal in regression analysis. *Annals of Statistics*, 4, no. 3, 621-624.

Bickel, P.J. and Rosenblatt, M. (1973). On some global measures of the deviation of density function estimators. *Annals of Mathematical Statistics*, 1, 1071-1095.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley.

Johnston, G. (1979). Smooth nonparametric regression analysis. Institute of Statistics Mimeo Series #1253, University of North Carolina.

Lai, S.L. (1977). Large sample properties of K-nearest neighbor procedures. Ph.D. Dissertation, Department of Mathematics, UCLA.

Parzen, E. (1962). On estimation of a probability density function. *Annals of Mathematical Statistics*, 33, 1065-1076.

Priestly, M.B. and Chao, M.T. (1972). Nonparametric function fitting. *Journal of the Royal Statistical Society, Series B*, 34, 385-392.

Revesz, P. (1976). On strong approximation of the multidimensional empirical process. *Annals of Probability*, 4, 729-743.

Rosenblatt, M. (1976). On the maximal deviation of a k-dimensional density estimator. *Annals of Probability*, 4, 1009-1015.

Schuster, E.F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Annals of Mathematical Statistics*, 43, 84-88.

Schuster, E.F. and Yokowitz, S. (1979). Contributions to the theory of nonparametric regression, with applications to system identification. *Annals of Statistics*, 7, 139-149.

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

Watson, G.S. and Leadbetter, M.R. (1963). On the estimation of the probability density, I. *Annals of Mathematical Statistics*, 34, 480-491.

Yang, S.S. (1977a). Linear functions of concomitants of order statistics. Technical Report No. 7, Department of Mathematics, M.I.T.

Yang, S.S. (1977b). General distribution theory of the concomitants of order statistics. *Annals of Statistics*, 6, 996-1002.

## Appendix A

We begin with two lemmas. Let W be Brownian motion on $(-\infty,\infty)$ and let K be a symmetric density which satisfies (A5).

Lemma A.1 (Bickel and Rosenblatt (1976)). Let $d_n$ and $\lambda(K)$ be as in Theorem 2.1 and let $\varepsilon_n = n^{-\delta}$ $(0 < \delta < \frac{1}{2})$. Define

$$V_n(t) = \varepsilon_n^{-\frac{1}{2}} \int K((t-x)/\varepsilon_n)\,dW(x) \ .$$

Then

$$P\left\{(2\delta \log n)^{\frac{1}{2}} \left\{\sup_{0\le t\le 1} |V_n(t)|/(\lambda(K))^{\frac{1}{2}} - d_n\right\} < x\right\} \to e^{-2e^{-x}} \ .$$

Lemma A.2 (Revesz (1976), Rosenblatt (1976)). Let $(\underline{X}_1,\ldots,\underline{X}_n,\ldots)$ be independent and uniformly distributed on $[0,1]^2$. One can construct a sequence of Brownian bridges $B_n$ such that

$$\sup\left\{\left|n^{1/2}\left(F_n(\underline{x}) - \prod_{j=1}^{2} x_j\right) - B_n(\underline{x})\right|\right\} = O\left(n^{-1/6}(\log n)^{3/2}\right) \ \text{(a.s.)} \ ,$$

where $\underline{x} = (x_1, x_2)$ and sup is over the set $0 \le x_1, x_2 \le 1$.

When Y is bounded, since K vanishes off an interval, the proof of Theorem 2.1 is an easy extension of Rosenblatt's (1976) result; the relevant change of variables formula is

$$\text{(A.1.1)} \qquad \int_{-A}^{A}\int_{-B}^{B} f(x,y)\,dg(x,y) = \int_{-A}^{A}\int_{-B}^{B} g(x,y)\,df(x,y)$$
$$+ \int_{-A}^{A} f(B,y)\,dg(B,y) - \int_{-A}^{A} f(-B,y)\,dg(B,y)$$
$$+ \int_{-B}^{B} g(x,A)\,df(x,A) - \int_{-B}^{B} g(x,-A)\,df(x,-A) \ .$$

Hence, for the case Y unbounded, we merely sketch the proof, pointing out where the various assumptions are used. Let $Z_n(x,y) = n^{\frac{1}{2}}(F_n(x,y)-F(x,y))$, so that

$$\text{(A.1.2)} \qquad Y_n(t) = [s(t)f(t)]^{-\frac{1}{2}}\, \varepsilon_n^{-\frac{1}{2}} \iint yK((t-x)/\varepsilon_n)\,dZ_n(x,y) \ .$$

We also make the definition

$$\text{(A.1.3)} \qquad Y_{0,n}(t) = [s(t)f(t)]^{-\frac{1}{2}}\, \varepsilon_n^{-\frac{1}{2}} \iint_{|y| \le a_n} yK((t-x)/\varepsilon_n)\,dZ_n(x,y) \ .$$

Let $||V(\cdot)|| = \sup\{|V(t)|: 0 \le t \le 1\}$ .

<u>Lemma A.3</u>. $||Y_n - Y_{0n}|| = o_p((\log n)^{-\frac{1}{2}})$ .

<u>Proof</u>. $||Y_n - Y_{0n}|| \le \varepsilon_n^{-\frac{1}{2}}\, ||g^{-\frac{1}{2}}||\ ||U_n||$, where $g(x) = f(x)s(x) = \int y^2 f(x,y)\,dy$ and

$$U_n(x) = \iint_{|y|>a_n} yK((t-x)/\varepsilon_n)dZ_n(x,y) \ .$$

By (A4), $||g^{-\frac{1}{2}}|| > 0$. It is easy to show by Markov's inequality and (A1) that $U_n(x) \xrightarrow{P} 0$ for any $0 \le x \le 1$. The lemma will follow if $U_n$ is tight on $D[0,1]$. By (A5) and the Schwarz inequality,

$$E|U_n(t) - U_n(t_1)| \ |U_n(t_2) - U_n(t)|$$
$$\le M_0(\log n)\varepsilon_n^{-3} \ |t_1 - t| \ |t_2 - t| \int_{|y|>a_n} y^2 f_2(y)dy \ ,$$

verifying tightness by (A1) and Theorem 15.6 of Billingsley (1968). □

Define

(A.1.4)
$$s_n(t) = E\left\{Y^2 I(|Y| \le a_n) \middle| X=t\right\}$$
$$Y_{1n}(t) = (s_n(t)/s(t))^{-\frac{1}{2}} \ Y_{0n}(t) \ .$$

Our next approximation is

<u>Lemma A.4</u>. $||Y_{0n} - Y_{1n}|| = o_p((\log n)^{-\frac{1}{2}})$ .

<u>Proof</u>. We will later prove that

$$(\log n)^{\frac{1}{2}} \left\{||Y_{1n}||\{\lambda(K)\}^{-\frac{1}{2}} - d_n\right\}$$

has a limit distribution. Since $d_n = O((\log n)^{\frac{1}{2}})$, this means $||Y_{1n}|| = O_p((\log n)^{\frac{1}{2}})$. By (A3), (A7) and (A6), recalling that $g_n(x) = f(x)s_n(x)$, we have

$$||(s_n/s)^{-\frac{1}{2}} - 1|| = o((\log n)) ,$$

completing the proof. □

Next let T be the transformation of (X,Y) to a uniform random variable on $[0,1]^2$ ((26),(27) of Rosenblatt (1976)). Define

$$Y_{2n}(t) = [s_n(t)f(t)]^{-\frac{1}{2}} \varepsilon_n^{-\frac{1}{2}} \iint_{|y|\le a_n} yK((t-x)/\varepsilon_n) dB_n(T(x,y)) .$$

$$Y_{3n}(t) = [s_n(t)f(t)]^{-\frac{1}{2}} \varepsilon_n^{-\frac{1}{2}} \iint_{|y|\le a_n} yK((t-x)/\varepsilon_n) dW_n(T(x,y)) ,$$

where $B_u(u,s) = W_n(u,s) - usW_n(1,1)$ ($W_n$ here is the two-dimensional Wiener process).

Lemma A.5.

$$||Y_{1n} - Y_{2n}|| = O_p\left(a_n \varepsilon_n^{-1/2} n^{-1/6} (\log n)^{3/2}\right) = o_p\left((\log n)^{-1/2}\right) \text{ (by (A2))}$$

and

$$||Y_{2n} - Y_{3n}|| = o_p((\log n)^{-\frac{1}{2}}) .$$

Proof. Using Lemma A.2, (A5) and the integration by parts formula (A1.1), extremely detailed calculations show

$$\varepsilon_n^{1/2} ||g_n||^{1/2} ||Y_{1n} - Y_{2n}|| = O_p\left(n^{-1/6} (\log n)^{3/2}\right)$$

$$\times \left\{4a_n \int_{-A}^{A} |K'(u)| du + 4a_n [K(A) + K(-A)]\right\} = O_p\left\{a_n n^{-1/6} (\log n)^{3/2}\right\} ,$$

completing the first part of the proof. Since the Jacobian of the transform T is f(x,y), we have

$$|Y_{2n}(t) - Y_{3n}(t)| = |(g_n(t))^{-\frac{1}{2}} \varepsilon_n^{-\frac{1}{2}} \iint_{|y|\le a_n} yK((t-x)/\varepsilon_n)f(x,y)dxdy| \cdot |W_n(1,1)| .$$

Thus,

$$||Y_{2n} - Y_{3n}|| \le |W_n(1,1)| \; ||g_n^{-\frac{1}{2}}|| \; \varepsilon_n^{-\frac{1}{2}} \times ||\int |y| f(x,y)dyK((t-x)/\varepsilon_n)dx|| .$$

By (A8) and (A4), $||Y_{2n} - Y_{3n}|| = O_p(\varepsilon_n^{\frac{1}{2}})$, completing the proof. □

Now define

$$Y_{4n}(t) = [s_n(t)f(t)]^{-\frac{1}{2}} \varepsilon_n^{-\frac{1}{2}} \int [s_n(x)f(x)]^{\frac{1}{2}} K((t-x)/\varepsilon_n)dW(x)$$

$$Y_{5n}(t) = \varepsilon_n^{-\frac{1}{2}} \int K((t-x)/\varepsilon_n)dW(x) .$$

Since $Y_{3n}$ and $Y_{4n}$ are Gaussian with the same covariance function, they have the same distribution. Thus, by Lemmas A.1, A.3, A.4 and A.5, we need merely prove

<u>Lemma A.6</u>. $||Y_{4n} - Y_{5n}|| = o_p((\log n)^{-\frac{1}{2}})$ .

<u>Proof</u>. First note that

$$|Y_{4n}(t) - Y_{5n}(t)| = \varepsilon_n^{-\frac{1}{2}} \left| \int_{-A}^{A} \{(g_n(t-u\varepsilon_n)/g_n(t))^{\frac{1}{2}} - 1\}K(u)dW(t-u\varepsilon_n) \right| .$$

Since by (A7)

$$\varepsilon_n^{-\frac{1}{2}} \sup_{0\le t\le 1} |(g_n(t\pm A\varepsilon_n)/g_n(t))^{\frac{1}{2}} - 1| = O(1) ,$$

using integration by parts and the assumptions that $g_n^{\frac{1}{2}}$ and K are absolutely continuous, we obtain

$$|Y_{4n}(t) - Y_{5n}(t)| \le \varepsilon_n^{-\frac{1}{2}} \left| \int_{-A}^{A} W(t-u\varepsilon_n) \frac{\partial}{\partial u} \left\{ \left\{ [g_n(t-u\varepsilon_n)/g_n(t)]^{\frac{1}{2}} - 1 \right\} K(u) \right\} du \right.$$

$$+ O_p(\varepsilon_n^{\frac{1}{2}}) = J_n(t) + O_p(\varepsilon_n^{\frac{1}{2}}) .$$

Note that $\varepsilon_n^{-1} \frac{\partial}{\partial u} \left\{ (g_n(t-u\varepsilon_n)/g_n(t))^{\frac{1}{2}} - 1 \right\}$ is uniformly bounded by (A4) and (A7), so that

$$\varepsilon_n^{-\frac{1}{2}} J_n(t) \le \varepsilon_n^{-1} \left| \int_{-A}^{A} W(t-u\varepsilon_n) K'(u) [\{g_n(t-u\varepsilon_n)/g_n(t)\}^{\frac{1}{2}} - 1] du + C_1 \int_{-A}^{A} |W(t-u\varepsilon_n)| du \right.$$

$$\le C_2 \int_{-A}^{A} |W(t-u\varepsilon_n)|\; uK'(u) du + C_1 \int_{-A}^{A} |W(t-u\varepsilon_n)| du ;$$

hence $\varepsilon_n^{-\frac{1}{2}} ||J_n|| = O_p(1)$ and $||Y_{4n} - Y_{5n}|| = O_p(\varepsilon_n^{\frac{1}{2}})$, which completes the proof. □

## Appendix B

Define

$$M_n^*(x) = (n\varepsilon_n)^{-1} \sum_{i=1}^{n} Y_i K((F_n(X_i) - F(x))/\varepsilon_n) .$$

We will prove Theorem 3.1 by showing

(B.1.1) $$(n\varepsilon_n \log n)^{\frac{1}{2}} \sup_{a\le u\le b} |M_n(u) - M_n^*(u)| \xrightarrow{P} 0$$

and

$$\text{(B.1.2)} \qquad (n\varepsilon_n \log n)^{\frac{1}{2}} \sup_{a \le u \le b} |M_n^*(u) - \overline{M}_n(u)| \xrightarrow{P} 0 \;.$$

We only prove (B.1.1) as (B.1.2) is similar.

The following lemma is very similar to Lemma 1 of Bhattacharyya (1967).

Lemma B.1. Assume that $g(u) = E[Y|X = F^{-1}(u)]$ has r continuous derivatives on [0,1], $r > 0$, and that K has bounded support and r bounded derivatives on the support. Then for a,b such that $0 < F(a) < F(b) < 1$,

$$\left|\varepsilon_n^{-(r+1)} \iint yK^{(r)}((F(x)-F(z))/\varepsilon_n)dF(x,y)\right| = O(1)$$

uniformly in $z \in [a,b]$ as $n \to \infty$ . □

Letting $Z_n(x,y) = F_n(x,y) - F(x,y)$, we see

$$M_n^*(u) - M(u) = \varepsilon_n^{-1} \iint y[K((F_n(x)-F_n(u))/\varepsilon_n) - K((F_n(x)-F_n(u))/\varepsilon_n)]$$
$$\times [dZ_n(x,y) + dF(x,y)] = J_1 + J_2 \;.$$

We first show $(n\varepsilon_n \log n)^{\frac{1}{2}} |J_2| \xrightarrow{P} 0$. Let $\xi_n(u) = (F_n(u)-F(u))/\varepsilon_n$. By (B5),

$$J_2 = \varepsilon_n^{-1} \xi_n(u) \iint y[K'(\xi_n(u)) + 1/2\, \xi_n(u)K''(\xi_n(u))$$
$$+ 1/6\, \xi_n(u)^2 K'''(\xi_n(u)+w_n(u)/\varepsilon_n)]dF(x,y)$$

$= J_2^{(1)} + J_2^{(2)} + J_2^{(3)}$, where $w_n(u)$ is between $F_n(u)$ and $F(u)$.

Recall that K has three bounded continuous derivatives on a compact support. This together with the fact that $\sup|F_n(x) - F(x)| = O_p(n^{-\frac{1}{2}})$ yields by a Taylor expansion

$$\text{(B.1.3)} \quad (n\varepsilon_n \log n)^{\frac{1}{2}} \sup\{|J_2^{(1)}|: a \le u \le b\} \le (n\varepsilon_n \log n)^{\frac{1}{2}} \varepsilon_n^{-2} O_p(n^{-\frac{1}{2}})$$

$$\times \sup_u \left\{ \left| \iint yK'((F(x)-F(u))/\varepsilon_n) dF(x,y) \right. \right.$$

$$\left. + \varepsilon_n^{-1} O_p(n^{-\frac{1}{2}}) \int_0^1 h(t) |K''((t-F(u))/\varepsilon_n) dt + \varepsilon_n^{-2} O_p(n^{-1}) E|Y| \right\} .$$

Applying Lemma B.1 shows that the first term on the right of (B.1.3) converges in probability to zero. Making a change of variable shows that the second term is $(n\varepsilon_n \log n)^{\frac{1}{2}} \varepsilon_n^{-2} O_p(n^{-1}) = o_p(1)$ by (B4). The third term is $(\log n)^{1/2} \varepsilon_n^{-3/2} O_p(n^{-1}) = o_p(1)$, also by (B4). Similar calculations apply to $J_2^{(2)}$ and $J_2^{(3)}$, so we have shown $(n\varepsilon_n \log n)^{\frac{1}{2}} \sup\{|J_2|: a \le u \le b\} \xrightarrow{P} 0$. We thus need only prove

$$\text{(B.1.4)} \qquad (n\varepsilon_n \log n)^{\frac{1}{2}} \sup\{|J_1|: a \le u \le b\} \xrightarrow{P} 0 .$$

Rewrite

$$J_1(u) = \varepsilon_n^{-1} \left[ \int_{|y|>a_n} + \int_{|y|<a_n} \right] (yG_n(x,u) Z_n(dx,dy)) ,$$

$$= J_1^{(1)} + J_1^{(2)} , \quad \text{where}$$

$$G_n(x,u) = K((F_n(x)-F(u))/\varepsilon_n) - K((F_n(x)-F_n(u))/\varepsilon_n) .$$

Define $Q_n(y) = F_n(y) - F(y)$ and use integration by parts to show

$$J_1^{(2)} = \varepsilon_n^{-1} \iint_{|y|\le a_n} Z_n(x,y) dy G_n(dx,u) + \varepsilon_n^{-1} \int_{-a_n}^{a_n} yG_n(\infty,u) dQ_n(y)$$

$$+ a_n \varepsilon_n^{-1} \int \{Z_n(x,a_n) + Z_n(x,-a_n)\} G_n(dx,u)$$

$$= (I_1+I_2+I_3+I_4)(u) .$$

Now, by the mean value theorem and the boundedness of K',

$$|I_2(u)| \le \varepsilon_n^{-2}\, O_p(n^{-\frac{1}{2}}) \int_{-a_n}^{a_n} |y|\, dQ_n(y) \ .$$

By Markov's inequality, $\int_{-a_n}^{a_n} |y|\, dQ_n(y) = O_p(a_n n^{-\frac{1}{2}})$, whence

$$(n\varepsilon_n \log n)^{\frac{1}{2}} \sup_u |I_2(u)| = O_p(a_n(n\varepsilon_n^2)^{-1}) = o_p(1) \ . \qquad \text{(B4)}$$

We deal only with $I_3(u)$, as $I_4(u)$ is similar. If $V[\cdot]$ denotes total variation,

$$|I_3(u)| \le a_n\varepsilon_n^{-1}\, O_p(n^{-\frac{1}{2}}) V[G_n(\cdot,u)]$$

$$= a_n\varepsilon_n^{-1}\, O_p(n^{-\frac{1}{2}})\{\varepsilon_n^{-1}\, O_p(n^{-\frac{1}{2}})\}$$

uniformly in $a \le u \le b$. Thus, by (B4), $(n\varepsilon_n \log n)^{\frac{1}{2}} \sup_u\{|I_3(u)| + |I_4(u)|\} \xrightarrow{P} 0$. Similarly,

$$(n\varepsilon_n \log n)^{\frac{1}{2}} \sup\{|I_1(u)|: a \le u \le b\} \le (n\varepsilon_n \log n)^{\frac{1}{2}}\, \varepsilon_n^{-1}\, O_p(n^{-\frac{1}{2}}) V[yG_n(x,u)]$$

$$= (n\varepsilon_n \log n)^{\frac{1}{2}}\, \varepsilon_n^{-1}\, O_p(n^{-\frac{1}{2}}) a_n\varepsilon_n^{-1}\, O_p(n^{-\frac{1}{2}})$$

$$= o_p(1) \ ,$$

where here V denotes total variation in $(x,y)$ over $R \times [-a_n,a_n]$.

Thus to verify (B.1.4) we must show $(n\varepsilon_n \log n)^{\frac{1}{2}} \sup|J_1^{(1)}| \xrightarrow{P} 0$. Routine calculations show

$$\varepsilon_n |J_1^{(1)}| \le \varepsilon_n^{-1}\, O_p(n^{-\frac{1}{2}}) \int_{|y|\ge a_n} |y|\, dF_n(y) + a_n \varepsilon_n^{-1}\, O_p(n^{-\frac{1}{2}}) ,$$

completing the proof by (B4).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | | |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (*and Subtitle*)<br>Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates | | 5. TYPE OF REPORT & PERIOD COVERED<br>TECHNICAL |
| | | 6. PERFORMING ORG. REPORT NUMBER<br>Mimeo Series No. 1271 |
| 7. AUTHOR(s)<br>Gordon J. Johnston | | 8. CONTRACT OR GRANT NUMBER(s)<br>Grant AFOSR-75-2796<br>Grant AFOSR-80-0080 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Air Force Office of Scientific Research<br>Bolling Air Force Base<br>Washington, DC 20332 | | 12. REPORT DATE<br>March 1980 |
| | | 13. NUMBER OF PAGES<br>18 |
| 14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*) | | 15. SECURITY CLASS. (*of this report*)<br>UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for Public Release - Distribution Unlimited

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

density estimates, nonparametric regression estimates, maximal deviations, Gaussian processes

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

Let (X,Y) have regression function $m(x) = E(Y|X=x)$, and let X have a marginal density $f_1(x)$. We consider two nonparametric estimates of m(x): the Watson estimate when $f_1$ is known and the Yang estimate when $f_1$ is known or unknown. For both estimates the asymptotic distribution of the maximal deviation from m(x) is proved, thus extending results of Bickel and Rosenblatt for the estimation of density functions.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)