RATES OF CONVERGENCE FOR WINDOW ESTIMATES OF A
REGRESSION FUNCTION

C. Spiegelman

Department of Statistics
The Florida State University
Tallahassee, Florida 32306

FSU Statistics Report No. 406
January 1977

ABSTRACT

It is shown here that rates of convergence for window estimates of a regression function are not highly dependent upon smoothness conditions.

INTRODUCTION

It is shown here that non parametric estimates of a regression function are to a certain extent a property of moments, and not necessarily a property of smoothness in the usual sense. Indeed no inequality in this paper requires the regression function of interest to be continuous. To some extent the availability of rates of convergence for irregular functions justifies, for a practical point of view, the use of window estimates of a regression function, because if the functions of interest are sufficiently
smooth then estimates such as Splines (for example see Feder (1975)), and the method of Sacks and Ylvisaker (1976) "appear' to be better choices for estimating the unknown regression function.

**GENERAL FRAMEWORK**

Let \((\Theta, X)\) be a random pair such that \(E\theta^2 < \infty\). For any estimate \(\delta(X)\) of \(\Theta\) the statistician incurs loss \(L(\Theta, \delta) = (\Theta - \delta)^2\). The risk to the statistician is defined as \(R(\Theta, \delta) = E[L(\Theta, \delta)]\). It is well known that the Bayes estimator \(\delta*(X) = E[\Theta|X]\) has the property that \(R(\Theta, \delta*) = \min \delta R(\Theta, \delta)\). Often the form of the joint distribution of \((\Theta, X)\) is not known, even approximately, but from prior experience, independent and identically distributed pairs of random variables \((\Theta_i, X_i), i = 1, \ldots, n\) having the same distribution as \((\Theta, X)\) are available. When the joint distribution of \((\Theta, X)\) is not known up to a finite number of parameters, the usual maximum likelihood procedures, and least square methods of estimation are not applicable. However it has been shown by Spiegelman (1976) and Stone (1975) that certain types of nonparametric estimates, \(\delta_n\) of \(\delta*\) have the property that \(E(\delta_n - \delta*)^2 \to 0\) as \(n \to \infty\), under weak assumptions.

In order to define a window estimate of \(\delta*\) some additional notation will be given.

Let \(b\) be a positive real number. Define \(W_i = \begin{cases} 1 & \text{if } |X_i - X| < b \\ 0 & \text{otherwise} \end{cases}\) and let \(\delta_n = \frac{\sum_{i=1}^n \Theta_i W_i}{\sum_{i=1}^n W_i}\), \(\delta_n\) is called a window estimate of \(\delta*\). Notice that \(E(\delta_n - \Theta)^2 = R(\Theta, \delta*) + E(\delta_n - \delta*)^2\). If \(E(\delta_n - \delta*)^2\) is small one may be confident that \(\delta_n\) is a good estimate of \(\delta*\), and that the lack of knowledge about the joint distribution of \((\Theta, X)\) is not serious.

**STATEMENT AND PROOFS OF THE LEMMA AND THEOREMS**

Lemma. \[E(\delta_n - \delta)^2 \leq 20 \max \left\{ E[(\delta*(X_i) - \delta*(X))^2|W_i = 1], \right.\]
\[
\left. \int_\infty^\infty \frac{E[\delta_i^2|W_i = 1, X = t] P(\Theta) dt}{1 + \sum_{i=1}^n P(|X_i - t| < b)} \right\}.
\]
Amplification: It was shown by Spiegelman (1976) that
\[ E[(\delta*(X_1) - \delta*(X))^2 | W_1 = 1] \rightarrow 0 \text{ as } n \rightarrow \infty, \] and it is not hard to show that
\[ \int_{-\infty}^{\infty} \frac{E[0^2 | W_1 = 1, X = t]}{1 \vee n \cdot P[|X_1 - t| < b]} \cdot P(dt) \rightarrow 0 \text{ as } n \rightarrow \infty. \] Although both terms on the right hand side of the above inequality appear to be complicated, various interpretable inequalities will follow from this result.

Proof of Lemma.

By straightforward computation

1) \[ E(\delta_n - \delta^*)^2 \leq 4 \left( E \left[ \frac{\sum_{i=1}^{n+1} (\delta*(X_1) - \delta*(X)) W_i}{1 \vee \sum_{i=1}^{n} W_i} \right]^2 \right) + \left( E \left[ \frac{\sum_{i=1}^{n} (\delta_1 - \delta*(X_1)) W_1}{1 \vee \sum_{i=1}^{n} W_i} \right]^2 \right) + E\delta^2(X)I_{\{0\}}[W_1], \]

The first term on the right hand side of 1) will be dealt with first.

By application of Cauchy-Schwartz

\[ E \left[ \frac{\sum_{i=1}^{n} (\delta*(X_1) - \delta*(X)) W_i}{1 \vee \sum_{i=1}^{n} W_i} \right]^2 \leq E \left[ \frac{\sum_{i=1}^{n} (\delta*(X_1) - \delta*(X))^2 W_i}{1 \vee \sum_{i=1}^{n} W_i} \right] \]

\[ = E \left[ \frac{\sum_{i=1}^{n} E[(\delta*(X_1) - \delta*(X))^2 | W_1 = 1, X = j \neq i, X] W_i}{1 \vee \sum_{i=1}^{n} W_i} \right] \]

\[ = E \left[ \sum_{i=1}^{n} E[(\delta*(X_1) - \delta*(X))^2 | W_1 = 1, X] W_i \right] \]

\[ \leq E[(\delta*(X_1) - \delta*(X))^2 | W_1 = 1]. \]

Now the second term on the right hand side of 1) will be dealt with.
\[
E\left(\frac{\sum_{i=1}^{n}(\Theta_i-\delta*(X_i))W_i}{1 + \sum_{i=1}^{n}W_i}\right)^2 
\leq E\left(\frac{E[(\Theta_i-\delta*(X_i))^2|W_i=1, X]W_i}{1 + \sum_{i=1}^{n}W_i}\right) 
\leq E\left(\frac{E[\Theta^2|W_i=1, X]}{1 + \sum_{i=1}^{n}W_i}\right) 
\leq 2\int_{-\infty}^{\infty} \frac{E[\Theta^2|W_i=1, X=t]}{1 + \sum_{i=1}^{n}W_i} \, dx \, dt, \text{ by application of an elementary binomial inequality found in the appendix to this paper.}
\]

Finally the term \(E\delta^2(X)I_{\{0\}}(\sum_{i=1}^{n}W_i)\) is dealt with:

\[
E\delta^2(X)I_{\{0\}}(\sum_{i=1}^{n}W_i) = E\delta^2(X) \left(\frac{1 + \sum_{i=1}^{n}W_i - \sum_{i=1}^{n}W_i}{1 + \sum_{i=1}^{n}W_i}\right) 
\leq E\delta^2(X) \left(\frac{1 + \sum_{i=1}^{n}W_i}{1 + \sum_{i=1}^{n}W_i}\right) 
\leq 2\int_{-\infty}^{\infty} \frac{E[\Theta^2|W_i=1, X=t]}{1 + \sum_{i=1}^{n}W_i} \, dx \, dt, \text{ by previously used arguments. Q.E.D.}
\]

Some moment inequalities are not given for the expression \(E[\delta*(X_1-\delta*(X))^2|W_i=1]\). Let \(F(t)\) be the right continuous distribution function of \(X\). Define \(\delta = \{t|F(t) - F(t-) > 0\}\) and

\[
W_1 = \begin{cases} 
W_i & \text{if } X \neq X_1 \\
0 & \text{if } X = X_1
\end{cases}
\]

**Theorem 1.** If \(\|\delta^*\|_1 = 1\), and \(\delta \neq \phi\), then

\[
E[(\delta*(X_1)-\delta*(X))^2|W_1=1] \leq \frac{4EW'}{E_t\delta(F(t)-F(t-))^2}
\]

**Proof:**

\[
E[(\delta*(X_1)-\delta*(X))^2|W_1=1] = \left\{\sum_{W_1=1} (\delta*(X_1)-\delta*(X))^2 \frac{x_1}{EW} \frac{P_{X_1}(dx_1)}{P_X(dx_1)}\right\}
\]

\[
\leq 4\int_{W_1=1} \frac{P_{X_1}(dx_1)P_{X_2}(dx_2)}{E_t\delta(F(t)-F(t-))^2}
\]

Q.E.D.
Let $\Delta(r)$ be a monotonically increasing function defined on the positive half line. Such that $\Delta(0) = 0$. Let $(\delta* \mathbb{C} X_1) - \delta(X))^2 = K(X_1, X) \Delta(|X_1 - X|)$

Theorem 2. If $E K^2(X, X_1) = 1$, then

$$E[(\delta* \mathbb{C} X_1) - \delta(X))^2 | W_1] \leq 6 \Delta(b)$$

Proof: $E[(\delta* \mathbb{C} X_1) - \delta(X))^2 | W_1=1] \leq \Delta(b) E[K(X_1, X) | W_1=1]$

$$= \Delta(b) \sum_{i=-\infty}^{\infty} E[K(X_1, X) | W_1=1, X \in U_i] P[X \in U_i | W_1=1]$$

where $\{U_i\}_{i=-\infty}^{\infty}$ is an interval partition of the line such that $U_{i-1}$ lies to the left of $U_i$ and each $U_i$ has length $b$.

$$E[K(X_1, X) | W_1=1, X \in U_i] \leq \sum_{j=-1}^{1} \int_{U_{i-j}}^{U_i} \frac{K(X_1, X)}{P(W_1=1, X \in U_i)} P(dx_1) P(dx)$$

$$\leq \sum_{j=-1}^{1} \int_{U_{i-j}}^{U_i} \frac{(1 \vee K(X_1, X))}{P(U_i)} P(dx_1) P(dx)$$

$$\leq \sum_{j=-1}^{1} \int_{U_{i-j}}^{U_i} \frac{(1 \vee K(X_1, X))^2}{P(U_i)} P(dx_1) P(dx),$$

by application of the Cauchy-Schwarz inequality. Combining the above one obtains

$$E[(\delta* \mathbb{C} X_1) - \delta(X))^2 | W_1=1] \leq \Delta(b) \sum_{i=-\infty}^{\infty} \sum_{j=-1}^{1} \int_{U_{i-j}}^{U_i} (1 \vee K(X_1, X))^2 P(dx_1) P(dx)$$

$$\leq 3 \Delta(b) E[(1 \vee K(X_1, X))^2] \leq 6 \Delta(b) Q \Delta(b)$$

and let $U_i$ be an interval partition of the line, such that $U_{i-1}$ lies to the left of $U_i$ and each $U_i$ has length $b$.

Suppose there exists a function $K^*(X)$ such that

$$\int_{\{X_1 \in U_i, X \in U_j\}} K(X_1, X) P(dx_1) P(dx) \leq P(X_1 \in U_i) \int_{\{X \in U_j\}} K(x, X) P(dx)$$
Theorem: If $E(\theta^2) < \infty$, and $K(X) < \infty$, and $X$ either has a density $\eta$ such that $||h||_\infty < \infty$ or $S$, the set of mass points, is nonempty then

$$E[(\delta^*(X_1) - \delta^*(X))^2]_{W_1=1} \leq \begin{cases} \frac{3||h||_\infty E(\theta^2)}{\Delta(b)} & \text{if } X \text{ has density } \eta \\ \frac{3E\Delta(b)}{\sum_{t \in S}(F(t) - F(t-))^2} & \text{if } S \neq \emptyset \end{cases}$$

Proof: $E[K(X_1, X) | W_1=1, X \in U_1] \leq \Delta(b) \sum_{i=-\infty}^{\infty} E[K(X_1, X) | W_1=1, X \in U_1] P[X \in U_1 | W_1=1]$.

Now we look at the term

$$E[K(X_1, X) | W_1=1, X \in U_1] \leq \frac{\sum_{i=-1}^{1} \int_{U_1} K^*(X) P(dx)}{P(W_1=1, X \in U_1)} \int_{U_1} K^*(X) P(dx).$$

Then

$$P[X \in U_1 | W_1=1, X \in U_1] E[K(X_1, X) | W_1=1, X \in U_1] \leq \frac{\sum_{i=-1}^{1} P(U_{i-1})}{E(W_1)} \int_{U_1} K^*(X) P(dx).$$

It is shown in the appendix to this paper that $EW \geq b$, and therefore if $X$ has a density $\eta$ such that

$$||h||_\infty < \infty \frac{\sum_{i=-1}^{1} P(U_{i-1})}{EW} \int_{U_1} K^*(X) P(dx) \leq 3||h||_\infty \int_{U_1} K^*(X) P(dx).$$

Also if $S \neq \emptyset$

$$\frac{\sum_{i=-1}^{1} P(U_{i-1})}{EW} \int_{U_1} K^*(X) P(dx) \leq \frac{3}{\sum_{t \in S}(F(t) - F(t-))^2} \int_{U_1} K^*(X) P(dx).$$

Finally the result is obtained by summing on $i$. Q.E.D.

Now the expression

$$\int_0^\infty E[\theta^2 | W_1=1, X=t] \frac{\eta}{\sum_{i \leq n} P(|X_1-t| < b)} P_X (dt)$$

is dealt with.

Theorem 3. If $||\theta||_\infty = 1$ and $E[X]^{1+t} = 1$ for some $t > 0$ then

$$\int_{-\infty}^\infty E[\theta^2 | W_1=1, X=t] \frac{\eta}{\sum_{i \leq n} P(|X_1-t| < b)} P_X (dt) \leq \left(\frac{2+s}{n} + \frac{1}{b^{1+t}} \int_0^\infty \frac{1}{\xi^{1+t}} d\xi \right) \text{ with }$$

$$s_0 = \frac{nt}{b^{1+t}}.$$

\[\]
Proof: Let \( \{U_i\}_{i=1}^{\infty} \) be an interval partition of \( \mathbb{R} \) such that each \( U_i \) has length \( b \) and \( U_{i-1} \) lies to the left of \( U_i \). Then one has

\[
\int_{-\infty}^{\infty} E[\theta_t^2 | W_i=1, X=t] \frac{1}{1 \vee n} P([X_1-t] < b) P_x(dt) \leq E_{i=\infty}^{\infty} \int_{U_i} \frac{P}{1 \vee n} P([X_1-t] < b) \frac{P_x(U_i)}{1 \vee n P_x(U_i)} ;
\]

Take \( \{U_i^*\}_{i=1}^{\infty} \) to be a reordering of \( \{U_i\}_{i=1}^{\infty} \) s.t. \( P(U_i^*) \geq P(U_i) \). If \( X^* \) is a random variable such that \( P((X-1)b \leq X^* \leq Kb) = P(U_i^*) \), then it is easy to show that \( E|X^*|^{t+1} \leq 1 \). Note the following elementary inequalities

\[
1 \geq E|X^*|^{1+t} \geq \int_{U_i^*} |s|^{1+t} dF(s) \geq ((1-1)b)^{1+t} P(U_i^*).
\]

Also one has \( \frac{X}{1 \vee nX} \leq \frac{1}{n} \) for \( X \geq 0 \). Then one has

1) \( E_{i=\infty}^{\infty} \frac{P_x(U_i)}{1 \vee n P_x(U_i)} = E_{i=1}^{\infty} \frac{P(U_i^*)}{1 \vee n P(U_i)} \frac{1}{n} + E_{i \geq s_0} P(U_i^*) \leq \frac{S_0}{n} + \frac{1}{b^{1+t}} \int_{s_0}^{\infty} \frac{1}{1+t} d\ell + \frac{2}{n}
\]

for all \( s_0 > 0 \). By calculus one obtains the minimum of the right hand side of 2 is obtained with \( s_0 = \left(\frac{nt}{b^{1+t}}\right)^{1/2} \).

Theorem 4. If \( E\theta_1=1 \), and \( E|X|^{1+t}=1 \), then

\[
\int_{-\infty}^{\infty} E[\theta_t^2 | W_i=1, X=t] \frac{1}{1 \vee n} P([X_1-t] < b) P_x(dt) \leq 3^{1/2} \left(\frac{2+1+b}{n} + \frac{1}{b^{1+t}} \int_{s_0}^{\infty} \frac{d\ell}{\ell^{1+t}} \right)^{1/2}
\]

with so = \( \left(\frac{nt}{b^{1+t}}\right)^{1/2+t} \).

Proof: By application of the Cauchy-Schwartz inequality

\[
\int_{-\infty}^{\infty} E[\theta_t^2 | W_i=1, X=t] \frac{1}{1 \vee n} P([X_1-t] < b) P_x(dt) \leq \left(\int_{-\infty}^{\infty} E[\theta_t^2 | W_i=1, X=t] P_x(dt)\right)^{1/2}
\]

\[
\int_{-\infty}^{\infty} (E[\theta_t^2 | W_i=1, X=t])^{1/2} \leq E^{1/2}[\theta_1^4 | W=1],
\]

it can be shown that
following the methodology of theorem 2 \(E[\mathbf{c}_1^4|W_1=1] \leq 3E \mathbf{c}_1^4\). Now

\[
\int_{-\infty}^{\infty} \left\{ \frac{P_X(t)}{1+\sqrt{n_2^2 P(X_1-t)^2}} \right\} \leq \int_{-\infty}^{\infty} \frac{P_X(t)}{1+\sqrt{n_2 P(X_1-t)^2}}
\]

\[
\leq \left[ \frac{2+s_0}{n} + \frac{1}{b^{1+t}} \int_{0}^{\infty} \frac{d\ell}{\ell^{1+t}} \right] \text{ with } s_0 = \left( \frac{nt}{b^{1+t}} \right)^{1/2+t},
\]

by the proof of theorem 3. Q.E.D.

**Examples:**

Let \(\delta^*(X) = \sum_{n=0}^{\infty} \frac{1}{2^n} \cos 3^n X\). It is well known that \(\delta^*(X)\) is nowhere differentiable.

**Example 1.** \(F_X(t) = \begin{cases} 0 & t < 0 \\ \frac{t}{6} & t = 0 \\ \frac{t}{2} + \frac{1}{3} & 0 \leq t \leq 1 \\ 1 & t > 1 \end{cases} \)

\[E[(\delta^*(X_1) - \delta^*(X))^2|W_1=1] \leq 12b \text{ by Theorem 1.} \]

**Example 2.** \(F_X(t) = \begin{cases} 0 & t < 0 \\ t & 0 \leq t \leq 1 \\ 1 & t > 1 \end{cases} \)

\[(\delta^*(X_1) - \delta^*(X))^2 \leq 2 \frac{|X_1-X|^1}{|X_1-X|^1-\varepsilon} \text{ for } 0 < \varepsilon < 1 \]

\[
\int |t_1-t_0| < b \frac{1}{|t_1-t_0|^{1-\varepsilon}} \ dt_1 \leq \left( \frac{2b}{\varepsilon} \right)^\varepsilon,
\]

and hence \(E[(\delta^*(X_1) - s^*(X))^2|W_1=1] \leq 7b \) by application of Theorem 2.

**COMMENT**

If \(\delta^*(X) = b(X)\) where \(X\) is Brownian motion note that

\[
b E \left[ \frac{(B(X_1) - B(X))^2}{|X_1-X|^{1/2}} \right] = 1 \geq E \left[ (B(X_1) - B(X))^2|W_1=1 \right] \] so that there
exists at least one Brownian path such that $E[(b(X_1) - b(X))^2 | \delta = b] \leq b$.

(B of course stands for a random Brownian process, independent of $X, X_1, \ldots, X_n$.)

**CONCLUSIONS**

To a large extent rates of convergence for window estimates of a regression function are a property of moments, and not of smoothness of the regression function.

**Appendix 1. An elementary Binomial inequality**

Let $I_n$ be the indicator function of the set $\Lambda$ and

$$\delta_n = 1 + \sum_{i=1}^{n-1} I_n(X_i).$$

Then $E \frac{1}{1 + \delta_n} \leq \frac{1}{1 + n} \Pr(X \in A)$.

**Proof:** $\delta_n \sim B(n-1, p)$ where $p = p(X \in A)$.

$$E \frac{1}{1 + \delta_n} = \sum_{j=0}^{n-1} \left( \frac{1}{j+1} \right) \binom{n-1}{j} p^j (1-p)^{n-1-j}$$

Now

$$\frac{p^j (n-1)}{j+1} \binom{n}{j+1} \frac{1}{p^{j+1}} = \frac{n}{np} \binom{n}{j} \frac{p^{j+1}}{np^j}.$$  

Then 4) equals $\frac{1}{np} \sum_{k=1}^{n} \binom{n}{k} p^{k} (1-p)^{n-k}$, letting $k = j+1 \leq \frac{1}{np}$.

Finally $E \frac{1}{1 + \delta_n} \leq 1$. Q.E.D.

**Appendix 2.** $EW \geq b$.

$$EW \geq \sum_{i=-\infty}^{\infty} p^2(U_i),$$

where $U_i$ are as in theorem 2. Now suppose $E_{i=-n}^{n} U_i = [-1, 1]$ then $n = \frac{2(b + 1)}{b}$ and we have $E_{i=-n}^{n} p^2(U_i) \geq b P[-1, 1]$ by repeating a similar argument over intervals $[\pm 3, \pm 5]$, etc. we obtain the result.

**ACKNOWLEDGMENT**

This paper was motivated by discussions with Jerome Sacks.

The author also thanks Jayaram Sethuraman for pointing out an over-statement in the original version of theorem 1.
BIBLIOGRAPHY


