Uniform Confidence Bounds for Regression Based on a Simple Moving Average
Author(s): Steinar Bjerve, Kjell A. Doksum, Brian S. Yandell
Source: *Scandinavian Journal of Statistics*, Vol. 12, No. 2 (1985), pp. 159-169
Published by: Blackwell Publishing on behalf of Board of the Foundation of the Scandinavian Journal of Statistics
Stable URL: http://www.jstor.org/stable/4615981
Accessed: 24/02/2009 09:58

Scand J Statist 12: 159–169, 1985

Uniform Confidence Bounds for Regression Based on a Simple Moving Average

STEINAR BJERVE[1], KJELL A. DOKSUM[2] and BRIAN S. YANDELL[3]
[1] *University of Oslo;* [2] *University of California, Berkeley;* [3] *University of Wisconsin, Madison*

ABSTRACT. Let $m(x)=E(Y|X=x)$ be the regression function of Y on x. Suppose that $Y_1, \ldots, Y_n$ are independent observations of Y at $x=x_1, \ldots, x_n$. We consider nearest neighbour estimates, $\hat{m}(x)$, and employ well-known inequalities to obtain finite sample size uniform confidence bounds for $E\hat{m}(x)$ and asymptotic uniform confidence bounds for $E\hat{m}(x)$ and $m(x)$ based on $\hat{m}(x)$. Finally we discuss bias and consistency properties of $\hat{m}(x)$.

Key words: non-parametric regression, confidence bands, nearest neighbour, consistency.

1. Introduction

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate population with distribution function $F(x, y)$. We are interested in constructing uniform confidence bounds and bands for the unknown regression function

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

without making parametric assumptions about either m or the distributional form of F. We will assume existence of the conditional variance function given by

$$\sigma^2(x)=\text{var}\ (Y|X=x).$$

The results we derive are conditional on the given x values, essentially reducing us to the usual regression situation.

For a recent review of some of the work that has been done in non-parametric regression see Collomb (1981). Asymptotic results based on strong approximations leading to simultaneous confidence bands have been obtained for histogram estimates by Major (1973), for kernel estimates by Révész (1979) and Liero (1982), and for nearest neighbour estimates by Révész (1977, 1979). Although most of these results assume that the variance does not depend on x, Liero (1982) derived simultaneous confidence bands with mild restrictions on the conditional variance. Wahba (1983) developed Bayesian confidence bands for smoothing spline estimates of the regression function. Asymptotic simultaneous confidence bands for censored survival problems, using ideas from nearest neighbour approaches, have been derived by Doksum & Yandell (1982).

Our construction will be based on the k-nearest neighbour estimator,

$$\hat{m}(x)=\sum Y_i/k,$$

where the summation is taken over the indices of the k values of X lying closest to x.

Section 2 provides some preliminary probability results based on inequalities. A Chebychev type inequality for regression and exact uniform confidence bounds for $E\hat{m}(x)$ are derived in section 3. Asymptotic distribution theory is used to obtain another set of bounds in section 4. Section 5 focuses on bounds based on non-overlapping neighbourhoods while bias and consistency is considered in section 6. Conditions under which the bounds and bands for $E\hat{m}(x)$ can be extended to $m(x)$ are given in section 7 and the problem of choosing the size of the neighbourhood is discussed in section 8. Finally, section 9 presents an illustration of the bounds.

2. Preliminaries

Let X_{ni} be the ith order statistic obtained from $X_1, X_2, \ldots, X_n$, and let Y_{ni} denote the ith induced order statistic of the Y-observations. That is, if Q_i is the anti-rank of X_i, $X_{Q_i}=X_{ni}$, then $Y_{ni}=Y_{Q_i}$. We assume that the distribution of X_i is continuous, in which case, conditional on $X_{ni}=x_{ni}$, $i=1,\ldots,n$, the Y_{ni} are independent (see e.g. Bhattacharya, 1974, lemma 1). Thus the distributional assumptions may be written as follows:

$$Y_{ni}=m(x_{ni})+\varepsilon_{ni}, \qquad i=1,\ldots,n, \tag{2.1}$$

where $\varepsilon_{n1},\ldots,\varepsilon_{nn}$ are independent,

$$x_{n1}\leqslant\ldots\leqslant x_{nn}, \quad E\varepsilon_{ni}=0 \quad \text{and} \quad \mathrm{Var}\,(\varepsilon_{ni})=\sigma^2(x_{ni}).$$

For convenience we will, until section 5, write Y_i for Y_{ni} and x_i for x_{ni}.

Our main estimate for $m(x)$ will be the k-nearest neighbour estimate, or simple moving average, defined by

$$\hat{m}(x)=\sum_{i\in I_{nk}(x)} Y_i/k, \tag{2.2}$$

where $I_{nk}(x)$ are the indices of the k values of $x_1,\ldots,x_n$ closest to x. In the case of a tie between two values x_s and x_{s+k} we adopt the convention of including only the smaller of the two indices, s, in $I_{nk}(x)$. Thus $\hat{m}(x)$ will be left continuous.

We also consider an estimate introduced by Révész (1979) which is a nearest neighbour estimate with the index set, $I'_{nk}(x)$, balanced around the point x. That is, the index set has the k nearest neighbours such that $x_i\leqslant x$ for at least $k/2$ indices i, and $x_i\geqslant x$ for at least $k/2$ indices, where k is assumed even. In other words, x is a median of the set $\{x_i: i\in I'_{nk}(x)\}$. We denote this estimate by

$$\tilde{m}(x)=\sum_{i\in I'_{nk}(x)} Y_i/k. \tag{2.3}$$

Again, in case of a tie between two indices, we include only the smaller one.

Let

$$J_i=\{x: I_{nk}(x)=\{i+1,\ldots,i+k\}\}, \qquad i=0,\ldots,n-k. \tag{2.4}$$

Denote $\hat{m}_i$ by

$$\hat{m}_i=\hat{m}(x) \quad \text{for} \quad x\in J_i, \qquad i=0,\ldots,n-k.$$

Thus we have

$$\hat{m}_i=(Y_{i+1}+\ldots+Y_{i+k})/k, \qquad i=0,\ldots,n-k. \tag{2.5}$$

We assume that the x's all are in the interval (a_n, b_n) where possibly $a_n=-\infty$, $b_n=\infty$. Then $J_0=(a_n, (x_1+x_{k+1})/2]$, $J_{n-k}=((x_{n-k}+x_n)/2, b_n)$ and

$$J_i=((x_i+x_{i+k})/2, \ (x_{i+1}+x_{i+k+1})/2], \qquad i=1,\ldots,n-k-1.$$

Let

$$S_i=\hat{m}_i-E\hat{m}_i=\sum_{j=i+1}^{i+k} W_j, \qquad i=0,\ldots,n-k,$$

where

$$W_j=[Y_j-m(x_j)]/k, \qquad j=1,\ldots,n.$$

Define $\tilde{S}_0=0$ and

$$\tilde{S}_i=\sum_{j=1}^{i} W_j, \qquad i=1,\ldots,n.$$

Then

$$S_i=\tilde{S}_{i+k}-\tilde{S}_i,\ i=0,\ldots,n-k.$$

Lemma
For each $t>0$

$$P(S_i\geq -t; \qquad \textit{all } i=0,\ldots,n-k)\geq 1-4(kt)^{-2}\left\{\sum_{i=1}^{n-k}\sigma^2(x_i)+\sum_{i=1}^{n}\sigma^2(x_i)\right\}.$$

Proof. Let

$$A=\{\tilde{S}_{k+i}\geq -t/2, \qquad \text{all } i=0,\ldots,n-k\}=\{\tilde{S}_i>-t/2, \qquad \text{all } i=k,\ldots,n\}$$

$$B=\{\tilde{S}_i\leq t/2, \qquad \text{all } i=1,\ldots,n-k\}$$

$$C=\{S_i\geq -t, \qquad \text{all } i=0,\ldots,n-k\}.$$

Then $A\cap B\subset C$ and by Bonferroni's inequality, we have

$$P(C)\geq P(A\cap B)\geq 1-P(A^c)-P(B^c)$$

Kolmogorov's inequality (e.g. Loève, 1963, p. 235), yields

$$P(A^c)=P\left(\max_{k\leq i\leq n}\{-\tilde{S}_i\}>t/2\right)\leq P\left(\max_{1\leq i\leq n}\{-\tilde{S}_i\}>t/2\right)\leq 4(kt)^{-2}\sum_{i=1}^{n}\sigma^2(x_i)$$

By a similar argument,

$$P(B^c)\leq 4(kt)^{-2}\sum_{i=1}^{n-k}\sigma^2(x_i)$$

and the result follows.

3. Exact uniform confidence bounds for $E\hat{m}(x)$

Since $\inf_{(x)}\{\hat{m}(x)-E\hat{m}(x)\}=\min\{S_i;\ i=0,\ldots,n-k\}$, then by the lemma,

$$P\left(\inf_x[\hat{m}(x)-E\hat{m}(x)]\geq -t\right)\geq 1-4(kt)^{-2}\left[\sum_{i=1}^{n-k}\sigma^2(x_i)+\sum_{i=1}^{n}\sigma^2(x_i)\right].$$

Thus if we set

$$t_\alpha=\{2/(k\alpha^{1/2})\}\left[\sum_{i=1}^{n-k}\sigma^2(x_i)+\sum_{i=1}^{n}\sigma^2(x_i)\right]^{1/2}$$

then $\hat{m}(x)+t_\alpha$ is a simultaneous upper confidence boundary for $E\hat{m}(x)$ with confidence coefficient at least $1-\alpha$.

Similarly, $\hat{m}(x)-t_\alpha$ is a simultaneous lower confidence boundary for $E\hat{m}(x)$ with confidence coefficient at least $1-\alpha$. Let $T^{\pm}=\sup_{(x)}\pm\{\hat{m}(x)-E\hat{m}(x)\}$ and $T=\max\{T^-, T^+\}$. Then $P(T\geq t)\leq P(T^+\geq t)+P(T^-\geq t)$. It follows that we have obtained the following uniform Chebychev type inequality for regression. For each $t>0$,

$$P\left\{\sup_x |\hat{m}(x)-E\hat{m}(x)|\leq t\right\}\geq 1-8(kt)^{-2}\left[\sum_{i=1}^{n-k}\sigma^2(x_i)+\sum_{i=1}^{n}\sigma^2(x_i)\right]. \tag{3.1}$$

Thus $\hat{m}(x)-E\hat{m}(x)$ converges in probability to 0 provided $k^{-2}\sum_{(i=1)}^{(n)}\sigma^2(x_i)$ tends to zero as $n\to\infty$ and $k\to\infty$. If $\sigma^2(x)$ is bounded, this follows if $(n/k^2)\to 0$ as $n\to\infty$.

Note that

$$\hat{m}(x)\pm 2^{1/2}t_\alpha. \tag{3.2}$$

is a level $(1-\alpha)$ simultaneous confidence band for $E\hat{m}(x)$. If we assume that $\sigma^2(x)=\sigma^2$ for all x, then

$$t_\alpha=(2\sigma/k)\{(2n-k)/\alpha\}^{1/2}$$

and the width of the confidence band is $(4\sigma/k)\{(4n-2k)/\alpha\}^{1/2}$. If we choose $k=n^{\Delta+1/2}$, $\Delta<½$, then the width is

$$8n^{-\Delta}\sigma\alpha^{-1/2}$$

minus a smaller order term. Thus the width tends to zero provided that $0<\Delta<½$.

To use the band, we need an estimate of σ^2. A natural estimate is the residual mean square

$$\hat{\sigma}^2=n^{-1}\sum_{i=1}^{n}\{Y_i-\hat{m}(x_i)\}^2. \tag{3.3}$$

This estimate of variance is consistent under the conditions of Révész (1979) if $\hat{m}(\cdot)$ is replaced by $\bar{m}(\cdot)$ of equation (2.3). We conjecture that the proposed estimate (3.3) is also consistent under the same conditions.

4. Bounds based on asymptotic distribution theory

We assume that $\sigma^2(x_i)=\sigma^2$ and we let $\tilde{S}_i$, A and B be defined as in the proof of the lemma of section 2. First note that

$$\begin{aligned}\max_{k\leq i\leq n}\{-\tilde{S}_i\}&=\max\{-\tilde{S}_k, -\tilde{S}_k-W_{k+1},\ldots,-\tilde{S}_k-W_{k+1}-\ldots-W_n\}\\&=-\tilde{S}_k+\max\{0,-W_{k+1},\ldots,-W_{k+1}-\ldots-W_n\}.\end{aligned}$$

Thus we can write

$$kn^{-1/2}\max_{k\leq i\leq n}\{-\tilde{S}_i\}=-(k/n)^{1/2}k^{1/2}\tilde{S}_k+\{1-(k/n)\}^{1/2}(n-k)^{-1/2}\max_{0\leq i\leq n-k}\{-\hat{S}_i\}$$

where $\hat{S}_0=0$ and $\hat{S}_i=k\sum_{(j=1)}^{(i)}W_{k+j}$, $i=1,\ldots,n-k$. In the above equation, the first term on the right-hand side converges to zero provided $(k/n)\to 0$ as $n\to\infty$. Thus $P(kn^{-1/2}\max_{(k\leq i\leq n)}\{-\tilde{S}_i\}\leq t)$ converges to $2\Phi(t/\sigma)-1$ by a result which gives the asymptotic distribution of the maximum of partial sums in terms of the distribution of the maximum of Brownian motion (e.g. Billingsley, 1968, p. 72). Thus, when $(k/n)\to 0$ as $n\to\infty$, then

$$P(A)=P\left(\max_{k\leq i\leq n}\{-\tilde{S}_i\}\leq t/2\right)\approx 2\Phi\{kt/(2\sigma n^{1/2})\}-1.$$

Table 1. *Relative widths of bands (3.2) and (4.1)*

α	0.1	0.05	0.01
(3.2)/(4.1)	2.28	2.82	5.02

A similar argument shows that the same approximation holds for $P(B)$. Using this and the arguments of section 3 we find that asymptotically a level $(1-\alpha)$ confidence band for $E\hat{m}(x)$ is

$$\hat{m}(x)\pm 2\sigma\Phi^{-1}(1-\alpha/4)n^{1/2}/k. \tag{4.1}$$

Choosing $k=n^{\Delta+1/2}$, the width of the band is

$$4\sigma\Phi^{-1}(1-\alpha/4)n^{-\Delta}.$$

Table 1 compares the relative widths of the confidence bands (3.2) and (4.1).

5. Bands based on non-overlapping neighbourhoods

The bounds and bands in sections 3 and 4 will be of use only for large data sets. In this section we develop a band which is much narrower, but it is simultaneous only for a sequence $t_{n1},\dots,t_{n\beta}$ of x-values. The model is

$$Y_{ni}=m(x_{ni})+\varepsilon_{ni}, \qquad x_{1n}<\dots<x_{nn}$$

where $\varepsilon_{n1},\dots,\varepsilon_{nn}$ are independent with $\operatorname{Var}(\varepsilon_{ni})=\sigma^2(x_{ni})$ and

$$\max_j |x_{n,j+k}-x_{nj}|=0(n^{\Delta-1/2}), \qquad 0<\Delta<½.$$

If we choose $t_{n1}<\dots<t_{n\beta}$ such that for some $c>0$ and some λ, $0\leqslant\lambda<½-\Delta$

$$\min_j(t_{n,j+1}-t_{nj})\geqslant cn^{-\lambda},$$

then for n large enough, there will be no overlap between the $k=n^{\Delta+1/2}$ nearest x-neighbours to the points $t_{n1},\dots,t_{n\beta}$. Thus if we define

$$T_{ni}=\hat{m}(t_{ni})-E\hat{m}(t_{ni}), \qquad i=1,\dots,\beta$$

then there exists N such that $T_{n1},\dots,T_{n\beta}$ are independent for all $n\geqslant N$. By Chebychev's inequality and (2.5), for $n\geqslant N$, $a>0$,

$$P\left(\max_i |T_{ni}|\leqslant a\right)\geqslant\prod_{i=1}^{\beta}\left\{1-(ak)^{-2}\sum_{j\in I_{nk}(t_{ni})}\sigma^2(x_{nj})\right\}.$$

If we assume $\sigma^2(x)=\sigma^2$ then

$$\hat{m}(x)\pm\sigma k^{-1/2}\{1-(1-\alpha)^{1/\beta}\}^{-1/2} \tag{5.1}$$

is a simultaneous confidence band for $E\hat{m}(x)$ valid for all $x\in\{t_{n1},\dots,t_{n\beta}\}$. The width of this band is of the order $0(n^{-\{\Delta+(1/2)\}/2})$.

By (2.5) and the Central Limit Theorem,

$$\lim_{n\to\infty} P(k^{1/2}T_{nj}\leqslant t)=\Phi(t/\sigma),$$

where $k=n^{\Delta+1/2}$, and j is fixed. Thus if we set

$$M_\beta=\max\{|T_{n1}|,\ldots,|T_{n\beta}|\}$$

where β is finite, then

$$\lim_{n\to\infty} P(k^{1/2}M_\beta\leq t)=\{2\Phi(t/\sigma)-1\}^\beta \tag{5.2}$$

and

$$\hat{m}(x)\pm\Phi^{-1}[\{1+(1-\alpha)^{1/\beta}\}/2]\sigma/k^{1/2} \tag{5.3}$$

is asymptotically a level $(1-\alpha)$ simultaneous confidence band for $E\hat{m}(x)$ valid for $x\in\{t_{n1},\ldots,t_{n\beta}\}$.

This band is also of order $0(n^{-\{\Delta+(1/2)\}/2})$. Note that (5.1) and (5.3) are considerably narrower than the bands (3.2) and (4.1).

We next derive approximations to (5.2) and (5.3) valid for large β. Let $V^\beta=\max_{(i)}\{T_{ni}\}$, $W_\beta=\min_{(i)}\{T_{ni}\}$, then $M_\beta=\max\{V_\beta,\ -W_\beta\}$ and

$$P(k^{1/2}M_\beta\leq t)=P(k^{1/2}V_\beta\leq t,\ -k^{1/2}W_\beta\leq t).$$

Using this and results on the asymptotic distribution of extreme order statistics (e.g. Galambos, 1978, pp. 65 and 106), we find that if

$$a_\beta=-\Phi^{-1}(1/\beta),\qquad \text{or } a_\beta=(2\log\beta)^{1/2}-\frac{(\log\log\beta+\log 4\pi)}{2(2\log\beta)^{1/2}}$$

and $b_\beta=(2\log\beta)^{-1/2}$, then

$$\lim_{\beta\to\infty}\left\{\lim_{n\to\infty} P(k^{1/2}M_\beta/\sigma\leq a_\beta+b_\beta z)\right\}=\exp(-2e^{-z}). \tag{5.4}$$

It follows that for large β, an approximation to (5.3) is

$$\hat{m}(x)\pm\sigma(a_\beta+b_\beta z_\alpha)k^{-1/2} \tag{5.5}$$

where $z_\alpha=-\log\{-½\log(1-\alpha)\}$.

The band (5.1) can be made asymptotically valid for all x provided that the t_{nj}'s are chosen suitably, that the bias is of smaller order than the widths of the bands (see section 6), and that $m(x)$ is uniformly continuous.

It would have been more elegant to take the limit in (5.4) as β and n simultaneously tend to ∞, say by setting $\beta=n^\gamma$, $0<\gamma\leq½-\Delta$. With $\gamma=½-\Delta$, this would lead to a band similar to that of Révész (1979). Theorem 1 of Révész (1979) holds for $\hat{m}(x)$ based on the sets $I_{nk}(x)$, yielding the confidence band

$$\hat{m}(x)\pm\sigma(a_s+b_s z_\alpha)k^{-1/2} \tag{5.6}$$

with $s=n^{1/2-\Delta}$. The width of this band is of order $0\{n^{-(\Delta+1/2)/2}(\log n)^{1/2}\}$ when $k=n^{\Delta+1/2}$.

6. Bias and consistency

The results of the previous sections are for $\hat{m}(x)$ centred at

$$\bar{m}(x)=E\hat{m}(x)=\sum_{i\in I_{nk}(x)} m(x_i)/\mathrm{k}.$$

In order to be able to centre at $m(x)$, we need to show that the bias

$$\bar{m}(x)-m(x)=\sum_{i\in I_{nk}(x)}\{m(x_i)-m(x)\}/k$$

is uniformly of small order.

Assume now that x is in an interval (a_n, b_n) with a_n and b_n finite and that the regression function satisfies the Lipschitz condition:

$$|m(x)-m(y)|\leq c|x-y|,\qquad x, y\in(a_n, b_n),\qquad \text{some } c>0. \tag{6.1}$$

Let $x\in J_i$, $i=0, 1, \dots, n-k$, with J_i as defined in section 2 and let $x_0=a_n$, $x_{n+1}=b_n$. Then

$$|\bar{m}(x)-m(x)|\leq c\sum_{j\in I_{nk}(x)}|x_j-x|/k$$

$$\leq c\max\{(x_{i+k}-x_i)/2,\quad (x_{i+k+1}-x_{i+1})/2\},\qquad x\in J_i$$

since $|x_j-x|\leq|x_j-x_{0j}|$, where x_{0j} is the point on the boundary of J_i furthest from x_j. Combining this with (3.1), we obtain the following uniform consistency result.

Theorem

Suppose that the assumptions of the model (2.1) hold, and that $m(\cdot)$ satisfies the Lipschitz condition (6.1). Then $\bar{m}(\cdot)$ is uniformly consistent in the sense that

$$P\left\{\sup_{a_n<x<b_n}|\bar{m}(x)-m(x)|\leq\varepsilon\right\}\to 1$$

provided

$$\sum_{i=1}^{n}\sigma^2(x_i)=o(k^2) \tag{6.2}$$

and

$$\max_{0\leq j\leq n-k+1}|x_{n,j+k}-x_{nj}|=o(1). \tag{6.3}$$

In the case where we can express $x_{n,j}$ as $\varphi_n\{i/(n+1)\}$ for some increasing function φ_n on (0, 1) satisfying the Lipschitz condition (6.1) uniformly in n, (6.3) is satisfied whenever $(k/n)\to 0$ as $n\to\infty$. Thus, typically, (6.2) holds when $(k^2/n)\to\infty$, while (6.3) holds when $(k/n)\to 0$. For $k=n^{1/2+\Delta}$, this means $0<\Delta<½$. Note that a_n and b_n need not be bounded. For instance, $x_{nj}=\Phi^{-1}\{j/(n+1)\}$, where Φ^{-1} is the inverse of the standard normal distribution function, satisfies (6.3).

Révész (1979) considers a different restriction on the regressor: suppose that X is random and that X is in the interval [0, 1]. Let X have a density f such that $f(x)\geq\varepsilon$, $x\in[0, 1]$, for some $\varepsilon>0$. Let k be such that

$$kn^{-2/3}\log n\to 0,\; k^{-1}(\log n)^3\to 0 \text{ as } n\to\infty. \tag{6.4}$$

Then (Révész, 1979, lemma 1)

$$\limsup_{n\to\infty}\left\{\frac{n}{k}\sup_{1\leq i\leq n-k}|X_i-X_{i+k}|\right\}\leq 2/\varepsilon\qquad\text{a.s.}$$

where $X_1, \dots, X_n$ are the order statistics of a sample of size n from a population with density f. From this and the Lipschitz condition (6.1), it follows that

$$\sup_x|\bar{m}(x)-m(x)|=0(k/n)\qquad\text{a.s.}$$

Thus we have shown consistency again under a different set of conditions than those of the theorem.

Under the conditions of the above paragraph, and assuming that $m(x)$ has a uniformly bounded derivative, Révész (1979) has shown (lemma 2) that

$$\sup_x |\bar{m}(x)-m(x)|=o\{(k \log n)^{-1/2}\} \qquad \text{a.s.}$$

Stone (1977) obtained general consistency results for nearest neighbour estimates in terms of L^r convergence, $r>1$.

Spiegelman & Sacks (1980) considered window estimates and showed the mean squared error to be of the order $0(n^{-2/3})$ when m satisfies the Lipschitz condition. Thus their bias is of order less than $n^{-1/3}$. Their bandwidth is of order $b_n=n^{-1/3}$, corresponding to $k=n^{2/3}$.

Lai (1977) and Mack (1981) considered the asymptomatic bias of nearest neighbour estimates and found the bias to be of order $(k/n)^2$ when f and m satisfy certain conditions including the condition that they are continuously differentiable up to second order.

7. Asymptotic confidence bands for m(x)

We want to extend the bands for $E\hat{m}(x)$ to $m(x)$. Note that if we write $b_{\text{MAX}}=\sup_{(x)}|\bar{m}(x)-m(x)|$ for the maximum bias, then

$$P\left(\sup_x |\hat{m}(x)-m(x)|\leq t\right)\geq P\left(\sup_x|\hat{m}(x)-\bar{m}(x)|\leq t-b_{\text{MAX}}\right).$$

Thus, from (3.1), when $\sigma^2(x)=\sigma^2$, a level $(1-\alpha)$ uniform confidence for band $m(x)$ is $\hat{m}(x)\pm 2^{1/2}t'_\alpha$, where

$$t'_\alpha=b_{\text{MAX}}+(2\sigma/k)\ \{(2n-k)/\alpha\}^{1/2}.$$

Note that when $b_{\text{MAX}}=o(n^{1/2}/k)$, then as $n\to\infty$, $t'_\alpha/t_\alpha\to 1$ and (3.2) is asymptotically valid as a confidence band for $m(x)$. Since typically $b_{\text{MAX}}=o(k/n)$, we require that k/n is of smaller order than $n^{1/2}/k$. When $k=n^{1/2+\Delta}$, this means $0<\Delta<\frac{1}{4}$.

Similarly, the band (4.1) is also asymptotically valid when $b_{\text{MAX}}=o(n^{1/2}/k)$.

Using the inequality

$$\max_j|\bar{m}(t_{nj})-m(t_{nj})|\leq b_{\text{MAX}} \tag{7.1}$$

we find that the bands (5.1) and (5.3) are asymptotically valid when $b_{\text{MAX}}=o(k^{-1/2})$. When $k=n^{1/2+\Delta}$ and $b_{\text{MAX}}=0(k/n)$, this corresponds to $0<\Delta<\frac{1}{6}$.

The above conditions for the bands (3.2), (4.1), (5.1) and (5.3) are equivalent to the condition that the bias is of smaller order than the widths of the bands. This condition does not work for the bands (5.5) and (5.6). By considering (5.4) and using the inequality (7.1), we see that a sufficient condition for the asymptotic validity of these bands as bands for $m(x)$ is

$$b_{\text{MAX}}=o\{(k \log \beta)^{-1/2}\}.$$

When $b_{\text{MAX}}=0(k/n)$, $\beta=n^\gamma$, $0<\gamma\leq\frac{1}{2}-\Delta$, and $k=n^{1/2+\Delta}$, this corresponds to $0<\Delta<\frac{1}{6}$.

Rosenblatt (1969) gives asymptotic results for kernel estimators of the regression function. He gets pointwise confidence-intervals of width $n^{-2/5}$ using a bandwidth of order $n^{-1/5}$ which corresponds to $k=n^{4/5}$ or $\Delta=0.3$.

8. Choice of neighbourhood size

The choice of neighbourhood size, k, involves a delicate trade-off between bias and variance. Specifically, with $k=n^{\Delta+1/2}$, as Δ approaches 0, the bias decreases and the variance increases. Alternatively, as Δ approaches ½, the bias increases and may become of larger order than the width of the bands. As more smoothing is imposed on the regression function $m(\cdot)$, bias becomes less important. However, this introduces restrictions on the shape of m which may not be desirable. An *ad hoc* procedure would be to choose an intermediate value of Δ, say 0.15, and see (a) how large k is relative to n for one's sample size, (b) what is the empirical trade-off between bias and variance, and (c) how smooth (visually) is the resulting non-parametric regression estimate. One hopes that methods, perhaps along the lines of penalized likelihood and cross-validation, will be developed to provide empirical guidelines for the choice of k.

9. An illustration

To get an idea of the accuracy of the bands, we computed the band (5.6) for data $(x_1, Y_1), \ldots, (x_{100}, Y_{100})$ generated from the model

$$Y_i=\exp(\gamma_1)\exp(\gamma_2 x_i)\,x_i^{\gamma_3}+\varepsilon_i \tag{9.1}$$

where $\gamma_1=5$, $\gamma_2=-½$, $\gamma_3=1$, and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $\mathcal{N}(0, 100)$, $x_i=i/25$, $i=1, \ldots, 100$. The significance coefficient is 0.90. $\Delta=0.15$, yielding a neighbourhood size of $k=20$. While this k is large relative to $n=100$, it seems to give a smooth but accurate estimate of the true curve. $\hat{\sigma}$ is computed from (3.3) ignoring data at the edge, that is, the first and last 10 points.

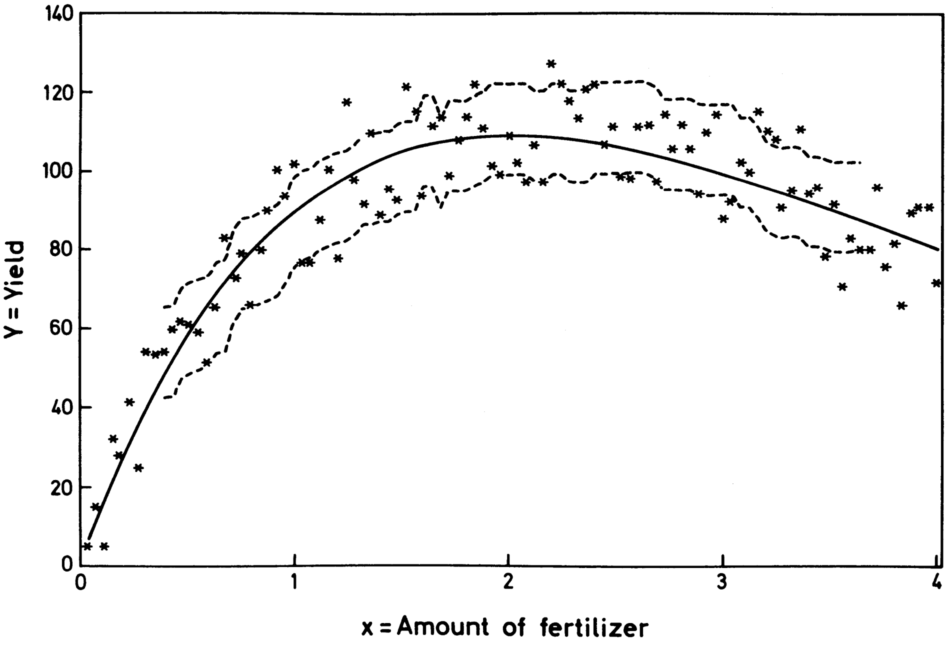

Fig. 1. Simultaneous confidence band. Solid line=true curve; dashed lines=asymptotic (Révész) band (5.6); stars=simulated data. Sample size 100. Band clipped to region of inference. Simultaneity for the five x-values, $x=0.4$, 1.2, 2, 2.8, 3.6.

Table 2. *Widths of the confidence bands for model (7.1)*

Band	(3.2)	(4.1)	(5.1)	(5.3)	(5.6)
Width	89.6	39.3	31.0	10.3	11.6

This model has been suggested for agricultural experiments where an amount x of fertilizer increases yield Y for low and moderate doses while it decreases yield at high doses.

The result is shown in Fig. 1 where the middle curve is the true function $m(x)=\exp(\gamma_1)\exp(\gamma_2 x)\, x^{\gamma_3}$ and the upper and lower curves define the band. The band is fairly accurate with the width being 11.6. Since the band is simultaneous, we can test model assumptions. For instance, a parabola does not fit in the band and a quadratic (in x) regression model is rejected.

We also computed the widths of the other bands (using $\sigma=10$ rather than $\hat{\sigma}$). The results are given in Table 2 using $k=20$ and $\beta=s=5$. Note that the widths of (5.5) and (5.6) are the same.

Note that since ε_i is normal, (5.3) is exactly a level 0.90 simultaneous confidence procedure for $\beta=5$.

Note also that the simultaneity is restricted to five x-values, and that the band is asymptotic as a band for $m(x)$.

Acknowledgement

We wish to thank Professor J. MacQueen for generously spending his time discussing the topic of this paper. We would also like to thank the reviewers for many helpful comments and suggestions.

Bjerve's work was done while the author was visiting the University of California at Berkeley on sabbatical leave from the University of Oslo.

Doksum's work was partially supported by National Science Foundation Grant MCS-81-02349.

References

Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *Ann. Statist.* **2,** 1034–1039.

Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

Collomb, G. (1981). Estimation non-parametrique de la regression: revue bibliographique. *Int. Statist. Rev.* **49,** 75–93.

Doksum, K. A. & Yandell, B. S. (1982). Properties of regression estimates on censored survival data. In *A Festschrift for Erich L. Lehmann* (ed. P. J. Bickel, K. A. Doksum & J. L. Hodges, Jr), pp. 140–156. Wadsworth Statistics Probability Series.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics.* Wiley, New York.

Lai, S. L. (1977). Large sample properties of k-nearest neighbour procedures. Ph.D. Dissertation, University of California, Los Angeles.

Liero, H. (1982). On the maximal deviation of the kernel regression function estimate. *Math. Operationsforsch. Statist., Ser. Statist.* **13,** 171–182.

Loève, M. (1963). *Probability theory,* 3rd edn. Academic Press, New York.

Mack, Y. P. (1981). Local properties of $k-NN$ regression estimates. *Siam J. Alg. Disc. Meth.* **2,** 311–323.

Major, P. (1973). On non-parametric estimation of the regression function. *Studia Sci. Math. Hung.* **8,** 347–361.

Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statist.* **8,** 119–126.

Révész, P. (1979). On the nonparametric estimation of the regression function. *Problems of Control and Inform Theor.* **8,** 297–302.

Rosenblatt, M. (1969). Conditional probability density and regression estimates. *International Symposium on Multivariate Analysis II, 25–31.*

Spiegelman, C. & Sacks, J. (1980). Consistent window estimation in non-parametric regression. *Ann. Statist.* **8,** 240–246.

Stone, C. J. (1977). Consistent non-parametric regression. *Ann. Statist.* **5,** 595–645.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validation smoothing spline. *J. R. Statist. Soc. B* **45,** 133–150.

Received March 1983, in final form January 1985

Steinar Bjerve, Matematisk Institutt, PO Box 1053, Blindern, Oslo 3, Norway