*The Annals of Statistics*
1989, Vol. 17, No. 3, 1001–1008

# HONEST CONFIDENCE REGIONS FOR NONPARAMETRIC REGRESSION[1]

BY KER-CHAU LI

*University of California, Los Angeles*

The problem of constructing honest confidence regions for nonparametric regression is considered. A lower rate of convergence, $n^{-1/4}$, for the size of the confidence region is established. The achievability of this rate is demonstrated using Stein's estimates and the associated unbiased risk estimates. Practical implications are discussed.

**1. Introduction.** In spite of the growing interest in the research and the application of nonparametric regression, the problem of setting a confidence region is not often addressed. The mathematical complexity may be one reason that discourages the development. Beyond that, however, there lies the more fundamental question of whether it is possible to have any useful honest confidence region or not. Here the word "honest" refers to the requirement that the minimum coverage probability over a rich class of (nonparametric) regression functions should be no less than the nominal confidence level.

Specifically, suppose we have $n$ independent observations, $y_1, \dots, y_n$, taken at sites $x_1, \dots, x_n$, in some compact region $C$ in $R^p$. Assume that

$$y_i = f(x_i) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $f$ is the unknown smooth function and $\varepsilon_i$ are normal with mean 0 and variance $\sigma^2$. There are many ways to quantify the degree of smoothness. For instance, when $p = 1$, it can be measured by the $L_2$ (or $L_\infty$) norm of the second (or the $k$th) derivative of $f$. But it is not easy to appropriately estimate this quantity or to give it a reasonable upper bound. One approach may be to estimate the derivative by, say, some kernel method. But the validity of this estimation in turn depends on the knowledge of some higher order derivatives. On the other hand, particularly for the purpose of choosing the smoothing parameter, we may use cross validation or other sample reuse techniques to bypass this difficulty without estimating any smoothness measure. In any case throughout this paper *we shall not assume any specific bound for the smoothness measure a priori* (so that we have an "honest" nonparametric setting).

To fix the idea, take $p = 1$, $C = [0, 1]$ and let the space of $f$ be $F = \{f\colon f$ is twice continuously differentiable in $C\}$. An ambitious goal may be to construct a confidence region for the entire function $f$ on $C$. But this requires interpolation or extrapolation and is easily seen to be impossible without a specified upper bound on a smoothness measure. Hence we shall only consider the problem of

Received July 1986; revised January 1988.

[1]Research supported by NSF Grants MCS-83-01587 and DMS-86-02018.

*AMS* 1980 *subject classifications*. 62G99, 62J99.

*Key words and phrases*. Adaptiveness, confidence region, convergence rate, nonparametric regression, Stein estimates, Stein's unbiased risk estimates.

setting confidence regions for the vector $f_n = (f(x_1), \dots, f(x_n))'$. We also assume that $\sigma^2$ is given and that all $x_i$ are distinct.

With these assumptions, it is clear that the space of $f_n$ is equal to the entire $R^n$. An obvious confidence region for $f_n$ is the $n$-dimensional ball with the center $\mathbf{y}_n = (y_1, \dots, y_n)'$ and the squared radius equal to $\sigma^2$ times the $(1-\alpha)$-quantile of the $\chi^2$ distribution with $n$ degrees of freedom. But this confidence region is almost useless for at least two obvious reasons: (i) the choice of the center implies that no smoothing is needed in estimating the smooth function and (ii) the radius is too large to extract any interesting features about $f$.

It seems easier to discuss (ii) in terms of the large sample theory. Let $\|\cdot\|$ be the Euclidean norm and $\|\cdot\|_n$ be $n^{-1/2}\|\cdot\|$. Consider any confidence region of the form

$$\{f_n: \|\hat{f}_n - f_n\|_n \le s_n\}, \tag{1.2}$$

where the center $\hat{f}_n$ and the (normalized) radius $s_n$ depend only on $\mathbf{y}_n$. For a large sample, we naturally expect to have a small radius $s_n$. Thus a basic requirement on the size of any useful confidence region is that, for a smooth function $f$, $s_n$ should converge to zero in probability as the sample size tends to the infinity and the sites $x_i$ get dense in $C$. However, the confidence region constructed before does not satisfy this requirement; $s_n$ always converges to $\sigma$ for any $f$. So the question of interest is whether or not there exists any honest confidence procedure satisfying this requirement. If yes, how fast can the convergence rate be?

In Section 2, we shall show that the best convergence rate for $s_n$ cannot be faster than $n^{-1/4}$. This result depends neither on the dimension $p$ nor on any smoothness measure of $f$. The only requirement is that the confidence region be asymptotically honest in the sense that

$$\lim_{n\to\infty} \inf_{f\in F} P\{\|\hat{f}_n - f_n\|_n \le s_n\} \ge 1 - \alpha, \tag{1.3}$$

where $F$ can be any class of functions such that the set $\{f_n: f \in F\}$ equals the entire $R^n$. The crucial point here is the ordering of lim and inf, which admits an honest $(1 - \alpha - \varepsilon)$ confidence region when the sample size is large. In Section 3, we shall demonstrate the achievability of the rate $n^{-1/4}$, using Stein's estimate and the associated unbiased risk estimate.

In view of the slowness of this convergence rate, one may doubt the wisdom of being fully honest in constructing a confidence region. This touches the philosophical issue about the practical application of confidence procedures. Even in the simple case of estimating the population mean, the usual confidence interval, sample mean $\pm z_{1-\alpha} \cdot$ sample standard deviation, is not asymptotically honest. In fact, Bahadur and Savage (1956) showed that if the class of error distributions is rich enough (say it contains all distributions with finite second moments), then the inf of the coverage probabilities is always zero, no matter how large the sample size is [see also Donoho (1984) for related results].

Before closing this section, we shall briefly review some related work. Knafl, Sacks and Ylvisaker (1982) introduced a method for finding model-robust confidence sets. The idea is to specify an upper bound on some smoothness measure of

$f$. Based on the given bound, a minimax linear estimate for $f(x_i)$ is used as the center for constructing confidence interval for each $f(x)$. Finally, they multiply the width for each interval by a suitable constant to ensure the probability of coverage for the entire function. This was carried out explicitly in the case that $p = 1$ and $f$ is twice differentiable with the supremum norm of the second derivative being bounded by a constant. Wahba (1983) constructed a pseudo Bayes confidence band for $f_n$, without specifying any bound for the smoothness measure. A novel notion about the coverage probability was introduced there. Instead of computing the probability that $f_n$ falls in the confidence band, she was more concerned about what percentage of intervals fail to capture their own target values $f(x_i)$. Eubank (1985) considered the jackknife method for constructing confidence intervals. During the revision of this paper, it was communicated to me by the Associate Editor that Hall and Titterington (1986) proposed another method for constructing confidence bands in nonparametric density and regression problems. Their method also depends on the bound of some smoothness measure of the unknown function to be estimated. A faster rate (depending on the smoothness condition) of convergence is obtained as expected. An implication of our result (see Remark 2.1), however, provides a warning to the unrestrained use of such procedures, since the minimum coverage probability is zero provided that the bound of smoothness measure is unknown and is estimated from the data.

**2. Lower rate of convergence.** It will be easier to present our results in the general framework of constructing confidence regions for a multivariate normal mean. Thus assume that $y_1, \dots, y_n$ are independent normal random variables with means $\mu_1, \dots, \mu_n$ and a common variance $\sigma^2$. Consider the confidence region for $\boldsymbol{\mu}_n = (\mu_1, \dots, \mu_n)'$ of the form

$$\left\{\boldsymbol{\mu}_n\colon \|\hat{\boldsymbol{\mu}}_n(\mathbf{y}_n) - \boldsymbol{\mu}_n\|_n \le s_n(\mathbf{y}_n)\right\},$$

where $\hat{\boldsymbol{\mu}}_n(\mathbf{y}_n)$ is a point estimate of $\boldsymbol{\mu}_n$ and $s_n(\cdot)$ is a measurable real function on $R^n$. When such a confidence region is honest asymptotically, we have

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\mu}_n \in R^n} P\left\{\|\hat{\boldsymbol{\mu}}_n(\mathbf{y}_n) - \boldsymbol{\mu}_n\|_n \le s_n(\mathbf{y}_n)\right\} \ge 1 - \alpha. \tag{2.1}$$

The following main theorem for this section provides the lower rate of convergence for the size of the confidence region as measured by $s_n$. The subscript of $P$ indicates the true mean of $\mathbf{y}_n$.

THEOREM 2.1. *Assume that* (2.1) *holds. Then for any sequence of* $\boldsymbol{\mu}_n^*$ *and for any sequence of positive numbers* $c_n$ *that converges to* 0, *we have*

$$\limsup_{n\to\infty} P_{\boldsymbol{\mu}_n^*}\left\{s_n(\mathbf{y}_n) \le c_n n^{-1/4}\right\} \le \alpha. \tag{2.2}$$

An immediate consequence of this theorem is that the size of any honest confidence region cannot converge to zero faster than the rate of $n^{-1/4}$, no matter how smooth the underlying function $f$ is. To see this, we simply take $\boldsymbol{\mu}_n^*$ as $(f(t_1), \dots, f(t_n))'$.

REMARK 2.1. It is now well-known that the optimum rate of convergence for estimating a nonparametric function can be made as close to the rate of root $n$ (the rate for parametric problems) as possible by simply assuming higher and higher order of differentiability about the unknown function $f$. See Stone (1982) and Speckman (1985) for the exact relationship between the smoothness assumption and the optimal rate of convergence. It seems therefore feasible to try some adaptive procedures to construct a confidence region with the optimal convergence rate. For instance, if we assume $f$ has a second derivative, we may try to construct a confidence region based on a cubic smoothing spline estimator and the corresponding mean squared error. Using Chebyshev's inequality, it is not hard to find a confidence region provided that an upper bound of the square root of the mean squared error is known. Such a bound is available [see Wahba (1978), for instance], which has rate of convergence $n^{-2/5}$, faster than $n^{-1/4}$. But this bound depends on the $L_2$ norm of the second derivative of $f$ and hence should be estimated. Fortunately a consistent estimate of the second derivative can be easily obtained and hence can be adapted to the construction of a confidence region. The resulting "confidence region" now enjoys a convergence rate $n^{-2/5}$ faster than $n^{-1/4}$. It satisfies (1.3) but with the "limit" and "inf" being interchanged. Now, applying Theorem 2.1, we see that the left side of (1.3) should be equal to 0. In other words, although these kinds of adaptive procedures may yield a smaller confidence region, the minimum probability of coverage can be very small as well. On the other hand, if one decides to ignore the asymptotic honesty, then the results of Nussbaum (1985), which give a sharp bound for the mean squared error, may help improve this adaptive procedure.

The rest of this section will be devoted to the proof of this theorem. Basic ideas come from testing the hypothesis and the standard prior-posterior argument. We first give an outline of the proof.

1. Assume that the conclusion (2.2) is false. Thus for some $\varepsilon > 0$, there exists some subsequence such that for large $n$,

$$P_{\mu_n^*}\{s_n(\mathbf{y}_n) \le c_n n^{-1/4}\} \ge \alpha + 2\varepsilon. \tag{2.3}$$

2. Put a suitable normal prior $\pi_n$ on $\mu_n$, so that from (2.3) it can be inferred that for large $n$,

$$P_{\pi_n}\{s_n(\mathbf{y}_n) \le c_n n^{-1/4}\} \ge \alpha + \varepsilon. \tag{2.4}$$

3. From (2.1) we get

$$P_{\pi_n}\{\|\hat{\mu}_n(\mathbf{y}_n) - \mu_n\|_n \le s_n(\mathbf{y}_n)\} \ge 1 - \alpha - \frac{\varepsilon}{2}.$$

   This together with (2.4) implies

$$P_{\pi_n}\{\|\hat{\mu}_n(\mathbf{y}_n) - \mu_n\|_n \le c_n n^{-1/4}\} \ge \frac{\varepsilon}{2}. \tag{2.5}$$

4. Show that even for the best choice of $\hat{\mu}_n(\mathbf{y}_n)$, (2.5) cannot hold. This leads to a contradiction and hence completes the proof.

The major steps are 2 and 4, which will be described in detail below.

2.1. *Step* 2. Consider a normal prior for $\mu_n$ with mean $\mu_n^*$ and covariance $b_n I_n$, where $I_n$ is the identity matrix and $b_n \geq 0$. Suppose we want to test $H_{0n}$: $b_n = 0$ against $H_{1n}$: $b_n = a_n n^{-1/2}\sigma^2$, where $\{a_n\}$ is a sequence of fixed positive numbers converging to 0.

Clearly under $H_{1n}$, $y_i$'s are independent normal random variables with means $\mu_i^*$ and a common variance $\sigma^2(1 + a_n^2 n^{-1/2})$. Hence the best test based on the Neyman–Pearson lemma is to reject $H_{0n}$ when $T_n = \sigma^{-2}\sum_{i=1}^n (y_i - \mu_i^*)^2$ is large. Under $H_{0n}$, $T_n$ is a $\chi^2$ with $n$ degrees of freedom, implying that

$$(T_n - n)/\sqrt{2n} \to N(0,1). \tag{2.6}$$

Likewise, under $H_{1n}$,

$$\left(T_n\left(1 + a_n n^{-1/2}\right)^{-1} - n\right)/\sqrt{2n} \to N(0,1),$$

which also implies (2.6) because of Slutsky's theorem. Thus $H_{0n}$ and $H_{1n}$ are asymptotically indistinguishable. Using the Neyman–Pearson lemma, we see that for any event $A$, in particular $A\{s_n(\mathbf{y}_n) \leq c_n n^{-1/4}\}$,

$$|P\{A|H_{0n}\} - P\{A|H_{1n}\}| \leq \sup_{t \in R} |P\{T_n \leq t|H_{0n}\} - P\{T_n \leq t|H_{1n}\}|. \tag{2.7}$$

With (2.6) satisfied under both $H_{0n}$ and $H_{1n}$, we see that the right-hand side of (2.7) converges to 0 due to a theorem of Polya [see Bickel and Doksum (1977), page 462]. Now (2.4) follows immediately from (2.3) when we take our prior $\pi_n$ according to $H_{1n}$.

2.2. *Step* 4. We shall use the standard prior-posterior argument. First, the posterior distribution of $\mu_n$, given $\mathbf{y}_n$, is easily seen to be normal with mean $\bar{\mu}_n(\mathbf{y}_n) = (\mu_n^* + a_n n^{-1/2}\mathbf{y}_n)/(1 + a_n n^{-1/2})$ and covariance $a_n n^{-1/2}\sigma^2/(1 + a_n n^{-1/2})I_n$. Clearly, $\bar{\mu}_n(\mathbf{y}_n)$ is the Bayes estimate of $\mu_n^*$ under squared error loss. Now, the left side in (2.5) equals

$$\begin{aligned} &E_{\pi_n} P\left\{\|\hat{\mu}(\mathbf{y}_n) - \mu_n\|_n \leq c_n n^{-1/4}|\mathbf{y}_n\right\} \\ &\quad \leq E_{\pi_n} P\left\{\|\bar{\mu}(\mathbf{y}_n) - \mu_n\|_n \leq c_n n^{-1/4}|\mathbf{y}_n\right\} \\ &\quad = P\left\{\frac{1}{n}\sum_{i=1}^n e_i^2 \leq c_n^2 a_n^{-1}\sigma^{-2}\left(1 + a_n n^{-1/2}\right)\right\}, \end{aligned} \tag{2.8}$$

where $e_i$'s denote i.i.d. standard normal random variables. The above inequality is based on the spherical symmetry and the unimodality properties of the normal distribution. Finally, choose $a_n$ suitably so that $a_n \to 0$ and $c_n^2 a_n^{-1} \to 0$. Then it is clear that the last expression in (2.7) should converge to 0, leading to a contradiction with (2.5) as desired.

**3. Stein estimates.** We shall demonstrate that using the Stein estimate and Stein's unbiased estimate (SURE), it is easy to construct a confidence region that achieves the lower rate of convergence.

For any nonidentity $n \times n$ matrix $M_n$, Li (1985) considered the following simplified version of the Stein estimate and SURE:

$$\text{(3.1)} \qquad \begin{aligned} \tilde{\mu}_n &= \mathbf{y}_n - \frac{\sigma^2 \operatorname{tr} A_n}{\|A_n \mathbf{y}_n\|^2} A_n \mathbf{y}_n, \\ \mathrm{SURE}_n &= \sigma^2 - \frac{\sigma^4 (\operatorname{tr} A_n)^2}{n \|A_n \mathbf{y}_n\|^2}, \end{aligned}$$

where $A_n = I - M_n$. It was shown that $\mathrm{SURE}_n$ is a uniformly consistent estimate of the true loss $n^{-1}\|\tilde{\mu}_n - \boldsymbol{\mu}_n\|^2$, regardless of what sequence $M_n$ is embedded in, although it may sometimes fail to estimate what it is supposed to estimate (namely, the expected value of the true loss). This result can be strengthened as follows. The proof will be given in the Appendix.

THEOREM 3.1. *For any $\alpha$, $0 < \alpha < 1$, there exists some constant $c(\alpha) > 0$, such that*

$$\lim_{n \to \infty} \sup_{\boldsymbol{\mu}_n \in R^n} P\left\{ \left| \mathrm{SURE}_n - n^{-1}\|\tilde{\mu}_n - \boldsymbol{\mu}_n\|^2 \right| \geq c(\alpha) n^{-1/2} \right\} \leq \alpha,$$

*for any sequence of $M_n$.*

REMARK 3.1. The normality assumption on the error distribution can be replaced by the conditions (A.1) and (A.2) in Li (1985), page 1358.

To apply this theorem in nonparametric regression, simply take $\boldsymbol{\mu}_n$ as $f_n$ and take $M_n$ as the matrix which will be used to form a good linear estimate, $\hat{\boldsymbol{\mu}}_n = M_n \mathbf{y}_n$, of $\mu_n$. Then use $\tilde{\mu}_n$ as the center $\hat{f}_n$ and $c(\alpha) n^{-1/2} + \mathrm{SURE}_n$ as $s_n^2$, the square of the normalized radius. We see that Theorem 3.1 implies the asymptotic honesty (1.3). It remains to show that $\mathrm{SURE}_n$ converges to 0 at rate no slower than $n^{-1/2}$. By Theorem 3.1 again, it suffices to find $M_n$ so that the rate of convergence for $n^{-1}\|\tilde{\mu}_n - \boldsymbol{\mu}_n\|^2$ is no slower than $n^{-1/2}$. On the other hand, the result of Li and Hwang (1984) shows that in most cases, Stein estimates and the associated linear estimates have the same rate of convergence. This means that we can use any linear estimate to begin with, provided that it converges to $f$ at rate no slower than $n^{-1/4}$. Such estimates are very easy to find, e.g., kernel estimates, nearest neighbor estimates, smoothing splines, etc., with the smoothing parameters appropriately chosen. Let us summarize this result by the following theorem. See Li and Hwang (1984) for examples and for more discussion of the conditions on $M_n$.

THEOREM 3.2. *For any sequence of linear estimators $M_n \mathbf{y}_n$ of $f_n$, the confidence region* (1.2) *with $\hat{f}_n = \tilde{\mu}_n$ and $s_n^2 = c(\alpha) n^{-1/2} + \mathrm{SURE}_n$ is asymptotically honest in the sense that* (1.3) *is satisfied for any class $F$. In addition, if*

$$\operatorname{tr} M_n M_n' \to \infty, \qquad \left(n^{-1} \operatorname{tr} M_n\right)^2 / n^{-1} \operatorname{tr} M_n M_n' \to 0,$$
$$\lambda(M_n M_n') / \operatorname{tr} M_n M_n' \to 0,$$

*where* $\lambda(\cdot)$ *denotes the maximal eigenvalue and* $n^{-1}\|\hat{\mu}_n - f_n\|^2 = O_p(n^{-1/2})$, *then the size of the confidence region* $s_n$ *converges to* 0 *in probability at rate* $n^{-1/4}$, *i.e.,* $s_n = O_p(n^{-1/4})$.

## APPENDIX

PROOF OF THEOREM 3.1. We shall assume $\lambda(A'A) = 1$, in view of the scale invariance property for $\tilde{\mu}_n$ and $\text{SURE}_n$. By the same argument as that given in the beginning paragraph of Section 7.1 of Li (1985), we need only to find some large $c(\alpha)$ such that

$$\limsup_{n\to\infty} P\Big\{\big|\sigma^2 - n^{-1}\|\varepsilon_n\|^2\big| \ge \tfrac{1}{3}c(\alpha)n^{-1/2}\Big\} \le \alpha/3, \tag{3.2}$$

$$\limsup_{n\to\infty} P\big\{n^{-1}|\operatorname{tr} A_n| \cdot |\langle \varepsilon_n, A_n\mu_n\rangle| \cdot \|A_n\mathbf{y}_n\|^{-2} \ge \tfrac{1}{6}c(\alpha)n^{-1/2}\big\} \le \alpha/3, \tag{3.3}$$

and

$$\limsup_{n\to\infty} P\Big\{n^{-1}|\operatorname{tr} A_n| \cdot \big|\langle \varepsilon_n, A_n\varepsilon_n\rangle - \sigma^2 \operatorname{tr} A_n\big| \cdot \|A_n\mathbf{y}_n\|^{-2} \ge \tfrac{1}{6}c(\alpha)n^{-1/2}\Big\} \le \alpha/3. \tag{3.4}$$

The central limit theorem guarantees (3.2). So we shall verify (3.3) only; (3.4) follows by a similar argument.

Consider an arbitrary sequence $\{\mu_n\}$, $\mu_n \in R^n$. Let $Q_n = E\|A_n\mathbf{y}_n\|^2 = \|A_n\mu_n\|^2 + \sigma^2 \operatorname{tr} A_nA_n'$. It suffices to consider two cases: (i) $\lim_{n\to\infty} Q_n < \infty$ and (ii) $\lim_{n\to\infty} Q_n = \infty$. For the first case (3.3) can be derived from

$$\limsup_{n\to\infty} P\big\{n^{-1}|\operatorname{tr} A_n| \cdot |\langle \varepsilon_n, A_n\mu_n\rangle| \ge \tfrac{1}{6}c(\alpha)n^{-1/2}\delta(\alpha)\big\} \le \alpha/6, \tag{3.5}$$

where $\delta(\alpha)$ is a positive number such that $P\{\chi_1^2 \ge \delta(\alpha)\} \ge 1 - \alpha/6$; $\chi_1^2$ denotes the $\chi^2$ random variable with one degree of freedom. To see this, simply observe that due to our assumption that the maximum eigenvalue of $A_n'A_n$ is 1, we have $\|A_n\mathbf{y}\|^2 \ge (\sum_{i=1}^n a_iy_i)^2$ for some $(a_1,\dots,a_n)'$ such that $\sum_{i=1}^n a_i^2 = 1$.

Now by Chebyshev's inequality and the condition (i) [in a way similar to the proof of (7.1.5) in Li (1985)], it is easy to check that (3.5) holds for some large $c(\alpha)$.

Next we consider the case (ii). By Chebyshev's inequality again, we can verify that

$$\limsup_{n\to\infty} P\big\{\|A_n\mathbf{y}_n\|^2 \le \tfrac{1}{2}Q_n\big\} \le \alpha/6. \tag{3.6}$$

[The inequality $\operatorname{Var}\|A_n\mathbf{y}_n\|^2 \le c(\sigma^2 \operatorname{tr} A_nA_n' + \|A_n\mathbf{y}_n\|^2)$ for some $c$ may be used, which can be derived from Theorem 2 of Whittle (1960).] Thus to obtain (3.3), we need only to prove

$$\limsup_{n\to\infty} P\big\{n^{-1}|\operatorname{tr} A_n| \cdot |\langle \varepsilon_n, A_n\mu_n\rangle| \ge \tfrac{1}{12}Q_nc(\alpha)n^{-1/2}\big\} \le \alpha/6,$$

which in turn follows easily from Chebyshev's inequality again. The proof is now complete. □

## REFERENCES

BAHADUR, R. R. and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27** 1115–1122.

BICKEL, P. and DOKSUM, K. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, San Francisco.

DONOHO, D. (1984). Nonparametric inference about functionals of a density. I. Unpublished.

EUBANK, R. L. (1985). Discussion of "Some aspects of the spline smoothing approach to nonparametric regression curve fitting," by B. W. Silverman. *J. Roy. Statist. Soc. Ser. B* **47** 35–36.

HALL, P. and TITTERINGTON, D. M. (1986). On confidence bands in nonparametric density estimation and regression. Unpublished.

KNAFL, G., SACKS, J. and YLVISAKER, D. (1982). Model robust confidence intervals. II. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **2** 87–102. Academic, New York.

LI, K.-C. (1985). From Stein's unbiased risk estimates to the mèthod of generalized cross-validation. *Ann. Statist.* **13** 1352–1377.

LI, K.-C. and HWANG, J. T. (1984). The data smoothing aspect of Stein estimates. *Ann. Statist.* **12** 887–897.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in $L_2$. *Ann. Statist.* **13** 984–997.

SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.

STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

WAHBA, G. (1983). Bayesian confidence intervals for the cross-validated smoothing splines. *J. Roy. Statist. Soc. Ser. B* **45** 133–150.

WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.* **5** 302–305.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024