SIMULTANEOUS CONFIDENCE BANDS FOR LINEAR REGRESSION AND SMOOTHING

Abbreviated Title: SIMULTANEOUS CONFIDENCE BANDS

BY JIAYANG SUN[1] AND CLIVE R. LOADER
The University of Michigan and AT&T Bell Laboratories
December 22, 1992

Abstract

Suppose we observe $Y_i = f(x_i) + \epsilon_i, i = 1, \ldots, n$. We wish to find approximate $1-\alpha$ simultaneous confidence regions for $\{f(x), x \in \mathcal{X}\}$. Our regions will be centered around linear estimates $\hat{f}(x)$ of parametric or nonparametric $f(x)$. There is a large amount of previous work on this subject. Substantial restrictions have been usually placed on some or all of the dimensionality of x, the class of functions f that can be considered, the class of linear estimates $\hat{f}$ and the region $\mathcal{X}$. The method we present is an approximation to the tube formula and can be used for multidimensional x and a wide class of linear estimates. By considering the effect of bias we are able to relax assumptions on the class of functions f which are considered. Simulations and numerical computations are used to illustrate the performance.

[1] Research is supported in part by the National Science Foundation under grant DMS-89-02188.
AMS 1991 *subject classifications.* Primary 62F25; Secondary 60G15, 62G07, 62J05.
Key Words and phrases. Linear Smoother; Regression; Simultaneous Confidence Regions; Tube Formula.

1 Introduction

In a standard regression model, we observe $(x_i, Y_i), i = 1, \ldots, n$ where $x_i \in \mathcal{R}^d$ are predictor variables and Y_i are reponses. We assume the relation $Y_i = f(x_i) + \epsilon_i$ for an unknown function $f(\,.\,)$ and independent residuals ϵ_i, distributed as $N(0, \sigma^2)$. In this paper, we are interested in constructing simultaneous confidence bands for the unknown function f.

A linear estimate has $f(x)$ is an estimate of the form

$$\hat{f}(x) = \sum_{i=1}^{n} l_i(x) Y_i = \langle l(x), Y \rangle$$

where $l(x) = (l_1(x), \ldots, l_n(x))^T$ and $Y = (Y_1, \ldots, Y_n)^T$. Many popular regression estimates are linear; for example least squares polynomial regression. Some nonparametric regression estimates, such as kernel methods, locally weighted polynomial regression and smoothing splines are also linear, at least after a smoothing parameter has been selected.

Confidence bands for $f(x)$ over a subset $\mathcal{X}$ of the predictor space may take the form

$$\{(\hat{f}(x) - c\sigma \|l(x)\|, \hat{f}(x) + c\sigma\|l(x)\|); x \in \mathcal{X}\} \tag{1.1}$$

for a suitable constant c. An estimate $\hat{\sigma}$ is used in place of σ when σ is unknown. The simultaneous coverage probability of the confidence bands is

$$1 - \alpha = \inf_{f \in \mathcal{F}} P_f \left\{ \hat{f}(x) - c\sigma\|l(x)\| \le f(x) \le \hat{f}(x) + c\sigma \|l(x)\|, \quad \forall\, x \in \mathcal{X} \right\} \tag{1.2}$$

where $\mathcal{F}$ is a suitably large class of functions. Setting confidence bands requires evaluation of the right hand side of (1.2).

Exact confidence bands have been obtained for a number of parametric regression models. The classic results of Scheffe (1959, p68ff) are applicable when f is assumed to be a plane in d dimensions and $\mathcal{X} = \mathcal{R}^d$. Other papers deriving exact bands include Bohrer and Francis (1972), Casella and Strawderman (1980), Uusipaikaa (1983) and Wynn (1984). However, for many models, exact evaluation of the coverage probabilities appears intractable, and approximate methods are required.

Knafl, Sacks and Ylvisaker (1985) apply upcrossing methods to approximate the probability (1.2). This approach is applicable to both parametric and nonparametric regression models, but is most suited to a ond dimensional predictor. Alternative methods of constructing confidence bands in nonparametric regression are discussed by Hall and Titterington (1988) and Härdle and Marron (1991).

A class $\mathcal{F}$ which is of natural interset is those functions for which $\hat{f}(x)$ is an unbiased estimator,

$$\mathcal{F} = \{f : f(x) = \langle l(x), \mu \rangle\,, \quad \forall\, x\} \tag{1.3}$$

and $\mu = (f(x_1), \ldots, f(x_n))^T$. In this case the probability in (1.2) is independent of f and

$$\alpha = P\left\{ \sup_{x \in \mathcal{X}} |\langle T(x), \epsilon \rangle| > c\sigma \right\} \tag{1.4}$$

where $T(x) = l(x)/\|l(x)\|$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$. Note that $U = \epsilon/\|\epsilon\|$ is uniformly distributed on the unit sphere S^{n-1} in $\mathcal{R}^n$, and is independent of $\|\epsilon\|$. Hence,

$$\alpha = \int_c^\infty P\left\{ \sup_{x \in \mathcal{X}} |\langle T(x), U \rangle| \ge \frac{c}{y} \right\} g(y, n) dy \tag{1.5}$$

where $g(y, n)$ is the density of a χ random variable with n degrees of freedom. Letting $\mathcal{M} = \{T(x) : x \in \mathcal{X}\}$, the probability in (1.5) is simply the area of a tubular neighborhood of $\mathcal{M} \cup -\mathcal{M}$, divided by the surface area of S^{n-1}. This type of problem has been studied by Hotelling (1939) for $d = 1$; Weyl

(1939) for $d \geq 1$ and more recently by Naiman (1986, 1990); Knowles and Siegmund (1989); Johansen and Johnstone (1990), Sun (1993) and others. In section 2, we apply these results to derive easily computed approximations to (1.4). These approximations involve certain geometric constants related to the manifold $\mathcal{M}$. Section 3 discusses the computation of these constants.

The class (1.3) of functions that are estimated without bias is generally fairly small. In practice, model selection procedures that trade off bias and variance are used; the selected model will result in an estimate $\hat{f}(x)$ that has a small but not negligible amount of bias. In section 4 we examine the effect of bias on the coverage probability, and discuss methods to adjust the confidence bands to account for the bias.

In section 5, the approximation formulae and bias corrections are compared with some simulation results.

2 The Approximation Formula

In this section we state approximation formulae for one, two and $d \geq 3$ dimensional predictors. A number of assumptions are made to keep the statements simple. We suppose $\mathcal{X}$ is a closed d dimensional rectangle. The manifold $\mathcal{M} = \{T(x); x \in \mathcal{X}\}$ is assumed to be C^3, with a positive critical radius. We also suppose $T : \mathcal{X} \rightarrow \mathcal{M}$ is 1-1, three times differentiable and there exists a vector λ with $\langle \lambda, T(x) \rangle > 0$ for all $x \in \mathcal{X}$. This last condition ensures tubes around $\mathcal{M}$ and $-\mathcal{M}$ do not intersect for sufficiently small radii. The vector $\lambda = (1, \ldots, 1)^T$ suffices in most regression models.

Proposition 1 (One dimensional) *Suppose $\mathcal{X} = [a, b]$. The length of $\mathcal{M}$ is $\kappa_0 = \int_a^b \|T'(x)\| dx$, and the boundary of $\mathcal{M}$ has size $\zeta_0 = 2$. If σ^2 is known, then*

$$\alpha = \frac{\kappa_0}{\pi} \exp(-c^2/2) + \zeta_0(1 - \Phi(c)) + o(e^{-c^2/2}) \tag{2.1}$$

as $c \rightarrow \infty$. If σ is estimated by $\hat{\sigma}$ with $\nu\hat{\sigma}^2/\sigma^2 \sim \chi^2_\nu$ and independent of $\langle T(x), \epsilon \rangle$, then

$$\alpha \approx \frac{\kappa_0}{\pi} \left(1 + \frac{c^2}{\nu}\right)^{-\nu/2} + \frac{\zeta_0}{2} \cdot P(|t_\nu| > c). \tag{2.2}$$

Here $\kappa_0 = \int_{x \in \mathcal{X}} \|T'(x)\| dx$.

Naiman (1986) shows (2.1) and (2.2) to be upper bounds for α. This can also be deduced from Knafl, Sacks and Ylvisaker (1985), where $\mathcal{X}$ is partitioned and discrete upcrossing methods used to approximate the coverage probabilities.

Proposition 2 (2 dimensional) *Suppose $\mathcal{X}$ is a rectangle in $\mathcal{R}^2$. Let κ_0 be the area of $\mathcal{M}$; ζ_0 be the length of the boundary of $\mathcal{M}$, and $E = 1$ be the Euler-Poincaré characteristic of $\mathcal{M}$ (cf. Kreyszig, 1968). Then*

$$\alpha = \frac{\kappa_0 c}{2^{1/2}\pi^{3/2}} e^{-c^2/2} + \frac{\zeta_0}{2\pi} e^{-c^2/2} + 2E \cdot (1 - \Phi(c)) + o(e^{-c^2/2}). \tag{2.3}$$

If σ is estimated by $\hat{\sigma}$ with $\nu\hat{\sigma}^2/\sigma^2 \sim \chi^2_\nu$,

$$\alpha \approx \frac{\kappa_0}{\pi^{3/2}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{c}{\sqrt{\nu}} \left(1 + \frac{c^2}{\nu}\right)^{-(\nu+1)/2} + \frac{\zeta_0}{2\pi} \left(1 + \frac{c^2}{\nu}\right)^{-\nu/2} + E \cdot P(|t_\nu| > c). \tag{2.4}$$

Proof: Suppose U is distributed uniformly over the unit sphere S^{n-1}. Knowles and Siegmund (1989), corollary 1, gives

$$\begin{aligned} P\left\{\sup_{x \in \mathcal{X}} |\langle T(x), U \rangle| \geq w\right\} &= \frac{\Gamma(n/2)\kappa_0 w(1-w^2)^{(n-3)/2}}{\pi^{3/2}\Gamma((n-1)/2)} + \frac{\zeta_0(1-w^2)^{(n-2)/2}}{2\pi} \\ &\quad + \frac{2\Gamma(n/2)E}{\pi^{1/2}\Gamma\{(n-1)/2\}} \int_w^1 (1-x^2)^{(n-3)/2} dx \end{aligned}$$

for $w_0 \le w \le 1$, where w_0 is the critical radius of $\mathcal{M} \cup -\mathcal{M}$. Substituting into (1.5) gives (2.3). □

For the d-dimensional case, approximations to the volumes of tubular neighborhoods can be carried out using similar ideas; that is, the tube is decomposed into a main region and boundary regions of various orders. These ideas have been developed in detail by Naiman (1990).

Let $A_k = 2\pi^{k/2}/\Gamma(k/2)$ be the area of the unit sphere embedded in $\mathcal{R}^k$, and for fixed n and d,

$$J_e(w) = \frac{A_{n-d+e-1}}{A_n} \int_w^1 (1-u^2)^{(n-d+e-3)/2} u^{d-e} du.$$

Then

$$P\left\{\sup_{x\in\mathcal{X}} \langle T(x), U\rangle \ge w\right\} = \kappa_0 J_0(w) + \frac{\zeta_0}{2} J_1(w) + \frac{\kappa_2 + \zeta_1 + m_0}{2\pi} J_2(w) + O((1-w^2)^{(n-d+2)/2})$$

as $w \to 1$. Here, κ_0 is the volume of the manifold $\mathcal{M}$, and ζ_0 is the area of the boundary of $\mathcal{M}$. κ_2 and ζ_1 are measures of the curvature of $\mathcal{M}$ and $\partial\mathcal{M}$ respectively, and m_0 measures the rotation angles in the regions $\partial^2\mathcal{M}$ where two faces of $\mathcal{M}$ meet. Explicit forms for these constants that are suitable for computation are given in Section 3.

Substituting into (1.5) leads to the following proposition.

Proposition 3 (d dimensional) *If $T(x)$ is a one to one C^3 map from $\mathcal{X}$, a rectangle in $\mathcal{R}^d$, to $\mathcal{S}^{n-1}$, then*

$$\begin{aligned} P\{\sup_{x\in\mathcal{X}} |\langle T(x), \epsilon\rangle| \ge c\sigma\} &= \kappa_0 \frac{\Gamma(\frac{d+1}{2})}{\pi^{(d+1)/2}} P(\chi^2_{d+1} > c^2) + \frac{\zeta_0}{2}\frac{\Gamma(\frac{d}{2})}{\pi^{d/2}} P(\chi^2_d > c^2) \\ &\quad + \frac{\kappa_2 + \zeta_1 + m_0}{2\pi} \frac{\Gamma(\frac{d-1}{2})}{\pi^{(d-1)/2}} P(\chi^2_{d-1} > c^2) + O(c^{d-4}e^{-c^2/2}). \end{aligned} \tag{2.5}$$

as $c \to \infty$. If σ is estimated by $\hat\sigma$ with $\nu\hat\sigma^2/\sigma^2 \sim \chi^2_\nu$,

$$\begin{aligned} \alpha &\approx \kappa_0 \frac{\Gamma(\frac{d+1}{2})}{\pi^{(d+1)/2}} P(F_{d+1,\nu} > \frac{c^2}{d+1}) + \frac{\zeta_0}{2}\frac{\Gamma(\frac{d}{2})}{\pi^{d/2}} P(F_{d,\nu} > \frac{c^2}{d}) \\ &\quad + \frac{\kappa_2 + \zeta_1 + m_0}{2\pi} \frac{\Gamma(\frac{d-1}{2})}{\pi^{(d-1)/2}} P(F_{d-1,\nu} > \frac{c^2}{d-1}) \end{aligned} \tag{2.6}$$

When $d = 2$, Knowles and Siegmund (1988) show $\kappa_0 + \kappa_2 + \zeta_1 + m_0 = 2\pi E$. Using this relationship and integrating the χ^2 density by parts, (2.5) can be shown to equal (2.3).

In the nonparametric regression setting, variance estimates usually have the form $Y^T\Lambda Y$ with Λ symmetric and positive definite, and $\mathrm{tr}(\Lambda) = 1$. Appropriate choices of Λ give rise to the normalized residual sum of squares (Cleveland and Devlin, 1988) or difference based estimates such as those considered by Hall, Kay and Titterington (1990). Writing $Y = \mu + \epsilon$,

$$Y^T\Lambda Y = \epsilon^T\Lambda\epsilon + 2\epsilon^T\Lambda\mu + \mu^T\Lambda\mu.$$

Usually, Λ is chosen so that $\Lambda\mu \approx 0$. It is easy to show $E(\epsilon^T\Lambda\epsilon) = \sigma^2\mathrm{tr}(\Lambda) = \sigma^2$ and $\mathrm{var}(\epsilon^T\Lambda\epsilon) = 2\sigma^4\mathrm{tr}(\Lambda^2)$. Letting $\nu = 1/\mathrm{tr}(\Lambda^2)$, this suggests the two moment approximation (Cleveland and Devlin, 1988):

$$\frac{\nu Y^T\Lambda Y}{\sigma^2} \approx \chi^2_\nu. \tag{2.7}$$

The simulations in Section 5 suggest this is a useful approximation in the confidence band setting. Cleveland and Devlin (1988) report on some extensive simulations using this approximation for certain F ratios.

Which terms of (2.5) are most important? As $c \to \infty$, the terms are arranged in decreasing order of magnitude; the lead term involving κ_0 dominates (2.5). In (2.6), all terms decay at a rate $c^{-\nu}$ as

$c \to \infty$ so there is no asymptotic dominance; this should not cause problems unless ν is small. For practical purposes, the importance of the terms depends on the relative sizes of κ_0, ζ_0 and $\kappa_2 + \zeta_1 + m_0$. In numerical examples in Section 5, we find the first two terms are most important, with the third term usually offering some improvement.

Many exact confidence bands derived in the literature correspond to problems with special manifolds $\mathcal{M}$ for which closed form evaluation of the volume of tubular neighborhoods is tractable at any radius. Scheffe's (1959) method corresponds to the case $\mathcal{M}$ is a hemisphere in $d+1$ dimensions; $\mathcal{M} \cup -\mathcal{M}$ is the full sphere. The results of Casella and Strawderman (1980) correspond to situations where $\mathcal{M} \cup -\mathcal{M} = \{u : u_1^2 + \ldots + u_{d+1}^2 = 1, u_1^2 + \ldots + u_r^2 \geq 1 - w\}$ with $r \leq d$ and $0 < w < 1$. Uusipaikaa (1983) considers cases when $\mathcal{M}$ is the subset of the unit circle in $\mathcal{R}^2$. Bohrer and Francis (1972) consider one sided bounds when $\mathcal{M}$ is a spherical triangle on S^2, and also when $\mathcal{M}$ is the positive orthant of S^{d+1}.

We conclude this section by discussing a relation between our methods and the simulation method proposed by Härdle and Marron (1991). Let $Y_i^* = \hat{f}(x_i) + \epsilon_i^*$, where ϵ_i^* are chosen from a distribution supposed to approximate the true residual distribution. Limits $L(x), U(x)$ are chosen to be the $\beta/2$ and $1-\beta/2$ quantiles of $\hat{f}^*(x) = \sum l_j(x) Y_j^*$, where $\beta < \alpha$ is chosen to obtain simultaneous coverage $1-\alpha$. Confidence bands for f then have the form

$$(\hat{f}(x) - \hat{b}(x) + L(x), \hat{f}(x) - \hat{b}(x) + U(x))$$

where $\hat{b}(x)$ is an estimate of bias to be discussed more later. Härdle and Marron choose ϵ_i^* from a certain two point distribution with scale factor ϵ_i, and estimate $L(x)$ and $U(x)$ by simulation. Suppose we instead choose $\epsilon_i^* \sim N(0, \epsilon_i^2)$. Then $-L(x) = U(x) = c(\sum \epsilon_i^2 l_i(x)^2)^{1/2}$ for an appropriate c. Propositions 1, 2 and 3 are easily modified when the observations have known but unequal variances, and this allows approximate values of c to be found without any simulation.

3 Computation of Constants

In this section we describe the geometric constants appearing in the approximation formulae, and give forms that are useful for numerical computation. Throughout this section, $(x_1, \ldots, x_d)$ will denote a point $x \in \mathcal{X}$, and $T_j(x) = \partial T(x)/\partial x_j$. For simplicity, we suppose $\mathcal{X} = [0,1]^d$. The interior of $\mathcal{X}$, denoted $I(\mathcal{X})$, is $(0,1)^d$. The boundary $\partial\mathcal{X}$ consists of those points x with exactly one component 0 or 1. The regions where two faces of $\partial\mathcal{X}$ meet are denoted $\partial^2\mathcal{X}$.

At a point $x \in \mathcal{X}$, there exist a set of $n-d-1$ orthogonal vectors $\eta_1, \ldots, \eta_{n-d-1}$ such that $\langle T(x), \eta_j \rangle = \langle T_i(x), \eta_j \rangle = 0$ for all i and j. Hence if $\langle T(x), u \rangle$ is maximized at a point X_0 in $I(\mathcal{X})$ and $\langle T(X_0), u \rangle > 1 - w$, then u can be represented as

$$u = \frac{T(X_0) + \sum_{j=1}^{n-d-1} t_j \eta_j}{(1 + \sum_{j=1}^{n-d-1} t_j^2)^{1/2}} \tag{3.1}$$

with $\sum_{j=1}^{n-d-1} t_j^2 \leq (1-w^2)/w^2$. The results of Weyl (1939) give the series

$$P(\sup_{x \in \mathcal{X}} |\langle T(x), U \rangle| > w; X_0 \in I(\mathcal{X})) = \sum_{\substack{e=0 \\ e \text{ even}}}^{d} \kappa_e (2\pi)^{-e/2} J_e(w)$$

for $w_0 \leq w \leq 1$. The leading coefficient κ_0 is the volume of the manifold $\mathcal{M}$:

$$\kappa_0 = \int_{\mathcal{X}} \det{}^{1/2}(A^T A) dx \tag{3.2}$$

For computational efficiency, note $\det^{1/2}(A^T A)$ can be obtained directly from the QR decomposition of A.

Let $S(x)$ denote the intrinsic scalar curvature of $\mathcal{M}$; see Kreyszig (1968) p310 and Sun (1993). The total scalar curvature is

$$\kappa_2 = \int_{\mathcal{X}} \{\frac{S(x)}{2} - \frac{d(d-1)}{2}\}\det{}^{1/2}(A^T A)dx. \tag{3.3}$$

Weyl (1939) expresses the scalar curvature S in terms of the metric tensor matrix $A^T A$ and certain curvature tensors; see also Sun (1992). However, these expressions are not suitable for fast numerical computation. A much more usable form is

$$S(x) = 2\sum_{j=2}^{d}\sum_{k=1}^{j-1}(\beta_{jj}^T\beta_{kk} - \beta_{jk}^T\beta_{kj}),$$

where

$$\beta_{k,j}^T = \beta_{k,j}^T(x) = e_k^T(A^T A)^{-1}\frac{\partial A^T}{\partial x_j}(I - A(A^T A)^{-1}A^T).$$

Evaluation of κ_0 and κ_2 can then now be carried out using any multidimensional numerical intgration routine.

Extending the ideas to boundary regions is intuitively straightforward but technically sophisticated; see Naiman (1990). At a point x on the boundary of $\mathcal{X}$, the vectors $\eta_1, \ldots, \eta_{n-d-1}$ are supplemented by a vector $U(x)$ that is tangent to the manifold $\mathcal{M}$ but normal to the boundary $\partial\mathcal{M}$, and oriented to point outwards from $\mathcal{M}$. For example, on the face where $x_d = 1$,

$$U(x) = c(I - A_*(A_*^T A_*)^{-1}A_*^T)T_d(x),$$

where $A_* = (\, T_1(x) \quad \ldots \quad T_{d-1}(x)\,)$ and c is a normalizing constant. The boundary region of a tube can then be represented as

$$\frac{T(x) + sU(x) + \sum_{j=1}^{n-d-1} t_j\eta_j}{(1 + s^2 + \sum_{j=1}^{n-d-1} t_j^2)^{1/2}}$$

with the restriction $s > 0$ and $s^2 + ||t||^2 \le (1-w^2)/w^2$. Following the derivation of Naiman (1990), one arrives at a series for the boundary region:

$$P(\sup_{x\in\mathcal{X}} |\langle T(x), U\rangle| > w; X_0 \in \partial\mathcal{X}) = \zeta_0 J_1(w) + \frac{\eta_1}{2\pi}J_2(w) + O(J_3(w))$$

as $w \to 1$. The coefficients ζ_0 is the volume of $\partial\mathcal{M}$ and ζ_1 is a measure of the curvature of $\partial\mathcal{M}$. For numerical computation, we use the forms

$$\begin{aligned} \zeta_0 &= \int_{\partial\mathcal{X}} \det{}^{1/2}(A_*^T A_*) \\ \zeta_1 &= \int_{\partial\mathcal{X}} \zeta_1(x)\det{}^{1/2}(A_*^T A_*)dx \end{aligned}$$

where

$$\zeta_1(x) = -\sum_{j=1}^{d-1} e_j^T(A_*^T A_*)^{-1}\frac{\partial A_*^T}{\partial x_j}U(x)$$

on the face where x_d is maximized, with similar definitions for $\zeta_1(x)$ on other faces.

At a point x at the meeting of the faces $x_{d-1} = 1$, $x_d = 1$, let $A_{**} = (\, T_1(x) \quad \ldots T_{d-2}(x)\,)$ and

$$m_0(x) = \cos^{-1}\langle U_{d-1}(x), U_d(x)\rangle\,,$$

and similarly at boundaries where other faces meet. Then define

$$m_0 = \int_{\partial^2\mathcal{X}} m_0(x)\det{}^{1/2}(A_{**}^T A_{**})dx.$$

Then

$$P(\sup_{x\in\mathcal{X}} |\langle T(x), U\rangle| > w; X_0 \in \partial^2\mathcal{X}) = m_0 J_2(w) + O(J_3(w)).$$

4 Bias Correction

The previous section gives approximations to the simultaneous coverage probability (1.2) when $\mathcal{F}$ is the class of functions which can be estimated without bias. As noted in the introduction, this class is fairly small and we would like to approximate the simultaneous coverage over larger classes of functions. We measure the bias by

$$m(x) = \frac{E\hat{f}(x) - f(x)}{\sigma||l(x)||} = \frac{\sum_{j=1}^{n} l_j(x)f(x_j) - f(x)}{\sigma||l(x)||}. \tag{4.1}$$

One can estimate $m(x)$ by substituting an estimate for f in (4.1). It is then tempting to shift the confidence bands by $\hat{m}(x)$, giving bands of the form

$$(\hat{f}(x) - (c + \hat{m}(x))\sigma||l(x)||, \hat{f}(x) + (c - \hat{m}(x))\sigma||l(x)||). \tag{4.2}$$

Unfortunately this technique cannot work well. If $\hat{f}(x)$ is a reasonably efficient estimate of f, subtracting $\hat{m}(x)$ will generally increase variance more than it reduces bias. In simulations reported in Section 5, this correction fails badly, sometimes being worse than no correction. A similar technique is used by Härdle and Marron (1991); their Table 1 also suggests this bias correction is unsatisfactory.

A more reasonable use of $\hat{m}(x)$ may be to estimate certain global functionals of the bias; the maximum bias is useful in the confidence band setting. The maxmimum bias may also be available from practical considerations; if one is able to specify bounds on derivatives of f then Taylor series expansions can be used to obtain bounds on the bias. See Knafl, Sacks and Ylvisaker (1985) and related ideas in Hall and Titterington (1988). For the remainder of this section we assume the maximum bias is known; the effect of estimation will be studied in simulations in Section 5.

Suppose $f \in \mathcal{F}_\delta$ where

$$\mathcal{F}_\delta = \{f : \sup_{x\in\mathcal{X}} |m(x)| \leq \delta\}.$$

An application of the triangle inequality gives

$$P\left\{|f(x) - \hat{f}(x)| \leq c\sigma||l(x)||, \quad \forall\, x \in \mathcal{X}\right\} \geq P\left\{|\langle l(x), \epsilon\rangle| \leq (c-\delta)\sigma||l(x)||, \quad \forall\, x \in \mathcal{X}\right\}.$$

This is similar to the bias corrections used in Knafl, Sacks and Ylvisaker (1984). If c is chosen so that

$$P\left(|\langle l(x), \epsilon\rangle| \leq c\hat{\sigma}||l(x)||, \quad \forall\, x \in \mathcal{X}\right) = 1 - \alpha$$

this leads to a lower bound for the true coverage probability of the form

$$\inf_{f\in\mathcal{F}_\delta} P_f(|\hat{f}(x) - f(x)| < c\hat{\sigma}||l(x)||, \quad \forall\, x \in \mathcal{X}) = 1 - \alpha - O(\delta) \tag{4.3}$$

as $\delta \to 0$. While this gives a lower bound on the true coverage probability, the infimum will be attained by functions f for which $m(x)$ oscillates rapidly between $-\delta$ and δ. Since we typically do not expect $m(x)$ to have this form, better results may be obtained by considering more restrictive classes of functions. Indeed, we show that over classes of functions with bounded derivatives, the error term in (4.3) can be improved to $O(\delta^2)$. We first give a straightforward example to illustrate the main ideas, then give a general result in Theorem 1.

Example: Suppose ϵ_1, ϵ_2 are independent $N(0,1)$, and $Z(x) = \epsilon_1\cos(x) + \epsilon_2\sin(x)$ for $0 \leq x < \pi$. Let $m(x) = \sin(x)$. Then

$$\begin{aligned}
P(\sup_{0\leq x<\pi} |Z(x) + \delta m(x)| < c) &= P(\sup_{0\leq x<2\pi} (Z(x) + \delta m(x)) < c) \\
&= P(\epsilon_1^2 + (\epsilon_2 + \delta)^2 < c^2) \\
&= P(\epsilon_1^2 + \epsilon_2^2 < c^2) - O(\delta^2)
\end{aligned}$$

as $\delta \to 0$. The final line can be derived from the bivariate normal integral.

More generally, suppose $m(x)$ has a bounded derivative, and $m(x) = -m(x+\pi)$. Then there exists m_1 such that

$$|m(x) - m(x')| \le m_1 \sin(|x - x'|)$$

whenever $\cos(x - x') > 0$. Symmetry arguments give

$$\sup_{0 \le x < \pi} |Z(x) + \delta m(x)| \stackrel{\mathcal{L}}{=} \sup_{0 \le x < \pi} |Z(x) + \delta V m(x)| = \sup_{0 \le x < 2\pi} (Z(x) + \delta V m(x)) \tag{4.4}$$

where $P(V = 1) = P(V = -1) = 0.5$. We can write $Z(x) = R\cos(x - X_0)$ for suitable R and X_0. For values of x with $\cos(x - X_0) > 0$,

$$\begin{aligned} Z(x) + \delta V m(x) &\le R\cos(x - X_0) + \delta V m(X_0) + \delta m_1 \sin(|x - X_0|) \\ &\le \delta V m(X_0) + (R^2 + \delta^2 m_1^2)^{1/2} \\ &\le R + \delta V m(X_0) + \frac{\delta^2 m_1^2}{2R}. \end{aligned}$$

For $\delta < c/2$ this implies

$$\begin{aligned} P(\sup_{0 \le x < 2\pi} |Z(x) + \delta m(x)| > c) &\le P(R + \delta V m(X_0) + \frac{\delta^2 m_1^2}{2R} > c) \\ &= P(R > c) + O(\delta^2) \end{aligned}$$

using Lemma 1 below.

Lemma 1 *Let X, V and R be random variables satisfying the following conditions:*

- *X has a density $f_X(\cdot)$ with a bounded derivative.*
- *There exists v_0 such that $P(|V| < v_0) = 1$; $G(v, x) = P(V > v | X = x)$ has a bounded partial derivative with respect to x, and $G(v, x) + G(-v, x) = 1$ for all v and x. That is, the conditional distribution of V given X is symmetric about 0.*
- *$P(R \ge 0) = 1$ and the conditional distribution of R given $X = x$ is continuous in x, and $E(1/R) < \infty$.*

Then for any constant c with $f_X(c) > 0$,

$$P(X + \delta V + \delta^2 R > c) = P(X > c) + O(\delta^2)$$

as $\delta \to 0$.

Proof:

$$P(X + \delta V + \delta^2 R > c) = P(X + \delta V > c) + P(X + \delta V < c, X + \delta V + \delta^2 R > c),$$

and

$$P(X + \delta V > c) = P(X > c) + P(X < c, X + \delta V > c) - P(X > c, X + \delta V > c).$$

Conditioning on X,

$$\begin{aligned} P(X < c, X + \delta V > c) &= \delta \int_0^{v_0} G(v, c - \delta v) f_X(c - \delta v) dv \\ P(X > c, X + \delta V < c) &= \delta \int_0^{v_0} G(v, c + \delta v) f_X(c + \delta v) dv. \end{aligned}$$

The differentiability conditions on G and f ensure both these integrals are equal to $\delta \int G(v,c) f_X(c) + O(\delta^2)$, and hence

$$P(X + \delta V > c) = P(X > c) + O(\delta^2).$$

Let g_δ be the density of $X + \delta V$. Then

$$\begin{aligned} P(X + \delta V < c, X + \delta V + \delta^2 R > c) &= \delta^2 \int_0^\infty P(R > r | X + \delta V = c - \delta^2 r) g_\delta(c - \delta^2 r) dr \\ &\sim \delta^2 \int_0^\infty P(R > r | X = c) f_X(c) dr \end{aligned}$$

provided $f_X(c) > 0$. The condition $E(1/R) < \infty$ ensures the intgral converges.

$\square$

The development of the preceding example can be extended to give the following theorem.

Theorem 1 *Let $\epsilon_1, \ldots, \epsilon_n$ be i.i.d. $\mathcal{N}(0,1)$ random variables and $\{T(x); x \in \mathcal{X}\}$ be a C^3 manifold embedded in S^{n-1}. Suppose $m : \mathcal{X} \to \mathcal{R} \in \mathcal{F}_{1,m_1}$ where*

$$\mathcal{F}_{1,m_1} = \{m : |m(x)| < 1, |m(x) - m(x')| < m_1 |x - x'| \|T'(x)\|, \quad \forall x, x' \in \mathcal{M}\}.$$

Let

$$Z_\delta(x) = \langle T(x), \epsilon \rangle + \delta m(x)$$

for $x \in \mathcal{X}$. Then

$$P\left\{\sup_{x \in \mathcal{X}} |Z_\delta(x)| < c\right\} = P\left\{\sup_{x \in \mathcal{X}} |Z_0(x)| < c\right\} - O(\delta^2).$$

as $\delta \to 0$. Moreover, the $O(\delta^2)$ term holds uniformly over the set $\mathcal{F}_{1,m_1}$.

Proof: Define $\mathcal{X}^*$ to be a translation of $\mathcal{X}$, with $T(x^*) = -T(x)$ and $m(x^*) = -m(x)$. Let $P(V = 1) = P(V = -1) = 0.5$, independent of ϵ. Similarly to (4.4), we have

$$\sup_{x \in \mathcal{X}} |Z(x) + \delta m(x)| \stackrel{\mathcal{L}}{=} \sup_{x \in \mathcal{X} \cup \mathcal{X}^*} (Z(x) + \delta V m(x)).$$

By the assumptions on $T(x)$, the third derivative of $Z(x)$ is bounded; $|Z'''(x)| < M$. Letting $X_0 = \text{argmax} Z(x)$,

$$Z(x) + \delta V m(x) \le Z(X_0) + \frac{Z''(X_0)}{2}(x - X_0)^2 + \frac{Mh^3}{6} + \delta V m(X_0) + \delta m_1 \|T'(X_0)\| |x - X_0|.$$

whenever $|x - X_0| < h$. The maximum occurs at $|x - X_0| = \delta m_1 \|T'(X_0)\| / |Z''(X_0)|$, and hence

$$\sup_{|x - X_0| < h} (Z(x) + \delta m(x)) \le Z(X_0) + \delta V m(X_0) + \frac{\delta^2 m_1^2 \|T'(X_0)\|^2}{2|Z''(X_0)|} + \frac{Mh^3}{6}. \tag{4.5}$$

Taking $h = \delta^{4/5}$,

$$P\left\{\sup_{x \in \mathcal{X} \cup \mathcal{X}^*} Z_\delta(x) > c\right\} \le P\left\{Z(X_0) + \delta V m(X_0) + \frac{\delta^2 m_1^2 \|T'(X_0)\|^2}{2|Z''(X_0)|} > c\right\} + O(\delta^2). \tag{4.6}$$

Applying Lemma 1 completes the result, provided $E(1/|Z''(X_0)|) < \infty$. For simplicity, suppose $\mathcal{X} = [0,1]$. Fix $h > 0$ and let $X_0(h) = \text{argmax } Z(kh)$. Then

$$P(X_0(h) = kh, |Z''(kh)| < \epsilon) = P(2h^{-1}|Z'(kh)| \le |Z''(kh)| \le \epsilon) + o(h)$$

as $h \to 0$. This implies

$$P(|Z''(X_0)| < \epsilon; X_0 \in I(\mathcal{X})) = O(\epsilon^2)$$

as $\epsilon \to 0$;

$$E(1/|Z''(X_0)|; X_0 \in I(\mathcal{X})) = \int_0^\infty P(1/|Z''(X_0)| > x; X_0 \in \mathcal{X})dx < \infty.$$

$\square$

An immediate corollary of Theorem 1 is

Corollary 1 *Suppose c is chosen so that*

$$P(|\langle l(x), \epsilon\rangle| < c\sigma\|l(x)\|, \quad \forall\, x \in \mathcal{X}) = 1 - \alpha$$

and $\{T(x); x \in \mathcal{X}\}$ satisfies the conditions of Theorem 1. Then the confidence region $(\hat{f}(x) - c\sigma\|l(x)\|, \hat{f}(x) + c\sigma\|l(x)\|)$ has simultaneous coverage probability $1 - \alpha - O(\delta^2)$ over the class

$$\mathcal{F}_{\delta,\delta_1} = \{f : |m(x)| < \delta, |m(x) - m(x')| < \delta_1|x - x'| \cdot \|T'(x)\|, \quad \forall\, x, x' \in \mathcal{X}\}$$

with $\delta_1 = O(\delta)$.

We now seek to modify the basic approximation formulae from Section 2. In the course of proving Theorem (1), we have identified several sources contributing to the $O(\delta^2)$ error term. Some of these sources are related to regions of overlap of the tube, and it is not possible to approximate these terms while retaining the simplicity of the original tube formulae.

Consider the representation (3.1) $\epsilon/\|\epsilon\|$, with the additional assumption that for each x, η_1 is chosen so that $T''(x)$ lies in the space spanned by $T(x)$, $T'(x)$ and η_1. That is,

$$T''(x) = -\|T'(x)\|^2 T(x) + \frac{\langle T''(x), T'(x)\rangle}{\|T'(x)\|^2} T'(x) + a(x)\eta_1(x)$$

for an appropriate $a(x) > 0$. This leads to

$$\langle T''(X_0), \epsilon\rangle = -\langle T(X_0), \epsilon\rangle (\|T'(X_0)\|^2 - a(X_0)t_1).$$

Rearranging (4.6),

$$\begin{aligned} (4.7) \qquad & P\left\{\sup_{x\in\mathcal{X}} |Z_\delta(x)| > c\right\} \\ \leq \quad & P\left\{Z(X_0) > c - \delta V m(X_0) - \frac{\delta_1^2\|T'(X_0)\|^2}{2c(\|T'(X_0)\|^2 - a(X_0)t_1)}\right\} + O(\delta^2). \end{aligned}$$

Analysis of (4.7) is made complicated by the $a(X_0)t_1$ term. For some special manifolds, such as in the example earlier in this section, $a(x) = 0$ for all x. More generally, interest is focused on values of ϵ such that t_1 is small, and little is lost by ignoring this term. This leads to the approximation

$$P\left\{\sup_{x\in\mathcal{X}} |Z(x) + \delta m(x)| > c\right\} \approx \frac{1}{2\pi}\int_{\mathcal{X}} (P(\chi_2^2 > c_1(x)^2) + P(\chi_2^2 > c_2(x)^2))\|T'(x)\|dx,$$

where $c_1(x) = c - \delta m(x) - \delta_1^2/2c$ and $c_2(x) = c + \delta m(x) - \delta_1^2/2c$. For large values of c, the χ^2 function is convex so $c - \delta_1^2/(2c)$ is usually conservative, leading to our final bias corrected approximation:

$$(4.8) \quad P(\sup_{x\in\mathcal{X}} |Z(x) + \delta m(x)| > c) \quad \approx \quad \frac{\kappa_0}{2\pi}\left[P(\chi_2^2 > (c-\delta)^2) + P(\chi_2^2 > (c + \delta - \min(\frac{\delta_1^2}{c}, 2\delta))^2)\right].$$

Similar corrections may be obtained for the boundary terms. The main ideas underlying this approximation extend to the case $d > 1$, with the definition of δ_1 extended to include all directional derivatives.

5 Simulations

The first example consists of a one dimensional design with $n = 50$ and $x_i = (i-1)/49$. Four fitting procedures were used; a quadratic regression and a local linear regression with three different bandwidths. The local linear regression assigns weights $w_i(x) = W((x - x_i)/h)$, and estimating $f(x)$ using weighted least squares; see Cleveland (1979) or Loader and Hastie (1993). We use the tricube weight function

$$W(u) = \begin{cases} (1-|u|^3)^3 & |u| < 1 \\ 0 & \text{otherwise} \end{cases} . \tag{5.1}$$

For each fitting procedure, a value of κ_0 is computed with $\mathcal{X} = [0,1]$; for $\alpha = 0.1, 0.05$ and 0.01, critical values c are computed by solving (2.2). Simulations of size 10^5 are then used to estimate the true coverage probabilities in the absence of bias; the results are shown in Table 1. As expected, the simulations indicate the true coverage slightly exceeds the nominal coverage in most cases. The approximation is most conservative when κ_0 is large. The normalized residual sum of squares is used as the variance estimate, with the distribution approximation (2.7).

Table 1 About Here

We show the results of a two dimensional simulation in Table 2. The design consisted of observations at the points $((i-1)/10, (j-1)/10); i, j = 1, \ldots, 10$; confidence regions are simulated over the unit square. We use a bivariate quadratic regression and local linear regression with three different bandwidths. When fitting at x, the observations receive weights $w_i(x) = W(\|x - x_i\|/h)$ with W defined by (5.1).

Table 2 About Here

Which terms are important? In Table 3, we compare the approximations obtained using one, two and three terms from (2.6). One term seems inadequate in most cases; the approximation may substantially overestimate the true coverage. The two and three term approximations are very similar in the two dimensional cases; the difference becomes slightly larger in higher dimensions.

Table 3 About Here

In Table 4 we show simulated coverage probabilities using a model with bias. The design is again 50 equally spaced points on $[0,1]$. The true mean function is $f(x) = 2\sin(\pi x/2)$ and $\sigma = 1$. Five different bandwidths are used; the risk, defined as $\sum_{i=1}^{n} E(\hat{f}(x_i) - f(x_i))^2$, indicate these range from too small through about right to too large.

Since the true mean function is known, exact values of δ and δ_1 may be calculated. The simulations are of size 10000.

Without any bias correction, the coverage is much less than the nominal 95% unless a small bandwidth is used. Using the true values of δ and δ_1, the bands are conservative in all cases. When estimated δ and δ_1 are used, the bands are conservative except for large bandwidths. Using $\hat{m}(x)$ as a point estimate in (4.2) is a failure; in many cases this is worse than doing no adjustment, and in all cases coverage is substantially less than the nominal 95%.

Table 4 About Here

We conclude with an application to the dataset of Snee (1977), studied in the simultaneous confidence band setting by several Authors. Naiman (1987) and Knafl, Sacks and Ylvisaker (1985) both fit a bivariate linear model; however, standard methods can be used to show quadratic terms are highly siginificant. Cubic terms appear less significant.

We show approximate values of c for 95% confidence and $\mathcal{X} = [1100, 1300] \times [5.3, 23]$ in Table 5. Note that when σ is known and bias is neglected, our method gives critical values within ± 0.01 of those obtained by Naiman (1987). However, we are unable to reconcile our results with those of Knafl, Sacks and Ylvisaker (1985) and believe their computations are in error.

We obtain approximate values of δ by considering the difference between the linear and quadratic fits in the linear case, and the difference between the quadratic and cubic fits in the quadratic case. We ignore the δ_1 term. The adjusted critical values are in both cases very much larger than the unadjusted critical values. The quadratic regression gives a much smaller confidence region.

Table 5 About Here

References

[1] Bohrer, R. and Francis, G. K. (1972). Sharp one-sided confidence bounds over positive regions. *Ann. Math. Statist.* 43, 1541-1548.

[2] Casella, G. and Strawderman, W. E. (1980). Confidence bands for linear regression with restricted predictor variables. *J. Amer. Statist. Assoc.* 75, 862-868.

[3] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829-836.

[4] Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83**, 596-610.

[5] Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27**, 228-254.

[6] Hall, P; Kay, J. W. and Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77**, 521-528.

[7] Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778-796.

[8] Hastie, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry. *Statist. Sci.* , to appear.

[9] Hotelling, H. (1939). Tubes and spheres in n-spaces, and a class of statistical problems. *Amer. J. Math.* **61**, 440-460.

[10] Johansen, S. and Johnstone, I. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.* **18**, 652-684.

[11] Knafl, G.; Sacks, J. and Ylvisaker, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80**, 683-691.

[12] Knowles, M. and Siegmund, D. (1989). On Hotelling's geometric approach to testing for a nonlinear parameter in Regression. *Internat. Statist. Rev.* **57**, 205-220.

[13] Kreyszig, E. (1968). *Introduction to Differential Geometry and Riemannian Geometry.* University of Toronto Press.

[14] Naiman, D. Q. (1986). Conservative confidence bands in curvilinear regression. *Ann. Statist.* **14**, 896-906.

[15] Naiman, D. Q. (1987). Simultaneous confidence bounds in multiple regression using predictor variable constraints. *J. Amer. Statist. Assoc.* **82**, 214-219.

[16] Naiman, D. Q. (1990). On volumes of tubular neighborhoods of spherical polyhedra and statistical inference. *Ann. Statist.* **18**, 685-716.

[17] Scheffe, H. (1959). *The Analysis of Variance.* John Wiley & Sons, New York.

[18] Snee, R. D. (1977). Validation of regression models: Methods and examples. *Technometrics* **19**, 415-428.

[19] Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.* , to appear.

[20] Uusipaikka, E. (1983). Exact confidence bands for linear regression over intervals. *J. Amer. Statist. Assoc.* **78**, 638-644.

[21] Weyl, H. (1939). On the volume of tubes. *Amer. J. Math.* **61**, 461-472.

[22] Wynn, H. P. (1984). An exact confidence band for one-dimensional polynomial regression. *Biometrika* **71**, 375-380.

Department of Statistics
1440 Mason Hall
The University Of Michigan
Ann Arbor, MI 48109

AT&T Bell Laboratories
Room 2C-279
600 Mountain Avenue
Murray Hill, NJ 07974

	κ_0	ν	$1-\alpha$ 0.90	0.95	0.99
Quadratic	3.9147	47	0.9068	0.9519	0.9899
$h = 0.10$	18.4906	38.649	0.9132	0.9556	0.9920
$h = 0.30$	6.7004	45.406	0.9088	0.9538	0.9913
$h = 0.50$	4.3422	46.768	0.9054	0.9528	0.9910
Simulation s.e.			0.0009	0.0007	0.0003

Table 1: Simulated coverage probabilities for a quadratic regression and locally weighted regression with 50 equally spaced points on $[0,1]$.

	κ_0	ζ_0	ν	$1-\alpha$ 0.90	0.95	0.99
Quadratic	9.6092	9.9055	94	0.9071	0.9526	0.9901
$h = 0.30$	46.9323	25.4649	77.785	0.9199	0.9602	0.9920
$h = 0.50$	19.1612	15.9367	88.968	0.9085	0.9550	0.9910
$h = 0.80$	8.6926	10.5536	93.702	0.9048	0.9534	0.9900
Simulation s.e.				0.0009	0.0007	0.0003

Table 2: Simulated coverage probabilities for a bivariate quadratic regression and a local linear smoother. The design consists of 100 points in a 10×10 grid on the unit square.

	c	ν	1 term	2 terms	3 terms	Sim.
Quadratic	3.1418	94	0.9633	0.9488	0.9500	0.9526
$h = 0.30$	3.6630	77.785	0.9554	0.9471	0.9500	0.9602
$h = 0.50$	3.3700	88.968	0.9598	0.9477	0.9500	0.9550
$h = 0.80$	3.1249	93.702	0.9652	0.9491	0.9500	0.9534
3 dim.	3.5178	115	0.9692	0.9501	0.9500	
4 dim.	3.8043	610	0.9721	0.9509	0.9500	

Table 3: Comparison of the approximation formula with one, two and three terms. The first four models correspond to the case of Table 2; the final two cases consist of fitting a quadratic regression to points on a 5^d grid on the unit cube, with $d = 3$ and $d = 4$ respectively.

	h				
	0.10	0.20	0.30	0.40	0.50
δ	0.084	0.465	1.244	2.441	5.737
δ_1	0.074	0.364	0.993	2.550	7.098
Risk	8.675	5.401	5.214	6.654	9.295
No correction	0.9488	0.9327	0.8489	0.5128	0.0900
δ only	0.9496	0.9674	0.9872	0.9903	0.9893
δ and δ_1	0.9497	0.9678	0.9872	0.9903	0.9893
Triangle	0.9587	0.9840	0.9937	0.9948	0.9947
δ (Estimate)	0.9906	0.9877	0.9772	0.9201	0.7101
δ, δ_1 (Est.)	0.9914	0.9891	0.9778	0.9204	0.7101
Triangle (Est.)	0.9957	0.9948	0.9870	0.9478	0.7853
Point Estimate	0.7860	0.8249	0.8481	0.8088	0.6241

Table 4: Simulated coverage probabilities in the presence of bias, with various bias corrections. The design consists of 50 equally spaced points on $[0,1]$; the mean function is $f(x) = 2\sin(\pi x/2)$. A local linear fit is used.

	Unadj.	Adjusted			
	c	$\hat{\sigma}$	δ	c	Volume
Linear	3.1618	3.624	3.986	6.7417	6.84×10^4
Quadratic	3.8184	1.066	1.059	4.4431	1.87×10^4

Table 5: Critical Values for Acetylene Data.