The Annals of Statistics
1984, Vol. 12, No. 4, 1199–1214

AVERAGE WIDTH OPTIMALITY OF SIMULTANEOUS CONFIDENCE BOUNDS[1]

By Daniel Q. Naiman

The Johns Hopkins University

Simultaneous confidence bounds for multilinear regression functions over subregions X of Euclidean space are defined to be μ-optimal in a class of bounds C, if they minimize average width with respect to μ over X, among all bounds in C with equal coverage probability. We show that for certain simultaneous confidence bounds we can find a measure μ relative to which the bounds are μ-optimal in C, where C is a large class of bounds. Such results are obtained for bounds over finite sets, and for bounds for simple linear regression functions over finite intervals.

1. Introduction. In (1973), Bohrer showed that Scheffé-type simultaneous confidence bounds for a multilinear regression function $f(\mathbf{x}) = \mathbf{x}'\beta$ have smallest average width with respect to Lebesgue measure among all simultaneous confidence bounds with equal coverage probability (in a certain class of bounds to be described below), if the region over which the regression function is to be bounded takes the form of an ellipsoid

$$X = \{\mathbf{x}: \mathbf{x}'(A'A)^{-1}\mathbf{x} \le a^2\},$$

where A is the design matrix. We generalize this result by showing that many simultaneous confidence bounds share such an optimality property when the class of competing bounds is restricted in a suitable way.

Consider a multilinear regression model in which we observe

$$\mathbf{Y} = A\beta + \mathbf{e}, \tag{1.1}$$

where β is an unknown k-vector, A is a known $n \times k$ matrix, and $\mathbf{e}$ has a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 I_n$, with σ^2 unknown.

For a given subset X of R^k, suppose we are interested in obtaining two-sided simultaneous confidence bounds for the regression function $f(\mathbf{x}) = \mathbf{x}'\beta$ over X, with coverage probability $1 - \alpha$, of the form

$$(\mathbf{x}'\hat{\beta} - S\phi(\mathbf{x}),\ \mathbf{x}'\hat{\beta} + S\phi(\mathbf{x})), \tag{1.2}$$

Received April 1983; revised May 1984.

[1] Section 5 of this paper was part of the author's Ph.D. dissertation, prepared at the University of Illinois/Urbana. This research was supported by the United States Department of Navy under Office of Naval Research Contract No. N00014-79-C-0801. Reproduction in whole or part is permitted for any purpose of the United States Government.

AMS 1980 *subject classifications*. 62J15, 62J10, 62C07.

Key words and phrases. Simultaneous confidence bounds, multilinear regression, analysis of variance.

or one-sided simultaneous confidence bounds of the form

$$(1.3)\qquad (-\infty, \mathbf{x}'\hat{\beta} + S\phi(\mathbf{x})),$$

where ϕ is a nonnegative measurable function on X, $\hat{\beta}$ is the least squares estimator of β and S^2 is the usual unbiased estimator of σ^2.

For any σ-finite measure μ on X we call the bounds determined by ϕ in (1.2) (resp. (1.3)), optimal with respect to μ, or μ-optimal, as two-sided (resp. one-sided) bounds, if among all bounds of the same form and with equal coverage probability, they minimize the average width over X with respect to μ, $\int_X \phi(\mathbf{x})\mu(d\mathbf{x})$.

The problem of finding μ-optimal bounds is motivated by the following practical problem. Suppose that one wants two-sided simultaneous confidence bounds for the unknown regression function $f(\mathbf{x})$ over X, of the form (1.2), with coverage probability $1 - \alpha$. Suppose further that the actual future points $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$, at which these bounds will be used, are random vectors which are independent and identically distributed according to the probability measure μ on X. The accuracy of the resulting intervals, as measured by average width, is proportional to $\sum_{i=1}^{n} \phi(\mathbf{x}_i)/n$, which converges a.s. to $\int_X \phi(\mathbf{x})\mu(d\mathbf{x})$ by the strong law of large numbers, provided this integral is finite. Consequently, μ-optimal simultaneous confidence bounds define simultaneous confidence intervals with optimal average accuracy. A similar remark can be made for one-sided bounds.

Bounds of the form (1.2) and (1.3) have appeared frequently in the literature. Working and Hotelling (1929) introduced the first such bounds, and Scheffé (1953, 1959) generalized them to multilinear regression. These bounds have the property that the probability of any of the intervals (1.2) (resp. (1.3)) covering $f(\mathbf{x})$ does not depend on $\mathbf{x}$. Bounds defined in this way are referred to as Scheffé-type bounds. Since 1959, many other bounds of the above form have been introduced to take into account restrictions on the regressors. See Miller (1977) for an excellent review of the literature on this subject.

To simplify notation we reparameterize the linear model (1.1) by expressing it as

$$\mathbf{Y} = A^*\beta^* + \mathbf{e},$$

where $A^* = AP^{-1}$, $\beta^* = P\beta$, and where P is the $k \times k$ symmetric square root of the matrix $A'A$. The linear transformation determined by P defines a one-to-one correspondence between simultaneous confidence bounds for the linear regression function $f(\mathbf{x}) = \mathbf{x}'\beta$ over X and simultaneous confidence bounds for $f^*(\mathbf{x}) = \mathbf{x}'\beta^*$ over $X^* = P^{-1}X$, and a one-to-one correspondence between σ-finite measures on X and X^*.

Henceforth, we will assume without loss of generality that the model has been transformed in this manner. We will use $f(\mathbf{x}) = \mathbf{x}'\beta$ to denote the regression function, and X will always denote the region over which it is to be bounded. We define the random k-vector $\mathbf{B} = S^{-1}(\beta - \hat{\beta})$. Since the model has been transformed as above, $\mathbf{B}$ has a spherically symmetric multivariate t-distribution. Furthermore, $\| \mathbf{B} \|$ is distributed as $\{kF_{k,\nu}\}^{1/2}$, where $\nu = n - k$. This distribution is unimodal with mode $m_0 = \{(k-1)\nu/(\nu+1)\}^{1/2}$. We use F to denote the distribution function of $\|\mathbf{B}\|$, and g to denote the density function for $\mathbf{B}$.

In terms of $\mathbf{B}$, the given subset X of R^k, and specified level α, a nonnegative measurable function ϕ on X defines two-sided simultaneous confidence bounds of the form (1.2) for the regression function $f(\mathbf{x})$, over X, if

$$P(|\mathbf{x}'\mathbf{B}| \le \phi(\mathbf{x}), \text{ all } \mathbf{x} \in X) = 1 - \alpha, \tag{1.4}$$

and ϕ defines one-sided simultaneous confidence bounds of the form (1.3) if

$$P(\mathbf{x}'\mathbf{B} \le \phi(\mathbf{x}), \text{ all } \mathbf{x} \in X) = 1 - \alpha. \tag{1.5}$$

Simultaneous confidence bounds of form (1.2) (resp. 1.3) will be referred to by the function ϕ on X that defines them. Given a σ-finite measure μ on X and a set C of bounds over X, we call an element ϕ of C μ-optimal in C, if ϕ minimizes average width with respect to μ among all bounds in C satisfying (1.4) (resp. (1.5)).

We can now state Bohrer's (1973) result as follows. If X is a ball centered at the origin in R^k, taking $\phi(\mathbf{x}) \propto \|\mathbf{x}\|$ yields a bound which is m-optimal, where m is Lebesgue measure in R^k restricted to X, provided α is sufficiently small. Bohrer's argument is modified easily to yield the μ-optimality of ϕ with respect to any finite measure μ on X which is spherically symmetric.

Bohrer's result does not imply optimality of Scheffé-type bounds when the multilinear regression model has an intercept parameter since the region X for bounding the regression function must be restricted to a hyperplane in R^k. The main result of Section 5 (Theorem 5.1) states that the Scheffé-type bound is m-suboptimal for bounding a simple linear regression function over a sufficiently large finite interval, where m is Lebesgue measure over the interval.

The paper is organized as follows. For convenience, results will be stated for one-sided bounds. All of the results of the paper have analogues for two-sided bounds. The necessary modifications for obtaining these results are summarized in Appendix 2.

In Section 2 we show that in a certain large class C of bounds for the regression function over an arbitrary subset X of R^k, local μ-optimality in C implies global μ-optimality in C. This follows from the fact that on the set C the coverage probability is concave in ϕ, and this leads to the following result (Theorem 2.2). If X is a finite set and ϕ is a taut bound in C (see Definition 2.2), then there exists a finite measure μ on X relative to which ϕ is optimal in C. If ϕ is in the interior of C, and X contains at most one point in any direction, then μ is unique up to a constant of proportionality. We show in Lemma 2.2 that C contains all bounds with sufficiently large coverage probability. Thus, if ϕ is μ-optimal in C, and the coverage probability of ϕ is sufficiently large, then ϕ is μ-optimal among all bounds.

Given a taut bound ϕ on an infinite set X, how do we find the measure μ relative to which ϕ is optimal? The answer is to find the measures for finite subsets of X and take limits. The results of Sections 3–5 refer to simple linear regression over a finite closed interval. Let ϕ be any bound which is taut. For any finite subset X_f of X, let μ_f be the measure constructed in Theorem 2.2 relative to which the restriction of ϕ to X_f is optimal. We show in Section 3 that there is a measure μ on X with the following property. For any dense sequence of points $\{\mathbf{x}_n\}$ in X, the measures μ_f corresponding to the sets $X_f = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, converge weakly to μ as $n \to \infty$. Theorem 3.2 states that ϕ is μ-optimal.

In Section 4 we derive this measure for some special cases. For the Bohrer and Francis (1972) Scheffé-type bound over a finite interval, and its two-sided analogue, the measure is a mixture of a measure with density proportional to $\{1 + (x - \bar{x})^2\}^{-3/2}$ on the interval, where $\bar{x}$ is the average of the design points, and a measure concentrated at the endpoints of the interval. For the trapezoidal bounds over finite intervals of Bowden and Graybill (1966), which generalize the uniform bounds of Gafarian (1964), the measure μ is shown to be one which is concentrated at the endpoints of the interval.

Given a measure μ on a set X, we would like to be able to find a bound ϕ which is μ-optimal. While the results given in this paper do not deal directly with this problem, they support the following indirect method. For a proposed solution ϕ of the μ-optimal bound problem, the methods herein show how to construct a measure relative to which ϕ is optimal. If this measure is sufficiently close to μ, then ϕ is close to being μ-optimal.

The constructed measure can also be used to indicate how a proposed solution can be modified to give a bound with smaller average width with respect to the given measure μ. In Sections 3 and 4 we construct a measure relative to which the Scheffé-type bound over a finite interval is optimal. The fact that the measure is approximately unimodal indicates why the modification of the bound given in Section 5 leads to an improvement in average width with respect to Lebesgue measure.

2. A general convexity result. Let X be an arbitrary subset of R^k and let ϕ be a (one-sided) bound over X. Recall that F denotes the distribution function for $\|\mathbf{B}\|$, and F is unimodal with mode m_0. Thus F is convex on $(-\infty, m_0]$, and concave on $[m_0, \infty)$. We let U denote the unit sphere in R^k.

We use $p(\phi)$ to denote the coverage probability (1.5) of ϕ. ϕ determines a convex set containing the origin in R^k,

$$A(\phi) = \{\mathbf{b} \in R^k : \mathbf{x}'\mathbf{b} \le \phi(\mathbf{x}), \text{ all } \mathbf{x} \in X\},$$

so that the coverage probability may be expressed as $p(\phi) = P[\mathbf{B} \in A(\phi)]$. For $\mathbf{u} \in U$, let $R_\phi(\mathbf{u})$ denote the distance to the boundary of $A(\phi)$ from the origin in the direction $\mathbf{u}$, so that

$$R_\phi(\mathbf{u}) = \inf\{\phi(\mathbf{x})/\mathbf{x}'\mathbf{u} : \mathbf{x} \in X, \mathbf{x}'\mathbf{u} > 0\},$$

where the infimum of the empty set is defined to be $+\infty$.

The following lemma is implicit in Bohrer (1973) and gives a convenient expression for the coverage probability of a general bound ϕ.

LEMMA 2.1. *The coverage probability $p(\phi)$ is given by*

$$E\{F[R_\phi(\mathbf{B}/\|\mathbf{B}\|)]\}. \tag{2.1}$$

PROOF. For an arbitrary bound ϕ,

$$p(\phi) = P\{\mathbf{B} \in A(\phi)\} = E\{P[\mathbf{B} \in A(\phi) \mid \mathbf{B}/\|\mathbf{B}\|]\},$$
$$= E\{P[\|\mathbf{B}\| \le R_\phi(\mathbf{B}/\|\mathbf{B}) \mid \mathbf{B}/\|\mathbf{B}\|]\},$$
$$= E\{F[R_\phi(\mathbf{B}/\|\mathbf{B}\|)]\},$$

where the last equality uses the independence of $\mathbf{B}$ and $\mathbf{B}/\|\mathbf{B}\|$. □

Let C be the set of bounds ϕ such that $R_\phi(\mathbf{u}) \ge m_0$ for every $\mathbf{u} \in U$. The main results of this paper give conditions for optimality of bounds in C. However, the following lemma shows that C consists of all bounds with sufficiently large coverage probability. Thus, optimality among bounds in C implies optimality among all bounds.

REMARK 2.1. By the Cauchy-Schwartz inequality, $\phi \in C$ if and only if $\phi(\mathbf{x}) \ge m_0 \|\mathbf{x}\|$ for every $\mathbf{x} \in X$.

LEMMA 2.2.

(a) *C contains all bounds with coverage probability greater than $1 - \alpha$ (i.e., $\{\phi: p(\phi) > 1 - \alpha\} \subseteq C$) if and only if $P(t_\nu \le m_0) \le 1 - \alpha$, where $\nu = n - k$, t_ν denotes a random variable with a t-distribution with ν degrees of freedom, and $m_0 = \{(k-1)\nu/(\nu+1)\}^{1/2}$.*

(b) *$\phi \notin C$ if $p(\phi) < F(m_0)$.*

PROOF. For (a), suppose $P(t_\nu \le m_0) \le 1 - \alpha$. If $\phi \notin C$ there exists $\mathbf{x}_0 \in X$ such that $\phi(\mathbf{x}_0) < m_0 \|\mathbf{x}_0\|$, by Remark 2.1. Using the fact that $\mathbf{B}$ has a spherically symmetric distribution and B_1, the first coordinate function of B, has a t-distribution with ν degrees of freedom, we obtain

$$p(\phi) = P(\mathbf{x}'\mathbf{B} \le \phi(\mathbf{x}), \text{ all } \mathbf{x} \in X)$$
$$\le P(\mathbf{x}_0'\mathbf{B} \le m_0 \|\mathbf{x}_0\|) = P(B_1 \le m_0) = P(t_\nu \le m_0) \le 1 - \alpha.$$

Thus, $p(\phi) > 1 - \alpha$ implies $\phi \in C$.

Conversely, if $P(t_\nu \le m_0) > 1 - \alpha$, let $\delta > 0$ be such that $P(t_\nu \le m_0 - \delta) > 1 - \alpha$. For any fixed $\mathbf{x}_0 \in X$ and $K > 0$, we can construct a bound ϕ_K as follows:

$$\phi_K(\mathbf{x}_0) = (m_0 - \delta)\|\mathbf{x}_0\|, \quad \phi_K(\mathbf{x}) = K, \quad \text{for} \quad \mathbf{x} \ne \mathbf{x}_0.$$

It follows that

$$\lim_{K \to \infty} p(\phi_K) = P(t_\nu \le m_0 - \delta) > 1 - \alpha,$$

so for K sufficiently large, $p(\phi_K) > 1 - \alpha$ and ϕ_K is not in C.

(b) follows immediately from Lemma 2.1. □

For two-sided bounds, (a) holds with $P(t_\nu \le m_0)$ replaced by $P(|t_\nu| \le m_0)$ and (b) holds as stated. Table 1 gives $P(t_\nu \le m_0)$, $P(|t_\nu| \le m_0)$ and $F(m_0)$ for various values of ν and k. For any value of the coverage probability smaller than the

TABLE 1
$P(t_\nu \le m_0)$, $P(|t_\nu| \le m_0)$, and $F(m_0)$ (see Lemma 2.2).

k	$P(t_\nu \le m_0)$	$P(\|t_\nu\| \le m_0)$	$F(m_0)$	$P(t_\nu \le m_0)$	$P(\|t_\nu\| \le m_0)$	$F(m_0)$
		$\nu = 1$			$\nu = \infty$	
1	.500	.000	.000	.500	.000	.000
2	.696	.392	.184	.841	.683	.393
3	.750	.500	.182	.921	.843	.428
4	.782	.564	.178	.958	.917	.442
5	.804	.608	.175	.977	.955	.451
6	.820	.641	.172	.987	.974	.456
7	.833	.667	.170	.993	.986	.460
8	.844	.687	.169	.996	.992	.463
9	.852	.705	.168	.998	.995	.466
10	.860	.720	.167	.999	.997	.468

value in the column $P(t_\nu \le m_0)$, C does not contain all bounds. Although this appears somewhat discouraging for large values of k and ν, we should note that the class C does contain the level $1 - \alpha$ Scheffé bound (when $X = R^k$), and hence nearby bounds, for $\alpha \le F(m_0)$.

LEMMA 2.3.

(a) *$R_\phi(\mathbf{u})$ is concave in ϕ for any $\mathbf{u} \in U$.*
(b) *The set C is convex.*

PROOF. Concavity of $R_\phi(\mathbf{u})$ follows immediately from the definition and (b) follows from (a). □

THEOREM 2.1. *The function p is concave in ϕ on the set C. Consequently, for any $\alpha \in (0, 1)$, the set of bounds $\phi \in C$ such that $p(\phi) \ge 1 - \alpha$ is a convex set.*

PROOF. Let ϕ and ψ be any two bounds in C and fix $\lambda \in [0, 1]$. Set $\rho = \lambda\phi + (1 - \lambda)\psi$. Lemma 2.3 gives

$$R_\rho(\mathbf{u}) \ge \lambda R_\phi(\mathbf{u}) + (1 - \lambda)R_\psi(\mathbf{u})$$

for any $\mathbf{u} \in U$. Using the monotonicity of F, the fact that ϕ and ψ are in C, and the concavity of F on $[m_0, \infty)$, we obtain

$$\begin{aligned} F[R_\rho(\mathbf{B}/\|\mathbf{B}\|)] &\ge F[\lambda R_\phi(\mathbf{B}/\|\mathbf{B}\|) + (1 - \lambda)R_\psi(\mathbf{B}/\|\mathbf{B}\|)] \\ &\ge \lambda F[R_\phi(\mathbf{B}/\|\mathbf{B}\|)] + (1 - \lambda)F[R_\psi(\mathbf{B}/\|\mathbf{B}\|)]. \end{aligned}$$

The proof is completed by taking expectations and using Lemma 2.1. □

From the point of view of optimality, a minimal requirement that any bound should satisfy is that of tautness, defined as follows.

DEFINITION 2.1. A function ϕ on X defines a taut set of inequalities $\{\mathbf{x}'\mathbf{b} \le \phi(\mathbf{x})\}_{\mathbf{x}\in X}$ if whenever another function ψ on X satisfies $\psi \le \phi$, with strict inequality holding for some $\mathbf{x} \in X$, there exists a solution $\mathbf{b}$ to $\{\mathbf{x}'\mathbf{b} \le \phi(\mathbf{x})\}_{\mathbf{x}\in X}$ which is not a solution to $\{\mathbf{x}'\mathbf{b} \le \psi(\mathbf{x})\}_{\mathbf{x}\in X}$.

DEFINITION 2.2. (Wynn and Bloomfield, 1971). *A bound ϕ is taut if ϕ defines a taut set of inequalities $\{\mathbf{x}'\mathbf{b} \le \phi(\mathbf{x})\}_{\mathbf{x}\in X}$.*

If a bound ϕ is not taut then there exists a bound with the same coverage probability but which defines intervals which are at least as accurate as the ones defined by ϕ, with at least one interval being strictly more accurate. For this reason we restrict our attention to taut bounds for the remainder of this paper.

Clearly, a bound ϕ over X is taut if and only if the hyperplane $\{\mathbf{x}'\mathbf{b} = \phi(\mathbf{x})\}$ is a support hyperplane for $A(\phi)$ for each $\mathbf{x} \in X$. For a given bound ϕ over X, define a bound ϕ^* over X by

$$\phi^*(\mathbf{x}) = \sup\{\mathbf{x}'\mathbf{y} : \mathbf{y} \in A(\phi)\}.$$

It is easily verified that the hyperplane $\{\mathbf{x}'\mathbf{b} = \phi^*(\mathbf{x})\}$ is a support hyperplane to $A(\phi)$, so that $A(\phi^*) = A(\phi)$. It follows that ϕ is taut if and only if $\phi^*(\mathbf{x}) = \phi(\mathbf{x})$ for every $\mathbf{x} \in X$.

We now give a characterization of taut bounds.

LEMMA 2.4. *Let X be any subset of R^k and let ϕ be any bound over X. If ϕ is taut then ϕ is convex and positively homogeneous on X. Conversely, if ϕ is convex and positively homogeneous on X and $\{(\mathbf{x}, \phi(\mathbf{x})) : \mathbf{x} \in X\}$ is a closed and bounded subset of R^{k+1},* then ϕ is taut.

PROOF. If ϕ is taut then $\phi^* = \phi$ and clearly ϕ^* is convex and positively homogeneous on X. The converse follows immediately from Rockafellar (1970) Theorem 17.3. □

REMARK 2.2. The following example shows that convexity and positive homogeneity are not sufficient conditions for tautness. Let $X = \{(1, x)' : 0 \le x \le 1\}$ and define $\phi((1, x)') = 1 + x^2$ for $0 < x \le 1$, and $\phi(0) = 2$. Then ϕ is convex and positively homogeneous on X but ϕ is not taut since we can take $\psi((1, x)') = 1 + x^2$ for $0 \le x \le 1$, so that $A(\psi) = A(\phi)$.

We define the equivalence relation $\sim$ on R^k by $\mathbf{x}_1 \sim \mathbf{x}_2$ if $\mathbf{x}_1 = c\mathbf{x}_2$ for some $c > 0$. Thus ϕ is positively homogeneous if and only if $\phi(\mathbf{x})/\|\mathbf{x}\|$ is constant on equivalence classes of X and $\phi(\mathbf{0}) = 0$ if $\mathbf{0} \in X$.

THEOREM 2.2. *If X is a finite set and ϕ is a taut bound over X which is in C and satisfies $p(\phi) > 0$, then we have the following.*

(a) *ϕ is μ-optimal in C for some probability measure μ on X.*

(b) *Let $X_1, \cdots, X_m$ be the equivalence classes of $X - \{\mathbf{0}\}$ under $\sim$. If μ' and μ are probability measures on X satisfying*

$$\sum_{\mathbf{x}\in X_i} \mu'(\{\mathbf{x}\})\,\|\mathbf{x}\| = \sum_{\mathbf{x}\in X_i} \mu(\{\mathbf{x}\})\,\|\mathbf{x}\|, \quad \text{for} \quad i = 1, \cdots, m, \tag{2.2}$$

and ϕ is μ-optimal in C, then ϕ is μ'-optimal in C.

(c) *If ϕ is in the interior of C, i.e., if $\phi(\mathbf{x}) > m_0 \|\mathbf{x}\|$, for all $\mathbf{x} \in X$, then for any probability measure μ relative to which ϕ is optimal, $\sum_{\mathbf{x}\in X_i} \mu(\{\mathbf{x}\}) \|\mathbf{x}\|$ is determined uniquely for $i = 1, \cdots, m$. In particular, if ϕ is in the interior of C and X contains at most one point in each ~-equivalence class then the probability measure in* (a) *is unique.*

PROOF. Fix a taut bound ϕ in C satisfying $p(\phi) > 0$. For (a) it suffices to show that there exists a finite measure μ on X such that ϕ is μ-optimal among all positively homogeneous bounds in C.

Let $X_1, \cdots, X_m$ denote the equivalence classes of $X - \{\mathbf{0}\}$ under $\sim$ and let $\mathbf{x}_i \in X_i$ for $i = 1, \cdots, m$. Let $V = \{\mathbf{v} \in R^k\colon v_i \geq m_0,\ i = 1, \cdots, m\}$. We can define a one-to-one correspondence between positively homogeneous bounds $\psi \in C$ and m-vectors $\mathbf{v} = \mathbf{v}(\psi) = (v_1, \cdots, v_m)' \in V$ by letting $v_i = \psi(\mathbf{x}_i)/\|\mathbf{x}_i\|$, for $i = 1, \cdots, m$. For $\mathbf{v}$ and ψ related in this way define $G(\mathbf{v}) = p(\psi)$ so that $G(\mathbf{v}) = P(\mathbf{x}_i'\mathbf{B} \leq v_i\|\mathbf{x}_i\|,\ i = 1, \cdots, m)$. Clearly G is differentiable.

For any measure μ on X concentrated on $\{\mathbf{x}_1, \cdots, \mathbf{x}_m\}$, if $\mathbf{v} = \mathbf{v}(\psi)$ then the average width of ψ under μ is

$$\int_X \psi(\mathbf{x})\mu(d\mathbf{x}) = \textstyle\sum_{i=1}^m v_i\mu(\{\mathbf{x}_i\})\|\mathbf{x}_i\|.$$

It follows that ϕ is μ-optimal if and only if $\mathbf{v} = \mathbf{v}(\phi)$ minimizes $\tau'\mathbf{v}$, among all $\mathbf{v} \in V$ satisfying $G(\mathbf{v}) \geq p(\phi)$, where

$$\tau_i = \mu(\{\mathbf{x}_i\})\|\mathbf{x}_i\|, \quad \text{for} \quad i = 1, \cdots, m. \tag{2.3}$$

Define $\tau = \nabla G(\mathbf{v})|_{\mathbf{v}=\mathbf{v}(\phi)}$. Since $G(\mathbf{v})$ is nondecreasing in each component v_i, τ has nonnegative components and defines a measure μ on X concentrated on $\{\mathbf{x}_1, \cdots, \mathbf{x}_m\}$ via (2.3). τ is nonzero since $p(\phi) > 0$ and the density function for $\mathbf{B}$ is nonvanishing.

We proceed to show ϕ is μ-optimal in C. Let $\mathbf{v} = \mathbf{v}(\phi)$ and suppose $\mathbf{w}$ in V has smaller average width than $\mathbf{v}$ (i.e., $\tau'\mathbf{w} < \tau'\mathbf{v}$). It suffices to show that $\mathbf{w}$ has smaller coverage probability (i.e., $G(\mathbf{w}) < G(\mathbf{v})$).

Since V is convex, $\mathbf{v} + h(\mathbf{w} - \mathbf{v}) \in V$ for all $0 \leq h \leq 1$. Since G is concave on V, it suffices to show $G(\mathbf{v} + h(\mathbf{w} - \mathbf{v})) < G(\mathbf{v})$ for all h sufficiently small and positive. This follows from the Taylor expansion

$$G(\mathbf{v} + h(\mathbf{w} - \mathbf{v})) = G(\mathbf{v}) + h\tau'(\mathbf{w} - \mathbf{v}) + o(h)$$

as $h \to 0^+$, so the proof of (a) is complete.

(b) follows from the fact that if ψ is positively homogeneous and μ and μ' satisfy (2.2) then we have

$$\int_X \psi(\mathbf{x})\mu'(d\mathbf{x}) = \int_X \psi(\mathbf{x})\mu(d\mathbf{x}).$$

To prove (c), assume $\mathbf{v}$ is in the interior of V and minimizes $\eta'\mathbf{v}$ among all $\mathbf{v}$ in V satisfying $G(\mathbf{v}) \geq p(\phi)$. We proceed to show that η is a nonnegative multiple of $\tau = \nabla \mathbf{G}(\mathbf{v})\,|_{\mathbf{v}=\mathbf{v}(\phi)}$. Suppose $\mathbf{w}$ is any m-vector such that $\tau'\mathbf{w} > 0$. By Taylor expansion,

$$G(\mathbf{v} + h\mathbf{w}) = G(\mathbf{v}) + h\tau'\mathbf{w} + o(h).$$

For all $h > 0$ sufficiently small it follows that $G(\mathbf{v} + h\mathbf{w}) > G(\mathbf{v})$ and $\mathbf{v} + h\mathbf{w} \in V$, hence $\eta'(\mathbf{v} + h\mathbf{w}) \geq \eta'\mathbf{v}$, i.e., $\eta'\mathbf{w} \geq 0$. Thus $\tau'\mathbf{w} > 0$ implies $\eta'\mathbf{w} \geq 0$. By Lemma 3 in Naiman (1984) it follows that η is a nonnegative multiple of τ. □

REMARK 2.3. The following description of the measure μ will be used in Section 3. Assume X contains at most one point in any direction $\mathbf{u} \in U$. For any $\mathbf{x} \in X$, the hyperplane $H = \{\mathbf{b}\colon \mathbf{x}'\mathbf{b} = \phi(\mathbf{x})\}$ is a support hyperplane to the region $A(\phi)$. Let F denote the face formed by the intersecting H with $A(\phi)$. Then $\mu(\{\mathbf{x}\}) = \int_F g(\mathbf{x})m(d\mathbf{x})/\|\mathbf{x}\|$, where m denotes $k-1$ dimensional Lebesgue measure on H, and g is the multivariate t-density.

3. Simple linear regression over a finite interval. In the remainder of this paper we give results for bounding a simple linear regression function with intercept $y = (1, x)'\beta$, for x in a finite closed interval. In this situation we can assume that the average of the design points is 0, since we can, if necessary, redefine the regressors by subtracting $\bar{x}$. We take $X = \{(1, x)'\colon a \leq x \leq b\}$.

The main result of this section concerns the μ-optimality of taut bounds. For any taut bound ϕ over X and for any finite subset X_0 of X, the restriction ϕ_0 of ϕ to X_0 is a taut bound over X_0. In Theorem 2.2 we constructed a measure μ_0 on X_0 such that ϕ_0 is μ_0-optimal.

Now suppose $\{(1, x_i)\}_{i=1}^{\infty}$ is a dense sequence of points in X and define $X_n = \{(1, x_1), \cdots, (1, x_n)\}$. Let μ_n be the measure from Theorem 2.2 relative to which ϕ_n, the restriction of ϕ to X_n, is optimal. μ_n defines a measure on X in the obvious way. In Theorem 3.1 we prove that μ_n converges weakly to a measure μ (not depending on the sequence $\{x_n\}$), and the main result of this section, Theorem 3.2, states that ϕ is μ-optimal if ϕ is continuous.

We introduce some notation. Fix a taut bound ϕ over X. For any subset Y of (a, b), let ϕ_Y denote the bound defined by restricting ϕ to $X_Y = \{(1, y)'\colon y \in Y\}$. Define the convex set $A_Y = \{\mathbf{b}\colon (1, y)\mathbf{b} \leq \phi((1, y)'),\text{ all } y \in Y\}$, so that $p(\phi_Y) = P(\mathbf{B} \in A_Y)$. Define

$$H_Y(t) = -\sup\{s\colon (s, t)' \in A_Y\},$$

for $t \in R$, so that

$$A_Y = \{(s, t)'\colon s \leq -H_Y(t)\}.$$

This description is possible because of the special form of X_Y. It is easy to see that H_Y is a convex function.

The following facts are consequences of the convexity of H_Y on closed intervals J. For a proof see Rockafellar (1970), Theorem 24.1. H_Y is absolutely continuous on J. The left- and right-hand derivatives D^-H and D^+H exist at each point of J and are equal to each other except on a countable set. The funtions D^-H and D^+H are monotone nondecreasing, D^+H is right continuous, D^-H is left continuous, and at each point $D^-H \leq D^+H$.

Let $\{y_n\}$ be a dense sequence of distinct points in (a, b), and take Y to be $Y_n = \{y_1, \cdots, y_n\}$. We use X_n to denote X_Y, ϕ_n to denote ϕ_Y, A_n to denote A_Y, and H_n to denote H_Y. We also let H denote $H_{(a,b)}$ and we let A denote $A_{(a,b)}$.

In Theorem 2.2 we constructed a measure μ_n on X_n relative to which ϕ_n is optimal. This measure can be described as follows. For each i, the face $F_{i,n}$ of A_n corresponding to the support line $\{(1, y_i)\mathbf{b} = \phi((1, y_i)')\}$ is a line segment in R^2 of the form $\{(-H_n(t), t): t_{i-1,n} \le t \le t_{i,n}\}$. Let $I_{i,n} = [t_{i-1,n}, t_{i,n})$ for $i = 1, \cdots, n$, so that the intervals $I_{i,n}$ partition R. The measure μ_n is given by

$$\mu_n\{(1, y_i)'\} = \int_{I_{i,n}} g(-H_n(t), t)\, dt$$

where g denotes the density function of $\mathbf{B}$. Furthermore, if t is in the interior of the interval $I_{i,n}$ then $y_i = DH_n(t)$. It follows that

$$\mu_n = \nu_n \circ D^+H_n^{-1} \circ \pi^{-1}, \tag{3.1}$$

where ν_n is the measure on R with density function $g(-H_n(t), t)$, and π is the mapping from $[a, b]$ to X defined by $\pi(x) = (1, x)'$.

We now define the limiting measure

$$\mu = \nu \circ D^+H^{-1} \circ \pi^{-1}, \tag{3.2}$$

where ν is the measure on R with density $g(-H(t), t)$ and $H = H_{(a,b)}$.

The following result plays an important role in the proof of Theorem 3.2, which is the main result of this section. A sketch of the proof appears in Appendix 1.

THEOREM 3.1. *μ_n converges weakly to μ as $n \to \infty$.*

LEMMA 3.1. *If ψ is any bound over X then $p(\psi + \delta) > p(\psi)$ for any $\delta > 0$.*

PROOF. Let $r = \inf\{\psi(\mathbf{x}): \mathbf{x} \in X\}$ and define

$$S = \{\mathbf{v} \in R^2: r + \delta/2 < \mathbf{x}'\mathbf{v} < r + \delta, \text{ for } \mathbf{x} = (1, a)', (1, b)'\}.$$

Fix $\mathbf{v} \in S$. If $\mathbf{x} = (1 - \lambda)(1, a)' + \lambda(1, b)'$ and $\lambda \in [0, 1]$, then $\mathbf{x}'\mathbf{v} < r + \delta \le \psi(\mathbf{x}) + \delta$, so $\mathbf{v} \in A(\psi + \delta)$. Furthermore, there exists $\mathbf{x}_0 \in X$ such that $\psi(\mathbf{x}_0) < r + \delta/2$. Since $\mathbf{x}_0 = (1 - \lambda)(1, a)' + \lambda(1, b)'$ for some $\lambda \in [0, 1]$, it follows that $\mathbf{x}_0'\mathbf{v} > r + \delta/2 > \psi(\mathbf{x}_0)$, so $v \notin A(\psi)$. Thus, $S \subseteq A(\psi + \delta) - A(\psi)$. Clearly, $A(\psi) \subseteq A(\psi + \delta)$ and it is easy to verify that $P(\mathbf{B} \in S) > 0$, so the result follows. □

LEMMA 3.2. *If ψ is any continuous bound over X then $p(\psi \mid_{X_n}) \to p(\psi)$ as $n \to \infty$.*

PROOF. Let ψ_n denote the restriction of ψ to X_n. For any $\mathbf{u} \in U$ the minimum of $\psi(\mathbf{x})/\mathbf{x}'\mathbf{u}$ is attained in the interior of $\{\mathbf{x} \in X: \mathbf{x}'\mathbf{u} > 0\}$ since ψ being bounded and continuous on X makes the ratio infinite on the boundary. It follows that $R_{\psi_n}(\mathbf{u}) \to R_\psi(\mathbf{u})$ as $n \to \infty$. The result follows from Lemma 2.1 and Lebesgue's dominated convergence theorem. □

THEOREM 3.2. *Let ϕ be a taut continuous bound over X in C and let μ be the measure defined by* (3.2). *Then ϕ is μ-optimal in C.*

PROOF. Suppose ϕ is not μ-optimal so that for some $\psi \in C$ and for some $\delta > 0$, $p(\psi) \geq p(\phi)$ and

$$\int_X \psi \, d\mu < \int_X \phi \, d\mu - \delta\mu(X).$$

We can assume ψ is taut since otherwise we can replace ψ by ψ^* and the above statements still hold. It follows from Lemma 2.4 that ψ is a convex function so that ψ is bounded, and by Theorem 10.2 in Rockafellar (1970), ψ is upper semicontinuous.

Since μ_n converges weakly to μ and ϕ is continuous and bounded on X, it follows that

$$\int_X \phi \, d\mu_n \to \int_X \phi \, d\mu \quad \text{as} \quad n \to \infty,$$

and using the upper semicontinuity and boundedness of $\psi + \delta$ on X,

$$\limsup_{n\to\infty} \int_X (\psi + \delta) \, d\mu_n \leq \int_X (\psi + \delta) \, d\mu < \int_X \phi \, d\mu.$$

This follows from Ash (1970) Theorem 4.5.1(b′). It follows that there exists N such that for all $n \geq N$

$$\int_X (\psi + \delta) \, d\mu_n \leq \int_X \phi \, d\mu_n.$$

We use ϕ_n to denote $\phi \,|_{X_n}$ and we use ψ_n to denote $(\psi + \delta)\,|_{X_n}$. It is easy to verify that $R_{\phi_n} \geq R_\phi$ so that $\phi_n \in C$ for every n. Since the measure μ_n is concentrated on X_n and ϕ_n is μ_n-optimal, $p(\phi_n) \geq p(\psi_n)$ for all $n \geq N$.

We have $p(\psi + \delta) \leq p(\psi_n) \leq p(\phi_n)$ for $n \geq N$, and by Lemma 3.2, $p(\phi_n) \to p(\phi)$ as $n \to \infty$, so it follows that

$$p(\psi + \delta) \leq p(\phi) \leq p(\psi).$$

This contradicts Lemma 3.1 and completes the proof. □

Thus far we have considered optimality among bounds whose width is finite at every point in X. It is reasonable to ask whether a given bound ϕ is optimal among bounds whose width can be infinite at certain points. Corollary 3.1 states that for many bounds this turns out to be the case. We let C^+ denote the set of measurable functions on X taking values in $[m_0, \infty]$. We continue to refer to bounds by the functions which define them.

COROLLARY 3.1. *For a given bound ϕ as in the statement of Theorem* 3.1 *define $F_a = A(\phi) \cap \{\mathbf{b}: (1, a)\mathbf{b} = \phi((1, a)')\}$ and*

$$F_b = A(\phi) \cap \{\mathbf{b}: (1, b)\mathbf{b} = \phi((1, b)')\}.$$

If either of the sets F_a or F_b consists of more than one point then

$$\mu\{(1, a)'\} + \mu\{(1, b)'\} > 0,$$

and ϕ is μ-optimal in C^+.

PROOF. The statement that $\mu\{(1, a)'\} + \mu\{(1, b)'\} > 0$ follows immediately from the definition of μ.

Suppose ϕ is suboptimal in C^+. There exists ψ in C^+ such that $p(\psi) \geq p(\phi)$ and

$$\int_X \psi \, d\mu < \int_X \phi \, d\mu < +\infty. \tag{3.3}$$

As in the proof of Theorem 3.2, we can assume ψ is taut and hence convex by Lemma 2.4. Note that the proof of Lemma 2.4 is still valid for bounds taking values in $[0, +\infty]$. If $\psi(\mathbf{x})$ is infinite for some $\mathbf{x}$ in X, then by convexity $\psi((1, a)') = +\infty$ or $\psi((1, b)') = +\infty$, but this implies $\int_X \psi \, d\mu = +\infty$, which contradicts (3.3), so ϕ must be in C and this contradicts Theorem 3.2. □

4. Examples. In this section we compute the measure μ given in (3.2), for some examples of bounds that have appeared in the literature for simple linear regression functions over a finite interval.

(i) *Scheffé-type bounds.* Bohrer and Francis (1972) introduced a one-sided analogue to the Working–Hotelling (1929) bound which is of the form $\phi((1, x)') \propto \{1 + x^2\}^{1/2}$. For this bound

$$H(t) = \begin{cases} \lambda(t_1) + \lambda'(t_1)(t - t_1) & t < t_1 \\ \lambda(t) & t_1 \leq t \leq t_2 \\ \lambda(t_2) + \lambda'(t_2)(t - t_2) & t > t_2, \end{cases}$$

where $\lambda(t) = -\{R^2 - t^2\}^{1/2}$, for some positive constant R, $t_1 = Ra\{1 + a^2\}^{-1/2}$, and $t_2 = Rb\{1 + b^2\}^{-1/2}$. See Bohrer and Francis (1972).

Since the distribution of $\mathbf{B}$ is symmetric about $\mathbf{0}$, $g(-H(t), t)$ is constant on $[t_1, t_2]$, so that the density function for ν (see (3.2)) is of the form

$$h(t) = \begin{cases} h_1(t) & t \leq t_1 \\ C & t_1 \leq t \leq t_2 \\ h_2(t) & t \geq t_2, \end{cases}$$

for some constant $C > 0$ and positive functions h_1 and h_2. Also

$$DH(t) = \begin{cases} a & t \leq t_1 \\ t/\{R^2 - t^2\}^{1/2} & t_1 \leq t \leq t_2 \\ b & t \geq t_2. \end{cases}$$

Since DH is constant on the intervals $(-\infty, t_1)$ and $(t_2, +\infty)$, the functional form of the h_i is not particularly important for determining the form of μ. It follows that μ assigns some positive measure to $\{a\}$ and to $\{b\}$, and on (a, b), μ is absolutely continuous with respect to Lebesgue measure, with density function proportional to $(1 + t^2)^{-3/2}$.

As $a \rightarrow -\infty$ and $b \rightarrow +\infty$, μ converges weakly to a distribution whose density is proportional to $(1 + t^2)^{-3/2}$. Hoel (1951) gave a heuristic argument for the optimality of the Working–Hotelling (1929) bound over R with respect to this measure, and the result may be obtained by an argument analogous to the one given in Bohrer (1973).

(ii) *Trapezoidal bounds.* Bowden and Graybill (1966) introduced two-sided bounds whose width is a linear function. These bounds generalize the uniform width bounds of Gafarian (1964). For such bounds $H(t)$ is a continuous function which is linear on $(-\infty, t_0)$ and $(t_0, +\infty)$ for some t_0, and DH takes only two values, a and b. Thus μ puts all of its mass at the endpoints of the interval. In the balanced case, when the center of the interval is the average of the design points, the measure for the Gafarian bound puts equal mass at each endpoint of the interval, but this needn't be true in the general case.

More generally, if the bound ϕ is continuous and piecewise linear, then μ is concentrated on the at most countably many points where ϕ' is discontinuous.

5. Suboptimality of Scheffé-type bounds. In this section we state a result giving the m-suboptimality of Scheffé-type bounds for simple linear regression functions over sufficiently large finite intervals, where m is Lebesgue measure. We obtain the result by constructing an improvement to the Scheffé-type bound.

It is assumed that we have a balanced design, i.e. the design matrix is such that the average of its columns is the center of the interval for bounding the regression function. After centering of the regressors about their mean, we can assume without loss of generality that the region X is given by $X_v = \{(1, x)': |x| \le v\}$, for some $v > 0$.

Consider the class of bounds over X_v of the form

$$\phi_{c,u,v}((1, x)') = \begin{cases} c\{1 + x^2\}^{1/2} & |x| \le u \\ c\{1 + u^2\}^{1/2} + cu\,|x - u|/\{1 + u^2\}^{1/2} & u \le |x| \le v, \end{cases}$$

where $0 \le u \le v$, and $c \ge 0$. This bound has the hyperbolic shape of the Scheffé-type bound on the subinterval $\{|x| \le u\}$, and is extended linearly outside this interval, so that ϕ is differentiable on X_v. Note that ϕ is constant for $u = 0$ and ϕ is a Scheffé-type bound for $u = v$. We can think of this new class of bounds as bridging a gap between the Scheffé-type and the constant width bounds.

The main result of this section is the following theorem. A proof is available from the author.

THEOREM 5.1. *For any fixed $u > 0$ and $c' > 0$, let $c(v)$ be the constant such that the Scheffé-type bound $\phi_{c(v),v,v}$ has coverage probability equal to that of the bound $\phi_{c',u,v}$ for each $v > u$, thus $p(\phi_{c(v),v,v}) = p(\phi_{c',u,v})$. Then there exists v' such that for all $v \ge v'$, the bound $\phi_{c',u,v}$ has smaller average width with respect to Lebesgue measure on X_v than has $\phi_{c(v),v,v}$.* □

The improvement to the Scheffé-type bound is hyperbolic in shape on a subinterval and becomes linear outside of the subinterval. Thus, the width of the

Scheffé-type bound grows too quickly toward the boundary of the interval. On the other hand, the constant width bound is improved by making the width grow by some degree.

We could have anticipated such a result since the Scheffé-type bound is optimal with respect to a measure which has approximate central tendency on the interval. In trying to optimize average width with respect to Lebesgue measure instead of a measure with central tendency, we are asking for more accuracy near the endpoints of the interval, and less at the center.

On the other hand, we may be interested in finding a bound which is optimal with respect to a truncated normal distribution. In this case, the above result suggests that for sufficiently large intervals we would obtain a smaller average width by using a bound which is less accurate than the Scheffé-type bound at the endpoints of the interval. Naiman (1982, Section 4.3) gives a class of bounds that may be useful for this purpose.

6. Acknowledgements. Many thanks are due to one of the referees and the Associate Editor, whose detailed and insightful comments led to a much improved version of the original paper.

APPENDIX 1

Sketch of the proof of Theorem 3.1. The proof of Theorem 3.1 uses a sequence of lemmas (A.1–A.8), which are stated below, and whose proofs are available from the author.

LEMMA A.1.

(i) *For each n there exist points $T_{i,n}$ for $i = 1, \cdots, n$ such that*
   (a) $(-H(T_{i,n}), T_{i,n}) \in F_{i,n}$
   (b) $t_{i-1,n} \leq T_{i,n} \leq t_{i,n}$
   (c) $H_n(T_{i,n}) = H(T_{i,n})$
   (d) $D^-H(T_{i,n}) \leq D^-H_n(T_{i,n}) \leq D^+H_n(T_{i,n}) \leq D^+H(T_{i,n})$.

(ii) *The sets $\{T_{i,n}, i = 1, \cdots, n\}$ may be chosen so that for some sequence of points $\{T_n\}$ we have $\{T_{i,n}, i = 1, \cdots, n\} = \{T_i, i = 1, \cdots, n\}$, for every n.*

DEFINITION A.1. A point t is a point of increase of DH if for every $u < t < v$ we have $D^+H(u) < D^-H(v)$. Note that we do not require that DH exist at t. We use Q to denote the set of points of increase of DH.

LEMMA A.2. *Any sequence $\{T_n\}$ satisfying the conditions of Lemma* A.1 *is dense in Q.*

LEMMA A.3. *If $t \in Q$ then $H_n(t) \to H(t)$ as $n \to \infty$.*

LEMMA A.4. *If $t \in Q$ then* $\lim \inf_{n\to\infty} D^-H_n(t) \geq D^-H(t)$ *and* $\lim \sup_{n\to\infty} D^+H_n(t) \leq D^+H(t)$. *Furthermore, if $t \in Q$ and $DH(t)$ exists then* $\lim \inf_{n\to\infty} D^-H_n(t) = \lim \sup_{n\to\infty} D^+H_n(t) = DH(t)$.

LEMMA A.5. *$H_n(t) \to H(t)$ as $n \to \infty$ for every t.*

LEMMA A.6. *$DH_n(t) \to DH(t)$ if $DH_n(t)$ exists for all n and $DH(t)$ exists.*

LEMMA A.7. *ν_n converges weakly to ν as $n \to \infty$.*

LEMMA A.8. *For any $u \in R$ if $m\{t: D^+H(t) = u\} = 0$ then*

$$m(\{t: D^+H(t) \le u\} \,\Delta\, \{t: D^+H_n(t) \le u\}) \to 0 \quad \text{as} \quad n \to \infty,$$

where m denotes Lebesgue measure.

PROOF OF THEOREM 3.1. The fact that $\mu_n(X) \to \mu(X)$ as $n \to \infty$ follows from Lemma A.7. Fix $u \in [a, b]$ such that $\mu(\{u\}) = 0$. It suffices to show

$$\nu_n \circ D^+H_n^{-1}[a, u] \to \nu \circ D^+H^{-1}[a, u] \quad \text{as} \quad n \to \infty.$$

Since g is nonvanishing $m\{t: D^+H(t) = u\} = 0$, and

$$m(\{t: D^+H(t) \le u\} \,\Delta\, \{t: D^+H_n(t) \le u\}) \to 0 \quad \text{as} \quad n \to \infty$$

by Lemma A.8. It follows that

$$\nu_n \circ D^+H_n^{-1}[a, u] - \nu_n \circ D^+H^{-1}[a, u] \to 0 \quad \text{as} \quad n \to \infty,$$

by Lebesgue's dominated convergence theorem. Thus

$$\begin{aligned} \nu_n \circ D^+H_n^{-1}[a, u] &- \nu \circ D^+H^{-1}[a, u] \\ &= \{\nu_n \circ D^+H_n^{-1}[a, u] - \nu_n \circ D^+H^{-1}[a, u]\} \\ &\quad + \{\nu_n \circ D^+H^{-1}[a, u] - \nu \circ D^+H^{-1}[a, u]\} \to 0 \quad \text{as} \quad n \to \infty, \end{aligned}$$

by Lemma A.7, and the proof is complete. □

APPENDIX 2

Results for two-sided bounds. We summarize the modifications necessary for results for two-sided bounds corresponding to the results for one-sided bounds given in Sections 2–5.

For subsets X of R^k over which two-sided bounds are defined, we will always require that $-\mathbf{x} \notin X$ whenever $\mathbf{x} \in X$. This is not an unreasonable assumption since for given $\mathbf{x} \in X$, the bounds (1.2) for $\mathbf{x}'\mathbf{b}$ define bounds for $-\mathbf{x}'\mathbf{b}$ automatically. Therefore, if necessary, we can replace a given X by a set which contains at most one of the points $\mathbf{x}$ and $-\mathbf{x}$ for each $\mathbf{x} \in R^k$. Henceforth, it will be assumed that X satisfies the above condition.

For a given X and a two-sided bound ϕ over X let $p'(\phi)$ denote the coverage probability (1.4) of ϕ. ϕ determines a convex set containing the origin in R^k,

$$A'(\phi) = \{\mathbf{b} \in R^k: |\,\mathbf{x}'\mathbf{b}\,| \le \phi(\mathbf{x}), \text{ all } \mathbf{x} \in X\},$$

so that $p'(\phi) = P\{\mathbf{B} \in A'(\phi)\}$. We can define a one-sided bound ϕ_1 over $X_1 = \{\mathbf{x} \in R^k: \mathbf{x} \in X \text{ or } -\mathbf{x} \in X\}$ by $\phi_1(\mathbf{x}) = \phi(\mathbf{x})$ if $\mathbf{x} \in X$ and $\phi_1(\mathbf{x}) = \phi(-\mathbf{x})$ if

$-\mathbf{x} \in X$. It is easy to verify that $A'(\phi) = A(\phi_1)$, where $A(\phi_1)$ is defined in Section 2.

The above correspondence between two-sided bounds ϕ over X and one-sided bounds ϕ_1 over X_1 leads to obvious analogues to the definitions and results of Section 2. The results of Sections 3–5 are valid after some minor modifications. For the two-sided case, the optimal measure μ of Section 3 is still of the form (3.2), but ν is the measure with density function $g(-H(t), t)I_{\{|t| \le T\}}$, for some $T > 0$. For each example of Section 4, the optimal measure μ is obtained from the one-sided analogue by removing some of the mass from the boundary of $[a, b]$. Theorem 5.1 remains valid for the two-sided case.

REFERENCES

ASH, R. (1972). *Real Analysis and Probability.* Academic, New York.

BOHRER, R. (1973). An optimality property of Scheffé bounds. *Ann. Statist.* **1** 766–772.

BOHRER, R. and FRANCIS, G. (1972). Sharp one-sided confidence bounds for linear regression over intervals. *Biometrika* **59** 99–107.

BOWDEN, D. C. and GRAYBILL, F. A. (1966). Confidence bands of uniform and proportional width for linear models. *J. Amer. Statist. Assoc.* **61** 182–198.

GAFARIAN, A. V. (1964). Confidence bands in straight line regression. *J. Amer. Statist. Assoc.* **59** 182–213.

HOEL, P. (1951). Confidence regions for linear regressions. *Proc. Second Berkeley Symp. Statist. Probab.* University of California Press.

MILLER, R. G., JR. (1977). Developments in multiple comparisons 1966–1976. *J. Amer. Statist. Assoc.* **72** 779–788.

NAIMAN, D. Q. (1982). Optimal simultaneous confidence bounds in regression. Ph.D. Thesis, University of Illinois/Urbana, Department of Mathematics.

NAIMAN, D. Q. (1984). Optimal simultaneous confidence bounds. *Ann. Statist.* **12** 702–715.

ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton University Press, New Jersey.

SCHEFFÉ, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40** 87–104.

SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.

WORKING, H. and HOTELLING, H. (1929). Application of the theory of error to the interpretation of tends. *J. Amer. Statist. Assoc.* Suppl. (Proc.) **24** 73–85.

WYNN, H. P. and BLOOMFIELD, P. (1971). Simultaneous confidence bounds in regression analysis. *J. Roy. Statist. Soc. B* **33** 202–217.

DEPARTMENT OF MATHEMATICAL SCIENCES
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218