The Annals of Statistics
1984, Vol. 12, No. 2, 702–715

OPTIMAL SIMULTANEOUS CONFIDENCE BOUNDS

BY DANIEL Q. NAIMAN

The Johns Hopkins University

The notion of a "simultaneous confidence bound" is redefined by requiring a bound on the expected coverage measure (ECM) instead of the coverage probability. This is analogous to a criterion introduced by Spjøtvoll for defining simultaneous tests of hypotheses. Bounds which minimize certain width functionals, subject to a bound on the ECM, are characterized. For bounds on a multilinear regression function over an arbitrary subset of Euclidean space, the bounds which minimize weighted average width, among all bounds with prescribed ECM, are expressed in closed form. As a special case, we give a weight function relative to which Scheffé-type bounds are optimal.

1. Introduction. Consider the multilinear regression model in which one observes

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

where A is a known $n \times k$ matrix, $\mathbf{b}$ is an unknown k-vector, and $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 I_n)$, with σ^2 unknown. Assume that constraints on the predictor variables have been specified and that one is interested in making inferences about the regression function $m(\mathbf{x}) = \mathbf{x}'\mathbf{b}$, for $\mathbf{x}$ restricted to a given subset X of $\mathbf{R}^k$.

A typical approach is to construct simultaneous confidence bounds for the regression function with coverage probability at least $1 - \alpha$, for some prescribed constant α. Bounds are usually taken to be of the form

$$J(\mathbf{x}) = (\mathbf{x}'\hat{\mathbf{b}}_{\text{ls}} - p(\mathbf{x})S,\ \mathbf{x}'\hat{\mathbf{b}}_{\text{ls}} + p(\mathbf{x})S), \quad \text{all} \quad \mathbf{x} \in X, \tag{1.2}$$

where $\hat{\mathbf{b}}_{\text{ls}}$ is the least squares estimator of $\mathbf{b}$, S^2 is the error sum of squares, SS_e, and p is a nonnegative function on X (which determines the shape of the bounds). The coverage probability (CP) is defined to be the probability that the intervals $J(\mathbf{x})$ cover $m(\mathbf{x})$ simultaneously, i.e.

$$\text{CP} = P(m(\mathbf{x}) \in J(\mathbf{x}), \text{ all } \mathbf{x} \in X).$$

In some situations, it may be appropriate to use a different notion of simultaneous confidence bound, where the coverage probability requirement is replaced by a lower bound on the expected μ-measure of the set where coverage takes place, i.e.

$$E\mu\{\mathbf{x} \in X \mid m(\mathbf{x}) \in J(\mathbf{x})\},$$

for a given finite measure μ. We refer to this as the expected coverage measure (ECM) of the bounds with respect to μ.

Received October 12, 1982; revised October 1983.

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

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

For example, suppose $\{\mathbf{x}_i, i = 1, \cdots, N\}$, is the sequence of future points where the bounds (1.2) will be used to give confidence intervals for the regression function. At the time the bounds are to be constructed the sequence of points $\{\mathbf{x}_i\}$ may not be known; however, it may be reasonable to assume that $\mathbf{x}_i$ are i.i.d. random vectors with distribution μ on X, independent of $\mathbf{Y}$. To measure the degree of coverage, we could use the fraction of intervals $\{J(\mathbf{x}_i), i = 1, \cdots, N\}$, that cover the regression function, and to control the degree of coverage we can require that the expected value of this fraction

$$E(N^{-1} \textstyle\sum_{i=1}^{N} I_{\{m(\mathbf{x}_i)\in J(\mathbf{x}_i)\}}),$$

be bounded below by $1 - \alpha$, for some prescribed α. The expected fraction converges to

$$EP\{m(\mathbf{x}_i) \in J(\mathbf{x}_i) \mid \mathbf{Y}\} = \text{ECM}$$

as $N \to \infty$, by the strong law of large numbers. The last equality follows from Fubini's theorem (see Lemma 1). Thus, a lower bound on the ECM is equivalent to a bound on the expected long run frequency of coverage.

If X is a finite set we can take μ to be the counting measure, and in this case the ECM is the expected number of points in X where coverage takes place. In this situation, requiring a lower bound on the ECM is analogous to the criterion introduced by Spjøtvoll (1972) for defining simultaneous tests of hypotheses. He suggested that instead of using a bound on the probability of at least one test giving a false rejection, it may be more appropriate to bound the expected number of false rejections. He argued that a lower bound on the expected number of false rejections leads to a lower bound on the probability of no false rejections, while a lower bound on the latter gives no information about the former.

In the confidence bound setting, when X is finite, Spjøtvoll's argument may be used to justify bounding the expected coverage measure instead of the coverage probability since we have the inequality

$$\text{CP} \geq \text{ECM} + 1 - |X|,$$

which follows from the Bonferroni inequality. The argument does not carry over to the case when X is infinite since the above inequality becomes useless. In fact, if X is a connected set, p is continuous, and μ is a probability measure, then a lower bound on the coverage probability does lead to a lower bound on the expected coverage measure since Markov's inequality gives

$$\text{CP} = P(\mu\{\mathbf{x} \in X \mid m(\mathbf{x}) \in J(\mathbf{x})\} \geq 1) \leq E\,\mu\{\mathbf{x} \in X \mid m(\mathbf{x}) \in J(\mathbf{x})\} = \text{ECM}.$$

For given X and μ the main results of this article characterize the bounds of the form (1.2) which minimize a given width functional of the form

$$\int_X l(\mathbf{x}, p(\mathbf{x}))\mu(d\mathbf{x}),$$

among all bounds whose expected coverage measure at least $1 - \alpha$. As a special case, when $l(\mathbf{x}, t) \equiv t$, the results characterize the bounds which minimize average width, and lead to closed form expressions for the optimal shape functions p.

The characterization of optimal bounds is obtained in a more general setting than the one described above. The function m to be bounded can be arbitrary and it is assumed that we have an estimator $\hat{m}(\mathbf{x})$ of $m(\mathbf{x})$, and a scale estimate S such that $(\hat{m}(\mathbf{x}) - m(\mathbf{x}))/S$ has a known distribution for each $\mathbf{x}$ in X. This generality allows for the characterization of optimal bounds in nonparametric regression.

The problem of finding bounds which are average width optimal among bounds with prescribed coverage probability is more difficult. Bohrer (1973) proved optimality of Scheffé-type bounds, i.e., bounds for which

$$p(\mathbf{x}) = \{\mathbf{x}'(A'A)^{-1}\mathbf{x}\}^{1/2},$$

when

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

the coverage probability is sufficiently large, and μ is Lebesgue measure restricted to X. Bohrer's result does not give optimality of Scheffé-type bounds when the linear model has an intercept parameter, in which case the design matrix contains a column of 1's and X is a subset of $\{\mathbf{x} \in \mathbf{R}^k \mid x_j = 1\}$, for some j. In fact Naiman (1982, 1983) has shown that Scheffé-type bounds are suboptimal for simple linear regression over a finite interval. In this situation optimal bounds have yet to be found and it appears that such bounds must be found numerically. Hoel (1951) and Naiman (1982) showed that for certain bounds one can construct measures relative to which the bounds are average width optimal.

Section 2.1 gives the general framework in which the results are obtained. Bounds that generalize the ones in (1.2), by allowing for asymmetry and randomization, are introduced. In Section 2.2 we give complete class theorems for symmetric nonrandomized bounds. Section 2.3 gives sufficient conditions for global optimality of nonrandomized bounds and Section 2.4 gives necessary conditions. Examples are given in Section 3 and we also discuss the optimality of some standard simultaneous confidence bounds. The results in Section 2 which have analogues for one-sided bounds are given in Section 4.

2. Main results.

2.1 *General framework.* We proceed to introduce the framework for the rest of this paper. Let $(\mathscr{Y}, \mathscr{G})$ be a measurable (observation) space and assume that we observe Y taking values in $\mathscr{Y}$. Let $(X, \mathscr{B})$ be a measurable space endowed with a probability measure μ. We use "a.e." to mean "almost everywhere with respect to μ". Let m be an unknown real-valued measurable function on X and assume that we have an estimator $\hat{m}(x, Y)$ of $m(x)$ for each x in X, where $\hat{m}$ is real-valued and product measurable on $X \times \mathscr{Y}$. Also assume that we have an estimate of scale $S(Y)$, where S is a nonnegative measurable function on $\mathscr{Y}$. We use $\hat{m}(x)$ to denote $\hat{m}(x, Y)$ and S to denote $S(Y)$.

Let $\{P_i,\ i \in I\}$ be a family of probability distributions on $(\mathscr{Y}, \mathscr{G})$ for the observation Y, and assume that for each x in X, the distribution function F_x of

$U_x = (m(x) - \hat{m}(x))/S$, under P_i doesn't depend on i. From this point on, whenever we refer to $P_i(A)$ for A in $\mathcal{G}$, this probability will not depend on i, so we denote it by $P(A)$. Similarly, the expected value of a random variable U under P_i will not depend on i, so we denote it by $E(U)$.

EXAMPLE 1. *Multilinear regression.* For the multilinear regression model (1.1) we can take $\mathcal{Y} = \mathbf{R}^n$ and $Y = \mathbf{Y}$. X can be any subset of $\mathbf{R}^k$ and for $\mathbf{x}$ in X define $m(\mathbf{x}) = \mathbf{x}'\mathbf{b}$, $\hat{m}(\mathbf{x}) = \mathbf{x}'\hat{\mathbf{b}}_{ls}$, and $S = (SS_e)^{1/2}$. The set of probability measures $\{P_i,\ i \in I\}$ is indexed by the set of all ordered pairs $(\mathbf{b}, \sigma)$ in $\mathbf{R}^k \times (0, \infty)$. Let $\| \mathbf{x} \|_A$ denote $\{\mathbf{x}'(A'A)^{-1}\mathbf{x}\}^{1/2}$, for $\mathbf{x}$ in X; then if

$$U_{\mathbf{x}} = (m(\mathbf{x}) - \hat{m}(\mathbf{x}))/S,$$

$U_{\mathbf{x}}/\| \mathbf{x} \|_A$ has a t-distribution with $\nu = n - k$ degrees of freedom, for every $\mathbf{x}$ in X and $(\mathbf{b}, \sigma)$.

EXAMPLE 2. *Nonparametric regression.* This example is a generalization of the previous one. Suppose that X is a subset of $\mathbf{R}^k$, and for some unknown function m on X, in a given class T, we observe $Y_i = m(\mathbf{x}_i) + \sigma e_i$, for $i = 1, \cdots, n$. Assume that we have an estimator $\hat{m}(\mathbf{x})$ and an estimator of S of σ such that $(m(\mathbf{x}) - \hat{m}(\mathbf{x}))/S$ has a known distribution for each $\mathbf{x}$ in X. We can take $\mathcal{Y} = \mathbf{R}^n$, $Y = \mathbf{Y}$, and the family of probability distributions is indexed by $T \times (0, \infty) \times \mathcal{F}$, where $\mathcal{F}$ denotes a class of distributions of the random vector $\mathbf{e}$.

Now we define bounds for the function m. Let p_i be real-valued, product-measurable functions on $X \times \mathcal{Y}$, for $i = 1, 2$. To simplify notation, we use $p_i(x)$ to denote $p_i(x, Y)$. We make the following assumptions.
(a) $p_1(x) \le p_2(x)$ for every x in X.
(b) $(p_1(x), p_2(x))$ is independent of U_x, under P_i, for each x in X and i in I.
The pair (p_1, p_2) defines (randomized) bounds for the function m, which take the form

$$(\hat{m}(x) + p_1(x)S,\ \hat{m}(x) + p_2(x)S), \tag{2.1}$$

for $x \in X$. For the remainder of this paper, we use (p_1, p_2) to refer to the bounds defined in (2.1), for p_i satisfying (a) and (b).

For any (p_1, p_2) and for any x in X, let $C(p_1, p_2)(x)$ denote the event that coverage occurs at x, i.e., that the interval (2.1) covers $m(x)$. Define $X(p_1, p_2)$ to be the (random) subset of X where coverage takes place. Thus x is in $X(p_1, p_2)$ if and only if $C(p_1, p_2)(x)$ occurs.

The expected coverage measure (ECM) of a bound is defined as

$$\mathrm{ECM}(p_1, p_2) = E\mu\{X(p_1, p_2)\}. \tag{2.2}$$

The following lemma, whch will be used in the sequel to characterize the optimal bounds, gives an alternative expression for the ECM of a bound.

LEMMA 1. *For any bound* (p_1, p_2)

$$\mathrm{ECM}(p_1, p_2) = \int_X P\{C(p_1, p_2)(x)\}\mu(dx)$$

$$= \int_X E\{F_x(p_2(x)) - F_x(p_1(x))\}\mu(dx).$$

PROOF. By Fubini's theorem

$$\mathrm{ECM}(p_1, p_2) = E\left[\int_X I_{\{C(p_1,p_2)(x)\}}\mu(dx)\right]$$

$$= \int_X E[I_{\{C(p_1,p_2)(x)\}}]\mu(dx) = \int_X P\{C(p_1, p_2)(x)\}\mu(dx).$$

For any x in X, by the independence of $(p_1(x), p_2(x))$ and U_x we have

$$P\{C(p_1, p_2)(x)\} = P\{\hat{m}(x) + p_1(x)S \le m(x) \le \hat{m}(x) + p_2(x)S\}$$

$$= P\{p_1(x) \le U_x \le p_2(x)\}$$

$$= E\{P(p_1(x) \le U_x \le p_2(x) \mid (p_1(x), p_2(x)))\}$$

$$= E\{F_x(p_2(x)) - F_x(p_1(x))\}. \quad \square$$

The lemma gives us an interpretation of the expected coverage measure which could prove useful. Let x be a random point in X distributed according to μ. The expected coverage measure of a bound is the probability that coverage occurs at the point x.

2.2 *Complete class theorems.* Let M denote the space of all measurable functions on X. A functional L on M is said to be order-preserving if for every f and g in M, $f \le g$ a.e. implies $L(f) \le L(g)$. If L is an order-preserving functional on M, and α is a fixed constant in (0, 1), our goal is to find a bound (p_1, p_2) which satisfies $\mathrm{ECM}(p_1, p_2) \ge 1 - \alpha$, and which minimizes $L\{E(p_2 - p_1)\}$.

For any two bounds (p_1, p_2) and (q_1, q_2) and for arbitrary x in X, we say that (p_1, p_2) dominates (q_1, q_2) at x if

$$E\{p_2(x) - p_1(x)\} \le E\{q_2(x) - q_1(x)\},$$

and

$$P\{C(p_1, p_2)(x)\} \ge P\{C(q_1, q_2)(x)\}.$$

We say that (p_1, p_2) L-dominates (q_1, q_2) if

$$L\{E(p_2 - p_1)\} \le L\{E(q_2 - q_1)\}$$

and

$$\mathrm{ECM}(p_1, p_2) \ge \mathrm{ECM}(q_1, q_2).$$

The next lemma follows from Lemma 1, and is useful in that it allows us to eliminate certain bounds from consideration on the basis of their pointwise behavior (in x).

LEMMA 2. *If* (p_1, p_2) *and* (q_1, q_2) *are bounds such that* (p_1, p_2) *dominates* (q_1, q_2) *a.e. then* (p_1, p_2) *L-dominates* (q_1, q_2). □

The main results of this section are the following complete class theorems.

THEOREM 1. *Suppose that for a.e. x in X,* F_x *has a unimodal density* f_x *with respect to Lebesgue measure with mode* $\theta(x)$, *where* θ *is a measurable function on X. Then for any bound* (q_1, q_2) *there is a nonrandomized bound* (p_1, p_2) *such that* $p_1 \leq \theta \leq p_2$ *a.e. and* (p_1, p_2) *L-dominates* (q_1, q_2).

PROOF. For any (q_1, q_2) define

$$p_1(x) = \begin{cases} \theta(x) & \text{if } q_1(x) > \theta(x) \\ q_1(x) & \text{if } q_1(x) \leq \theta(x) \leq q_2(x) \\ q_1(x) + \theta(x) - q_2(x) & \text{if } q_2(x) < \theta(x), \end{cases}$$

and

$$p_2(x) = \begin{cases} q_2(x) + \theta(x) - q_1(x) & \text{if } q_1(x) > \theta(x) \\ q_2(x) & \text{if } q_1(x) \leq \theta(x) \leq q_2(x) \\ \theta(x) & \text{if } q_2(x) < \theta(x). \end{cases}$$

It is trivial to verify that $p_1 \leq \theta \leq p_2$, (p_1, p_2) is a well-defined bound, and $p_2 - p_1 = q_2 - q_1$. Furthermore $P\{C(p_1, p_2)(x)\} \geq P\{C(q_1, q_2)(x)\}$ whenever f_x is unimodal so that (p_1, p_2) dominates (q_1, q_2) a.e. By Lemma 2 it follows that (p_1, p_2) L-dominates (q_1, q_2).

Now suppose (q_1, q_2) is any bound satisfying $q_1 \leq \theta \leq q_2$ a.e. Define $q_i(x) = E\{p_i(x)\}$ for $i = 1, 2$, so that (q_1, q_2) is a nonrandomized bound. Trivially $E\{p_2(x) - p_1(x)\} \leq E\{q_2(x) - q_1(x)\}$ for every x in X. Since F_x is convex on $(-\infty, \theta(x)]$ a.e., Jensen's inequality gives $F_x\{p_1(x)\} \leq E[F_x\{q_1(x)\}]$ a.e. The concavity of F_x on $[\theta(x), \infty)$ a.e. implies $F_x\{p_2(x)\} \geq E[F_x\{q_2(x)\}]$ a.e. It follows that $P\{C(p_1, p_2)(x)\} \geq P\{C(q_1, q_2)(x)\}$ a.e. and by Lemma 2 (p_1, p_2) L-dominates (q_1, q_2). □

One consequence of the next result is that in cases when the density is symmetric, the symmetric bounds form a complete class.

THEOREM 2. *Suppose* f_x *is unimodal and continuous with mode* $\theta(x)$ *for a.e. x in X. For any nonnegative function g on X, there exist functions* p_1 *and* p_2 *on X satisfying*

(a) $p_1 \leq \theta \leq p_2$ *a.e.*,

(*b*) $p_2 - p_1 = g$ *a.e.*,

and

(c) $f_x(p_2(x)) = f_x(p_1(x))$ *for a.e. x in X.*

If measurable functions p_i *can be found satisfying* (a), (b), *and* (c), *then* (p_1, p_2) *is a bound which L-dominates all bounds satisfying* (b).

PROOF. The first claim is elementary. For the second, if f_x is unimodal it is easy to see that conditions (a) and (c) imply that the integral

$$\int_{I(u)} f_x(t)\,dt,$$

where $I(u) = (p_1(x) + u, p_2(x) + u)$, is maximized in u at $u = 0$. If (q_1, q_2) satisfies (b) it follows that $P\{C(p_1, p_2)(x)\} \geq P\{C(q_1, q_2)(x)\}$ a.e. so (p_1, p_2) dominates (q_1, q_2) a.e. and the conclusion of the theorem follows from Lemma 2. □

REMARK 1. When f_x is symmetric about $\theta(x)$, condition (c) states that the bound is symmetric about θ, i.e., $\frac{1}{2}(p_1 + p_2) = \theta$ a.e.

REMARK 2. It is clear from the proof of Theorem 2 that if (p_1, p_2) is any bound which fails to satisfy condition (c) then there exists a strict improvement to (p_1, p_2), i.e., a bound (q_1, q_2) with equal ECM and with a width function $q_2 - q_1$ which satisfies $q_2 - q_1 \leq p_2 - p_1$ a.e. and $q_2 - q_1 < p_2 - p_1$ on a set of positive μ-measure. We will use this fact in the proof of Theorem 4.

2.3 *Global optimality* (*sufficient conditions*). The main result of this section, Theorem 3, gives sufficient conditions for global optimality of bounds when the functional L to be minimized is of a certain form, and under some mild regularity conditions. We list the relevent assumptions here.

ASSUMPTION (D). U_x has a continuously differentiable and unimodal density f_x for a.e. x in X, with mode $\theta(x)$, where θ is a measurable function.

For the remainder of this article we restrict our attention to nonrandomized bounds. We use B to denote the class of bounds (p_1, p_2) such that $p_1 \leq \theta \leq p_2$ a.e. and $\mathrm{ECM}(p_1, p_2) > 0$.

ASSUMPTION (L1). The functional L is of the form

$$L(h) = \int_X l_x(h(x))\mu(dx),$$

where $l_\cdot(\cdot)$ is a nonnegative product-measurable function on $X \times [0, \infty)$ and $l_x(\cdot)$ is twice continuously differentiable for a.e. x in X.

ASSUMPTION (L2). $l_x'(\cdot) \geq 0$, and $l_x''(\cdot) \geq 0$, for a.e. x in X.

The problem is to find a bound (p_1, p_2) minimizing

$$L(p_2 - p_1) = \int_X l_x(p_2(x) - p_1(x))\mu(dx),$$

subject to $\mathrm{ECM}(p_1, p_2) \geq 1 - \alpha$.

EXAMPLES. We can take $l_x(t) \equiv t$, so that $L(p_2 - p_1)$ is proportional to the average width of (p_1, p_2). More generally, we can consider average squared width, or width to any power $t \geq 1$. We can also multiply any of these functions by an arbitrary weight function $w(x)$. In Section 3 we give an example that leads to optimality of Scheffé-type bounds.

The following condition on a bound (p_1, p_2) will be shown in Theorems 3 and 4 to be necessary and sufficient for global optimality.

CONDITION (O). For some $c \geq 0$ and for a.e. x in X either

$$p_1(x) = \theta(x) = p_2(x),$$

or

$$f_x(p_i(x)) = c l_x'(p_2(x) - p_1(x)) \quad \text{for} \quad i = 1, 2.$$

THEOREM 3. *If (p_1, p_2) satisfies Condition (0) for some $c > 0$, then for any (q_1, q_2) in B satisfying* $\mathrm{ECM}(q_1, q_2) \geq \mathrm{ECM}(p_1, p_2)$, *we have $L(p_2 - p_1) \leq L(q_2 - q_1)$.*

PROOF. Using Lemma 1 and Taylor expansion, we have

$$\begin{aligned}
0 &\leq \mathrm{ECM}(q_1, q_2) - \mathrm{ECM}(p_1, p_2)\\
&= \int_X \{F_x(q_2(x)) - F_x(q_1(x))\} - \{F_x(p_2(x)) - F_x(p_1(x))\}\mu(dx)\\
&= \int_X \{F_x(q_2(x)) - F_x(p_2(x))\} - \{F_x(q_1(x)) - F_x(p_1(x))\}\mu(dx)\\
&= \int_X \{f_x(p_2(x))(q_2(x) - p_2(x)) + \tfrac{1}{2}f_x'(r_2(x))(q_2(x) - p_2(x))^2\\
&\qquad - f_x(p_1(x))(q_1(x) - p_1(x)) - \tfrac{1}{2}f_x'(r_1(x))(q_1(x) - p_1(x))^2\}\mu(dx),
\end{aligned}$$

where r_i is between q_i and p_i for $i = 1, 2$. Since (p_1, p_2) and (q_1, q_2) are in B we have $r_1 \leq \theta \leq r_2$ a.e., so that $f_x'(r_2(x)) \leq 0 \leq f_x'(r_1(x))$ a.e. It follows that

$$\int_X \{f_x(p_2(x))(q_2(x) - p_2(x))\} - \{f_x(p_1(x))(q_1(x) - p_1(x))\}\mu(dx) \geq 0.$$

Using Condition (O), the nonnegativity of l_x' a.e. and the fact that c is positive we obtain

$$\int_X l_x'(p_2(x) - p_1(x))\{(q_2(x) - q_1(x)) - (p_2(x) - p_1(x))\}\mu(dx) \geq 0. \tag{2.3}$$

By Taylor expansion, for some $u(x)$ between $p_2(x) - p_1(x)$ and $q_2(x) - q_1(x)$

a.e., we have

$$L(q_2 - q_1) - L(p_2 - p_1)$$

$$= \int_X \{l_x(q_2(x) - q_1(x)) - l_x(p_2(x) - p_1(x))\}\mu(dx)$$

$$= \int_X l_x'(p_2(x) - p_1(x))\{(q_2(x) - q_1(x)) - (p_2(x) - p_1(x))\}$$

$$+ \tfrac{1}{2} l_x''(u(x))\{(q_2(x) - q_1(x)) - (p_2(x) - p_1(x))\}^2\mu(dx).$$

Using (2.3) and the fact that l_x'' is nonnegative a.e., we conclude that $L(q_2 - q_1) - L(p_2 - p_1)$ is nonnegative. □

2.4 *Global optimality* (*necessary conditions*). The main result of this section, Theorem 4, which is a converse to Theorem 3, gives necessary conditions for global optimality of bounds. We continue to make the assumptions (D), (L1) and (L2).

The proof of Theorem 4 uses the following elementary lemma for (real) Hilbert spaces.

LEMMA 3. *Let g and h be elements of a* (*real*) *Hilbert space H, with inner product denoted by* $(\,,)$. *Suppose that for every k in H, $(g, k) > 0$ implies $(h, k) \geq 0$. Then there exists a nonnegative constant c such that $h = cg$.*

PROOF. The claim is trivial for $g = 0$, so assume $g \neq 0$. Write $h = h_1 + h_2$, where $(g, h_1) = 0$ and $h_2 = cg$ for some real constant c. For any $v > 0$, if we let $k = -h_1 + vg$ then $(g, k) = v(g, g) > 0$ and it follows that $0 \leq (h, k) = -(h_1, h_1) + v(h_2, g)$. If we let $v \to 0$ we see that $(h_1, h_1) \leq 0$, so that $h_1 = 0$ and $h = h_2 = cg$. It follows easily that $c \geq 0$. □

We use L^2 to denote the Hilbert space $L^2(X, \mu)$, $\| h \|$ to denote the norm of h, and (h_1, h_2) to denote the inner product for functions in L^2.

THEOREM 4. *Let (p_1, p_2) be any bound in B and suppose that the following conditions hold*:

(a) *l_x is strictly increasing and $l_x(0) = 0$ for a.e. x in X.*

(b) *$p_1, p_2, f.(p_2(\cdot))$ and $l.((p_2(\cdot) - p_1(\cdot)))$ are in L^2.*

(c) *There exists a constant C such that for all h in some L^2 neighborhood of p_2*

$$\left| \int_X (h(x) - p_2(x))^2 f_x'(h(x))\mu(dx) \right| \leq C \| h - p_2 \|^2,$$

and

$$\left| \int_X (h(x) - p_2(x))^2 l_x''(h(x))\mu(dx) \right| \leq C \| h - p_2 \|^2.$$

If (p_1, p_2) *is optimal in the sense that* $L(p_2 - p_1) \le L(q_2 - q_1)$ *whenever* $\mathrm{ECM}(q_1, q_2) \ge \mathrm{ECM}(p_1, p_2)$, *then* (p_1, p_2) *satisfies Condition* (O).

PROOF. Define

$$g(x) = f_x(p_2(x)) I_{\{p_2(x) > \theta(x)\}}$$

and

$$h(x) = l_x'(p_2(x) - p_1(x)) I_{\{p_2(x) > \theta(x)\}}.$$

We proceed to show

$$(g, k) > 0 \quad \text{implies} \quad (h, k) \ge 0, \tag{2.4}$$

for all k in L^2. From this, Lemma 3, and assumption (b), it follows that $h = cg$ a.e. for some $c \ge 0$. By (a) and Remark 2 of Section 2.2, $f_x(p_1(x)) = f_x(p_2(x))$ a.e. so (p_1, p_2) satisfies Condition (O).

We proceed to prove (2.4), so fix k in L^2 and assume $(g, k) > 0$. For $u > 0$ define

$$X_u = \{x \in X \mid p_2(x) + uk(x) > \theta(x) \quad \text{and} \quad p_2(x) > \theta(x)\}$$

and

$$q_{2,u}(x) = \begin{cases} p_2(x) + uk(x) & \text{if } x \in X_u, \\ p_2(x) & \text{if } x \notin X_u. \end{cases}$$

$(p_1, q_{2,u})$ is in B, and

$$\mathrm{ECM}(p_1, q_{2,u}) - \mathrm{ECM}(p_1, p_2) = \int_{X_u} \{F_x(p_2(x) + uk(x)) - F_x(p_2(x))\}\mu(dx).$$

Using Taylor expansion,

$$F_x(p_2(x) + uk(x)) - F_x(p_2(x)) = uk(x)f_x(p_2(x)) + \tfrac{1}{2} u^2 k(x)^2 f_x'(r_{2,u}(x)),$$

a.e., where $r_{2,u}(x)$ is between $p_2(x)$ and $q_{2,u}(x)$. By (c)

$$\mathrm{ECM}(p_1, q_{2,u}) - \mathrm{ECM}(p_1, p_2) = u \int_{X_u} k(x) f_x(p_2(x))\mu(dx) + O(u^2),$$

as $u \to 0^+$. Since

$$\lim_{u \to 0^+} I_{X_u}(x) = I_{\{p_2(x) > \theta(x)\}}(x)$$

for every x in X, Lebesgue's dominated convergence theorem gives

$$\lim_{u \to 0^+} \int_{X_u} k(x) f_x(p_2(x))\mu(dx) = \int_X k(x) g(x)\mu(dx) > 0.$$

For sufficiently small $u > 0$ it follows that

$$\mathrm{ECM}(p_1, q_{2,u}) > \mathrm{ECM}(p_1, p_2).$$

By the optimality of (p_1, p_2) it follows that

$$\begin{aligned} 0 &\le L(p_1, q_{2,u}) - L(p_1, p_2) \\ &= \int_{X_u} l_x(q_{2,u}(x) - p_1(x)) - l_x(p_2(x) - p_1(x))\mu(dx), \end{aligned}$$

for $u > 0$ sufficiently small. Using Taylor expansion, and (c) we have

$$0 \le u \int_{X_u} l_x'(p_2(x) - p_1(x))k(x)\mu(dx) + O(u^2)$$

as $u \to 0^+$. By (b) and the dominated convergence theorem it follows that

$$0 \le u \int_X h(x)k(x)\mu(dx) + \grave{O}(u^2),$$

as $u \to O^+$ and this implies $(h, k) \ge 0$ which was to be shown. □

3. Comments and examples.

REMARK. It is interesting to note that for f and L fixed, even though we have introduced the probability measure μ to define the ECM of a bound, the family of optimal bounds defined by Condition (O) depends only on the support of the measure μ.

For the examples of Section 2.1 we now characterize the globally optimal bounds.

EXAMPLE 1. *Multilinear regression.* (See Section 2.1.) For this example the density function $f_{\mathbf{x}}$ is given by

$$f_{\mathbf{x}}(t) = K_\nu \| \mathbf{x} \|_A^{-1}\{1 + (t/\| \mathbf{x} \|_A)^2/\nu\}^{-(\nu+1)/2},$$

for each $\mathbf{x}$ in X, where $K_\nu = \Gamma(½(\nu + 1))/\{\nu^{1/2}\Gamma(½\nu)\Gamma(½)\}$.

Suppose $l_{\mathbf{x}}(t) = w(\mathbf{x})t$ where w is nonnegative and measurable on X, so that $L(p_1, p_2)$ is proportional to the weighted average width of (p_1, p_2). We have $l_{\mathbf{x}}'(t) = w(\mathbf{x})$ for all $\mathbf{x}$ in X. It is a simple matter to verify that Condition (O) is equivalent to the statement that for some $c \ge 0$ and for a.e. $\mathbf{x}$ in X, either $p_i(\mathbf{x}) = 0$ or $p_2(\mathbf{x}) = p_1(\mathbf{x}) = a(\mathbf{x})$, where

$$a(\mathbf{x}) = \| \mathbf{x} \|_A\{(cw(\mathbf{x}) \| \mathbf{x} \|_A)^{-2/(\nu+1)} - 1\}^{1/2}. \tag{2.5}$$

Note that for fixed α, a globally optimal bound is not uniquely determined since we can vary the constant c and the set where $p_1(\mathbf{x}) = p_2(\mathbf{x}) = 0$. However if w is continuous then it is easy to check that the optimal continuous bound is uniquely determined by α. In fact $p_2(\mathbf{x}) = -p_1(\mathbf{x}) = a(\mathbf{x})$ if $cw(\mathbf{x}) \| \mathbf{x} \|_A \le 1$, and $p_i(\mathbf{x}) = 0$ for $i = 1, 2$, otherwise.

Analogous statements hold in the variance-known case, when S is replaced by σ in the definition of the bounds, and f_x is a normal density for each $\mathbf{x}$ in X. The function $a(\mathbf{x})$ is replaced by

$$a(\mathbf{x}) = \| \mathbf{x} \|_A[-2 \log\{cw(\mathbf{x}) \| \mathbf{x} \|_A\}]^{1/2}.$$

An important case is when $w(\mathbf{x}) \propto \| \mathbf{x}_A \|^{-1}$, since the condition for optimality states that $p_2(\mathbf{x}) = p_1(\mathbf{x}) = C \| \mathbf{x} \|_A$, for some $C \geq 0$, and the bounds (2.1) are of the Scheffé-type. Thus we have proved an optimality property of Scheffé-type bounds.

Constant-width bounds, i.e., bounds (p_1, p_2) for which $p_2 - p_1$ is a constant function, were first introduced by Gafarian (1964) for simple linear regression over a finite interval. These bounds are known to minimize the maximum width among all bounds with equal coverage probability, and it can also be shown that they minimize maximum width among all bounds with ECM $\geq 1 - \alpha$. It is of interest to ask whether constant width bounds can be optimal in the sense of minimizing weighted average width for some weight function w, among all bounds with equal ECM. For any constant $a \geq 0$, we can solve the equation $a(\mathbf{x}) \equiv a$ for w, where $a(\mathbf{x})$ is given by (2.5), to obtain the weight function relative to which the bound with $p_i = a$ is average width optimal. In contrast to the situation for Scheffé-type bounds, the weight function depends on the number of degrees of freedom and the constant a, so that it is not possible to make general statements about what weight functions correspond to constant-width bounds.

For other functions $l_{\mathbf{x}}$, (see the examples in Section 2.3) we can characterize the optimal bounds using condition b of Theorem 4 (resp. Theorem 5), but closed form expressions for the optimal bounds have not been obtained. The defining equation is harder to solve because $p(\mathbf{x})$ appears on both sides of the equation when $l'_{\mathbf{x}}(t)$ depends on t.

In Example 2 (Nonparametric regression) of Section 2.1 we obtain optimal bounds of the same form as above when the distribution of $U_{\mathbf{x}}$ has either a normal distribution or a t-distribution. If the distributions are asymptotically normal or t, we obtain approximations to optimal bounds for large samples.

4. Summary of results for one-sided bounds. We define one-sided bounds for m as follows. Let p be a nonnegative product measurable function on $X \times \mathscr{Y}$. To simplify notation, $p(x)$ is used to denote $p(x, Y)$. We assume $p(x)$ is independent of U_x, under P_i, for each x in X and i in I. p defines (randomized) bounds for the function m which take the form

$$(-\infty, \hat{m}(x) + p(x)S), \quad \text{for} \quad x \in X. \tag{2.1'}$$

For any one-sided bound p and for any x in X, let $C(p)(x)$ denote the event that coverage occurs at x, that is, that the interval (2.1′) covers $m(x)$. Define $X(p)$ to be the subset of X where coverage takes place, and define

$$\text{ECM}(p) = E\mu\{X(p)\}.$$

The proofs of the results below are similar to the corresponding results for two-sided bounds.

LEMMA 1′. *For any bound p*

$$\mathrm{ECM}(p) = \int_X P\{C(p)(x)\}\mu(dx) = \int_X E\{F_x(p(x))\}\mu(dx). \quad \square$$

Let M^+ denote the space of nonnegative measurable functions on X. If L is an order-preserving functional on M^+, and α is a fixed constant in (0, 1), our goal is to find a bound p such that $\mathrm{ECM}(p) \geq 1 - \alpha$, which minimizes $L\{E(p)\}$. For one-sided bounds, we have obvious analogues to the notions of domination and L-domination given in Section (2.2). Furthermore, Lemma 2 holds for one-sided bounds and this leads to the following result.

THEOREM 1′. *Suppose that for a.e. x in X, F_x has a unimodal density f_x with respect to Lebesgue measure with mode $\theta(x)$. If p is any bound satisfying $p \geq \theta$ a.e., then $E(p)$ defines a nonrandomized bound which L-dominates p.* □.

For the remainder of this section we consider only nonrandomized bounds. For the results corresponding to Theorems 3 and 4 we define B^+ to be the set of (nonrandomized) bounds satisfying $p \geq \theta$ a.e. We introduce the following conditions.

ASSUMPTION (Ll′). The functional L is of the form

$$L(h) = \int_X l_x(h(x))\mu(dx),$$

where $l.(\cdot)$ is a nonnegative product-measurable function on $X \times (-\infty, \infty)$ and $l_x(\cdot)$ is twice continuously differentiable for a.e. x in X.

ASSUMPTION (L2′). $l_x'(t) \geq 0$, and $l_x''(t) \geq 0$, for $t \geq \theta(x)$, for a.e. x in X.

CONDITION (O′). For some $c \geq 0$ and for a.e. x in X either

$$p(x) = \theta(x)$$

or

$$f_x(p(x)) = c l_x'(p(x)).$$

THEOREM 3′. *If p satisfies Condition (O') for some $c > 0$, then for any q in B^+ satisfying* $\mathrm{ECM}(q) \geq \mathrm{ECM}(p)$, *we have* $L(p) \leq L(q)$. □

THEOREM 4′. *Let p be any bound in B^+ and suppose that the following conditions hold.*
(a) *l_x is strictly increasing and $l_x(0) = 0$ for a.e. x in X.*
(b) *p, $f.(p(\cdot))$ and $l.(p(\cdot))$ are in L^2.*
(c) *There exists a constant C such that for all h in some L^2 neighborhood of p*

$$\left| \int_X (h(x) - p(x))^2 f_x'(h(x))\mu(dx) \right| \leq C \| h - p \|^2,$$

and

$$\left| \int_X (h(x) - p(x))^2 l_x''(h(x))\mu(dx) \right| \le C \| h - p \|^2.$$

If p is optimal in the sense that $L(p) \le L(q)$ *whenever* $\mathrm{ECM}(q) \ge \mathrm{ECM}(p)$, *then p satisfies Condition* (O'). □

5. Concluding remarks. We have introduced a definition of "simultaneous confidence bound" which replaces the requirement of a lower bound on the coverage probability by a bound on the expected coverage measure. This is analogous to a criterion introduced by Spjøtvoll for defining simultaneous tests of hypotheses. In this setting we have derived optimal simultaneous confidence bounds and we see that careful consideration of the desired definition of a simultaneous confidence bound may lead to bounds which are better solutions to the problem at hand, and which may be easier to find.

6. Acknowledgements. The author wishes to express his deepest gratitude to his advisor Robert Bohrer for his strong encouragement. Thanks also go to Robert Wijsman and Stephen Portnoy for their helpful suggestions. The author would also like to thank the editor and the referees, whose comments led to a much improved version of the original paper This paper is based partly on the author's Ph.D. dissertation which was prepared at the University of Illinois. The research was supported in part by the U.S. Department of the Navy under Office of Naval Research Contract No. N00014-79-C-0801.

REFERENCES

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

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

HOEL, P. G. (1951). Confidence regions for a linear relation. *Proc. 2nd Berkeley Symp. Math. Statist. Probab.* 75–81. University of Calif. Press, Berkeley.

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

NAIMAN, D. Q. (1983). Average width optimality of simultaneous confidence bounds. Technical report #380 (ONR #83-1), Department of Mathematical Sciences, Johns Hopkins University.

SPJØTVOLL, E. (1972). On the optimality of some multiple comparisons procedures. *Ann. Math. Statist.* **43** 398–411.

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