SIMULTANEOUS CONFIDENCE REGIONS FOR PREDICTIONS

E. Carlstein

University of North Carolina, Chapel Hill, NC 27514, USA

Abstract: After observing n independent responses at n corresponding design points in a linear regression setting, we wish to make a confidence statement about future responses that will apply simultaneously to all possible design points. Two appropriate prediction regions are derived using normal theory.

AMS Subject Classification: 62F25, 62J05.

Key Words: Simultaneous confidence, prediction, linear regression.

Supported by NSF Grant DMS-8400602.
1. Introduction.

Consider a linear regression situation in which the response for the \( i \)th individual follows the model:

\[
Y_i = Y_i(x) = x' \beta + \varepsilon_i \quad \forall x \in \mathbb{R}^p,
\]

where \( x \) is a vector of "independent" variables, \( \beta \in \mathbb{R}^p \) is a vector of unknown parameters, and \( \varepsilon_i \) is a random error associated with the \( i \)th individual.

We are able to observe \( n \) independent responses \( \{Y_i: 1 \leq i \leq n\} \) at the corresponding known design points \( \{x_i: 1 \leq i \leq n\}, x_i \in \mathbb{R}^p \). Based on this data, we would like to make a simultaneous confidence statement about the future realizations \( Y_{n+1}(x) \) at all possible design points \( x \in \mathbb{R}^p \).

Example: The output (\( Y \)) of a certain type of machine is a linear function of: the control settings (\( x \)), the unknown parameters (\( \beta \)), and a random error (\( \varepsilon_i \)) which is unobservable and is specific to the \( i \)th machine. Having sampled the outputs (\( Y_i \)) of \( n \) such machines--each at one setting (\( x_i \))--we now want a simultaneous confidence band for the performance of a new machine at all possible settings.

No solution to this simultaneous prediction interval problem appears in the literature. We shall obtain two solutions to a more general problem (involving \( k \) future individuals) by applying normal theory.

2. Results.

Assume model (1) holds for each \( i \in \{1, 2, \ldots, n+k\} \), with \( k \geq 1 \) and \( n > p \). Write \( X = (x_1, x_2, \ldots, x_n)' \) and \( Y = (Y_1(x_1), Y_2(x_2), \ldots, Y_n(x_n))' \) for the observed data.

Assume \( \text{rank}(X) = p \) and denote: \( D = (X'X)^{-1} \), \( \hat{\beta} = DX'Y \), \( s^2 = |Y-X\hat{\beta}|^2/(n-p) \).

Assume \( \{\varepsilon_i: 1 \leq i \leq n+k\} \) are iid \( N(0, \sigma^2) \) with \( \sigma^2 > 0 \) unknown.

Theorem 1: For \( \alpha \in (0, 1) \),

\[
P(Y_i(x) \leq x'\hat{\beta} \pm s((p+k)(1+x'Dx)F(1-\alpha; p+k, n-p))/2) \leq 1 - \alpha,
\]

\( \forall x \in \mathbb{R}^p \) and \( \forall i \in \{n+1, n+2, \ldots, n+k\} \),

where \( F(y; N, M) \) is the \( y \)-percentile of the F-distribution with \( N \) and \( M \) degrees of freedom.
of freedom in the numerator and denominator respectively.

Notice that the confidence region is centered at the usual point estimates. This region is analogous to the Scheffé-type simultaneous confidence band on $x' \beta$, $\forall x \in \mathbb{R}^p$. It is also analogous to Lieberman's (1961) simultaneous prediction interval on $\{Y_i: n+1 \leq i \leq n+k\}$, which applies to a fixed set $\{x_i: n+1 \leq i \leq n+k\}$. Both of these methods are discussed in detail by Miller (1981, Chapter 3). In a sense, our simultaneous prediction region combines the features of these two methods. Our proof extends the Scheffé F-projection technique (Miller, 1981, Chapter 2, Section 2) to the case of a $(p+k)$-dimensional linear space and a total of $n+k$ random variables $Y_i$.

Proof of Theorem 1: Denote $b' = (\hat{\beta}' - \beta', \varepsilon_{n+1}, \varepsilon_{n+2}, \ldots, \varepsilon_{n+k})$, $\tilde{D} = \begin{bmatrix} D & 0 \\ 0 & I_k \end{bmatrix}$, and $Q = (\hat{\beta} - \beta)'\tilde{D}^{-1}(\hat{\beta} - \beta)$. By the generalized Cauchy-Schwarz inequality (Rao, 1973, eq. 1e.1.4):

$$\max \{(b'a)^2/a'Da: a \in \mathbb{R}^{p+k}\} = b'\tilde{D}^{-1}b$$

$$= Q + \sum_{j=1}^{k} \varepsilon_{n+j}^2$$

$$\leq \chi^2_{(p+k)\sigma^2}$$

since $Q \sim \chi^2_{(p)\sigma^2}$ (Rao, 1973, p. 188) and $Q$ is independent of the $\varepsilon_{n+j}$'s. It is well known that $(n-p)s^2 \sim \chi^2_{(n-p)\sigma^2}$, and furthermore $s^2$ is independent of $Q$ and also of the $\varepsilon_{n+j}$'s. Hence $b'\tilde{D}^{-1}b/(p+k)s^2 \sim F(p+k,n-p)$, so that by (2):

$$P\{|b'a| \leq s((p+k)a'\tilde{D}F(1-\alpha; p+k,n-p))^{1/2}: a \in \mathbb{R}^{p+k}\} = 1-\alpha.$$ 

Now consider only those $a \in \mathbb{R}^{p+k}$ s.t. $a' = (x' | \delta_1, \delta_2, \ldots, \delta_k)$, where $x \in \mathbb{R}^p$ is arbitrary and the $\delta_j$'s are all zero except for a single $\delta_1 = -1$. Then $b'a = x'\hat{\beta} - Y_{n+1}(x)$, concluding the proof. \(\square\)
An alternative approach is to break up $Y_{n+1}(x)$ into its components $x'B$ and $e_{n+1}$, and to separately determine confidence intervals on these components. The separate intervals may then be combined into a confidence interval on $Y_{n+1}(x)$ via the Bonferroni inequality. This approach is formalized by

**Theorem 2:** For $\alpha \in (0,1)$ and $\bar{\alpha} \in (0,\alpha)$,

$$P\{Y_{n+1}(x) \leq x'B \pm s((px'Dx)(1-\alpha;p,n-p))^{1/2} + (kF(1-\alpha+\bar{\alpha};k,n-p))^{1/2}\} \geq 1-\alpha.$$  

**Proof:** The standard Scheffé-type simultaneous confidence statement for $x'B$ is:

$$P\{x'B \leq x'B \pm s((px'Dx)(1-\alpha;p,n-p))^{1/2} \forall x \in R^p\} = 1-\alpha.$$  

Denote $\varepsilon = (\varepsilon_{n+1}, \varepsilon_{n+2}, \ldots, \varepsilon_{n+k})'$. Since $\max\{\varepsilon'a\}^2/a'k = \varepsilon'\varepsilon/k^2 F(k,n-p)$, we have:

$$1-\alpha+\bar{\alpha} = P\{|\varepsilon'a| \leq s(a'kF(1-\alpha+\bar{\alpha};k,n-p))^{1/2} \forall a \in R^k\}$$

$$\leq P\{|\varepsilon_{n+1}^2| \leq s(kF(1-\alpha+\bar{\alpha};k,n-p))^{1/2} \forall i \in \{1,2,\ldots,k\}\}.$$  

Applying the inequality $P\{A \cap B\} \geq P\{A\} + P\{B\} - 1$ establishes the Theorem.  

3. Comparison.

Neither confidence region is uniformly superior to the other. Consider, for example, the case of simple linear regression with an intercept ($p=2$), with $k=1$ future individual to be predicted. We shall compare the widths of the confidence regions at the "center" of the data, i.e. at $x' = (1, \bar{x})$, where $\bar{x}$ denotes the average of the second coordinates of the design points $x_1, x_2, \ldots, x_n$. Here the relevant comparison is between:

$$w_1(n, \alpha) = (3(1+n^{-1})F(1-\alpha; 3, n-2))^{1/2}$$ and

$$w_2(n, \alpha, \bar{\alpha}) = (2n^{-1}F(1-\alpha; 2, n-2))^{1/2} + (F(1-\alpha+\bar{\alpha}; 1, n-2))^{1/2}.$$  

On the one hand, if $n=10$ and $\alpha=0.01=2\bar{\alpha}$ we find $w_1=5.00$ and $w_2=5.32$. On the other hand, as $n \to \infty$:

$$w_1(n, \alpha) \to (\chi^2(1-\alpha; 3))^{1/2}, \text{ while}$$

$$w_2(n, \alpha, \bar{\alpha}) \to (\chi^2(1-\alpha; 1))^{1/2}.$$
provided that \( \{ \tilde{a}_n : n \geq 1 \} \) is chosen so that \( \tilde{a}_n \to 0 \) and \( F(1-\tilde{a}_n; 2, n-2)/n \to 0 \). In practice, this suggests that for large sample sizes there is an \( \tilde{a}_n \in (0, \alpha) \) for which the second method provides a narrower band than the first, for \( x \) within a neighborhood of \( \bar{x} \). For example, if \( n=122 \) and \( \alpha=0.01=2\tilde{a} \) then \( w_1=3.46 \) and \( w_2=3.16 \).

Acknowledgment.

I thank Professor David Ruppert for his advice on the presentation of these results.