Available at www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Statistics & Probability Letters 64 (2003) 403–406
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Scheffe's method for constructing simultaneous confidence intervals subject to cone constraints☆

Alexander Shapiro*

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA

Received February 2003; received in revised form June 2003

Abstract

We discuss in this paper Scheffe's method for constructing simultaneous confidence intervals which hold for all linear combinations of the parameters subject to the weight vector being restricted to a convex cone.

Keywords: Scheffe's method; Confidence intervals; Confidence band; Chi-bar-squared distributions; Cone constraints

1. Introduction

Let $\hat{\gamma}$ be an estimator of the true value $\gamma \in \mathbb{R}^p$ of a parameter vector. Suppose that $\hat{\gamma}$ has a normal distribution with mean vector γ and nonsingular covariance matrix $\sigma^2 V$, known up to the coefficient σ^2. For example, $\hat{\gamma}$ can be the least squares estimator of the standard linear model

$$Y = X\gamma + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, \sigma^2 I_n). \tag{1.1}$$

In that case $V = (X'X)^{-1}$, of course (we discuss this further in Example 2.1).

Suppose, moreover, that an estimator S^2 of σ^2 is available, such that S^2 is independent of $\hat{\gamma}$ and vS^2/σ^2 has a chi-squared distribution with v degrees of freedom. Then $(\hat{\gamma}-\gamma)'V^{-1}(\hat{\gamma}-\gamma)/pS^2 \sim F_{p,v}$, and by the well known Scheffe's procedure,

$$a'\hat{\gamma} \pm (pF_{\alpha,p,v})^{1/2} S(a'Va)^{1/2} \tag{1.2}$$

gives $100(1-\alpha)\%$ confidence interval for any linear function $a'\gamma$ of the parameter vector (e.g., Seber 1977, pp. 128–130).

☆ Supported by the National Science Foundation under grant DMS-0073770.
* Tel.: +1-404-894-6544; fax: +1-404-894-2301.
E-mail address: ashapiro@isye.gatech.edu (A. Shapiro).

0167-7152/$ - see front matter
doi:10.1016/S0167-7152(03)00205-0

Suppose now that one would like to impose linear restrictions on the weight vector a. That is, we would like to construct confidence intervals for all a in a certain convex cone $C \subset \mathbb{R}^p$. For example, we may be interested in $a \in \mathbb{R}^p$ with all components being nonnegative, i.e., we may restrict a to the nonnegative orthant $\mathbb{R}^p_+ := \{x \in \mathbb{R}^p\colon x_i \geqslant 0,\ i = 1, \dots, p\}$.

2. Derivations

Our analysis is based on the following derivations. Choose $b \in \mathbb{R}^p$ and consider the optimization problem

$$\max_{x \in \mathbb{R}^p} \frac{b'x}{(x'Vx)^{1/2}} \quad \text{subject to} \quad x \in C, \quad x \neq 0. \tag{2.1}$$

Recall that V is a $p \times p$ symmetric positive definite matrix and C is a closed convex cone. Let us observe that the objective function $f(x) := b'x/(x'Vx)^{1/2}$ of problem (2.1) has the following property: $f(tx) = f(x)$ for any $t > 0$ and $x \neq 0$. Therefore we can normalize vector x in (2.1) by adding the constraint $x'Vx = 1$. It follows that problem (2.1) is equivalent to the optimization problem

$$\max_{x \in \mathbb{R}^p} b'x \quad \text{subject to} \quad x'Vx \leqslant 1, \quad x \in C. \tag{2.2}$$

Indeed, since V is positive definite, the constraint $x'Vx \leqslant 1$ defines a bounded subset of $\mathbb{R}^p$. Consequently the feasible set of problem (2.2) is bounded and clearly is closed, and hence is compact. It follows that problem (2.2) has an optimal solution. Moreover, since the objective function of (2.2) is linear, the constraint $x'Vx \leqslant 1$ is active at an optimal solution of (2.2). That is, the constraint $x'Vx \leqslant 1$ can be replaced by the equality constraint $x'Vx = 1$. We obtain that the optimal values of problems (2.1) and (2.2) are equal to each other.

With problem (2.2) is associated the Lagrangian

$$L(x, \lambda) := b'x - \lambda(x'Vx - 1), \quad \lambda \in \mathbb{R}.$$

We have

$$\inf_{\lambda \geqslant 0} L(x, \lambda) = \begin{cases} b'x & \text{if } x'Vx - 1 \leqslant 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Therefore problem (2.2) can be formulated in the following min–max form:

$$\max_{x \in C} \min_{\lambda \geqslant 0} L(x, \lambda). \tag{2.3}$$

The (Lagrangian) dual of problem (2.2) is obtained by interchanging the order of min and max operators. That is, the dual of (2.2) is the problem

$$\min_{\lambda \geqslant 0} \max_{x \in C} L(x, \lambda). \tag{2.4}$$

By the standard duality theory we have that the optimal value of problem (2.3), and hence of problem (2.2), is less than or equal to the optimal value of problem (2.4). Moreover, problem (2.2) is convex and the Slater condition holds. That is, there exists $x_* \in C$ such that $x_*' V x_* - 1 < 0$ (e.g., take $x_* = 0$). It follows then that there is no duality gap between problems (2.2) and (2.4), i.e., the optimal values of the problem (2.2) and its dual (2.4) are equal to each other (see, e.g., Rockafellar, 1970).

Let us represent the dual problem (2.4) in a more explicit form. For $\lambda > 0$ we have

$$L(x, \lambda) = \lambda d' V d - \lambda (d - x)' V (d - x) + \lambda,$$

where $d := (2\lambda)^{-1} V^{-1} b$. Consequently, problem (2.4) can be written in the form

$$\min_{\lambda > 0} \left\{ (4\lambda)^{-1} b' V^{-1} b + \lambda - (4\lambda)^{-1} \min_{x \in C} (V^{-1} b - x)' V (V^{-1} b - x) \right\}. \tag{2.5}$$

Note that vector x in (2.4) is rescaled to $(2\lambda)^{-1} x$ in (2.5). Since C is a cone, such rescaling does not change the optimal value of (2.5). Furthermore, we have that

$$\kappa := b' V^{-1} b - \inf_{x \in C} (V^{-1} b - x)' V (V^{-1} b - x) \geqslant 0$$

and hence

$$\inf_{\lambda > 0} \{ \lambda + (4\lambda)^{-1} \kappa \} = \kappa^{1/2}.$$

We obtain that the optimal value of the dual problem, and hence of problem (2.1), is equal to

$$\left[h' V h - \inf_{x \in C} (h - x)' V (h - x) \right]^{1/2}, \tag{2.6}$$

where $h := V^{-1} b$. Note finally that the optimal value of (2.6) is equal to $(\bar{x}' V \bar{x})^{1/2}$, where $\bar{x} = \bar{x}(h)$ is the optimal solution of (2.6). This holds true since we have by the first-order optimality conditions that $(h - \bar{x})' V \bar{x} = 0$.

Let us take now $b := \gamma - \hat{\gamma}$ and denote by Z the common optimal value of (2.1) and (2.6). We have then that

$$h = V^{-1}(\gamma - \hat{\gamma}) \sim \mathrm{N}(0, \sigma^2 V^{-1}).$$

It follows from (2.6) that Z^2/σ^2 has a chi-bar-squared distribution. That is, for $c > 0$,

$$\Pr\left(\frac{Z^2}{\sigma^2} \geqslant c \right) = \sum_{i=0}^{p} w_i \Pr(\chi_i^2 \geqslant c), \tag{2.7}$$

where $w_i = w_i(p, V^{-1}, C)$, $i = 0, \ldots, p$, are nonnegative weights, summing up to one, known as level probabilities. The chi-bar-squared distributions and the weights $w_i(p, V^{-1}, C)$ are discussed in detail in Sen and Silvapulle, 2002 and Shapiro, 1988, for example. (Note that, by the definition, $\chi_0^2 \equiv 0$.) Moreover, we then have that

$$\Pr\left(\frac{Z^2}{S^2} \geqslant c \right) = \sum_{i=1}^{p} w_i \Pr(\mathrm{i} f_{i,v} \geqslant c), \tag{2.8}$$

where $f_{i,v}$ are random variables having F-distribution with i and v degrees of freedom.

Let c_α be the critical value, at significance level $1-\alpha$, associated with the distribution given in the right-hand side of (2.8), i.e.,

$$\sum_{i=1}^{p} w_i \Pr(i f_{i,v} \geqslant c_\alpha) = \alpha. \tag{2.9}$$

We obtain then that

$$a'\hat{\gamma} + c_\alpha^{1/2} S (a'Va)^{1/2} \tag{2.10}$$

provides an upper $100(1-\alpha)\%$ confidence bound for $a'\gamma$ which holds for all $a \in C$. In a similar way, by taking $b := \hat{\gamma} - \gamma$, one can obtain a lower $100(1-\alpha)\%$ bound

$$a'\hat{\gamma} - c_\alpha^{1/2} S (a'Va)^{1/2} \tag{2.11}$$

for $a'\gamma$ which holds for all $a \in C$.

If C is a linear subspace of $\mathbb{R}^p$ of dimension k, then $w_k = 1$ and all other weights are zero. In that case $c_\alpha = kF_{\alpha,k,v}$. In particular, if $C = \mathbb{R}^p$, then $c_\alpha = pF_{\alpha,p,v}$ and (2.10) coincides with (1.2). Note that in this case the confidence interval is two-sided, simply because a and $-a$ belong to $C = \mathbb{R}^p$ for all a. On the other hand, if the cone C is pointed (i.e., $a \in C$ and $-a \in C$ imply that $a = 0$), then the bounds (2.10) and (2.11) are one sided by the nature of cone constraints. In particular, if $C := \{x : x = ta,\ t \geqslant 0\}$ is the ray generated by a vector $a \in \mathbb{R}^p$, then $w_0 = w_1 = 1/2$ and all other weights are zero. In that case $c_\alpha = F_{2\alpha,1,v}$ and (2.10) is the usual upper $100(1-\alpha)\%$ confidence bound for $a'\gamma$.

Example 2.1. Consider the linear regression model (1.1) with k predictors, $n \times p$ design matrix X, where $p = k+1$, and $\gamma' = (\gamma_0, \gamma_1, \ldots, \gamma_k) \in \mathbb{R}^p$ (as usual, the first column of X is column of ones). Suppose that X has full column rank p. Let $\hat{\gamma} := (X'X)^{-1}X'Y$ be the least squares estimator of γ and $S^2 := (n-p)^{-1}\sum_{i=1}^n e_i^2$ be the standard (unbiased) estimator of σ^2. By the theory of linear regression (e.g., Seber, 1977) we have that $\hat{\gamma} \sim \mathrm{N}(\gamma, \sigma^2 V)$, S^2 is independent of $\hat{\gamma}$ and $vS^2/\sigma^2 \sim \chi_v^2$, where $V := (X'X)^{-1}$ and $v := n-p$. Then $100(1-\alpha)\%$ confidence intervals for $a'\gamma$, simultaneously for all $a' = (a_0, a_1, \ldots, a_k) \in \mathbb{R}^p$, are given by formula (1.2). Suppose now that we want to consider simultaneous confidence intervals for all a of the form $a_0 = 1$ and $a_i \geqslant 0$, $i = 1, \ldots, k$. That is, we want to construct simultaneous confidence intervals for responses at prediction points with nonnegative coordinates. Let us observe that $a'\gamma$ belongs to an interval $L \subset \mathbb{R}$ for some $a \in \mathbb{R}_+^p$, with $a_0 > 0$, if and only if $a_0^{-1} a \in a_0^{-1} L$. Therefore, constructing simultaneous confidence intervals for $a'\gamma$ with $a \in \mathbb{R}_+^p$ and $a_0 = 1$ is the same as constructing simultaneous confidence intervals for $a'\gamma$ with $a \in \mathbb{R}_+^p$. The developed theory can be then applied in a straightforward way for the cone $C := \mathbb{R}_+^p$.

References

Rockafellar, R.T., 1970. Convex Analysis. Princeton University Press, Princeton, NJ.

Seber, G.A.F., 1977. Linear Regression Analysis. Wiley, New York.

Sen, P.K., Silvapulle, M.J., 2002. An appraisal of some aspects of statistical inference under inequality constraints. J. Statist. Plann. Inference 107, 3–43.

Shapiro, A., 1988. Towards a unified theory of inequality constrained testing in multivariate analysis. Internat. Statist. Rev. 56, 49–62.