

IMPROVED CONFIDENCE ESTIMATORS FOR THE MULTIVARIATE NORMAL CONFIDENCE SET

Hsiuying Wang

Academia Sinica, Taipei

Abstract: Traditionally, the constant coverage probability estimator, confidence coefficient, is used to report the confidence of a multivariate normal confidence set. Robinson (1979), Lu and Berger (1989) and Robert and Casella (1994) all showed that there are certain estimators better than the confidence coefficient when the number of unknown parameters is greater than 4. In this paper some other better estimators are provided.

Key words and phrases: Admissibility, confidence coefficient, coverage function, the usual constant coverage probability estimator.

1. Introduction

Let $X = (X_1, \dots, X_p)'$ be a random vector with distribution $N(\theta, I_p)$, where $\theta = (\theta_1, \dots, \theta_p)'$. The usual $1 - \alpha$ confidence set for θ is

$$C_X = \{\theta : |X - \theta| \leq c\}, \quad (1)$$

where c satisfies $P(\theta \in C_X) = 1 - \alpha$. The coverage function of this confidence set is defined to be

$$I(\theta \in C_X) = \begin{cases} 1, & \text{if } \theta \in C_X, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Traditionally we use the confidence coefficient $1 - \alpha$, the constant coverage probability estimator, to report the confidence of C_X . However, $1 - \alpha$ is a data-independent confidence report. Kiefer (1977) pointed out that a better approach is to provide a data-dependent confidence report if there exists one which is better than the confidence coefficient. In the following, we give an example to explain why a data-independent estimator is not feasible in some situations.

Assume that the possible values of random variable Y are $\eta - 1$ and $\eta + 1$, where η is an unknown parameter, and the probabilities assigned to the possible values of Y are

$$p(Y = \eta - 1) = p(Y = \eta + 1) = \frac{1}{2}.$$

Suppose we have two observations y_1 and y_2 of Y . A usual confidence interval of η based on these two observations is

$$C_y = \left\{ \eta : \left| \frac{y_1 + y_2}{2} - \eta \right| < \frac{1}{2} \right\}.$$

When $y_1 \neq y_2$, then $(y_1 + y_2)/2 = \eta$, which implies that C_y covers the parameter η and $I(\eta \in C_y) = 1$. In this case, the confidence report should be 1. On the other hand, when $y_1 = y_2$, then $(y_1 + y_2)/2 = \eta + 1$ or $\eta - 1$, which implies that C_y does not cover the parameter η and $I(\eta \in C_y) = 0$. Hence, a reasonable confidence report in this case is 0. We see that the confidence report for C_y should depend on the observations, and a data-dependent estimator is better for estimating the confidence than a data-independent one. Moreover, this simple example not only explains why we should consider data-dependent confidence reports but also demonstrates that reporting the confidence of (1) is equivalent to estimating (2).

For the confidence set (1), Robinson (1979) showed that $1 - \alpha + d/(1 + |X|^2)$ is better than $1 - \alpha$ under squared error loss if $p = 5$ and d is small enough. That is, when $p = 5$ and d is a sufficiently small positive constant,

$$E\left(1 - \alpha + \frac{d}{1 + |X|^2} - I(\theta \in C_X)\right)^2 < E\left(1 - \alpha - I(\theta \in C_X)\right)^2$$

for all θ . Lu and Berger (1989) showed that when $p \geq 5$, the estimator

$$1 - \alpha + \frac{a}{b + |X|^2} \tag{3}$$

is better than $1 - \alpha$ under squared error loss for b large enough and a small enough. Robert and Casella (1994) showed that $1 - \alpha$ is inadmissible for $p \geq 5$ under squared error loss by an approach different from the above two papers, and used statistical simulation to show that

$$\beta(X) = 1 - \alpha + \frac{e}{|X|^2}$$

is better than $1 - \alpha$, where e is some positive constant.

When $p \leq 4$, Brown and Hwang (1990) showed that $1 - \alpha$ is an admissible estimator. Wang (1998) proved a similar result that the constant coverage probability estimator is admissible for reporting certain confidence intervals when the number of unknown parameters is smaller than 5. Moreover, Wang (1997) showed that the constant coverage probability estimator is inadmissible in regression models when the number of slope parameters is greater than 5.

In this paper, we consider estimators

$$r_n(X) = 1 - \alpha + \sum_{i=1}^n \frac{a_i}{(b + |X|^2)^i}, \quad n \geq 1, \tag{4}$$

where a_i and b , $i = 1, \dots, n$, are some positive constants. We show that $r_n(X)$ is better than $r_{n-1}(X)$, $n \geq 2$, under squared error loss, provided $p > 2(n + 1)$.

2. Main Results

The fact that the estimator (3) is better than $1 - \alpha$ can be explained from the empirical Bayes point of view. For $X \sim N(\theta, I_p)$, a conjugate prior for θ is

$$\theta|\tau \sim N(0, \tau^2 I_p),$$

where $\tau \in R$. Using $|X|^2/\zeta$ to estimate $1 + \tau^2$, the empirical Bayes estimator of $I(\theta \in C_X)$ with respect to the prior is

$$P(\theta \in C_X) = F_{p, \frac{\zeta^2}{|X|^2 - \zeta}}(c + \frac{\zeta c}{|X|^2 - \zeta}), \tag{5}$$

where ζ is some positive constant and $F_{p,g}(\cdot)$ is the distribution function of the non-central chi-square random variable with p degrees of freedom and noncentrality parameter g . A Taylor expansion gives

$$P(\theta \in C_x) \simeq 1 - \alpha + \frac{k_1}{|X|^2 - \zeta} + \frac{k_2}{(|X|^2 - \zeta)^2} + \frac{k_3}{(|X|^2 - \zeta)^3} + \dots, \tag{6}$$

where k_i are some constants. Based on the first order Taylor expansion, (3) is obtained. Similarly by using n terms, we derive the estimators in (4).

Theorem 1. For $n \geq 2, p > 2(n + 1)$,

$$a_1 < nE\{(X_1 - \theta_1)^2[1 - \alpha - I(|X - \theta| \leq c)]\}[p - 2(n + 1)] \tag{7}$$

and $\sum_{i=2}^n a_i < \infty$, there exists a b_0 such that if $b > b_0$,

$$E_\theta[r_n(X) - I(|X - \theta| \leq c)]^2 < E_\theta[r_{n-1}(X) - I(|X - \theta| \leq c)]^2, \quad \text{for all } \theta.$$

Before we prove Theorem 1, we need the following lemma.

Lemma 1.

$$\begin{aligned} (b + |z + \theta|^2)^{-i} &= (b + |\theta|^2)^{-i} - 2i \sum_{j=1}^p z_j (b + |\theta|^2)^{-i-1} \theta_j \\ &\quad + 2i(i + 1) \sum_{j,k=1}^p z_j z_k (b + |\theta|^2)^{-i-2} \theta_j \theta_k - i \sum_{j=1}^p z_j^2 (b + |\theta|^2)^{-i-1} \\ &\quad + o((b + |\theta|^2)^{-i-1}), \end{aligned}$$

where $z = (z_1, \dots, z_p)'$ and $\theta = (\theta_1, \dots, \theta_p)'$.

Proof. By Taylor expansion ($b \rightarrow \infty$).

Proof of Theorem 1. By a straightforward calculation,

$$\begin{aligned} & E[r_{n-1}(X) - I(|X - \theta| \leq c)]^2 - E[r_n(X) - I(|X - \theta| \leq c)]^2 \\ &= -2E[r_{n-1}(X) - I(|X - \theta| \leq c)] \frac{a_n}{(b + |X|^2)^n} - E\left(\frac{a_n^2}{(b + |X|^2)^{2n}}\right). \end{aligned}$$

Replacing $X - \theta$ by $Z = (Z_1, \dots, Z_p)'$ where $Z \sim N(0, I_p)$, the last expression is equal to

$$\begin{aligned} & -2E[1 - \alpha - I(|Z| \leq c)] \frac{a_n}{(b + |Z + \theta|^2)^n} - 2E \sum_{i=1}^{n-1} \left(\frac{a_i a_n}{(b + |Z + \theta|^2)^{i+n}}\right) \\ & - E\left(\frac{a_n^2}{(b + |Z + \theta|^2)^{2n}}\right). \end{aligned} \tag{8}$$

By Lemma 1 together with $E\{Z_i[1 - \alpha - I(|Z| \leq c)]\} = 0$, $E\{Z_i Z_j [1 - \alpha - I(|Z| \leq c)]\} = 0$ ($i \neq j$), and $E\{Z_i^2 [1 - \alpha - I(|Z| \leq c)]\} > 0$ (both functions are increasing in Z_i^2 and hence the covariance is positive), (8) becomes

$$\begin{aligned} & \frac{2a_n}{(b + |\theta|^2)^{n+2}} \{n|\theta|^2[kp - 2(n + 1)k] + knpb - a_1(b + |\theta|^2)\} \\ & + a_n \left(\sum_{i=2}^n a_i + 1\right) o\left(\frac{1}{(b + |\theta|^2)^{n+1}}\right) \\ & = \frac{2a_n}{(b + |\theta|^2)^{n+2}} \{|\theta|^2\{nk[p - 2(n + 1)] - a_1\} + b(knp - a_1)\} \\ & + a_n \left(\sum_{i=2}^n a_i + 1\right) o\left(\frac{1}{(b + |\theta|^2)^{n+1}}\right), \end{aligned} \tag{9}$$

where $k = E\{Z_1^2 [1 - \alpha - I(|Z| \leq c)]\}$. The fact that the error term is $o((b + |\theta|^2)^{-n-1})$ follows from an argument similar to the one that was used to derive (2.9) of Hwang and Brown (1991). Therefore, there exists a b_0 such that the expression at (9) is positive for all θ if $b \geq b_0$, $p > 2(n + 1)$, $a_1 < nk[p - 2(n + 1)]$ and $\sum_{i=2}^n a_i < \infty$.

When $n = 1$, we define $r_0(X) = 1 - \alpha$. Then by a similar argument,

$$\begin{aligned} & E[r_0(X) - I(|X - \theta| \leq c)]^2 - E[r_1(X) - I(|X - \theta| \leq c)]^2 \\ &= \frac{2a_1}{(b + |\theta|^2)^3} \{k[|\theta|^2(p - 4) + pb] - a_1(b + |\theta|^2)\} + o\left(\frac{a_1}{(b + |\theta|^2)^2}\right). \end{aligned} \tag{10}$$

Hence when $p \geq 5$, $r_1(X)$ is better than $r_0(X)$ if a_1 is small enough and b is large enough. From Theorem 1, (7) provides an upper bound for a_1 in r_n . Also from (10), $k(p - 4)$ is an upper bound for a_1 in r_1 . Thus, by straightforward calculation, for $p \geq 8$, the upper bound for a_1 in r_2 is larger than that for a_1 in r_1 . Some upper bounds for a_1 are provided in Table 1. The confidence set in Table 1 is chosen such that the confidence level $1 - \alpha$ is 0.8.

Table 1.

p	8	10	12	15
upper bound for a_1 in r_1	0.61	0.79	1.01	1.09
upper bound for a_1 in r_2	0.61	1.05	1.51	1.78

3. Simulation Results

Although Robinson and Lu and Berger showed that $1 - \alpha + d/(1 + |X|^2)$ and (3) are better than the confidence coefficient in estimating (2), respectively, they did not specify the values of d , a and b in their estimators. Robert and Casella specified the value of e in $\beta(X)$. I have done some simulations(not reported here) to compare the estimator (3) with $\beta(X)$. These results reveal that $\beta(X)$ is better than (3). Hence, in the following, we only compare the risks of our estimator, when $n = 2$, with those of Robert and Casella and with $1 - \alpha$.

Let $R(r(X), \theta)$ denote the risk $E[r(X) - I(|X - \theta| \leq c)]^2$, where $r(X)$ is an estimator of (2). Table 2 compares the risks of $r_2(X)$ with those of $\beta(X)$ and $1 - \alpha$ for $p = 8, c = 3.33$ ($1 - \alpha = 0.8$), where a_1, a_2 and b in $r_2(X)$ are 0.1, 9 and 4, respectively, and e in $\beta(X)$ is 0.443. Note that the value of e in $\beta(X)$ is suggested by Robert and Casella.

Table 2.

$ \theta $	0	1	2	5	10
$\frac{R(r_2(X), \theta)}{R(\beta(X), \theta)}$	0.980	0.982	0.988	0.990	1.000
$\frac{R(r_2(X), \theta)}{R(1 - \alpha, \theta)}$	0.930	0.936	0.958	0.990	1.000

(100,000 replications)

Table 3 gives the results when $p = 15$ and $c = 4.42$ ($1 - \alpha = 0.8$), where a_1, a_2 and b in $r_2(X)$ are 0.1, 27 and 6, respectively and e in $\beta(X)$ is 0.919. The value of e is also suggested by Robert and Casella.

Table 3.

$ \theta $	0	1	2	5	10
$\frac{R(r_2(X),\theta)}{R(\beta(X),\theta)}$	0.980	0.982	0.984	0.990	1.000
$\frac{R(r_2(X),\theta)}{R(1-\alpha,\theta)}$	0.943	0.953	0.962	0.990	1.000

(100,000 replications)

Acknowledgements

I wish to thank the Editor, Prof. C. F. Jeff Wu, and the referees for their suggestions and comments.

References

- Brown, L. D. and Hwang, J. T. Gene (1990). Admissibility of confidence estimators. In *Proceedings, 1990 Taipei Symposium in Statistics* (Edited by M. T. Chao and P. E. Cheng). Academia Sinica, Taiwan.
- Hwang, J. T. and Brown, L. D. (1991). Estimated confidence under the validity constraint. *Ann. Statist.* **19**, 1964-1977.
- Kiefer, J. (1977). Conditional confidence statements and confidence estimators. *J. Amer. Statist. Assoc.* **72**, 789-827.
- Lu, K. L. and Berger, J. O. (1989). Estimated confidence procedures for multivariate normal means. *J. Statist. Plann. Inference* **23**, 1-19.
- Robinson, G. K. (1979). Conditional properties of statistical procedures for location and scale parameters. *Ann. Statist.* **7**, 756-771.
- Robert, C. and Casella, G. (1994). Improved confidence estimators for the usual multivariate normal confidence set. In *Statistical Decision Theory and Related Topics V* (Edited by S. S. Gupta and J. O. Berger), 351-368, Academic, New York.
- Wang, H. (1998). Admissibility of the constant coverage function of certain confidence interval. *Statist. Probab. Lett.* **36**, 365-372.
- Wang, H. (1996). Brown's paradox in the estimated confidence approach. *Ann. Statist.* **27**, 610-626.

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.

(Received December 1997; accepted June 1999)