Canad. Math. Bull. Vol. 15 (4), 1972

ON UNBIASED ESTIMATION OF A VECTOR PARAMETER

BY
K. L. MEHRA AND P. V. RAMACHANDRAMURTY

1. **Summary.** It is shown in this paper that Rao's criterion of comparing two unbiased estimators on the basis of definiteness of the difference between their dispersion matrices is equivalent to Cramer's criterion based on their concentration ellipsoids. When the estimators have normal distributions it is shown that both the criteria have a desirable property in terms of the probabilities of the estimators lying in ellipsoids with the parameter point as the center.

2. **Criteria for comparison.** Let $\mathbf{X}=(X_1, \ldots, X_n)'$ be a random vector with distribution function $F_\lambda(\mathbf{x}')=F_\lambda(x_1, \ldots, x_n)$ an element of the family $\{F_\lambda : \lambda \in \Lambda\}$, where Λ may be an abstract index set, and consider the problem of estimating a vector parameter $\boldsymbol{\theta}=\theta(\lambda)=(\theta_1, \ldots \theta_p)'(p\geq 1)$ on the basis of $\mathbf{X}$. For the case $p=1$, namely, when θ is real, one criterion of a "good" estimator commonly used is that it be unbiased and of uniformly minimum variance. (If for a given problem a uniformly minimum variance unbiased estimator does not exist, then of any two unbiased estimators the one with the smaller variance is preferred.) The former has obvious justification and the latter is justified on the following two grounds: Firstly from the decision theoretic point of view variance is the risk corresponding to the "squared error" loss function. Secondly, in the important case when two unbiased estimators have normal distributions the one with the smaller variance has larger probability of lying in any interval containing the true parameter value θ. In the same strain, for comparing unbiased estimators of a vector parameter $\boldsymbol{\theta}$ (i.e., $p>1$), three criteria have been commonly employed in literature, which will be referred to in the sequel as Cramer's Criterion (CC), Rao's Criterion (RC) and Wilk's Criterion (WC): Let $\boldsymbol{\delta}_1=(\delta_{11}, \delta_{12}, \ldots \delta_{1p})'$ and $\boldsymbol{\delta}_2=(\delta_{21}, \delta_{22}, \ldots \delta_{2p})'$ be two unbiased estimators for $\boldsymbol{\theta}$ with finite dispersion matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively. Suppose that both $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ have nonsingular distributions. Let $\succ(\succcurlyeq)$ be read as "better than" ("at least as good"), $|\boldsymbol{\Sigma}|$ as the determinant of $\boldsymbol{\Sigma}$ and let $E_i=\{\mathbf{x}:(\mathbf{x}-\boldsymbol{\theta})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\theta})\leq p+2\}$, $i=1, 2$. Following Cramer [1], we shall call $(\mathbf{x}-\boldsymbol{\theta})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\theta})=p+2$, the concentration ellipsoid of $\boldsymbol{\delta}_i$, $i=1, 2$. (If $K>0$ and $\mathbf{M}$ is a positive definite (p.d.) matrix, then $(\mathbf{x}-\boldsymbol{\theta})'\mathbf{M}(\mathbf{x}-\boldsymbol{\theta})=K$ always defines an ellipsoid (see [4, p. 407]). E_1 will be said to be *strictly contained* in E_2 if $\mathbf{x} \in E_1$ implies that $\mathbf{x}$ belongs to the interior of E_2.

Received by the editors January 4, 1971.

https://doi.org/10.4153/CMB-1972-096-3 Published online by Cambridge University Press

CC. *Cramer's Criterion* $\boldsymbol{\delta}_1 \geqslant \boldsymbol{\delta}_2 (\boldsymbol{\delta}_1 > \boldsymbol{\delta}_2)$ iff E_1 is contained (strictly contained) in E_2. (See [1, pp. 283–285].)

RC. *Rao's Criterion* $\boldsymbol{\delta}_1 \geqslant \boldsymbol{\delta}_2 (\boldsymbol{\delta}_1 > \boldsymbol{\delta}_2)$ iff $(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)$ is positive semi-definite (positive definite). (See [2].)

WC. *Wilk's Criterion* $\boldsymbol{\delta}_1 \geqslant \boldsymbol{\delta}_2$ $(\delta_1 > \delta_2)$ iff $|\boldsymbol{\Sigma}_1| \leq |\boldsymbol{\Sigma}_2| (|\boldsymbol{\Sigma}_1| < |\boldsymbol{\Sigma}_2|)$. (See [5].) ($|\boldsymbol{\Sigma}_1|$ is known as the generalized variance of $\boldsymbol{\delta}_1$ etc.) whereas it is common knowledge that both Cramer's and Rao's criteria imply Wilk's criterion, it seems that the equivalence of Cramer's and Rao's criteria has been overlooked. We prove this fact in

THEOREM 1. *If $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are nonsingular, then*

(i) $\boldsymbol{\delta}_1 \overset{R}{\geqslant} \boldsymbol{\delta}_2 \Leftrightarrow \boldsymbol{\delta}_1 \geqslant \boldsymbol{\delta}_2 \Rightarrow \boldsymbol{\delta}_1 \geqslant \boldsymbol{\delta}_2$, *and*

(ii) *these assertions remain true if $\geqslant$ is replaced by $>$ throughout.*

Proof. Assume without loss of generality that $\boldsymbol{\theta} = \mathbf{0}$. Then

$$
\begin{aligned}
\delta_1 \overset{R}{\geqslant} \delta_2 &\Rightarrow (\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1) && \text{is positive semi-definite (p.s.d.)} \\
&\Rightarrow (\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}) && \text{is p.s.d. (See [3, p. 56, Exercise 9].)} \\
&\Rightarrow \mathbf{x}'\boldsymbol{\Sigma}_1^{-1}\mathbf{x} \geq \mathbf{x}'\boldsymbol{\Sigma}_2^{-1}\mathbf{x} && \text{for all } \mathbf{x}.
\end{aligned}
$$

Thus if $\mathbf{x}^*$ lies on the boundary of E_1, it follows from $p+2 = \mathbf{x}^{*\prime}\boldsymbol{\Sigma}_1^{-1}\mathbf{x}^* \geq \mathbf{x}^{*\prime}\boldsymbol{\Sigma}_2^{-1}\mathbf{x}^*$ that $\mathbf{x}^* \in E_2$. It follows that E_1 is contained in E_2, so that RC$\Rightarrow$CC. Conversely, suppose $\boldsymbol{\delta}_1 \overset{C}{\geqslant} \boldsymbol{\delta}_2$. Since by assumption $\boldsymbol{\Sigma}_1(\boldsymbol{\Sigma}_2)$ and consequently $\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\Sigma}_2^{-1})$ is p.d., for all $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}'\boldsymbol{\Sigma}_1^{-1}\mathbf{x} > 0$. Let $C = [(p+2)/(\mathbf{x}'\boldsymbol{\Sigma}_1^{-1}\mathbf{x})]^{1/2}$. Then $(C\mathbf{x})'\boldsymbol{\Sigma}_1^{-1}(C\mathbf{x}) = p+2$ implies that $(C\mathbf{x}) \in E_1 \subset E_2$, so that

$$
\begin{aligned}
(C\mathbf{x})\boldsymbol{\Sigma}_2^{-1}(C\mathbf{x}) &\leq p+2 \\
&\Rightarrow C^2\{\mathbf{x}'\boldsymbol{\Sigma}_1^{-1}\mathbf{x} - \mathbf{x}'\boldsymbol{\Sigma}_2^{-1}\mathbf{x}\} \geq 0 \quad \text{for all } \mathbf{x} \neq 0 \\
&\Rightarrow (\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}) \quad \text{is p.s.d.} \\
&\Rightarrow (\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1) \quad \text{is p.s.d.}
\end{aligned}
$$

Thus CC$\Leftrightarrow$RC. That $\boldsymbol{\delta}_1 \overset{R}{\geqslant} \delta_2 \Rightarrow \boldsymbol{\delta}_1 \overset{W}{\geqslant} \boldsymbol{\delta}_2$ follows from the fact that if $(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)$ is p.s.d., $|\boldsymbol{\Sigma}_2| \geq |\boldsymbol{\Sigma}_1|$. This completes the proof of part (i). Part (ii) can be proved similarly by retracing steps in the above argument with obvious modifications. This completes the proof of the theorem.

It is well known and easy to see that $(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)$ is positive semi-definite if and only if the variance $(\boldsymbol{l}'\boldsymbol{\delta}_2) \geq$variance $(\boldsymbol{l}'\boldsymbol{\delta}_1)$ for every p-vector $\boldsymbol{l}$. In fact as far as the authors are aware this is the only justification given for Rao's criterion. The justification for Cramer's criterion is that the "smaller" the concentration ellipsoid of a vector estimate $\boldsymbol{\delta}$, the larger is its (probabilistic) "concentration" around the true parameter value $\boldsymbol{\theta}$ (see [1]). We see from the above theorem that both CC and RC have both the desirable properties mentioned above.

https://doi.org/10.4153/CMB-1972-096-3 Published online by Cambridge University Press

3. **Probability concentration and dispersion matrices.** As remarked in §2, for the real parameter case when the two unbiased estimators have normal distributions, the one with the smaller variance has higher probability of falling in any interval containing the true parameter value. We prove in the following theorem that a similar desirable property is also satisfied for the case $p>1$ when the two unbiased estimators have multivariate normal distributions.

THEOREM 2. *Suppose* $\boldsymbol{\delta}_1$ *is* $N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_1)$ *and* $\boldsymbol{\delta}_2$ *is* $N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_2)$ *with* $\boldsymbol{\Sigma}_1$ *and* $\boldsymbol{\Sigma}_2$ *nonsingular and* $(\boldsymbol{\Sigma}_2-\boldsymbol{\Sigma}_1)$ *is* p.s.d. *Then*

(i) *For any sphere* $S=\{\mathbf{x}:(\mathbf{x}-\boldsymbol{\theta})'(\mathbf{x}-\boldsymbol{\theta})\le K\}$

$$P[\boldsymbol{\delta}_1 \in S] \geq P[\boldsymbol{\delta}_2 \in S]$$

(ii) *For any ellipsoid* $E=\{\mathbf{x}:(\mathbf{x}-\boldsymbol{\theta})'\mathbf{M}(\mathbf{x}-\boldsymbol{\theta})\le K\}$ *where* $\mathbf{M}$ *is* p.d.,

$$P[\boldsymbol{\delta}_1 \in E] \geq P[\boldsymbol{\delta}_2 \in E].$$

If $(\boldsymbol{\Sigma}_2-\boldsymbol{\Sigma}_1)$ *is* p.d. *then the above inequalities are strict.*

Proof. Without loss of generality assume that $\boldsymbol{\theta}=\mathbf{0}$. Being dispersion matrices of nonsingular distributions, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are symmetric p.d. Hence there exist symmetric p.d. matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{A}^2=\boldsymbol{\Sigma}_2$ and $\mathbf{B}^2=\boldsymbol{\Sigma}_1$ (see [3, p. 55, Problem 3.5]). Since $(\mathbf{A}\mathbf{B}^{-1})\boldsymbol{\Sigma}_1(\mathbf{A}\mathbf{B}^{-1})'=\mathbf{A}\mathbf{B}^{-1}\mathbf{B}^2\mathbf{B}^{-1}\mathbf{A}=\mathbf{A}^2=\boldsymbol{\Sigma}_2$, it follows that $\mathbf{A}\mathbf{B}^{-1}\boldsymbol{\delta}_1$ has the same distribution as $\boldsymbol{\delta}_2$ viz., $N(\mathbf{0}, \boldsymbol{\Sigma}_2)$. Therefore

$$\begin{aligned} P[\boldsymbol{\delta}_2'\boldsymbol{\delta}_2 \leq K] &= P[(\mathbf{A}\mathbf{B}^{-1}\boldsymbol{\delta}_1)'(\mathbf{A}\mathbf{B}^{-1}\boldsymbol{\delta}_1) \leq K] \\ &= P[\boldsymbol{\delta}_1'\mathbf{B}^{-1}\mathbf{A}\mathbf{A}\mathbf{B}^{-1}\boldsymbol{\delta}_1 \leq K] \\ &= P[\boldsymbol{\delta}_1'\mathbf{B}^{-1}\boldsymbol{\Sigma}_2\mathbf{B}^{-1}\boldsymbol{\delta}_1 \leq K]. \end{aligned} \tag{3.1}$$

Now

$$\mathbf{B}(\mathbf{B}^{-1}\boldsymbol{\Sigma}_2\mathbf{B}^{-1}-\mathbf{I})\mathbf{B} = \boldsymbol{\Sigma}_2-\mathbf{B}^2 = \boldsymbol{\Sigma}_2-\boldsymbol{\Sigma}_1,$$

so that, $\mathbf{B}$ being symmetric and nonsingular and $(\boldsymbol{\Sigma}_2-\boldsymbol{\Sigma}_1)$ being p.s.d., it follows that $(\mathbf{B}^{-1}\boldsymbol{\Sigma}_2\mathbf{B}^{-1}-\mathbf{I})$ is p.s.d. (see [3, p. 31, (i)]). Hence

$$\begin{aligned} P[\boldsymbol{\delta}_1'\boldsymbol{\delta}_1 \leq K] &= P[\boldsymbol{\delta}_1'\mathbf{I}\boldsymbol{\delta}_1 \leq K] \\ &\geq P[\boldsymbol{\delta}_1'\mathbf{B}^{-1}\boldsymbol{\Sigma}_2\mathbf{B}^{-1}\boldsymbol{\delta}_1 \leq K] \\ &= P[\boldsymbol{\delta}_2'\boldsymbol{\delta}_2 \leq K], \end{aligned} \tag{3.2}$$

where the last equality follows from (3.1). This completes the proof of part (i). For part (ii), note that since M is p.d., there exists a nonsingular matrix $\mathbf{R}$ such that $\mathbf{R}'\mathbf{M}\mathbf{R}=\mathbf{I}$. Let $\boldsymbol{\xi}_1=\mathbf{R}^{-1}\boldsymbol{\delta}_1$, $\boldsymbol{\xi}_2=\mathbf{R}^{-1}\boldsymbol{\delta}_2$ so that $\boldsymbol{\delta}_1=\mathbf{R}\boldsymbol{\xi}_1$, $\boldsymbol{\delta}_2=\mathbf{R}\boldsymbol{\xi}_2$. Then

$$\begin{aligned} P[\boldsymbol{\delta}_2'\mathbf{M}\boldsymbol{\delta}_2 \leq K] &= P[\boldsymbol{\xi}_2'\mathbf{R}'\mathbf{M}\mathbf{R}\boldsymbol{\xi}_2 \leq K] \\ &= P[\boldsymbol{\xi}_2'\boldsymbol{\xi}_2 \leq K]. \end{aligned}$$

Similarly $P[\boldsymbol{\delta}_1'\mathbf{M}\boldsymbol{\delta}_1\le K]=P[\boldsymbol{\xi}_1'\boldsymbol{\xi}_1\le K]$. Since $\boldsymbol{\xi}_i$ is $N(\mathbf{0}, \boldsymbol{\Lambda}_i)$ where $\boldsymbol{\Lambda}_i=\mathbf{R}^{-1}\boldsymbol{\Sigma}_i(\mathbf{R}^{-1})'$, $i=1, 2$ and $(\boldsymbol{\Lambda}_2-\boldsymbol{\Lambda}_1)$ is p.d. because $(\boldsymbol{\Sigma}_2-\boldsymbol{\Sigma}_1)$ is so, it follows from (i) of this

https://doi.org/10.4153/CMB-1972-096-3 Published online by Cambridge University Press

theorem that $P[\boldsymbol{\xi}_2'\boldsymbol{\xi}_2 \leq K] \leq P[\boldsymbol{\xi}_1\boldsymbol{\xi}_1 \leq K]$ and hence $P[\boldsymbol{\delta}_2'\mathbf{M}\boldsymbol{\delta}_2 \leq K] \leq P[\boldsymbol{\delta}_1'\mathbf{M}\boldsymbol{\delta}_1 \leq K]$. This completes the proof of part (ii). That the inequalities in (i) and (ii) are strict if $(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)$ is p.d., follows because in this case the continuity properties of normal distributions imply that the appropriate inequality in (3.2) is strict.

In view of Theorem 2, if one takes relative concentration of $\boldsymbol{\delta}_1$- and $\boldsymbol{\delta}_2$-probabilities in (appropriate) regions containing the true value $\boldsymbol{\theta}$ as a criterion for the comparison of estimators, it makes little sense to compare dispersion matrices in complete disregard of whether or not the distributions of $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are normal. However, since asymptotically, many estimators have limiting normal distributions, the comparison of dispersion matrices in the sense of RC (or CC) is justifiable in such situations. The above considerations would apply, for example, to the obvious multivariate analogues of Neyman's Best Asymptotically Normal (BAN) estimators.

References

1. Harold Cramer, *Mathematical methods of statistics*, Princeton Univ. Press, Princeton, N.J., 1946.

2. C. R. Rao, *Minimum variance and the estimation of several parameters*, Proc. Cambridge Philos. Soc. **43** (1947), 280–283.

3. ———, *Linear statistical inference and its applications*, Wiley, New York, 1965.

4. Henry Scheffe, *Analysis of variance*, Wiley, New York, 1959.

5. S. S. Wilks, *Certain generalisations in the analysis of variance*, Biometrika **24** (1932), 471–494.

University of Alberta,
Edmonton, Alberta

https://doi.org/10.4153/CMB-1972-096-3 Published online by Cambridge University Press