The Annals of Statistics
1989, Vol. 17, No. 2, 920–928

EQUIVARIANT ESTIMATION IN A MODEL WITH AN ANCILLARY STATISTIC[1]

BY TAKEAKI KARIYA

Hitotsubashi University

This paper reformulates a result of Hora and Buehler on best equivariant estimators to treat a model admitting an ancillary statistic. The approach itself was established by Pitman, Girshick and Savage and Kiefer, and expanded by Zidek. The model considered in this paper is assumed to be generated as an orbit under a group acting on the parameter space. The general result obtained here is applied to a model in the Nile problem, a model with a known variation coefficient, a circle model and the GMANOVA model, and best equivariant estimators (BEE's) are derived. In the first two models, the BEE's dominate the MLE's uniformly.

1. Introduction. As is well known, an ancillary statistic is defined to be a statistic which is a part of a minimal sufficient statistic and whose marginal distribution is parameter-free. In this paper, in line with Pitman (1938), Girshick and Savage (1951), Kiefer (1957), Hora and Buehler (1966), Zidek (1969), Berger [(1980), page 245], Eaton [(1983), Proposition 7.12] and Lehmann [(1983), Chapter 3] among others, we formulate an invariance approach to estimation in a model admitting an ancillary statistic. A model in our formulation is assumed to be generated as an orbit under an induced group acting on the parameter space and there an ancillary statistic is realized as a maximal invariant. The approach here provides a systematic method for finding a best equivariant estimator (BEE) in such a model. A feature of our formulation will lie in the usefulness in applications rather than in its novelty. In fact, in Sections 3–5 the approach is applied to finding the BEE's in (1) a model of the Nile problem [see Fisher (1973)], (2) a model with a known variation coefficient [see Efron (1975), Cox and Hinkley (1974) and Hinkley (1977)], (3) a circle model [see Fisher (1973) and Amari (1982)] and (4) the GMANOVA (general MANOVA) model [see Gleser and Olkin (1970) and Kariya (1978, 1985)]. Since an MLE is equivariant under a mild condition in general [see Eaton (1983)], a BEE dominates the MLE uniformly unless the MLE is a BEE. The MLE's in the cases of (1) and (2) are not the BEE's and hence inadmissible.

Our results are also related to the work of Efron (1975, 1978) and Barndorff-Nielsen (1980).

2. A main result. Let z be a random variable taking on values in a measurable space $\mathscr{X}$ and let $\mathscr{P}(\tilde{\Theta}) = \{P_\theta | \theta \in \tilde{\Theta}\}$ be a class of probability

---

Received July 1983; revised September 1988.

[1]Supported by the Ministry of Education of Japan under General C59540104.

AMS 1980 *subject classifications*. Primary 62F10; secondary 62A05.

Key words and phrases. Invariance, best equivariant estimator, sufficient statistic, ancillary statistic, curved model, MLE, the Nile problem, normal model with a known variation coefficient, circle model, GMANOVA model.

measures on $\mathscr{X}$, where the parameter space $\tilde{\Theta}$ is a measurable space such that for $\theta_1 \neq \theta_2$ in $\tilde{\Theta}$, $P_{\theta_1} \neq P_{\theta_2}$. Further let

$$\Theta = \{\theta \in \tilde{\Theta} | \theta = \psi(\eta),\ \eta \in \Upsilon\} \tag{2.1}$$

be a subspace of $\tilde{\Theta}$ parametrized by η, where Υ is a measurable space and $\psi\colon \Upsilon \to \tilde{\Theta}$ is a known bimeasurable bijection from Υ onto its image $\psi(\Upsilon) \subset \tilde{\Theta}$. In applications, it is often the case that $\tilde{\Theta} \subset R^p$ and $\Upsilon \subset R^q$ with $q < p$ so that Θ is regarded as a "surface" of $\tilde{\Theta}$.

ASSUMPTION 2.1. There exists a topological group $\mathscr{G}$ acting measurably on $\mathscr{X}$ such that $\mathscr{P}(\tilde{\Theta})$ is invariant under $\mathscr{G}$ [i.e., $g\mathscr{P}(\tilde{\Theta}) = \mathscr{P}(\tilde{\Theta})$ with $gP_\theta = P_\theta \circ g^{-1}$ for all $g \in \mathscr{G}$] and such that for a maximal invariant parameter $\lambda(\theta)$ under the induced group $\bar{\mathscr{G}}$ acting on $\tilde{\Theta}$, the subspace Θ in (2.1) is expressed as

$$\Theta = \{\theta \in \tilde{\Theta} | \lambda(\theta) = \lambda_0\}, \tag{2.2}$$

where λ_0 is a known value in the space of $\lambda(\theta)$ and the map $g \to \bar{g}$ is measurable.

The condition (2.2) means that the subspace Θ in (2.1) is realized as the orbit of a point in $\tilde{\Theta}$ under the induced group $\bar{\mathscr{G}}$. Since (2.2) together with (2.1) implies $\bar{g}\psi(\eta) \in \Theta$ for any $\bar{g} \in \bar{\mathscr{G}}$ and $\eta \in \Upsilon$ or equivalently $\bar{g}\Theta = \Theta$ for all $\bar{g} \in \bar{\mathscr{G}}$ and since $\psi(\cdot)$ is injective, we can define a group acting on Υ by

$$\tilde{\mathscr{G}} = \{\tilde{g} | \tilde{g} = \psi^{-1}\bar{g}\psi \text{ for } \bar{g} \in \bar{\mathscr{G}}\}. \tag{2.3}$$

Then $\tilde{\mathscr{G}}$ is a homomorphic image of $\bar{\mathscr{G}}$ and so of $\mathscr{G}$ and hence the subfamily $\mathscr{P}(\Theta) = \{P_{\psi(\eta)} | \eta \in \Upsilon\}$ is invariant under $\tilde{\mathscr{G}}$. Further it is easily shown that the action of $\tilde{\mathscr{G}}$ on Υ is transitive.

Next, let $u = u(z)$ be a maximal invariant under $\mathscr{G}$. Then the distribution of u depends on $\theta \in \tilde{\Theta}$ only through $\lambda(\theta)$ and hence from (2.2) it depends only on the known value λ_0 for the subfamily $\mathscr{P}(\Theta)$. Consequently $u(z)$ is an ancillary statistic if z is minimally sufficient for $\mathscr{P}(\Theta)$.

ASSUMPTION 2.2. There is a bijective bimeasurable map π from $\mathscr{X}$ onto $\mathscr{G} \times \mathscr{U}$ such that if $\pi(z) = (h(z), u(z))$, then $\pi(gz) = (gh(z), u(z))$, where $\mathscr{U}$ is a measurable space.

Under this assumption, the map $h(z)$ from $\mathscr{X}$ onto $\mathscr{G}$ is measurable and equivariant, i.e., $h(gz) = gh(z)$ and $u(z)$ is regarded as a maximal invariant.

Now we consider the problem of estimating η with an invariant measurable loss $L\colon \Upsilon \times \Upsilon \to R$ satisfying

$$L(\tilde{g}a, \tilde{g}\eta) = L(a, \eta) \quad \text{for any } \tilde{g} \in \tilde{\mathscr{G}} \text{ and } a, \eta \in \Upsilon, \tag{2.4}$$

where $L(a, \eta) \geq 0$. Hence the problem is left invariant. Define the risk of an estimator $\hat{\eta}$ as

$$R(\hat{\eta}, \eta) = E_{\psi(\eta)} L(\hat{\eta}(z), \eta) \tag{2.5}$$

and let $\mathscr{D}_E$ denote the class of equivariant estimators of η. Regarding an estimator as a function of $\pi(z)$ by Assumption 2.2, $\hat{\eta} \in \mathscr{D}_E$ if and only if

$$\hat{\eta}(gh(z), u(z)) = \tilde{g}\hat{\eta}(h(z), u(z)) \quad \text{for all } g \in \mathscr{G} \text{ and } z \in \mathscr{Z}. \tag{2.6}$$

Since $h(z) \in \mathscr{G}$, (2.6) implies

$$\hat{\eta}(h(z), u(z)) = \tilde{h}(z)\hat{\eta}(e, u(z)) \quad \text{for } \hat{\eta} \in \mathscr{D}_E, \tag{2.7}$$

where e is the unit element of $\mathscr{G}$. Note that by (2.3), $\tilde{h}(z) = \psi^{-1}\bar{h}(z)\psi$ and so $\tilde{h}(z)$ does not depend on $\hat{\eta}$.

THEOREM 2.1. *Under Assumptions 2.1 and 2.2, a BEE, when it exists, is given by $\hat{\eta}_1(h(z), u(z)) = \tilde{h}(z)\hat{\eta}_1(e, u(z))$ with $\hat{\eta}_1(e, u(z))$ minimizing the conditional expectation*

$$E_{\psi(\eta_0)}\left[L(\tilde{h}(z)\hat{\eta}(e, u(z)), \eta_0)|u(z)\right], \tag{2.8}$$

where η_0 is any fixed value in Υ for which $R(\hat{\eta}, \eta_0) < \infty$ for some $\hat{\eta} \in \mathscr{D}_E$.

PROOF. Since $\tilde{\mathscr{G}}$ acts transitively on Υ, for a given $\hat{\eta} \in \mathscr{D}_E$ and for any $\eta \in \Upsilon$, $R(\hat{\eta}, \eta) = R(\hat{\eta}, \eta_0)$. Hence the minimum of $R(\hat{\eta}, \eta)$ is attained by $\hat{\eta} \in \mathscr{D}_E$ which minimizes (2.8), completing the proof. □

COROLLARY 2.1. *Suppose $\Upsilon \subset R^1$. Then under Assumptions 2.1 and 2.2, the estimator*

$$\hat{\eta}_1(z) = \tilde{h}(z) - E_0\left[\tilde{h}(z)|u(z)\right] \qquad (\in R^1) \tag{2.9}$$

is a BEE for the loss $L(a, \eta) = (a - \eta)^2$, and the estimator

$$\hat{\eta}_1(z) = \tilde{h}(z)E_1\left[\tilde{h}(z)|u(z)\right]/E_1\left[\tilde{h}(z)^2|u(z)\right] \qquad (\in R^1) \tag{2.10}$$

is a BEE for the loss $L(a, \eta) = (a - \eta)^2/\eta^2$.

These results are regarded as a modification of Hora and Buehler (1966) for a model with an ancillary statistic. The arguments used above are rather similar to Girshick and Savage (1951) and Kiefer (1957). In Kiefer (1957), Condition NR on page 579 states that $\bar{\mathscr{G}}$ rather than $\tilde{\mathscr{G}}$ acts transitively on $\tilde{\Theta}$. The case $\Theta = \tilde{\Theta}$ with ψ the identity is reduced to Kiefer's result.

Two remarks follow. First, the existence of a bimeasurable bijection between $\mathscr{Z}$ and $\mathscr{G} \times \mathscr{U}$ in Assumption 2.2 means that $\mathscr{G}$ acts freely on $\mathscr{Z}$ (i.e., $g \neq e$ implies $gz \neq z$ for any $z \in \mathscr{Z}$). Sometimes a model admits a larger group whose action is not free. In such a case, it is usually possible to choose a smaller group so that Assumptions 2.1 and 2.2 will be satisfied. It is noted that the smaller the group is, the larger the class of equivariant estimators $\mathscr{D}_E$. Second, as shown in Eaton (1983), the MLE is an equivariant estimator under a mild condition on the pdf. Hence the above approach provides a systematic method for making a uniform improvement on the MLE.

3. Applications.

3.1. *The Nile Problem.* Let (x_i, y_i)'s be iid from the pdf $(\eta e^{-\eta x})(\eta^{-1} e^{-y/\eta})$, $i = 1, \dots, n$, where $\eta > 0$ and $x, y > 0$. This is the model in the Nile problem considered by Fisher [see Fisher (1973) or Buehler (1980)]. Clearly the mean $(\bar{x}, \bar{y}) = (z_1, z_2) = z$ is a minimal sufficient statistic with

$$E(z) = (\eta^{-1}, \eta) = (\psi_1(\eta), \psi_2(\eta)) \equiv \psi(\eta) = (\theta_1, \theta_2) \equiv \theta. \tag{3.1}$$

The pdf of z is given by

$$\begin{aligned} P_\theta(dz) &= c\theta_1^{-n}\theta_2^{-n}(nz_1)^{n-1}(nz_2)^{n-1} \\ &\quad \times \exp[-(nz_1/\theta_1) - (nz_2/\theta_2)]\, dz_1\, dz_2. \end{aligned} \tag{3.2}$$

Here let $\mathscr{X} = \tilde{\Theta} = (R_+)^2$, $\Upsilon = R_+$ and

$$\Theta = \{\theta \in \tilde{\Theta} | \theta = \psi(\eta), \eta \in \Upsilon\}, \tag{3.3}$$

where $R_+ = \{a > 0\}$. Then $\mathscr{P}(\Theta) = \{P_\theta | \theta \in \Theta\}$ is a subfamily or curved model of $\mathscr{P}(\tilde{\Theta}) = \{P_\theta | \theta \in \tilde{\Theta}\}$. In fact, Θ is a hyperbola: $\Theta = \{\theta \in \tilde{\Theta} | \theta_1\theta_2 = 1\}$.

To apply our approach to the estimation of η, let the group $\mathscr{G} = R_+$ act on $\mathscr{X}$ by $z = (z_1, z_2) \to gz = (g^{-1}z_1, gz_2)$, which in turn induces the group $\bar{\mathscr{G}} = R_+$ acting on $\tilde{\Theta}$ by $\theta = (\theta_1, \theta_2) \to \bar{g}\theta = (\bar{g}^{-1}\theta_1, \bar{g}\theta_2)$. Then $\lambda(\theta) = \theta_1\theta_2$ is a maximal invariant under $\bar{\mathscr{G}}$ and so Θ is realized as an orbit: $\Theta = \{\theta \in \tilde{\Theta} | \lambda(\theta) = 1\}$. Further with $\psi(\eta)$ in (3.1), $\tilde{\mathscr{G}} = \{\tilde{g} | \tilde{g} = \psi^{-1}\bar{g}\psi,\ \bar{g} \in \bar{\mathscr{G}}\}$ acts on Υ by $\eta \to \tilde{g}\eta$. Hence Assumption 2.1 is satisfied. Next, observe that the map

$$\pi(z) = (h(z), u(z)) \equiv ((z_2/z_1)^{1/2}, (z_1 z_2)^{1/2}), \tag{3.4}$$

which is in one-one correspondence with z, gives a measurable isomorphism between $\mathscr{X}$ and $\mathscr{G} \times \mathscr{U}$: $\mathscr{X} = \mathscr{G} \times \mathscr{U}$. In fact, $h(z)$ satisfies $h(gz) = \tilde{g}h(z)$ and $u(z)$ is a maximal invariant. Hence Assumption 2.2 is satisfied. Therefore, taking the loss $L(a, \eta) = (a - \eta)^2/\eta^2 = (a/\eta - 1)^2$, by Corollary 2.1 the unique BEE is given by (2.10), and using the conditional pdf of h given u [see Fisher (1973)], it is evaluated as

$$\hat{\eta}_1 = h(z)K_1(2nu)/K_2(2nu) \tag{3.5}$$

since $E_1[h^i | u] = K_i(2nu)/K_0(2nu)$, $i = 1, 2$, where

$$K_\lambda(w) = \int_0^\infty x^{\lambda - 1} \exp[-\tfrac{1}{2}w(x + x^{-1})]\, dx$$

is the modified Bessel function of the third kind with index λ [see Abramowitz and Stegun (1965)].

The MLE is given by $h(z)$ in (3.4), which belongs to $\mathscr{D}_E$, and hence it is dominated by the BEE in (3.5). Further the bias-corrected MLE is given by

$$\hat{\eta}_{0b} = h(z) - b(h(z)) = \{2 - [\Gamma(n - \tfrac{1}{2})\Gamma(n + \tfrac{1}{2})/\Gamma(n)^2]\}h(z)$$

since the bias of the MLE $h(z)$ is computed as

$$b(\eta) = \left\{\left[\Gamma\left(n - \tfrac{1}{2}\right)\Gamma\left(n + \tfrac{1}{2}\right)/\Gamma(n)^2\right] - 1\right\}\eta.$$

Hence $\hat{\eta}_{0b} \in \mathscr{D}_E$ and it is uniformly dominated by the BEE $\hat{\eta}_1$.

3.2. *Normal model with known variational coefficient.* Let $x_1, \dots, x_n$ be iid from $N(\eta, a\eta^2)$ with $\eta \in \Upsilon = R_+$, where $a > 0$ is known and so it is assumed to be one below. This example was originally considered by Fisher and is treated in Efron (1975), Amari (1982), Cox and Hinkley (1974) and Hinkley (1977). Clearly $z = (z_1, z_2) = (\bar{x}, \Sigma(x_i - \bar{x})^2)$ is sufficient for

$$\mathscr{P}(\tilde{\Theta}) = \left\{N(\theta_1, \theta_2) | \theta = (\theta_1, \theta_2) \in \tilde{\Theta}\right\} \quad \text{with } \tilde{\Theta} = (R_+)^2$$

and for $\mathscr{P}(\Theta)$ as well, where

$$\Theta = \left\{\theta \in \tilde{\Theta} | \theta = \psi(\eta) = (\eta, \eta^2),\ \eta \in \Upsilon\right\}. \tag{3.6}$$

Then group $\mathscr{G} = R_+$ acts on $\mathscr{X} = R \times R_+$ by $(z_1, z_2) \to (gz_1, g^2 z_2)$, $u(z) = z_1/\sqrt{z_2}$ is a maximal invariant under $\mathscr{G}$ and $\lambda(\theta) = \theta_1^2/\theta_2$ is a maximal invariant parameter under the group $\bar{\mathscr{G}} = \mathscr{G}$ induced on $\tilde{\Theta}$. Consequently Θ in (3.6) is expressed as $\Theta = \{\theta \in \tilde{\Theta} | \lambda(\theta) = 1\}$ and $\tilde{\mathscr{G}} = \{\tilde{g} | \tilde{g} = \psi^{-1}\bar{g}\psi,\ \bar{g} \in \bar{\mathscr{G}}\} = \mathscr{G}$, which acts on Υ by $\eta \to \tilde{g}\eta$. Hence Assumption 2.1 is satisfied. Further, taking $h(z) = \sqrt{z_2}$, $\pi(z) = (h(z), u(z))$ gives a homeomorphism from $\mathscr{X}$ onto $\mathscr{G} \times \mathscr{U}$ and Assumption 2.2 is satisfied. Therefore taking the loss $L(a, \eta) = (a - \eta)^2/\eta^2$, by Corollary 2.1 the unique BEE is given by (2.10) and it is directly shown to be evaluated as

$$\hat{\eta}_1 = \sqrt{z_2}\, E_1\left[\sqrt{z_2} | u\right] / E_1[z_2 | u], \tag{3.7}$$

where the conditional expectations are given by

$$E_1\left[z_2^{i/2} | u\right] = \left[\sum_{j=0}^{\infty} \Gamma\left(\frac{n + j + i}{2}\right) \frac{v^{j+i}}{j!}\right] \Big/ \left[\sum_{j=0}^{\infty} \Gamma\left(\frac{n + j}{2}\right) \frac{v^j}{j!}\right], \quad i = 1, 2, \tag{3.8}$$

with $v = nu2^{1/2}/(nu^2 + 1)^{1/2}$.

On the other hand, the MLE is given by

$$\hat{\eta}_0 = -(\bar{x}/2) + \left[s^2 + 5\bar{x}^2/4\right]^{1/2} \quad \text{with } s^2 = z_2/n \tag{3.9}$$

and it is equivariant. Hence the MLE is uniformly dominated by the BEE. Similarly the bias-corrected MLE and the dual MLE proposed by Amari (1982) are inadmissible. Hinkley (1977) investigated some properties of the model.

3.3. *Circle model.* As an example for a non-location-scale model, we here consider the circle model treated in Fisher (1973). Let (x_i, y_i)'s be iid from the normal distribution $N(\theta, I_2)$, $i = 1, \dots, n$, where

$$\theta = (\theta_1, \theta_2)' \equiv \psi(\eta) = (\cos\eta, \sin\eta)' \quad \text{with } \eta \in \Upsilon \equiv (-\pi, \pi]. \tag{3.10}$$

Clearly $z = (z_1, z_2)' = (\bar{x}, \bar{y})'$ is sufficient. Let $\mathscr{X} = R^2 - \{0\}$, $\tilde{\Theta} = R^2$ and

$$\Theta = \left\{\theta \in \tilde{\Theta} | \theta_1^2 + \theta_2^2 = 1\right\}. \tag{3.11}$$

Naturally $\mathscr{P}(\tilde{\Theta}) = \{N(\theta, (1/n)I_2) | \theta \in \tilde{\Theta}\}$. In this setup we consider the problem of estimating η in (3.10). As is well known, the MLE of η is given by

$$\hat{\eta}_0(z) = \begin{cases} \tan^{-1}(z_2/z_1) \in (0, \pi] & \text{for } z_1 > 0, \\ \tan^{-1}(z_2/z_1) \in (-\pi, 0] & \text{for } z_1 < 0, \\ \pi/2 & \text{for } z_1 = 0 \text{ and } z_2 > 0, \\ -\pi/2 & \text{for } z_1 = 0 \text{ and } z_2 < 0. \end{cases} \tag{3.12}$$

It is noted that $\hat{\eta}_0(z)$ is a bijection from $\mathscr{X}$ onto $\Upsilon = (-\pi, \pi]$. A group $\mathscr{G}$ yielding Θ in (3.11) as an orbit under its induced group $\bar{\mathscr{G}}$ is the special orthogonal group $\mathscr{O}_+(2)$,

$$\mathscr{G} = \left\{g | g = \begin{pmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{pmatrix}, \tau \in (-\pi, \pi]\right\}. \tag{3.13}$$

Then $\mathscr{G}$ acts homeomorphically on $\mathscr{X}$ by $z \to gz$, $\bar{\mathscr{G}} = \mathscr{G}$ and $\tilde{\mathscr{G}} = \{\tilde{g} | \tilde{g} = \psi^{-1}\bar{g}\psi, \bar{g} \in \bar{\mathscr{G}}\}$, which is identified with the group $(-\pi, \pi]$ acting on Υ by $\eta \to \eta + \tau$ [mod(2π)]. Hence $\tilde{\mathscr{G}}$ is denoted by $\tilde{\mathscr{G}} = (-\pi, \pi]$. Further, $u(z) = (z_1^2 + z_2^2)^{1/2}$ is a maximal invariant under $\mathscr{G}$ and $\lambda(\theta) = (\theta_1^2 + \theta_2^2)^{1/2}$ is a maximal invariant parameter under $\bar{\mathscr{G}}$. Consequently Θ in (3.11) is expressed as $\Theta = \{\theta \in \tilde{\Theta} | \lambda(\theta) = 1\}$ and so Assumption 2.1 is satisfied.

Next, with $h(z) \equiv \hat{\eta}_0(z)$ in (4.3) and $\mathscr{U} = R_+$, $\pi(z) \equiv (h(z), u(z))$ provides a homeomorphism from $\mathscr{X}$ onto $(-\pi, \pi] \times \mathscr{U}$, which is identified with $\mathscr{O}_+(2) \times \mathscr{U}$. Since $gz = (z_1 \cos\tau - z_2 \sin\tau, z_1 \sin\tau + z_2 \cos\tau)'$,

$$h(gz) = \tan^{-1}[\tan(\hat{\eta}_0 + \tau)] = \hat{\eta}_0 + \tau, \quad \text{mod}(2\pi). \tag{3.14}$$

Therefore Assumption 2.2 is satisfied.

Now taking the loss $L(a, \eta) = (a - \eta)^2$, mod($2\pi$) by Theorem 2.1 with $\eta_0 = 0$, the unique BEE is given by

$$\hat{\eta}_1(z) = \hat{\eta}_0(z) - E_0[\hat{\eta}_0(z) | u(z)].$$

To evaluate $E_0[\hat{\eta}_0(z) | u(z)]$, note that the conditional pdf of $\hat{\eta}_0$ given u is given by

$$q(\hat{\eta}_0 | u, \eta) = \exp[u \cos(\hat{\eta}_0 - \eta)] \Big/ \int_{-\pi}^{\pi} \exp[u \cos(\hat{\eta}_0 - \eta)]\, d\hat{\eta}_0,$$

which is known as the pdf of the von Mises distribution. Then $E_\eta[\hat{\eta}_0 | u] = \eta$ for any u (a.e.), implying $E_0[\hat{\eta}_0 | u] = 0$. Thus in this model the MLE $\hat{\eta}_0$ in (4.3) is the BEE. Furthermore, since $\mathscr{G}$ is compact, not only is it minimax in the class $\mathscr{D}$ of all estimators but also it is admissible in $\mathscr{D}$ [see Ferguson (1967), pages 154–157].

The reader may be referred to Efron (1978) and Amari (1982) for some recent studies on this model.

4. Applications: The GMANOVA model. A canonical form of the GMANOVA model with sufficiency reduction is derived by Gleser and Olkin (1970) as

$$(4.1)\qquad Z=(Z_1,Z_2)\sim N((\mu_1,0),I_m\otimes\Sigma)\quad\text{and}\quad V\sim W_{p_1+p_2}(\Sigma,n),$$

where Z: $m\times(p_1+p_2)$ and V: $(p_1+p_2)\times(p_1+p_2)$ are independent and $W_b(\Sigma,n)$ denotes the b-dimensional Wishart distribution with mean $n\Sigma$ and degrees of freedon n. Since $E(Z_2)=0$, the model is incomplete but (Z,V) is minimally sufficient. To find an ancillary statistic, let $V=(V_{ij})$ with V_{ij}: $p_i\times p_j$, $\Sigma=(\Sigma_{ij})$ with Σ_{ij}: $p_i\times p_j$, $i,j=1,2$, and $E(Z)=(\mu_1,\mu_2)=\mu$ with $\mu_2=0$. Then the model (4.1) is rewritten as

$$(4.2)\qquad\begin{aligned}&Z_1\ \text{ given }\ Z_2\sim N\big(\mu_1+Z_2\Sigma_{22}^{-1}\Sigma_{21},I_m\otimes\Sigma_{11\cdot2}\big),\\&Z_2\sim N(0,I_m\otimes\Sigma_{22}),\\&V_{11\cdot2}\equiv V_{11}-V_{12}V_{22}^{-1}V_{21}\sim W_{p_1}(\Sigma_{11\cdot2},n-p_2),\\&B\equiv V_{12}V_{22}^{-1}\ \text{ given }\ V_{22}\sim N\big(\beta,V_{22}^{-1}\otimes\Sigma_{11\cdot2}\big)\ \text{ with }\ \beta=\Sigma_{12}\Sigma_{22}^{-1},\\&V_{22}\sim W_{p_2}(\Sigma_{22},n),\end{aligned}$$

where $\Sigma_{11\cdot2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and $V_{11\cdot2}$ is independent of all the other statistics. To make the model fit the situation described in Section 2, let $\mathscr{X}=\tilde{\Theta}=R^{m(p_1+p_2)}\times\mathscr{S}(p_1+p_2)$ and $\Theta=\{\theta\in\tilde{\Theta}|\theta=(\mu,\Sigma),\ \mu_2=0\}$, where $\mathscr{S}(q)$ denotes the set of $q\times q$ positive definite matrices. Further let $\mathscr{G}=\mathscr{A}\times R^{mp_1}$ with

$$\mathscr{A}=\left\{A=\begin{pmatrix}A_{11}&A_{12}\\0&A_{22}\end{pmatrix}\in GU(p_1+p_2)\right\},$$

where $GU(q)$ denotes the group of $q\times q$ upper triangular matrices with positive diagonal elements. Then $\mathscr{G}$ acts on (Z,V) by $(Z,V)\to(ZA'+(F,0),AVA')$ or on $(Z_1,Z_2,V_{11\cdot2},B,V_{22})$ by

$$(4.3)\qquad\begin{aligned}&Z_1\to Z_1A_{11}'+Z_2A_{12}'+F,\qquad Z_2\to Z_2A_{22}',\qquad V_{11\cdot2}\to A_{11}V_{11\cdot2}A_{11}',\\&B\to A_{11}BA_{22}^{-1}+A_{12}A_{22}^{-1}\quad\text{and}\quad V_{22}\to A_{22}V_{22}A_{22}',\end{aligned}$$

where $(A,F)\in\mathscr{G}$. Then a maximal invariant under $\mathscr{G}$ is $u=u(Z,V)=Z_{22}S_{22}'$, where $V_{22}^{-1}=S_{22}'S_{22}$, where $S_{22}\in GU(p_2)$ and a maximal invariant parameter under $\tilde{\mathscr{G}}=\mathscr{G}$ is $\lambda(\theta)=\mu_2\psi_{22}'$, where $\Sigma_{22}^{-1}=\psi_{22}'\psi_{22}$ with $\psi_{22}\in GU(p_2)$. Hence $\Theta=\{\theta\in\tilde{\Theta}|\lambda(\theta)=0\}$ and u is ancillary. Further it is easy to see that $(Z_1,u,V_{11\cdot2},B,S_{22})$ is in one-one correspondence with (Z,V) and so it is minimally sufficient. Note that $\mathscr{G}$ acts on S_{22} by $S_{22}\to S_{22}A_{22}^{-1}$.

Now consider the problem of estimating μ_1. The loss function we adopt here is matrix-valued,

$$(4.4)\qquad L(\hat{\mu}_1,\mu_1)=(\hat{\mu}_1-\mu_1)\Sigma_{11\cdot2}^{-1}(\hat{\mu}_1-\mu_1)'\colon m\times m.$$

Then it is easy to see that the problem is left invariant under $\mathscr{G}$. From $\theta=\psi(\mu_1,\Sigma)=((\mu_1,0),\Sigma)\in\tilde{\Theta}$, $\tilde{\mathscr{G}}=\mathscr{G}$ acts transitively on $\Upsilon=\{(\mu_1,\Sigma)\}$. Further letting

$$h(Z, V) = \left(\begin{pmatrix} R' & V_{12}S_{22}' \\ 0 & S_{22}' \end{pmatrix}, Z_1 - Z_2B' \right),$$

where $V_{11\cdot 2}^{-1} = R'R$ with $R \in GU(p_1)$, $\pi(Z, V) = (h(Z, V), u(Z, V))$ gives a bimeasurable bijection from $\mathscr{X}$ onto $\mathscr{G} \times \mathscr{U}$ where $\mathscr{U}$ is the space of u and Assumptions 2.1 and 2.2 are satisfied. Therefore a BEE is an estimator of the form

$$\hat{\mu}_1^*(h, u) = \tilde{h}\hat{\mu}_1^*(e, u) = \hat{\mu}_1^*(e, u)R' + Z_1 - Z_2B', \tag{4.5}$$

which minimizes the conditional risk

$$E_{(0, I)}[(Z_1 - Z_2B' + \hat{\mu}_1(e, u)R')(Z_1 - Z_2B' + \hat{\mu}_1(e, u)R')'|u].$$

Using $E_{(0, I)}[Z_1 - Z_2B'|u] = 0$, this is minimized in the ordering of nonnegative definiteness if and only if $\hat{\mu}_1(e, u) = 0$. Thus,

$$\hat{\mu}_1^*(Z, V) = Z_1 - Z_2B' = Z_1 - Z_2V_{22}^{-1}V_{21} \tag{4.6}$$

is the unique BEE. In fact, it is the MLE. Consequently in the problem of estimating μ_1, the MLE is the BEE under the matrix loss (4.4).

Next, let us consider the problem of estimating Σ_{22}. For simplicity we assume $p_2 = 1$ and write $V_{22} = v_{22}$ and $\Sigma_{22} = \sigma_{22}$. Then under the loss $L(\hat{\sigma}_{22}, \sigma_{22}) = (\hat{\sigma}_{22} - \sigma_{22})^2/\sigma_{22}^2$, the problem is left invariant under $\mathscr{G}$. Arguing similarly, an equivariant estimator is shown to be of the form

$$\hat{\sigma}_{22}(Z, V) = v_{22}a(u). \tag{4.7}$$

Then by checking Assumptions 2.1 and 2.2, it follows from Corollary 2.1 that a BEE is given by

$$\hat{\sigma}_{22}^* = v_{22}E[v_{22}|u]/E[v_{22}^2|u], \tag{4.8}$$

which is shown to be different from the MLE

$$\tilde{\sigma}_{22} = (Z_2'Z_2 + v_2)/(n + m). \tag{4.9}$$

Since $\tilde{\sigma}_{22}$ is equivariant, the BEE $\hat{\sigma}_{22}^*$ dominates the MLE uniformly. The evaluation of (4.8) is similar to the case of the model $N(\eta, a\eta^2)$ in Section 3 and is omitted here. A corresponding result for a general $p_2 > 1$ is obtained if an invariant loss is adopted.

A special case of some interest is the model with $p_1 = p_2 = 1$ and $\mu_1 = [\alpha, \dots, \alpha]'$, where $\alpha \in R$. This model is regarded as a canonical form of the model appearing in discriminant analysis with covariates [see, e.g., Kariya (1985)]. The above results are of course effective in this model.

Acknowledgment. The author is grateful to referees for their kind and invaluable comments.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1965). *Handbook of Mathematical Functions*. Dover, New York.

AMARI, S. (1982). Differential geometry of curved exponential families—Curvature and information loss. *Ann. Statist.* **10** 357–385.

BARNDORFF-NIELSEN, O. (1980). Conditionality resolution. *Biometrika* **67** 293–310.

BERGER, J. (1980). *Statistical Decision Theory*. Springer, New York.

BUEHLER, R. (1980). Fiducial inference. *R. A. Fisher: An Appreciation. Lecture Notes in Statist.* **1** 109–118. Springer, Berlin.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

EATON, M. L. (1983). *Multivariate Statistical Analysis*. Wiley, New York.

EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency) (with discussion). *Ann. Statist.* **3** 1189–1242.

EFRON, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376.

FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic, New York.

FISHER, R. A. (1973). *Statistical Methods and Scientific Inference*. Hafner, New York.

GIRSHICK, M. A. and SAVAGE, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 53–74. Univ. California Press.

GLESER, L. J. and OLKIN, I. (1970). Linear models in multivariate analysis. In *Essays in Probability and Statistics* (R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao and K. J. C. Smith, eds.) 267–292. Wiley, New York.

HINKLEY, D. V. (1977). Conditional inference about a normal mean with known coefficient of variation. *Biometrika* **64** 105–108.

HORA, R. B. and BUEHLER, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Statist.* **37** 643–656.

KARIYA, T. (1978). The general MANOVA problem. *Ann. Statist.* **6** 200–214.

KARIYA, T. (1985). *Testing in the Multivariate General Linear Model*. Kinokuniya, Tokyo.

KIEFER, J. (1957). Invariance, minimax sequential estimation and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

PITMAN, E. J. G. (1938). The estimation of location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391–421.

ZIDEK, J. V. (1969). A representation of Bayes invariant procedures in terms of Haar measure. *Ann. Inst. Statist. Math.* **21** 291–308.

INSTITUTE OF ECONOMIC RESEARCH
HITOTSUBASHI UNIVERSITY
KUNITACHI-SHI, TOKYO 186
JAPAN