ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS BASED ON CONDITIONAL SPECIFICATION[1]

By Pranab Kumar Sen

University of North Carolina, Chapel Hill

Along with the asymptotic distribution, expressions for the asymptotic bias and asymptotic dispersion matrix of the preliminary test maximum likelihood estimator for a general multi-sample parametric model (when the null hypothesis relating to the restraints on the parameters may not hold) are derived and compared with the parallel expressions for the unrestricted and restricted maximum likelihood estimators. This study reveals the robustness property of the preliminary test estimator when the assumed restraints may not hold.

1. **Introduction.** In a parametric model, assuming that the underlying distributions are of specified forms and the parameter (vector) θ belongs to a suitable *parameter space* Ω, the *(unrestricted) maximum likelihood estimator* (MLE) $\tilde{\theta}$ of θ is obtained by maximizing (over $\theta \in \Omega$) the *likelihood function* of the sample observations. Under appropriate regularity conditions, $\tilde{\theta}$ is an (asymptotically) optimal estimator of θ. In certain problems, ω, a proper subspace of Ω, can be identified from extraneous considerations and a *restricted* MLE $\hat{\theta}$ of θ can be derived by maximizing the likelihood function subject to the restraint that $\theta \in \omega$. When $\theta \in \omega$, $\hat{\theta}$ is (asymptotically) a better estimator than $\tilde{\theta}$. But, if, contrary to this assumed restrain, actually $\theta \notin \omega$, then $\hat{\theta}$ may not only lose its optimality but also may be a biased (or even an inconsistent) estimator. This lack of validity-robustness of $\hat{\theta}$ may be of some concern in a class of problems arising in applied statistics, where $\omega (\subset \Omega)$ can be suggested from certain practical considerations, but, there may not be sufficient a priori evidence of $\theta \in \omega$ so as to warrant the use of $\hat{\theta}$ without any reservation. In such a case, a compromise between $\hat{\theta}$ and $\tilde{\theta}$ based on a *conditional specification* appears to be appealing: a *preliminary test* for $H_0 : \theta \in \omega$ is made and the *preliminary test estimator* (PTE) θ^* is then taken to be $\hat{\theta}$ or $\tilde{\theta}$ according as H_0 is tenable or not. In view of the asymptotic optimality of the *likelihood ratio test*, the preliminary test for H_0 is generally based on the likelihood ratio statistics. Thus, the PTE θ^* depends on the restricted and unrestricted MLE's as well as on the likelihood ratio test for $H_0 : \theta \in \omega$.

For a variety of specific problems, mostly relating to univariate and multivariate normal distributions, various workers have considered various PTE's; we may refer to Kitagawa (1963) and a recent bibliography by Bancroft and Han (1977). The object of the present investigation is to study the asymptotic properties of the

---

1019

PTMLE θ^* under the classical regularity conditions pertaining to the asymptotic theory of $\hat{\theta}$ or $\tilde{\theta}$ [viz., Aitchison and Silvey (1958)]. Huntsberger (1955) has shown that for normal populations, the actual mean square error of θ^* can not be smaller than that of $\hat{\theta}$ when $H_0 : \theta \in \omega$ holds. Generally, θ^* is not (asymptotically) optimal or unbiased when $\theta \in \omega$; nevertheless, it has good asymptotic properties when $\theta \in \omega$. For $\theta \notin \omega$, θ^* may perform better than either of $\tilde{\theta}$ and $\hat{\theta}$. Indeed, $\hat{\theta}$ may lose its optimality and its bias may push up its variability when H_0 does not hold and θ^* may perform better than $\hat{\theta}$. For θ close to the boundary of ω, expressions for the asymptotic bias and dispersion matrix have been derived here for each of $\tilde{\theta}$, $\hat{\theta}$ and θ^* and these are incorporated in the study of the comparative performances of these estimators. These results cast light on the asymptotic superiority of θ^* to $\tilde{\theta}$ or $\hat{\theta}$ when H_0 may not hold.

Along with the preliminary notions, these estimators are introduced in Section 2. Section 3 deals with their asymptotic distributions when $H_0 : \theta \in \omega$ holds. Parallel results for the nonnull case are presented in Section 4. The last section is devoted to the asymptotic comparisons of these estimators.

2. Preliminary notions. Since, in a PTE problem, typically, a multi-sample situation may be involved, we conceive of k ($\geqslant 1$) independent samples. Let $X_{i1}, \cdots, X_{in_i}$ be n_i independent and identically distributed random vectors (i.i.d. rv) with a distribution function (df) $F_i(x, \theta)$, for $i = 1, \cdots, k$, where $x \in E^p$, the $p (\geqslant 1)$-dimensional Euclidean space and $\theta = (\theta_1, \cdots, \theta_t)' \in \Omega \subset E^t$, for some $t \geqslant 1$. Actually, for every i ($= 1, \cdots, k$), F_i may not depend on all the parameters $\theta_1, \cdots, \theta_t$; rather, each element of θ is associated with at least one df. Further, we assume that for each $\theta \in \Omega$ and i ($= 1, \cdots, k$), $F_i(x, \theta)$ admits a density function $f_i(x, \theta)$ (with respect to some sigma-finite measure μ). Then, the (log-) likelihood function is defined by

$$(2.1) \qquad \log L_n(\mathbf{X}_n, \theta) = \sum_{i=1}^k \sum_{j=1}^{n_i} \log f_i(X_{ij}, \theta), \qquad \theta \in \Omega,$$

where $n = n_1 + \cdots + n_k$ and $\mathbf{X}_n = (X_{11}, \cdots, X_{kn_k})$ is the sample point ($\in E^{pn}$). The true parameter θ_0 ($\in \Omega$) is not known. An unrestricted MLE $\tilde{\theta}_n$ is an element of Ω such that

$$(2.2) \qquad \log L_n(\mathbf{X}_n, \tilde{\theta}_n) = \sup_{\theta \in \Omega} \log L_n(\mathbf{X}_n, \theta).$$

Suppose now that θ_0, though unknown, belongs to a subset ω, where

$$(2.3) \qquad \omega = \{\theta : \mathbf{h}(\theta) = (h_1(\theta), \cdots, h_r(\theta)) = 0\} \text{ for some } r < t.$$

Then, a restricted MLE $\hat{\theta}_n$ is an element of ω such that

$$(2.4) \qquad \log L_n(\mathbf{X}_n, \hat{\theta}_n) = \sup_{\theta \in \omega} \log L_n(\mathbf{X}_n, \theta).$$

For testing $H_0 : \theta \in \omega$, the classical likelihood ratio statistic is

$$(2.5) \qquad \mathcal{L}_n = -2 \log\{[\sup_{\theta \in \omega} L_n(\mathbf{X}_n, \theta)]/[\sup_{\theta \in \Omega} L_n(\mathbf{X}_n, \theta)]\}$$

$$= -2 \log(L_n(\mathbf{X}_n; \hat{\theta}_n)/L_n(\mathbf{X}_n; \tilde{\theta}_n)).$$

Let $l_{n,\alpha}$ be a real number such that

$$(2.6) \qquad P\{\mathcal{L}_n \geq l_{n,\alpha}|H_0\} \geq \alpha > P\{\mathcal{L}_n > l_{n,\alpha}|H_0\},$$

where $\alpha(0 < \alpha < 1)$ is the desired level of significance of the test. Then the likelihood ratio test consists in rejecting $H_0 : \theta \in \omega$ when $\mathcal{L}_n > l_{n,\alpha}$ and accepting H_0, otherwise. The PTMLE θ_n^* is defined by

$$(2.7) \qquad \theta_n^* = \hat{\theta}_n \quad \text{if} \quad \mathcal{L}_n \leq l_{n,\alpha}$$
$$= \tilde{\theta}_n \quad \text{if} \quad \mathcal{L}_n > l_{n,\alpha}.$$

Our primary concern is to study the asymptotic properties of $\{\theta_n^*\}$ and compare them with those of $\{\hat{\theta}_n\}$ and $\{\tilde{\theta}_n\}$, when $H_0 : \theta \in \omega$ may or may not hold.

For our study, we make the following assumptions:

[A1] Ω is a convex, compact subspace of E^t, and for every $\theta_1 \neq \theta_2$, (both $\in \Omega$), for at least one $i (= 1, \cdots, k)$,

$$(2.8) \qquad f_i(x, \theta_1) \neq f_i(x, \theta_2), \quad \text{at least on a set of measure nonzero.}$$

[A2] For every $\theta \in \Omega$ and every i $(= 1, \cdots, k)$, $Z_i(\theta) = \int_{E^p} \log f_i(x, \theta) \, dF_i(x, \theta_0)$ exists. In fact, for the ith density, the Kullback-Leibler information is

$$(2.9) \qquad I_i(\theta, \theta_0) = \int_{E^p} \log\{f_i(x, \theta_0)/f_i(x, \theta)\} \, dF_i(x, \theta_0) = Z_i(\theta_0) - Z_i(\theta)$$

where for every $\theta \in \Omega$, $I_i(\theta, \theta_0) \geq 0$ with the strict equality only when $f_i(x, \theta) = f_i(x, \theta_0)$ almost everywhere (a.e.)

[A3] For every $\theta \in \Omega$ and $i (= 1, \cdots, k)$, $\log f_i(x, \theta)$ is (a.e.) thrice differentiable with respect to θ and

$$(2.10) \qquad |(\partial^s/\partial\theta_a^{s_1}\partial\theta_b^{s_2}\partial\theta_c^{s_3})\log f_i(x, \theta)| \leq G_s(x), \qquad \forall x \in E^p, \theta \in \Omega$$

where $s_j \geq 0, j = 1, 2, 3, s_1 + s_2 + s_3 = s = 1, 2, 3$ and $1 \leq a, b, c \leq t$, and where

$$(2.11) \qquad \int_{E^p} G_s(x) \, dF_i(x, \theta_0) < \infty \quad \text{for} \quad i = 1, \cdots, k \quad \text{and} \quad s = 1, 2, 3.$$

(It is possible to eliminate the third order derivatives conditions in (2.10)–(2.11) by imposing the following:

$$(2.12) \qquad \lim_{\delta\downarrow 0} \max_{i,j,l}\left\{\left[\left[\sup_{\theta : \|\theta - \theta_0\| < \delta}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_l}\log f_i(x, \theta)\right|\right]_\theta - \frac{\partial^2}{\partial\theta_j\partial\theta_l}\log f_i(x, \theta)\right]_{\theta_0}\right|\right\} = 0.$$

It is also possible to avoid both the second and third order derivatives conditions in (2.10)–(2.11) by those in Huber (1967) and Inagaki (1973). But, these alternative conditions, in turn, require extra conditions on the first and second order moments of

$$\sup_{\theta : \|\theta - \theta_0\| < \delta}\|(\partial/\partial\theta) \log f_i(x, \theta)]_\theta$$
$$(2.13) \qquad - (\partial/\partial\theta) \log f_i(x, \theta)]_{\theta_0}\|$$

for small δ (> 0). In the sequel we shall deal with (2.10)–(2.11) only—though towards the end of Section 3, we shall make certain comments on these alternative conditions.)

[A4] For every i ($= 1, \cdots, k$) and $\theta \in \Omega$,

$$(2.14) \qquad \int_{E^p} (\partial^2/\partial\theta_j\partial\theta_l) f_i(x, \theta) \, d\mu(x) = 0, \qquad \forall j, l = 1, \cdots, t.$$

Let us define for each i ($= 1, \cdots, k$)

$$(2.15) \quad \mathbf{B}_\theta^{(i)} = \left(\left(\int_{E^p} (\partial/\partial\theta_j) \log f_i(x,)(\partial/\partial\theta_l) \log f_i(x, \theta) \, dF_i(x, \theta) \right) \right)_{j, l = 1, \cdots, t}.$$

[A5] $\mathbf{B}_\theta^{(1)}, \cdots, \mathbf{B}_\theta^{(k)}$ are all continuous in θ in some neighbourhood of θ_0 and

$$(2.16) \qquad B_{\theta_0}^* = \Sigma_{i=1}^k (n_i/n) \mathbf{B}_{\theta_0}^{(i)} \qquad \text{is positive definite.}$$

[A6] $\mathbf{h}(\theta)$ possesses continuous first and second order derivatives with respect to θ, $\forall\theta \in \Omega$. Let then

$$(2.17) \qquad \mathbf{H}_\theta = (((\partial/\partial\theta)\mathbf{h}(\theta))) \qquad \text{(of order } t \times r\text{).}$$

[A7] $\mathbf{H}_{\theta_0}$ is of rank r ($< t$).

[A8] The following matrix (of order $(r + t) \times (r + t)$)

$$(2.18) \qquad \begin{pmatrix} \mathbf{B}_{\theta_0}^* & -\mathbf{H}_{\theta_0} \\ -\mathbf{H}_{\theta_0}' & \mathbf{0} \end{pmatrix}$$

is of full-rank and we denote by

$$(2.19) \qquad \begin{bmatrix} \mathbf{B}_{\theta_0}^* & -\mathbf{H}_{\theta_0} \\ -\mathbf{H}_{\theta_0}' & 0 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{P}_{\theta_0}^* & \mathbf{Q}_{\theta_0}^* \\ \mathbf{Q}_{\theta_0}^{*\prime} & \mathbf{R}_{\theta_0}^* \end{bmatrix}.$$

Note that $\mathbf{B}_{\theta_0}^*$ (and hence, $\mathbf{P}_{\theta_0}^*$, $\mathbf{Q}_{\theta_0}^*$ and $\mathbf{R}_{\theta_0}^*$) may depend on n through $n_1, \cdots, n_k$. We make the final assumption:

[A9] $\lim_{n\to\infty} \frac{1}{n} n_i = \rho_i$ and ($0 < \rho_i < 1$) exists, $\forall 1 \leqslant i \leqslant k$ and $\Sigma_{i=1}^k \rho_i = 1$.

Under [A9], $\mathbf{B}_{\theta_0}^*$ converges to

$$(2.20) \qquad \overline{\mathbf{B}}_{\theta_0} = \Sigma_{i=1}^k \rho_i \mathbf{B}_{\theta_0}^{(i)}$$

and in (2.19), on replacing $\mathbf{B}^*$ by $\overline{\mathbf{B}}$, the corresponding matrices on the right hand side (rhs) are denoted by $\overline{\mathbf{P}}_{\theta_0}$, $\overline{\mathbf{Q}}_{\theta_0}$ and $\overline{\mathbf{R}}_{\theta_0}$ respectively; these do not depend on n. Note that, by definition,

$$(2.21) \qquad \begin{bmatrix} \overline{\mathbf{P}}_{\theta_0} & \overline{\mathbf{Q}}_{\theta_0} \\ \overline{\mathbf{Q}}_{\theta_0}' & \overline{\mathbf{R}}_{\theta_0} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{B}}_{\theta_0} & -\mathbf{H}_{\theta_0} \\ -\mathbf{H}_{\theta_0}' & 0 \end{bmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

Note that $\overline{\mathbf{B}}_{\theta_0}$, $\overline{\mathbf{P}}_{\theta_0}$ and $\overline{\mathbf{R}}_{\theta_0}$ are all symmetric matrices. For later use, we also define the vectors,

$$(2.22) \qquad \Lambda_n(\theta) = n^{-\frac{1}{2}} (\partial/\partial\theta)\log L_n(\mathbf{X}_n, \theta), \qquad \Lambda_n^0 = \Lambda_n(\theta_0),$$

and for vectors or matrices use the notations $\mathbf{o}_p$ and $\mathbf{0}_p$ (or $\mathbf{o}$ and $\mathbf{0}$) in the sense that these orders apply to the individual elements of them.

3. **Asymptotic distribution theory under** $H_0 : \theta_0 \in \omega$. First, consider the case of the unrestricted MLE $\tilde{\theta}_n$. Under the assumptions made in Section 2, $\tilde{\theta}_n$ exists (a.e.), it almost surely (a.s.) converges to θ_0 and further [viz., Silvey (1959)],

(3.1) $$n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) = \bar{\mathbf{B}}_{\theta_0}^{-1}\Lambda_n^0 + o_p(1) \quad \text{as} \quad n \to \infty.$$

Also, by a direct application of the multivariate central limit theorem,

(3.2) $$\Lambda_n^0 \to_{\mathfrak{D}} \mathfrak{N}_t(\mathbf{0}, \bar{\mathbf{B}}_{\theta_0}), \quad \text{as} \quad n \to \infty.$$

Consequently, from (3.1) and (3.2), we have

(3.3) $$n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \to_{\mathfrak{D}} \mathfrak{N}_t(\mathbf{0}, \bar{\mathbf{B}}_{\theta_0}^{-1}) \quad \text{as} \quad n \to \infty.$$

For the restricted MLE $\hat{\theta}_n$, consider the equations

(3.4) $$\begin{aligned} n^{-\frac{1}{2}}\Lambda_n(\theta) + \mathbf{H}_\theta\lambda &= \mathbf{0} \\ \mathbf{h}(\theta) &= \mathbf{0} \end{aligned}$$

where λ ($\in E'$) is a Lagrangian multiplier vector; the solutions for θ and λ are $\hat{\theta}_n$ and $\hat{\lambda}_n$, respectively. From the results in Section 7 of Silvey (1959), we conclude that under $H_0 : \theta_0 \in \omega$ and the regularity conditions of Section 2, as $n \to \infty$,

(3.5) $$n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) = \bar{\mathbf{P}}_{\theta_0}\Lambda_n^0 + o_p(1),$$

(3.6) $$n^{\frac{1}{2}}\hat{\lambda}_n = \bar{\mathbf{Q}}_{\theta_0}'\Lambda_n^0 + o_p(1)$$

and, further, for the likelihood ratio statistic $\mathcal{L}_n$ in (2.5)

(3.7) $$\mathcal{L}_n = n(\hat{\theta}_n - \tilde{\theta}_n)'\bar{\mathbf{B}}_{\theta_0}(\hat{\theta}_n - \tilde{\theta}_n) + o_p(1).$$

Note that by (2.21), $\bar{\mathbf{B}}_{\theta_0}\bar{\mathbf{P}}_{\theta_0} = \mathbf{I} + \mathbf{H}_{\theta_0}\bar{\mathbf{Q}}_{\theta_0}'$, $\bar{\mathbf{B}}_{\theta_0}\bar{\mathbf{Q}}_{\theta_0} = \mathbf{H}_{\theta_0}\bar{\mathbf{R}}_{\theta_0}$, $-\mathbf{H}_{\theta_0}'\bar{\mathbf{P}}_{\theta_0} = \mathbf{0}$ and $-\mathbf{H}_{\theta_0}'\bar{\mathbf{Q}}_{\theta_0} = \mathbf{I}$. Thus, noting that $\bar{\mathbf{P}}_{\theta_0}' = \bar{\mathbf{P}}_{\theta_0}$, we have

(3.8) $$\begin{aligned} \bar{\mathbf{P}}_{\theta_0}\bar{\mathbf{B}}_{\theta_0}\bar{\mathbf{P}}_{\theta_0}' &= \bar{\mathbf{P}}_{\theta_0}\bar{\mathbf{B}}_{\theta_0}\bar{\mathbf{P}}_{\theta_0} = \bar{\mathbf{P}}_{\theta_0} + \bar{\mathbf{P}}_{\theta_0}\mathbf{H}_{\theta_0}\bar{\mathbf{Q}}_{\theta_0}' \\ &= \bar{\mathbf{P}}_{\theta_0} + \bar{\mathbf{P}}_{\theta_0}'\mathbf{H}_{\theta_0}\bar{\mathbf{Q}}_{\theta_0}' = \bar{\mathbf{P}}_{\theta_0}. \end{aligned}$$

Hence, from (3.2), (3.5) and (3.8), we have under $H_0 : \theta_0 \in \omega$

(3.9) $$n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \to_{\mathfrak{D}} \mathfrak{N}_t(\mathbf{0}, \bar{\mathbf{P}}_{\theta_0}), \quad \text{as} \quad n \to \infty.$$

Also, from (3.1), (3.5), (3.6) and (3.7), we have (on using the identities presented before (3.8)) under $H_0 : \theta_0 \in \omega$,

(3.10) $$\begin{aligned} \mathcal{L}_n &= \Lambda_n^{0\prime}(\bar{\mathbf{P}}_{\theta_0} - \bar{\mathbf{B}}_{\theta_0}^{-1})\bar{\mathbf{B}}_{\theta_0}(\bar{\mathbf{P}}_{\theta_0} - \bar{\mathbf{B}}_{\theta_0}^{-1})\Lambda_n^0 + o_p(1) \\ &= \Lambda_n^{0\prime}\bar{\mathbf{Q}}_{\theta_0}\mathbf{H}_{\theta_0}'\bar{\mathbf{B}}_{\theta_0}^{-1}\mathbf{H}_{\theta_0}\bar{\mathbf{Q}}_{\theta_0}'\Lambda_n^0 + o_p(1) \\ &= -\hat{\lambda}_n'\bar{\mathbf{R}}_{\theta_0}^{-1}\hat{\lambda}_n + o_p(1) = -\Lambda_n^{0\prime}\bar{\mathbf{Q}}_{\theta_0}\bar{R}_{\theta_0}^{-1}\bar{\mathcal{Q}}_{\theta_0}'\Lambda_n^0 + o_p(1), \end{aligned}$$

where, by using (2.21), it is easy to show that

$$(3.11) \qquad \overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1}\overline{\mathbf{Q}}'_{\theta_0}\overline{\mathbf{B}}_{\theta_0}\overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1}\overline{\mathbf{Q}}'_{\theta_0} = -\overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1}\overline{\mathbf{Q}}'_{\theta_0},$$

$$(3.12) \qquad \text{Rank of } \left[-\overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1}\overline{\mathbf{Q}}'_{\theta_0} \right] = r(< t).$$

Hence, by (3.2), (3.10), (3.11), (3.12) and the Cochran theorem on quadratic forms in (asymptotically) normally distributed random vectors, we obtain that under $H_0 : \theta_0 \in \omega$,

$$(3.13) \qquad \mathfrak{L}_n \to_{\mathfrak{D}} \chi_r^2,$$

and let $\chi_{r,\alpha}^2$ be the upper $100\alpha\%$ point of the chi-square df with r degrees of freedom (DF). Then, from (2.6) and (3.13), we have

$$(3.14) \qquad l_{n,\alpha} \to \chi_r^2, \qquad \text{as} \quad n \to \infty.$$

Let us now consider the case of $\{\theta_n^*\}$. By (2.7), we have for every $\mathbf{y} \in E'$,

$$(3.15) \quad P\left\{ n^{\frac{1}{2}}(\theta_n^* - \theta_0) \leqslant \mathbf{y} | H_0 \right\} = P\left\{ n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \leqslant \mathbf{y}, \, \mathfrak{L}_n \leqslant l_{n,\alpha} | H_0 \right\}$$
$$+ P\left\{ n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \leqslant \mathbf{y}, \, \mathfrak{L}_n > l_{n,\alpha} | H_0 \right\}.$$

By (3.2), (3.5), (3.6), (3.10) and (3.14), the first term on the right-hand side of (3.15) converges to

$$(3.16) \qquad P\left\{ \overline{\mathbf{P}}_{\theta_0}\Lambda_n^0 \leqslant \mathbf{y}, \, -n\hat{\lambda}'_n\overline{\mathbf{R}}_{\theta_0}^{-1}\hat{\lambda}_n \leqslant \chi_{r,\alpha}^2 | H_0 \right\}.$$

Note that by (2.21), $\overline{\mathbf{Q}}'_{\theta_0}\overline{\mathbf{B}}_{\theta_0}\overline{\mathbf{P}}_{\theta_0} = \overline{\mathbf{Q}}'_{\theta_0} + \overline{\mathbf{Q}}'_{\theta_0}\mathbf{H}_{\theta_0}\overline{\mathbf{Q}}_{\theta_0} = \overline{\mathbf{Q}}_{\theta_0} - \overline{\mathbf{Q}}_{\theta_0} = \mathbf{0}$, so that $\overline{\mathbf{P}}_{\theta_0}\Lambda_n^0$ and $n^{\frac{1}{2}}\hat{\lambda}_n$ are asymptotically independent, and hence, (3.16) reduces (asymptotically) to

$$(3.17) \qquad P\left\{ \overline{\mathbf{P}}_{\theta_0}\Lambda_n^0 \leqslant \mathbf{y} | H_0 \right\} P\left\{ \mathfrak{L}_n \leqslant \chi_{r,\alpha}^2 | H_0 \right\} \to (1 - \alpha)G_t\left(\mathbf{y}; \mathbf{0}, \overline{\mathbf{P}}_{\theta_0}\right),$$

where $G_t(\mathbf{y}; \mu, \Sigma)$ stands for a t-variate multinormal df with mean vector μ and dispersion matrix Σ. Let us also denote by

$$(3.18) \qquad E_1 = \left\{ \mathbf{x} \in E^r : -\mathbf{x}'\overline{\mathbf{R}}_{\theta_0}^{-1}\mathbf{x} > \chi_{r,\alpha}^2 \right\}.$$

Then by (3.1), (3.10) and (3.14), the second term on the right-hand side of (3.15) is asymptotically equivalent to

$$(3.19) \qquad P\left\{ \overline{\mathbf{B}}_{\theta_0}^{-1}\Lambda_n^0 \leqslant \mathbf{y}, \, n^{\frac{1}{2}}\hat{\lambda}_n \in E_1 | H_0 \right\}$$

where by (3.2) and (3.6), under H_0, $(\overline{\mathbf{B}}_{\theta_0}^{-1}\Lambda_n^0, n^{\frac{1}{2}}\hat{\lambda}_n)$ has asymptotically a $(t + r)$-variate normal df with $\mathbf{0}$ mean and dispersion matrix

$$(3.20) \qquad \begin{bmatrix} \overline{\mathbf{B}}_{\theta_0}^{-1} & \overline{\mathbf{Q}}_{\theta_0} \\ \overline{\mathbf{Q}}'_{\theta_0} & -\mathbf{R}_{\theta_0} \end{bmatrix},$$

so that the conditional df of $\overline{\mathbf{B}}_{\theta_0}^{-1}\boldsymbol{\Lambda}_n^0$ given $n^{\frac{1}{2}}\hat{\boldsymbol{\lambda}}_n = \mathbf{z}$ is asymptotically multinormal with mean vector $\overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1}\mathbf{Z}$ and dispersion matrix $\overline{\mathbf{B}}_{\theta_0}^{-1} + \overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}\overline{\mathbf{Q}}_{\theta_0}' = \overline{\mathbf{P}}_{\theta_0}$. Hence, (3.19) converges to

$$(3.21) \qquad \int_{E_1} G_t\left(\mathbf{y} - \overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1}\mathbf{Z}; \, 0, \, \overline{\mathbf{P}}_{\theta_0}\right)dG_r\left(\mathbf{Z}; \, 0, \, -\overline{\mathbf{R}}_{\theta_0}\right).$$

From (3.15), (3.17), and (3.21), we obtain, that under $H_0 : \theta_0 \in \omega$,

$$\lim_{n\to\infty} P\left\{ n^{\frac{1}{2}}(\theta_n^* - \theta_0) \leqslant \mathbf{y}\right\}$$

$$(3.22) \qquad \qquad = (1 - \alpha)G_t\left(\mathbf{y}; \, 0, \, \overline{\mathbf{P}}_{\theta_0}\right)$$

$$+ \int_{E_1} G_t\left(\mathbf{y} + \overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1}\mathbf{Z}; \, 0, \, \overline{\mathbf{P}}_{\theta_0}\right)dG_r\left(\mathbf{Z}; \, 0, \, -\overline{\mathbf{R}}_{\theta_0}\right),$$

$$\forall \, \mathbf{y} \in E'.$$

Thus, the asymptotic distribution of $n^{\frac{1}{2}}(\theta_n^* - \theta_0)$ is, in general, nonnormal. In particular, if $\overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1} = 0$ (which also implies that $\overline{\mathbf{P}}_{\theta_0} = \overline{\mathbf{B}}_{\theta_0}^{-1}$), (3.22) reduces to

$$(3.23) \qquad G_t\left(\mathbf{y}; \, 0, \, \overline{\mathbf{P}}_{\theta_0}\right)\left(\text{or } G_t\left(\mathbf{y}; \, 0, \, \overline{\mathbf{B}}_{\theta_0}^{-1}\right)\right)$$

so that all the three estimators have the same limiting normal distribution.

REMARK. Our (3.1), (3.5) and (3.6), as adapted from Silvey (1959), rest on the assumptions made in Section 2. As mentioned in Section 2, (2.12) may replace (2.10)–(2.11) for $s = 3$. In such a case, our (3.1), (3.5) and (3.6), would follow from the results of Feder (1968). Also, we may proceed as in Inagaki (1973) and show that (3.1), (3.5) and (3.6) follow, provided (2.13) satisfies appropriate growth conditions. The rest of the formulae in this section remain the same irrespective of the particular approach we choose.

4. Asymptotic nonnull distribution theory. Note that if $H_0 : \theta_0 \in \omega$ does not hold (i.e., $\mathbf{h}(\theta_0) \neq 0$), then there exists a $\theta^*(\in\omega)$, such that

$$(4.1) \qquad (\partial/\partial\theta)Z(\theta) + \mathbf{H}_\theta\boldsymbol{\lambda}^*\big|_{\theta=\theta^*} = 0 \qquad \text{and} \quad \mathbf{h}(\theta^*) = 0$$

where $\boldsymbol{\lambda}^*(\in E')$ is a Lagrangian multiplier and $\theta^* \neq \theta_0$ (by assumptions [A1] and [A2]). In this case, $\tilde{\theta}_n$ in (2.2) stochastically converges to θ_0 while $\hat{\theta}_n$ in (2.4) converges stochastically to θ^* and, hence, $\mathcal{L}_n$ in (2.5) tends (in probability) to ∞ as $n \to \infty$. Consequently, by (2.7) and (3.14),

$$(4.2) \qquad \lim_{n\to\infty} P\left\{\theta_n^* \neq \theta_n|\theta_0 \notin \omega\right\} \leqslant \lim_{n\to\infty} P\left\{\mathcal{L}_n \leqslant l_{n,\alpha}|\theta_0 \notin \omega\right\} = 0,$$

and, hence, noting that (3.3) does not depend on $H_0 : \theta_0 \in \omega$ being true or not, we have, from (3.3) and (4.2), for every $\mathbf{y} \in E^t$,

$$(4.3) \quad \lim_{n\to\infty} P\left\{ n^{\frac{1}{2}}(\theta_n^* - \theta_0) \leqslant \mathbf{y}|\theta_0 \notin \omega\right\}$$

$$= \lim_{n\to\infty} P\left\{ n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \leqslant \mathbf{y}|\theta_0 \notin \omega\right\} = G_t\left(\mathbf{y}; \, 0, \, \overline{\mathbf{B}}_{\theta_0}^{-1}\right).$$

Thus, for any (fixed) alternative, $n^{\frac{1}{2}}(\theta_n^* - \theta_0)$ and $n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0)$ are asymptotically equivalent in probability and have the same (asymptotic) multi-normal distribu-

tion. The situation becomes different when $\boldsymbol{\theta}_0$ lies *near* the boundary of ω. For this study, we conceive of the following sequence $\{K_n\}$ of local alternatives:

(4.4) $$K_n : \mathbf{h}(\boldsymbol{\theta}_0) = n^{-\frac{1}{2}}\boldsymbol{\gamma}, \qquad \boldsymbol{\gamma} \text{ real-vector } (\in E^r),$$

and consider the asymptotic distributions of the estimators under $\{K_n\}$.

First, (3.3) holds irrespective of H_0 or $\{K_n\}$, and hence

(4.5) $$\lim_{n\to\infty} P\left\{ n^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \leqslant \mathbf{y}|K_n \right\} = G_t\left(\mathbf{y}; \mathbf{0}, \overline{\mathbf{B}}_{\boldsymbol{\theta}_0}^{-1}\right), \qquad \forall \, \mathbf{y} \in E^t.$$

Also, under K_n the solutions $\boldsymbol{\theta}^*$, $\boldsymbol{\lambda}^*$ in (4.1) depend on n and are denoted by $\boldsymbol{\theta}^*_{(n)}$, $\boldsymbol{\lambda}^*_{(n)}$, respectively. Note that $\mathbf{h}(\boldsymbol{\theta}^*_{(n)}) = \mathbf{0}$. Hence, under (4.4) and the assumptions of Section 2, we have

(4.6) $$\mathbf{H}'_{\boldsymbol{\theta}_0}\left\{ n^{\frac{1}{2}}(\boldsymbol{\theta}^*_{(n)} - \boldsymbol{\theta}_0) \right\} = \boldsymbol{\gamma} + o(1),$$

(4.7) $$\mathbf{H}_{\boldsymbol{\theta}_0}\left\{ n^{\frac{1}{2}}\boldsymbol{\lambda}^*_{(n)} \right\} = \overline{\mathbf{B}}_{\boldsymbol{\theta}_0}\left\{ n^{\frac{1}{2}}(\boldsymbol{\theta}^*_{(n)} - \boldsymbol{\theta}_0) \right\} + o(1).$$

From (4.6) and (4.7), we conclude that

(4.8) $$\lim_{n\to\infty} n^{\frac{1}{2}}(\boldsymbol{\theta}^*_{(n)} - \boldsymbol{\theta}_0) = \boldsymbol{\gamma}^* \quad \text{and} \quad \lim_{n\to\infty} n^{\frac{1}{2}}\boldsymbol{\lambda}^*_{(n)} = \boldsymbol{\lambda}^*$$

both exist, and

(4.9) $$\boldsymbol{\gamma} = \mathbf{H}'_{\boldsymbol{\theta}_0}\boldsymbol{\gamma}^*, \, \mathbf{H}_{\boldsymbol{\theta}_0}\boldsymbol{\lambda}^* = \overline{\mathbf{B}}_{\boldsymbol{\theta}_0}\boldsymbol{\gamma}^*\left(\Rightarrow \boldsymbol{\lambda}^* = -\overline{\mathbf{R}}'_{\boldsymbol{\theta}_0}\overline{\mathbf{H}}'_{\boldsymbol{\theta}_0}\boldsymbol{\gamma}^*, \, \boldsymbol{\gamma}^* = \overline{\mathbf{B}}_{\boldsymbol{\theta}_0}^{-1}\mathbf{H}_{\boldsymbol{\theta}_0}\boldsymbol{\lambda}^*\right).$$

From (3.4), (4.4), (4.6) through (4.9) and under the assumptions of Section 2, we obtain that under $\{K_n\}$,

(4.10) $$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \boldsymbol{\gamma}^* + \overline{\mathbf{P}}_{\boldsymbol{\theta}_0}\boldsymbol{\Lambda}_n^0 + o_p(1),$$

(4.11) $$n^{\frac{1}{2}}\hat{\boldsymbol{\lambda}}_n = \boldsymbol{\lambda}^* + \overline{\mathbf{Q}}'_{\boldsymbol{\theta}_0}\boldsymbol{\Lambda}_n^0 + o_p(1),$$

where $\hat{\boldsymbol{\theta}}_n$ is the restricted MLE and $\hat{\boldsymbol{\lambda}}_n$ is the Lagrangian multiplier in (3.4). Comparing (3.5) and (4.10) and using the same arguments as in (3.8) and (3.9), we obtain that for every $\mathbf{y} \in E^t$,

(4.12) $$\lim_{n\to\infty} P\left\{ n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \leqslant \mathbf{y}|K_n \right\} = G_t\left(\mathbf{y} - \boldsymbol{\gamma}^*; \mathbf{0}, \overline{\mathbf{P}}_{\boldsymbol{\theta}_0}\right).$$

It follows similarly that (3.7) continues to hold under $\{K_n\}$, where by (3.1) and (4.10), we have

(4.13) $$\begin{aligned} n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n) &= \boldsymbol{\gamma}^* + \left(\overline{\mathbf{P}}_{\boldsymbol{\theta}_0} - \overline{\mathbf{B}}_{\boldsymbol{\theta}_0}^{-1}\right)\boldsymbol{\Lambda}_n^0 + o_p(1) \\ &= \overline{\mathbf{B}}_{\boldsymbol{\theta}_0}^{-1}\mathbf{H}_{\boldsymbol{\theta}_0}\boldsymbol{\lambda}^* + \overline{\mathbf{Q}}_{\boldsymbol{\theta}_0}\mathbf{H}'_{\boldsymbol{\theta}_0}\overline{\mathbf{B}}_{\boldsymbol{\theta}_0}^{-1}\boldsymbol{\Lambda}_n^0 + o_p(1) \\ &= \overline{\mathbf{Q}}_{\boldsymbol{\theta}_0}\overline{\mathbf{R}}_{\boldsymbol{\theta}_0}^{-1}\left\{ n^{\frac{1}{2}}\hat{\boldsymbol{\lambda}}_n \right\} + o_p(1). \end{aligned}$$

Thus, under $\{K_n\}$, as $n \to \infty$

(4.14) $$\mathcal{L}_n = -n\hat{\boldsymbol{\lambda}}'_n\overline{\mathbf{R}}_{\boldsymbol{\theta}_0}^{-1}\hat{\boldsymbol{\lambda}}_n + o_p(1).$$

On noting that, by (2.21), $\overline{\mathbf{B}}_{\boldsymbol{\theta}_0}\overline{\mathbf{Q}}_{\boldsymbol{\theta}_0} = \mathbf{H}_{\boldsymbol{\theta}_0}\overline{\mathbf{R}}_{\boldsymbol{\theta}_0}$ and $\mathbf{H}_{\boldsymbol{\theta}_0}\overline{\mathbf{Q}}_{\boldsymbol{\theta}_0} = -\mathbf{I}$, we conclude from

(4.11) and (4.14) that under $\{K_n\}$, $\pounds_n$ has asymptotically a noncentral chi-square df with r DF and noncentrality parameter

$$(4.15) \qquad \Delta^* = \boldsymbol{\lambda}^{*\prime}\mathbf{H}_{\theta_0}'\bar{\mathbf{B}}_{\theta_0}^{-1}\mathbf{H}_{\theta_0}\boldsymbol{\lambda}^* = -\boldsymbol{\lambda}^{*\prime}\bar{\mathbf{R}}_{\theta_0}^{-1}\boldsymbol{\lambda}^*;$$

we denote this df by $\Pi_r(x; \Delta^*)$. Then, from (2.7) and (3.14),

$$(4.16) \qquad \begin{aligned} P\Big\{n^{\frac{1}{2}}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_0) \leqslant \mathbf{y}|K_n\Big\} \\ = P\Big\{n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \leqslant \mathbf{y}, \ \pounds_n \leqslant l_{n,\alpha}|K_n\Big\} \\ + P\Big\{n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \leqslant \mathbf{y}, \ \pounds_n > l_{n,\alpha}|K_n\Big\}, \\ \forall \ \mathbf{y} \in E^t. \end{aligned}$$

Since by (2.21), $\bar{\mathbf{P}}_{\theta_0}'\bar{\mathbf{B}}_{\theta_0}\bar{\mathbf{Q}}_{\theta_0} = \bar{\mathbf{P}}_{\theta_0}'\mathbf{H}_{\theta_0}\bar{\mathbf{R}}_{\theta_0} = \mathbf{0}$, we conclude from (4.10) and (4.11) that under K_n, $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and $n^{\frac{1}{2}}\hat{\boldsymbol{\lambda}}_n$ (and hence, by (4.14), $\pounds_n$) are asymptotically independent, so that by (3.2), (4.10), (4.11), (4.14) and (4.15), the first term on the right-hand side of (4.16) converges to

$$(4.17) \qquad G_t\big(\mathbf{y} - \boldsymbol{\gamma}^*; \ \mathbf{0}, \bar{\mathbf{P}}_{\theta_0}\big)\Pi_r(\chi_{r,\alpha}^2; \Delta^*).$$

Also, let

$$(4.18) \qquad E^*(\mathbf{c}) = \Big\{\mathbf{y} \in E^r : -(\mathbf{y} + \mathbf{c})'\bar{\mathbf{R}}_{\theta_0}^{-1}(\mathbf{y} + \mathbf{c}) > \chi_{r,\alpha}^2\Big\}, \qquad \forall \mathbf{c} \in E^r.$$

Then, by (3.1), (3.2), (4.11), (4.14) [and the fact that by (2.21), $\bar{\mathbf{Q}}_{\theta_0}\bar{\mathbf{R}}_{\theta_0}^{-1}\mathbf{Q}_{\theta_0}' = \bar{\mathbf{P}}_{\theta_0} - \bar{\mathbf{B}}_{\theta_0}^{-1}$], we obtain, for the second term on the right-hand side of (4.16),

$$(4.19) \quad \begin{aligned} P\Big\{\bar{\mathbf{B}}_{\theta_0}^{-1}\boldsymbol{\Lambda}_n^0 \leqslant \mathbf{y}, \ n^{\frac{1}{2}}\hat{\boldsymbol{\lambda}}_n \in E_1(\mathbf{0})|K_n\Big\} \\ = \int_{E_1(\mathbf{0})} G_t\big(\mathbf{y} + \mathbf{Q}_{\theta_0}\bar{\mathbf{R}}_{\theta_0}^{-1}(\mathbf{z} - \boldsymbol{\lambda}^*); \ \mathbf{0}, \bar{\mathbf{P}}_{\theta_0}\big)dG_r\big(\mathbf{z}; \boldsymbol{\lambda}^*, -\bar{\mathbf{R}}_{\theta_0}\big) + o(1) \\ = \int_{E_1(\boldsymbol{\lambda}^*)} G_t\big(\mathbf{y} + \bar{\mathbf{Q}}_{\theta_0}\bar{\mathbf{R}}_{\theta_0}^{-1}\mathbf{x}; \ \mathbf{0}, \bar{\mathbf{P}}_{\theta_0}\big)dG_r\big(\mathbf{x}; \mathbf{0}, -\bar{\mathbf{R}}_{\theta_0}\big) + o(1). \end{aligned}$$

From (4.16), (4.17) and (4.19), we arrive at the following.

THEOREM 4.1. *Under* $\{K_n\}$ *in* (4.4) *and the assumptions of Section* 2,

$$\begin{aligned} \lim_{n\to\infty} P\Big\{n^{\frac{1}{2}}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_0) \leqslant \mathbf{y}\Big\} = G_t\big(\mathbf{y} - \boldsymbol{\gamma}^*; \ \mathbf{0}, \bar{\mathbf{P}}_{\theta_0}\big)\Pi_r(\chi_{r,\alpha}^2; \Delta^*) \\ + \int_{E_1(\boldsymbol{\lambda}^*)} G_t\big(\mathbf{y} + \bar{\mathbf{Q}}_{\theta_0}\bar{\mathbf{R}}_{\theta_0}^{-1}\mathbf{x}; \ \mathbf{0}, \bar{\mathbf{P}}_{\theta_0}\big)dG_r\big(\mathbf{x}; \mathbf{0}, -\bar{\mathbf{R}}_{\theta_0}\big) \end{aligned}$$

$$(4.20) \qquad\qquad = G_t^*(\mathbf{y}; \boldsymbol{\gamma}), \ say, \ (\mathbf{y} \in E^t).$$

Here also, if $\bar{\mathbf{Q}}_{\theta_0}\bar{\mathbf{R}}_{\theta_0}^{-1} = \mathbf{0}$, (4.20) reduces to

$$(4.21) \quad G_t\big(\mathbf{y} - \boldsymbol{\gamma}^*; \ \mathbf{0}, \bar{\mathbf{P}}_{\theta_0}\big)\Pi_r(\chi_{r,\alpha}^2; \Delta^*) + \big[1 - \Pi_r(\chi_{r,\alpha}^2; \Delta^*)\big]G_t\big(\mathbf{y}; \mathbf{0}, \bar{\mathbf{P}}_{\theta_0}\big)$$

(that is a mixture of two multinormal df's). But, in general, it is nonnormal. For later use, we denote the probability density functions (pdf) corresponding to G_t and

G_t^* by g_t and g_t^*, respectively. Then,

$$(4.22) \qquad g_t^*(y, \gamma) = g_t\left(y - \gamma^*; 0, \overline{P}_{\theta_0}\right)\Pi_r\left(\chi_{r,\alpha}^2; \Delta^*\right)$$

$$+ \int_{E_1(\lambda^*)} g_t\left(y + \overline{Q}_{\theta_0}\overline{R}_{\theta_0}^{-1}x; 0, \overline{P}_{\theta_0}\right)g_r\left(x; 0, -\overline{R}_{\theta_0}\right)dx, \ \forall \ y \in E^t.$$

5. Asymptotic comparison of the estimators.

Let $\{T_n\}$ be a sequence of estimators of θ_0 such that $n^{\frac{1}{2}}(T_n - \theta_0)$ has a limiting distribution with finite second moments. Then the *mean vector and dispersion matrix (about the origin)* of this limiting df are taken as the *asymptotic bias and asymptotic dispersion matrix* (a.d.m.) of $n^{\frac{1}{2}}(T_n - \theta_0)$. In this section, we study the asymptotic bias and a.d.m. of each of the three estimators considered in earlier sections and compare them. We confine ourselves to the sequence $\{K_n\}$ of alternative hypotheses in (4.4), so that the null hypothesis case follows by letting $\gamma = 0$.

It follows from (3.1), (3.2) and (3.3) that

$$(5.1) \qquad \beta_1(\gamma) = \text{asymptotic bias of } n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \text{ when } \{K_n\} \text{ holds}$$

$$= \int_{E^t} y \, dG_t\left(y; 0, \overline{B}_{\theta_0}^{-1}\right) = 0;$$

$$(5.2) \qquad \nu_1(\gamma) = \text{a.d.m of } n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \text{ when } \{K_n\} \text{ holds}$$

$$= \int_{E^t} y \, y' \, dG_t\left(y; 0, \overline{B}_{\theta_0}^{-1}\right) = \overline{B}_{\theta_0}^{-1}.$$

Similarly, from (3.2) and (4.10), we have

$$(5.3) \qquad \beta_2(\gamma) = \text{asymptotic bias of } n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \text{ when } \{K_n\} \text{ holds},$$

$$= \gamma^* = \overline{Q}_{\theta_0}\overline{R}_{\theta_0}^{-1}\lambda^*;$$

$$(5.4) \qquad \nu_2(\gamma) = \text{a.d.m. of } n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \text{ when } \{K_n\} \text{ holds}$$

$$= \gamma^*\gamma^{*\prime} + \overline{P}_{\theta_0}\overline{B}_{\theta_0}\overline{P}'_{\theta_0} = \gamma^*\gamma^{*\prime} + \overline{P}_{\theta_0}.$$

At this stage, we note that for a multinormal df $G_p(x; 0, D)$,

$$(5.5) \qquad \int_{(x+a)'D^{-1}(x+a)>c} x \, dG_p(x; 0, D) = a\left[\Pi_p(c, \delta) - \Pi_{p+2}(c, \delta)\right],$$

$$\forall \ a \in E^p, c \geqslant 0;$$

$$\int_{(x+a)'D^{-1}(x+a)>c} xx' \, dG_p(x; 0, D) = \left\{1 - \Pi_{p+2}(c; \delta)\right\}D$$

$$(5.6) \qquad\qquad - aa'\left\{\Pi_p(c; \delta) - 2\Pi_{p+2}(c; \delta) + \Pi_{p+4}(c; \delta)\right\}$$

where $\delta = a'D^{-1}a$. Thus, from (4.20), (4.21) and (5.5), we have

$$\beta^*(\gamma) = \text{asymptotic bias of } n^{\frac{1}{2}}(\theta_n^* - \theta_0) \text{ when } \{K_n\} \text{ holds}$$

$$(5.7) \qquad = \gamma^*\Pi_r\left(\chi_{r,\alpha}^2; \Delta^*\right) - \overline{Q}_{\theta_0}\overline{R}_{\theta_0}^{-1}\lambda^*\left[\Pi_r\left(\chi_{r,\alpha}^2; \Delta^*\right) - \Pi_{r+2}\left(\chi_{r,\alpha}^2; \Delta^*\right)\right]$$

$$= \overline{Q}_{\theta_0}\overline{R}_{\theta_0}^{-1}\lambda^*\Pi_{r+2}\left(\chi_{r,\alpha}^2; \Delta^*\right), \ \left(\text{as } \gamma^* = \overline{Q}_{\theta_0}\overline{R}_{\theta_0}^{-1}\lambda^*\right);$$

from (4.20), (4.21) and (5.6), we have

$$
\begin{aligned}
\nu^*(\gamma) &= \text{a.d.m. of } n^{\frac{1}{2}}(\theta_n^* - \theta_0) \text{ when } \{K_n\} \text{ holds,} \\
&= \left(\gamma^*\gamma^* + \overline{P}_{\theta_0}\right)\Pi_r(\chi^2_{r,\alpha}; \Delta^*) + \overline{P}_{\theta_0}\{1 - \Pi_r(\chi^2_{r,\alpha}; \Delta^*)\} \\
&\quad + \overline{Q}_{\theta_0}\overline{R}_{\theta_0}^{-1}\Big\{-\big[1 - \Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*)\big]\overline{R}_{\theta_0} - \lambda^*\lambda^{*'}\big[\Pi_r(\chi^2_{r,\alpha}; \Delta^*) \\
&\quad - 2\Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*) + \Pi_{r+4}(\chi^2_{r,\alpha}; \Delta^*)\big]\Big\}\overline{R}_{\theta_0}^{-1}\overline{Q}_{\theta_0}'
\end{aligned}
$$

$$
\begin{aligned}
(5.8) \qquad &= \overline{P}_{\theta_0} - \big[1 - \Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*)\big]\overline{Q}_{\theta_0}\overline{R}_{\theta_0}^{-1}\overline{Q}_{\theta_0}' \\
&\quad + \gamma^*\gamma^{*'}\{2\Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*) - \Pi_{r+4}(\chi^2_{r,\alpha}; \Delta^*)\} \\
&= \overline{B}_{\theta_0}^{-1} + \left(\overline{P}_{\theta_0} - \overline{B}_{\theta_0}^{-1}\right)\Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*) + \gamma^*\gamma^{*'}\{2\Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*) \\
&\quad - \Pi_{r+4}(\chi^2_{r,\alpha}; \Delta^*)\}.
\end{aligned}
$$

We now proceed to compare the asymptotic bias and a.d.m. of the three estimators. First, consider the null hypothesis case where $\theta_0 \in \omega$, so that $\gamma^* = 0$, $\gamma = 0$ and $\lambda^* = 0$. From (5.1), (5.3) and (5.7), we obtain that

$$(5.9) \qquad \gamma^* = 0 \Rightarrow \beta_1(0) = \beta_2(0) = \beta^*(0) = 0,$$

so that all these estimators are asymptotically unbiased. Also, from (5.2), (5.4) and (5.8), we have for $\gamma = 0$,
(5.10)

$$\nu_1(0) = \overline{B}_{\theta_0}^{-1}, \ \nu_2(0) = \overline{P}_{\theta_0} \quad \text{and} \quad \nu^*(0) = \overline{B}_{\theta_0}^{-1} - \left(\overline{B}_{\theta_0}^{-1} - \overline{P}_{\theta_0}\right)\Pi_{r+2}(\chi^2_{r,\alpha}; 0)$$

where

$$(5.11) \qquad 0 < \Pi_{r+2}(\chi^2_{r,\alpha}; 0) < \Pi_r(\chi^2_{r,\alpha}; 0) = 1 - \alpha < 1.$$

Also, by the identities in (3.8) and prior to it, we note that both $\overline{B}_{\theta_0}\overline{P}_{\theta_0}$ and $-H_{\theta_0}\overline{Q}_{\theta_0}'$ are idempotent matrices and $\overline{B}_{\theta_0}$ is nonsingular. Hence, it follows that

$$(5.12) \qquad \overline{B}_{\theta_0}^{-1} - \overline{P}_{\theta_0} \text{ is positive semidefinite (p.s.d.).}$$

From (5.10), (5.11) and (5.12), we conclude that

$$(5.13) \quad \nu_1(0) - \nu_2(0), \ \nu_1(0) - \nu^*(0) \qquad \text{and} \quad \nu^*(0) - \nu_2(0) \qquad \text{are all p.s.d.}$$

In the multiparameter case, the relative efficiency may be judged by the generalized variance (D-optimality) or the trace of the covariance matrix (A-optimality) criterion. In view of the fact that by (2.18)–(2.20), both $\overline{B}_{\theta_0}$ and $\overline{P}_{\theta_0}$ are of full-rank, we have $\nu_1(0)$, $\nu_2(0)$ and $\nu^*(0)$ also of full rank. Hence, we have no difficulty in applying the first criterion. Similar results hold for the second criterion too. We define the *asymptotic generalized variance* as the tth root of the determinant of the a.d.m. In this light, the asymptotic relative efficiency (A.R.E.) of $\{\theta_n^*\}$ with respect to $\{\tilde{\theta}_n\}$ when $H_0 : \theta_0 \in \omega$ holds is

$$(5.14) \qquad e_0(\theta^*, \tilde{\theta}) = \{|\nu_1(0)|/|\nu^*(0)|\}^{1/t}.$$

Similarly, the A.R.E. of $\{\theta_n^*\}$ with respect to $\{\hat{\theta}_n\}$ is

$$(5.15) \qquad e_0(\theta^*, \hat{\theta}) = \{|\nu_2(0)|/|\nu^*(0)|\}^{1/t}.$$

Now, by (5.13), we have

$$(5.16) \qquad |\nu_1(0)| \geqslant |\nu^*(0)| \geqslant |\nu_2(0)|,$$

where the equality sign (in both places) holds when $\overline{\mathbf{B}}_{\theta_0}^{-1} = \overline{\mathbf{P}}_{\theta_0}$ i.e., $\mathbf{H}_{\theta_0}\overline{\mathbf{Q}}'_{\theta_0} = \mathbf{0}$ or equivalently, $\overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1} = \mathbf{0}$. This leads to the following.

THEOREM 5.1. *Under the assumptions made in Section 2, when $H_0 : \theta_0 \in \omega$ holds,*

$$(5.17) \qquad e_0(\theta^*, \hat{\theta}) \leqslant 1 \leqslant e_0(\theta^*, \tilde{\theta}),$$

where in both places the equality sign holds when $\overline{\mathbf{Q}}_{\theta_0}\overline{\mathbf{R}}_{\theta_0}^{-1} = \mathbf{0}$.

From Theorem 5.1 we conclude that under H_0, $\{\theta_n^*\}$ may not perform as well as $\{\hat{\theta}_n\}$; nevertheless, it is asymptotically at least as good as $\{\tilde{\theta}_n\}$, and hence, the PTMLE $\{\theta_n^*\}$ may be recommended as a replacement for the unrestricted MLE $\{\tilde{\theta}_n\}$.

Let us next consider the general case when $\{K_n\}$ in (4.4) holds. It follows from (5.1), (5.3) and (5.7) that

$$(5.18) \qquad \beta_1(\gamma) = 0, \beta_2(\gamma) = \gamma^* \qquad \text{and} \qquad \beta^*(\gamma) = \Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*)\beta_2(\gamma),$$

where $0 < \Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*) < \Pi_{r+2}(\chi^2_{r,\alpha}; 0) < \Pi_r(\chi^2_{r,\alpha}; 0) = 1 - \alpha < 1$. Thus, on noting that

$$(5.19) \qquad \lim_{\delta\uparrow\infty}\delta\Pi_{r+2}(\chi^2_{r,\alpha}; \delta) = 0,$$

we conclude that whereas $\tilde{\theta}_n$ is asymptotically unbiased, $\hat{\theta}_n$ and θ_n^* are not so, and, moreover, θ_n^* has smaller asymptotic bias than $\hat{\theta}_n$. Also, the asymptotic bias of $\hat{\theta}_n$ goes to ∞ if $\|\gamma\| \to \infty$, where as $\|\gamma\| \to \infty$ ($\Rightarrow \Delta^* \to \infty$), the asymptotic bias of $\theta_n^* \to 0$. Thus, θ_n^* has an edge over $\hat{\theta}_n$ with respect to the asymptotic bias. From (5.2) and (5.4), the A.R.E. of $\{\tilde{\theta}_n\}$ with respect to $\{\hat{\theta}_n\}$ is given by

$$(5.20) \qquad \begin{aligned} e(\tilde{\theta}, \hat{\theta}|\gamma^*) &= \left\{|\mathbf{P}_{\theta_0} + \gamma^*\gamma^{*\prime}|/|\overline{\mathbf{B}}_{\theta_0}^{-1}|\right\}^{1/t} \\ &= \left\{|\overline{\mathbf{B}}_{\theta_0}\overline{\mathbf{P}}_{\theta_0}||I + \overline{\mathbf{P}}_{\theta_0}^{-1}\gamma^*\gamma^{*\prime}|\right\}^{1/t} \\ &= e(\tilde{\theta}, \hat{\theta}|0)\left\{|\mathbf{I} + \overline{\mathbf{P}}_{\theta_0}^{-1}\gamma^*\gamma^{*\prime}|\right\}^{1/t} \\ &= e(\tilde{\theta}, \hat{\theta}|0)\left\{1 + \gamma^{*\prime}\overline{\mathbf{P}}_{\theta_0}^{-1}\gamma^*\right\}^{1/t}, \end{aligned}$$

as $\gamma^*\gamma^{*\prime}$ is of rank 1, so that $|\mathbf{I} + \mathbf{P}_{\theta_0}^{-1}\gamma^*\gamma^{*\prime}|$ = product of the characteristic roots of $\mathbf{I} + \mathbf{P}_{\theta_0}^{-1}\gamma^*\gamma^{*\prime} = 1 + $ largest root of $\overline{\mathbf{P}}_{\theta_0}^{-1}\gamma^*\gamma^{*\prime} = 1 + \gamma^{*\prime}\overline{\mathbf{P}}_{\theta_0}^{-1}\gamma^*$. Thus, if we let

$$(5.21) \qquad S_1 = \left\{\gamma^* : 1 + \gamma^{*\prime}\overline{\mathbf{P}}_{\theta_0}^{-1}\gamma^* \geqslant 1/e^t(\tilde{\theta}, \hat{\theta}|0) = |\overline{\mathbf{B}}_{\theta_0}\overline{\mathbf{P}}_{\theta_0}|^{-1}\right\},$$

then, from (5.20) and (5.21) we conclude that

$$(5.22) \qquad e(\tilde{\theta}, \hat{\theta}|\gamma^*) \geqslant 1, \qquad \forall \gamma^* \in S_1.$$

Similarly, it follows from (5.4) and (5.8) that

$$(5.23) \qquad e(\theta^*, \hat{\theta}|\gamma^*) = \left\{ \frac{|\overline{P}_{\theta_0} + \gamma^*\gamma^{*\prime}|}{|\overline{P}_{\theta_0} + a(\Delta^*)(\overline{B}_{\theta_0}^{-1} - \overline{P}_{\theta_0}) + b(\Delta^*)\gamma^*\gamma^{*\prime}|} \right\}^{1/t},$$

where

$$(5.24) \qquad a(\Delta^*) = 1 - \Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*) \qquad \text{and}$$

$$b(\Delta^*) = 2\Pi_{r+2}(\chi^2_{r,\alpha}; \Delta^*) - \Pi_{r+4}(\chi^2_{r,\alpha}; \Delta^*).$$

Now $a(\Delta^*) > \alpha$, $\forall \Delta^* \geq 0$ and it converges to 1 as $\Delta^* \to \infty$. Also, $0 \leq b(\Delta^*) \leq \Pi_r(\chi^2_{r,\alpha}; \Delta^*) \to 0$ as $\Delta^* \to \infty$. Thus, by (5.12), (5.23) and (5.24), we conclude that for γ^* close to $\mathbf{0}$, (5.24) is less than 1, while it exceeds 1 for all γ^* for which Δ^* is $\geq \Delta_0^*$ where

$$(5.25) \quad a(\Delta_0^*)\left[\overline{B}_{\theta_0}^{-1} - \overline{P}_{\theta_0}\right] = \left[1 - b(\Delta_0^*)\right]\gamma_0^*\gamma_0^{*\prime} \qquad \text{and} \qquad \gamma_0^{*\prime}\overline{B}_{\theta_0}\gamma_0^* = \Delta_0^*.$$

Thus, if we let

$$(5.26) \qquad S_2 = \left\{\gamma^* : \gamma^{*\prime}\overline{B}_{\theta_0}\gamma^* \geq \Delta_0^*\right\},$$

then, from (5.23) through (5.26), we conclude that

$$(5.27) \qquad e(\theta^*, \hat{\theta}|\gamma^*) \geq 1, \qquad \text{for every} \quad \gamma^* \in S_2.$$

We also note that if $\overline{B}_{\theta_0}^{-1} = \overline{P}_{\theta_0}$, then for every $\gamma^* \neq \mathbf{0}$, $b(\Delta^*)$ being bounded from above by 1, (5.23) is greater than 1; it tends to ∞ as $\Delta^* \to \infty$. In this case, S_2 constitutes the entire set of permissible values of γ^*. This leads us to the following.

THEOREM 5.2. *Under the assumptions made in Section 2, when $\{K_n\}$ in (4.4) holds,*

$$(5.28) \qquad e(\theta^*, \hat{\theta}|\gamma^*) \geq 1 \qquad \text{for every} \quad \gamma^* \in S_2,$$

where S_2 is defined by (5.26). Further, if $\overline{B}_{\theta_0}^{-1} = \overline{P}_{\theta_0}$, then the strict inequality sign in (5.28) holds for every $\gamma^* \neq \mathbf{0}$. Moreover, the A.R.E. in (5.28) tends to $+\infty$ as γ^* moves away from $\mathbf{0}$ i.e., Δ^* goes to $+\infty$.

From (5.2), (5.8) and (5.24), we obtain that

$$(5.29) \quad e(\theta^*, \tilde{\theta}|\gamma^*) = \left\{|\overline{B}_{\theta_0}^{-1} - \{1 - a(\Delta^*)\}\{\overline{B}_{\theta_0}^{-1} - \overline{P}_{\theta_0}\} + b(\Delta^*)\gamma^*\gamma^{*\prime}\|\overline{B}_{\theta_0}|\right\}^{-1/t}$$

$$= \left\{|\mathbf{I} + b(\Delta^*)\overline{B}_{\theta_0}\gamma^*\gamma^{*\prime} - (1 - a(\Delta^*))(I - \overline{B}_{\theta_0}\overline{P}_{\theta_0})|\right\}^{-1/t}.$$

Thus, for γ^* close to $\mathbf{0}$, (5.29) exceeds one, while, it is less than or equal to one when γ^* is away from $\mathbf{0}$ (in the sense that $\Delta^* \to \infty$). Note that when $\overline{B}_{\theta_0}^{-1} = \overline{P}_{\theta_0}$, the third term in the determinant on the right hand side of (5.29) drops out and as $b(\Delta^*) \geq 0$ for every $\Delta^* > 0$, (5.29) can not exceed 1 for any $\gamma^* \neq \mathbf{0}$. Usually $b(\Delta^*)$ is small and (5.29) remains close to 1. Further, $b(\Delta^*)$ converges to 0 as $\Delta^* \to \infty$ while $a(\Delta^*)$ goes to 1 as $\Delta^* \to \infty$, and hence, (5.29) converges to 1 as γ^* moves away from $\mathbf{0}$ (i.e., $\Delta^* \to \infty$). Thus, for large Δ^*, $\{\theta_n^*\}$ and $\{\tilde{\theta}_n\}$ have similar

performances (better than $\{\hat{\theta}_n\}$), while for smaller values of Δ^*, $\{\hat{\theta}_n\}$ appears to have an edge over the others. In particular, if $\overline{\mathbf{B}}_{\theta_0}^{-1} = \overline{\mathbf{P}}_{\theta_0}$ then the restricted MLE has no better performance even if $H_0 : \theta_0 \in \omega$ holds, and hence, it should not be recommended as a substitute for the unrestricted one. Combining these with (5.9), (5.17) and (5.18), we conclude that the restricted MLE, $\hat{\theta}_n$, though is asymptotically optimal when $H_0 : \theta_0 \in \omega$ holds, is (asymptotically) biased and its a.d.m. becomes larger (in the sense of the generalized variance or the trace) when γ^* moves away from $\mathbf{0}$ and this makes it less efficient when $H_0 : \theta_0 \in \omega$ may not hold. On the other hand, the unrestricted MLE $\tilde{\theta}_n$ remains asymptotically unbiased for θ_0 irrespective of $H_0 : \theta_0 \in \omega$, but it is usually not optimal when H_0 holds. As a compromise, the PTMLE θ_n^* performs better than $\tilde{\theta}_n$ when γ^* is small and has uniformly a lower order of bias than $\hat{\theta}_n$. For large γ^*, it performs better than $\hat{\theta}_n$ and very similar to $\tilde{\theta}_n$. Hence, it can be recommended on the grounds of robustness against any deviation from $H_0 : \theta_0 \in \omega$. If, however, $\overline{\mathbf{B}}_{\theta_0}^{-1} = \overline{\mathbf{P}}_{\theta_0}$, then the unrestricted MLE can be recommended without any reservation and there is no point in using either a restricted MLE or a PTMLE. Nonoptimality of the PTMLE under H_0 has been studied by Huntsberger (1955) for the special case of normal distributions. His conclusions do not remain valid when H_0 does not hold. We define

$$(5.30) \qquad S_3 = \left\{ \gamma^* : \left[1 - a(\Delta^*) \right] \left[\overline{\mathbf{B}}_{\theta_0}^{-1} - \overline{\mathbf{P}}_{\theta_0} \right] - b(\Delta^*) \gamma^* \gamma^{*\prime} \text{ is p.s.d.} \right\};$$

$$(5.31) \qquad S^* = S_2 \cap S_3,$$

where S_2 is defined by (5.26). Note that if $\overline{\mathbf{B}}_{\theta_0}^{-1} = \overline{\mathbf{P}}_{\theta_0}$, then S^* is an empty set. From (5.27), (5.29), (5.30) and (5.31), we arrive at the following.

THEOREM 5.3. *Under* $\{K_n\}$ *in* (4.4) *and the regularity conditions of Section 2, whenever* S^* *is a nonempty set,*

$$(5.32) \qquad e(\theta^*, \tilde{\theta}|\gamma^*) \geqslant 1 \quad \text{and} \quad e(\theta^*, \hat{\theta}|\gamma^*) \geqslant 1, \qquad \forall \, \gamma^* \in S^*.$$

The last theorem reveals the superiority of the PTMLE near the boundary of ω when n is large. We conclude this section with the following remark. The asymptotic distribution theory of the PTMLE and the A.R.E. of all the three MLE's studied in this paper rest on the asymptotic multinormality of the unrestricted and the restricted MLE's as well as on the asymptotic (noncentral) chi-square distribution of L_n, defined in (2.5). In some problems, arising mostly in testing against one-sided or restricted alternatives, a variant form of L_n may arise, which does not have asymptotically a noncentral chi-square distribution. For example, if $k = 2$ and F_1, F_2 are both normal with a common variance σ^2 and means μ_1 and μ_2, respectively, then, for testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 > \mu_2$ (or $\mu_1 < \mu_2$), the one-sided optimal test is based on the Student t-statistic which has asymptotically a normal distribution. In that case, the effect of such a one-sided test will be a somewhat different asymptotic distribution of the PTMLE where (5.7) may not be equal to 0 when H_0 holds. This asymptotic bias depends on the level of significance

of the preliminary test but is usually small. This bias of the PTMLE may also affect (5.17) to a certain extent. In this simple case, the situation is similar to the nonparametric PTMLE studied by Saleh and Sen (1978) and Sen and Saleh (1979). But, in a general framework of preliminary tests against restricted alternatives, expressions for the asymptotic bias and a.d.m. of the PTMLE's are quite involved and different from the parallel ones considered here. As a result, conclusions about the A.R.E. of the PTMLE may also be different from the ones obtained here.

REFERENCES

1. AITCHISON, J. and SILVEY, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29** 813–828.
2. BANCROFT, T. A. and HAN, C. P. (1977). Inference based on conditional specification: a note and a bibliography. *Review Internat. Statist.* **45** 117–128.
3. FEDER, P. I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is near the boundaries of the hypothesis region. *Ann. Math. Statist.* **39** 2044–2055.
4. HUBER, P. J. (1967). The behavior of maximum likelihood estimators under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probability* **1** 221–234.
5. HUNTSBERGER, D. V. (1955). A generalization of a preliminary test procedure for pooling data. *Ann. Math. Statist.* **26** 734–743.
6. INAGAKI, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. *Ann. Inst. Statist. Math.* **25** 1–26.
7. KITAGAWA, T. (1963). Estimation after preliminary tests of significance. *Univ. of Calif. Publ. Statist.* **3** 147–186.
8. SALEH, A. K. MD. E. and SEN, P. K. (1978). Nonparametric estimation of location parameter after a preliminary test on regression. *Ann. Statist.* **6** 154–168.
9. SEN, P. K. and SALEH, A. K. MD. E. (1979). Nonparametric estimation of location parameter after a preliminary test on regression in the multivariate case. *J. Multivariate Anal.* **9**.
10. SILVEY, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.* **30** 389–407.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514