Ann. Inst. Statist. Math.
Vol. 46, No. 3, 593–604 (1994)

ON FISHER INFORMATION INEQUALITIES IN THE PRESENCE OF NUISANCE PARAMETERS

VASANT P. BHAPKAR* AND CIDAMBI SRINIVASAN**

Department of Statistics, University of Kentucky, 817 Patterson Office Tower, Lexington, KY 40506-0027, U.S.A.

(Received August 7, 1991; revised November 12, 1993)

Abstract. The existence of a generalized Fisher information matrix for a vector parameter of interest is established for the case where nuisance parameters are present under general conditions. A matrix inequality is established for the information in an estimating function for the vector parameter of interest. It is shown that this inequality leads to a sharper lower bound for the variance matrix of unbiased estimators, for any set of functionally independent functions of parameters of interest, than the lower bound provided by the Cramér-Rao inequality in terms of the full parameter.

Key words and phrases: Information matrix, partial sufficiency, partial ancillarity, estimating functions.

1. Introduction

Suppose X has probability distribution P_θ with parameter θ in some open interval Θ of $\mathbb{R}_1$. Under the standard regularity assumptions for unbiased estimation of parametric function $f(\theta)$, the Cramér-Rao inequality

$$V_\theta(T) \geq \frac{[f'(\theta)]^2}{I(\theta)} \tag{1.1}$$

provides a lower bound for the variance of unbiased estimators $T = T(X)$, $I(\theta)$ being the Fisher information concerning parameter θ in X (or in the distribution P_θ of X).

If we consider the *score* function $l(x;\theta) = \partial \log p(x;\theta)/\partial\theta$, where p is the probability density function (p.d.f.), then the basic optimality of the score function l as an *estimating function* for θ was established by Godambe (1960). Suppose $g = g(x;\theta)$ is to be used as an estimating function to provide unbiased estimating equation $g(x;\theta) = 0$, then Godambe's result may be stated in the form

$$I_g(\theta) \leq I_l(\theta) = I(\theta), \tag{1.2}$$

* Supported in part by NSF Grant MCS-8806233.

** Supported in part by NSF Grants RII-8610671, ATM-9108177 and DMS-9204380.

for all *regular unbiased* estimating functions g, where I_g is the *information* concerning θ *in the estimating function* g (Bhapkar (1972))

$$I_g(\theta) = \frac{E_\theta^2[\partial g(X;\theta)/\partial\theta]}{\sigma_g^2(\theta)}, \tag{1.3}$$

and $\sigma_g^2(\theta) = V_\theta g(X;\theta)$. For $g(X;\theta) = T - f(\theta)$, the inequality (1.1) follows from (1.2).

More generally, for $\boldsymbol{\theta}$ in an open interval of $\mathbb{R}_k$, we have the matrix form of the Cramér-Rao inequality under regularity assumptions (see, e.g., Rao (1965))

$$\boldsymbol{V}_\theta(\boldsymbol{T}) \geq \boldsymbol{F}(\boldsymbol{\theta})\boldsymbol{I}^{-1}(\boldsymbol{\theta})\boldsymbol{F}'(\boldsymbol{\theta}), \tag{1.4}$$

where $\boldsymbol{F}(\boldsymbol{\theta})$ is the $k \times k$ matrix of partial derivatives $\partial f_i(\theta)/\partial\theta_j$, $\boldsymbol{I}(\boldsymbol{\theta})$ is the Fisher information matrix and $\boldsymbol{T}$ is unbiased estimator of $\boldsymbol{f}(\boldsymbol{\theta})$.

The matrix form of (1.2) is

$$\boldsymbol{I}_g(\boldsymbol{\theta}) \leq \boldsymbol{I}_l(\boldsymbol{\theta}) = \boldsymbol{I}(\boldsymbol{\theta}), \tag{1.5}$$

where the information in the estimating function $\boldsymbol{g} = \boldsymbol{g}(x;\boldsymbol{\theta})$ is

$$\boldsymbol{I}_g(\boldsymbol{\theta}) = \boldsymbol{G}'(\boldsymbol{\theta})\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\theta})\boldsymbol{G}(\boldsymbol{\theta}), \tag{1.6}$$

$\boldsymbol{G}(\boldsymbol{\theta})$ is the $k \times k$ matrix $E_{\boldsymbol{\theta}}[\partial g_i(X;\boldsymbol{\theta})/\partial\theta_j]$, and $\boldsymbol{\Sigma}_g(\boldsymbol{\theta}) = \boldsymbol{V}_{\boldsymbol{\theta}}(\boldsymbol{g}(X;\boldsymbol{\theta}))$ (see, Kale (1962), Bhapkar (1972)). Again (1.4) follows from (1.5) by taking $\boldsymbol{g}(X;\boldsymbol{\theta}) = \boldsymbol{T} - \boldsymbol{f}(\boldsymbol{\theta})$.

In Section 2 the basic notation and terminology is given for the case of parameter of interest θ in the presence of nuisance parameters ϕ. Existence of the generalized Fisher information matrix for θ is established in Section 3. Inequalities are considered in Section 4 for information in estimating functions in the presence of nuisance parameters. The inequality (4.4) is shown to lead to a sharper inequality than the Cramér-Rao inequality for unbiased estimation of parametric functions in the presence of nuisance parameters. Finally, Section 5 develops the inequality concerning generalized information matrices for X and any given statistic $S = S(X)$ with some examples.

2. Notation and terminology

Suppose that the random variable X with values x in the sample space χ has probability distribution P_ω with the density function (p.d.f.) $p(x;\omega)$ with respect to a σ-finite measure μ. Assume that the parameter $\omega = (\theta, \phi)$, where θ is the parameter of interest, and ϕ is the nuisance parameter variation independent of θ in the sense that the corresponding parametric spaces satisfy the relation $\Omega = \Theta \times \boldsymbol{\Phi}$. It is also assumed that Ω is an open subset of d-dimensional Euclidean space and θ, ϕ have dimensionalities d_1 and d_2, respectively.

Let $\boldsymbol{l}'_\omega(x) = \partial \log p(x;\omega)/\partial\omega$ denote the row-vector of partial derivatives and

$$\boldsymbol{I}(\omega) = E_\omega \boldsymbol{l}_\omega(X)\boldsymbol{l}'_\omega(X)$$

is then the usual Fisher information matrix.

Along with $\boldsymbol{l}_\omega(x)$, the *score function* for the distribution of X, we call its components $\boldsymbol{l}_\theta$, $\boldsymbol{l}_\phi$ the θ-score and ϕ-score, respectively.

We assume the following regularity conditions on the probability distribution P_ω:

$\boldsymbol{R}$: (i) P_ω has p.d.f. $p(x;\omega)$ with respect to measure μ and the partial derivatives up to second order, of $\int pd\mu$ can be obtained by differentiating with respect to ω under the integral sign for all $\omega \in \Omega$.

(ii) $\boldsymbol{I}(\omega) = E_\omega \boldsymbol{l}_\omega(X)\boldsymbol{l}'_\omega(X)$ is positive-definite (p.d.) for all ω.

Now the Fisher information matrix for parameter of interest θ is defined as

$$\boldsymbol{I}(\theta;\omega) = \min_{\boldsymbol{N}} E_\omega[\boldsymbol{m}(\boldsymbol{N})\boldsymbol{m}'(\boldsymbol{N})] \tag{2.1}$$

where $\boldsymbol{m}(\boldsymbol{N}) = \boldsymbol{l}_\theta(X) - \boldsymbol{N}'(\omega)\boldsymbol{l}_\phi(X)$ for any $d_2 \times d_1$ matrix $\boldsymbol{N}(\omega)$, and min denotes the *minimal* matrix $\boldsymbol{M}^*_\omega$ in the class $\mathcal{M}_\omega$ of non-negative definite (n.n.d.) matrices $\boldsymbol{M}_\omega(\boldsymbol{N}) = E_\omega[\boldsymbol{m}(\boldsymbol{N})\boldsymbol{m}'(\boldsymbol{N})]$ in the sense $\boldsymbol{M}^*_\omega \le \boldsymbol{M}_\omega$, i.e., $\boldsymbol{M} - \boldsymbol{M}^*$ is n.n.d., for all $\boldsymbol{M}_\omega$ in $\mathcal{M}_\omega$.

Such a minimal matrix exists and we have

$$\boldsymbol{I}(\theta;\omega) = \boldsymbol{I}_{11} - \boldsymbol{B}'\boldsymbol{I}_{22}\boldsymbol{B}, \tag{2.2}$$

for any $\boldsymbol{B}$ satisfying $\boldsymbol{I}_{22}\boldsymbol{B} = \boldsymbol{I}_{21}$. Here $\boldsymbol{I}_{ij}$ denote the partitioned submatrices of $\boldsymbol{I}(\omega)$ by d_1, d_2 rows and d_1, d_2 columns. Note that the expression (2.2) is invariant under any choice of $\boldsymbol{B}$, and $\boldsymbol{I}(\theta;\theta) = \boldsymbol{I}(\theta)$. If $\boldsymbol{I}_{22}$ is p.d., as under condition (ii) in $\boldsymbol{R}$, then $\boldsymbol{I}(\theta;\omega) = \boldsymbol{I}_{11} - \boldsymbol{I}_{12}\boldsymbol{I}_{22}^{-1}\boldsymbol{I}_{21}$.

Now we turn our attention to the generalization of the above information function due to Godambe (1984) and the matrix form discussed by Bhapkar (1989, 1991).

Let $\mathcal{G}$ be the class of real-valued functions $g(x;\theta)$ over $\mathcal{X} \times \Theta$ satisfying the following regularity condition:

$\boldsymbol{R}_\mathcal{G}$: $E_\omega g(X;\theta) = 0$, $E_\omega g^2 = \int g^2 pd\mu < \infty$ for all $\omega \in \Omega$.

Now consider the class of real-valued functions $u(x;\omega)$ such that for all $\omega \in \Omega$

$$\begin{aligned} &\text{(i)} \quad E_\omega u^2(X;\omega) < \infty, \\ &\text{(ii)} \quad E_\omega u(X;\omega)g(X;\theta) = 0, \end{aligned} \tag{2.3}$$

for all $g \in \mathcal{G}$.

The *generalized* Fisher information matrix is defined as

$$\boldsymbol{I}_G(\theta;\omega) = \min_{\boldsymbol{u}} E_\omega \boldsymbol{n}(\boldsymbol{u})\boldsymbol{n}'(\boldsymbol{u}), \tag{2.4}$$

where $\boldsymbol{n}(\boldsymbol{u}) = \boldsymbol{l}_\theta(X) - \boldsymbol{u}(X;\omega)$, $\boldsymbol{u}$ is a vector with elements satisfying (2.3) and the min denotes the minimal matrix in the class of n.n.d. matrices $\boldsymbol{M}(\boldsymbol{u}) = E_\omega(\boldsymbol{n}(\boldsymbol{u})\boldsymbol{n}'(\boldsymbol{u}))$.

If $\boldsymbol{u}$ is a vector with elements satisfying condition ((2.3)), let $E_\omega \boldsymbol{u}(X;\omega) = \boldsymbol{v}(\omega)$ and $\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{v}$. Then elements of $\boldsymbol{w}$ satisfy condition ((2.3)). Furthermore,

$$E_\omega \boldsymbol{n}(\boldsymbol{u})\boldsymbol{n}'(\boldsymbol{u}) = E_\omega \boldsymbol{n}(\boldsymbol{w})\boldsymbol{n}'(\boldsymbol{w}) + \boldsymbol{v}\boldsymbol{v}' \ge E_\omega \boldsymbol{n}(\boldsymbol{w})\boldsymbol{n}'(\boldsymbol{w}).$$

Hence, without loss of generality, we could confine attention to $\boldsymbol{u}$ with zero mean and elements satisfying (2.3) in order to define $\boldsymbol{I}_G(\theta;\omega)$.

Let then $\mathcal{U}$ be the space of real-valued functions $u = (x;\omega)$ such that $u(X;\omega)$ has zero mean, finite variance and it is uncorrelated with $g(X;\theta)$ for all $\omega \in \Omega$ and every $g \in \mathcal{G}$.

The following lemma is easy to prove by showing that $\boldsymbol{u} = \boldsymbol{0}$ gives the minimal matrix.

LEMMA 2.1. *If $\omega = \theta$ and ϕ is absent, then $\boldsymbol{I}_G(\theta;\theta) = \boldsymbol{I}(\theta)$.*

Although the existence of the matrix $\boldsymbol{I}_G(\theta;\omega)$ was established under some special conditions where suitable p-ancillary, or p-sufficient statistics exist for θ (see Bhapkar (1989, 1991)), its general existence under regularity conditions $\boldsymbol{R}$ and $\boldsymbol{R}_\mathcal{G}$ alone still remained unproven.

In the next section we establish the existence of the *integrated* (or global) version of generalized Fisher information matrix for θ from which the existence of the *local* version $\boldsymbol{I}_G$ follows.

3. Generalized information matrix for θ

Let $\mathcal{P}$ be a family of probability measures P_ω, $\omega \in \Omega$ over the measurable space $(\chi, \mathcal{A})$, dominated by a σ-finite measure μ, and suppose $p(x;\omega)$ is the p.d.f. $dP_\omega/d\mu$. Assume that p satisfies the Cramér-Rao regularity conditions $\boldsymbol{R}$. Assume also that II is the class of all probability measures π over the measurable space (Ω, β).

Before establishing existence of the minimal matrix $\boldsymbol{I}_G$, defined by (2.4), we first establish existence of its integrated version $\boldsymbol{J}_\pi$ as the minimal matrix in a certain class of n.n.d. matrices for any given probability measure π in II.

Let 'E' denote the overall expectation with respect to p (or μ) as well as π, while E_ω denotes the conditional expectation $E(\cdot \mid \omega)$. Consider the space $\mathcal{C}$ of functions $c = c(x;\omega)$ from $\chi \times \Omega \to R_1$ such that

$$
\begin{aligned}
&E_\omega(c) \equiv \int c(x;\omega)p(x;\omega)d\mu(x) = 0, \qquad E_\omega(c^2) < \infty \quad \text{for all } \omega \in \Omega, \\
&E(c^2) \equiv \int c^2(x;\omega)p(x;\omega)d\mu(x)d\pi(\omega) < \infty.
\end{aligned} \tag{3.1}
$$

Then $\mathcal{C}$ is a Hilbert space with the inner product defined by

$$
\langle c_1, c_2 \rangle = \int c_1(x;\omega)c_2(x;\omega)p(x;\omega)d\mu(x)d\pi(\omega). \tag{3.2}
$$

Let $\mathcal{G}^*$ be the subspace of functions $g = g(x;\theta)$ in $\mathcal{C}$ which depend on ω only through θ. Denote by $\mathcal{U}^*$ the orthogonal complement of $\mathcal{G}^*$ in $\mathcal{C}$.

Consider now the class $\mathcal{J}$ of n.n.d. matrices

$$
\boldsymbol{J}(\boldsymbol{u}) = E\boldsymbol{n}(\boldsymbol{u})\boldsymbol{n}'(\boldsymbol{u}), \tag{3.3}
$$

where $\boldsymbol{n}(\boldsymbol{u}) = \boldsymbol{l}_\theta(X) - \boldsymbol{u}$, and $\boldsymbol{u} = \boldsymbol{u}(x;\omega)$ is a vector with all elements belonging to $\mathcal{U}^*$, assuming that all elements of the θ-score function $\boldsymbol{l}_\theta$ belong to $\mathcal{C}$.

The integrated version of Fisher information function for parameter of interest θ is now defined as

$$\boldsymbol{J}_\pi = \min_{\boldsymbol{u}} \boldsymbol{J}(\boldsymbol{u}), \tag{3.4}$$

where the minimum is for the class $\mathcal{J}$ with $\boldsymbol{u}$ having elements in $\mathcal{U}^*$.

The existence of the minimal matrix can now be demonstrated. By the projection theorem (see, e.g., Friedman (1982)) that for each real-valued $c = c(x;\omega)$ in $\mathcal{C}$ there is a unique decomposition $c = g^* + u^*$, where $g^* = g^*(x;\theta) \in \mathcal{G}^*$ and $u^* = u^*(x;\omega) \in \mathcal{U}^*$.

If $\boldsymbol{l}_\theta$ has all elements in $\mathcal{C}$, then we have the unique decomposition $\boldsymbol{l}_\theta(X) = \boldsymbol{g}^*(X;\theta) + \boldsymbol{u}^*(X;\omega)$ where all elements of $\boldsymbol{g}^*$, $\boldsymbol{u}^*$ belong to subspaces $\mathcal{G}^*$, $\mathcal{U}^*$, respectively.

Hence, from (3.3), we have

$$\boldsymbol{J}(u) = E\boldsymbol{g}^*\boldsymbol{g}^{*\prime} + E(\boldsymbol{u}^* - \boldsymbol{u})(\boldsymbol{u}^* - \boldsymbol{u})'; \tag{3.5}$$

since the cross-product terms $E\boldsymbol{g}^*(\boldsymbol{u}^* - \boldsymbol{u})'$ vanish. From (3.4) and (3.5), it now follows that

$$\boldsymbol{J}_\pi = E\boldsymbol{g}^*\boldsymbol{g}^{*\prime}, \tag{3.6}$$

since the minimum is attained at $\boldsymbol{u} = \boldsymbol{u}^*$. Thus we have

PROPOSITION 3.1. *If all elements of $\boldsymbol{l}_\theta(X)$ belong to $\mathcal{C}$, and $\mathcal{G}^*$ is the subspace of functions depending on ω only through θ, then the integrated version $\boldsymbol{J}_\pi$ exists and it is given by* (3.6) *where $\boldsymbol{g}^*$ is the projection of $\boldsymbol{l}_\theta(X)$ onto the subspace $\mathcal{G}^*$.*

Considering now the one-point degenerate distribution π at ω we have

THEOREM 3.1. *Assume conditions $\boldsymbol{R}$ and $\boldsymbol{R}_\mathcal{G}$. Then the generalized information matrix $\boldsymbol{I}_G(\theta;\omega)$, defined by* (2.4), *exists; it is given by*

$$\boldsymbol{I}_G(\theta;\omega) = E_\omega \boldsymbol{g}^*(X;\theta)\boldsymbol{g}^{*\prime}(X;\theta), \tag{3.7}$$

where $\boldsymbol{g}^*$ is the projection of $\boldsymbol{l}_\theta(X)$ onto the subspace $\mathcal{G}$.

We note here that $\mathcal{G}^* = \mathcal{G}$ and $\mathcal{U}^* = \mathcal{U}$, with $\mathcal{G}$, $\mathcal{U}$ defined in Section 2, for one-point distribution π at ω.

Suppose now that the functions g in $\mathcal{G}$ are futher restricted by the *regularity* assumption (see Godambe (1984), Bhapkar (1989, 1991)).

$\boldsymbol{R}_\mathcal{G}^*$: In addition to $\boldsymbol{R}_\mathcal{G}$, the function $g = g(X;\theta)$ has partial derivatives with respect to θ and, furthermore, the equality $\int g p d\mu = 0$ can be differentiated with respect to elements of ω under the integral sign.

Hereafter we thus assume that $\mathcal{G}$ is the class of *regular* real-valued functions $g = g(x;\theta)$ which satisfy conditions $\boldsymbol{R}^*_{\mathcal{G}}$, and $\mathcal{U}$ is the class of real-valued functions $u = u(x;\omega)$ with zero mean which are uncorrelated with g in $\mathcal{G}$.

Let $\mathcal{G}^*$ denote now the *closed* linear subspace in $\mathcal{C}$ spanned by functions g in $\mathcal{G} \cap \mathcal{C}$. As before, let $\mathcal{U}^*$ denote the complement of $\mathcal{G}^*$ in $\mathcal{C}$. For $\boldsymbol{J}_\pi$, defined by (3.4) with $\mathcal{U}^*$ defined as here, the proof of Proposition 3.1 remains valid.

More specifically, for the one-point distribution π at ω, we have

THEOREM 3.2. *Assume conditions $\boldsymbol{R}$, and suppose $\mathcal{G}^*$ is the closed linear subspace spanned by functions g which satisfy $\boldsymbol{R}^*_{\mathcal{G}}$. Then $\boldsymbol{I}_G(\theta;\omega)$ exists and it is given by* (3.7) *where $\boldsymbol{g}^*$ is the projection of $\boldsymbol{l}_\theta(X)$ onto $\mathcal{G}^*$.*

Note here that for a one-point distribution π at ω, $\mathcal{G} \subset \mathcal{G}^*$ since not every g^* in $\mathcal{G}^*$ is differentiable with respect to θ. Hence $\mathcal{U}^* \subset \mathcal{U}$ and, thus, $\boldsymbol{I}_G(\theta;\omega)$ is the minimal matrix at $\boldsymbol{u} = \boldsymbol{u}^*$ (in the proof of Proposition 3.1).

4. Inequalities with nuisance parameters

We now generalize the inequalities in Section 1 to the case of parameter of interest θ in the presence of nuisance parameters ϕ.

Assume that the elements of estimating function $\boldsymbol{g}$ for θ satisfy the regularity conditions $\boldsymbol{R}^*_{\mathcal{G}}$ in Section 3. We have then

$$E_\omega[\boldsymbol{g}(X;\theta)\boldsymbol{l}'_\theta(X;\omega)] = -\boldsymbol{G}(\omega), \tag{4.1}$$

where $\boldsymbol{G}(\omega)$ is the $d_1 \times d_1$ matrix $E_\omega[\partial\boldsymbol{g}/\partial\theta]$. Using now the terminology of Section 3, we have then for probability measure π over Ω

$$\begin{aligned}\boldsymbol{E}[\boldsymbol{g}(X;\theta)\boldsymbol{l}'_\theta(X;\omega)] &= -E\left[\frac{\partial\boldsymbol{g}(X;\theta)}{\partial\theta}\right]\\ &= -\boldsymbol{H}, \qquad \text{say.}\end{aligned}$$

Therefore for any vector $\boldsymbol{u}$ with elements in $\mathcal{U}^*$ we have

$$\boldsymbol{E}[\boldsymbol{g}(X;\theta)\{\boldsymbol{l}_\theta(X;\omega) - \boldsymbol{u}(X;\omega)\}'] = -\boldsymbol{H}.$$

Arguing now as in Rao ((1965), p. 266) we see that, for every u in $\mathcal{U}^*$ the matrix

$$\begin{vmatrix}\boldsymbol{\Sigma}_g & -\boldsymbol{H}\\ \boldsymbol{H}' & \boldsymbol{J}(\boldsymbol{u})\end{vmatrix}$$

is n.n.d.; here $\boldsymbol{\Sigma}_g = \boldsymbol{E}(\boldsymbol{g}\boldsymbol{g}')$ and $\boldsymbol{J}(\boldsymbol{u})$ is defined by (3.3). If $\boldsymbol{\Sigma}_g$ is non-singular, we have $\boldsymbol{J}(\boldsymbol{u}) - \boldsymbol{H}'\boldsymbol{\Sigma}_g^{-1}\boldsymbol{H}$ *n.n.d.* for every $\boldsymbol{u}$. In view of (3.4), we have then

$$\boldsymbol{H}'\boldsymbol{\Sigma}_g^{-1}\boldsymbol{H} \le \boldsymbol{J}_\pi. \tag{4.2}$$

DEFINITION 4.1. The information concerning θ with regard to π in the regular unbaised estimating function $\boldsymbol{g}$, with *p.d.* covariance matrix $\boldsymbol{\Sigma}_g$ is defined as

$$\boldsymbol{J}_g = \boldsymbol{H}'\boldsymbol{\Sigma}_g^{-1}\boldsymbol{H}.$$

Since (4.2) is true for every vector $\boldsymbol{u}$ with elements in $\mathcal{U}^*$, we have

PROPOSITION 4.1. *For every regular estimating function $\boldsymbol{g}$ for θ with p.d. covariance matrix*

$$\boldsymbol{J}_g \leq \boldsymbol{J}_\pi. \tag{4.3}$$

The inequality (4.3) is a generalization of the inequality that one obtains for a one-point distribution π at ω, viz.

$$\boldsymbol{I}_g(\theta;\omega) \leq \boldsymbol{I}_G(\theta;\omega), \tag{4.4}$$

where $\boldsymbol{I}_g = \boldsymbol{G}'(\omega)\boldsymbol{\Sigma}_g^{-1}(\omega)\boldsymbol{G}(\omega)$ is the *local* information for θ in estimating function $\boldsymbol{g}$, and $\boldsymbol{\Sigma}_g(\omega) = E_\omega[\boldsymbol{gg}']$; see Ferreira (1982) for an equivalent definition of $\boldsymbol{I}_g$. The relation (4.4) was established in Bhapkar (1989) *provided* the minimal matrix $\boldsymbol{I}_G$ exists. In view of the existence Theorem 3.2, such a qualifying statement is no longer necessary. The inequality (4.4) is the vector extension of the scalar result by Godambe (1984). It is now seen that the inequality (4.4) is the generalization of the outside inequality in (1.5) to the case where nuisance parameters are present.

A weaker inequality

$$\boldsymbol{I}_g(\theta;\omega) \leq \boldsymbol{I}(\theta;\omega) \tag{4.5}$$

was established by Chandrasekar and Kale (1984), where $\boldsymbol{I}(\theta;\omega) = [\boldsymbol{I}^{11}(\theta;\omega)]^{-1}$ is the information concerning θ in P_ω, as defined by (2.1) and (2.2), and

$$\boldsymbol{I}^{-1}(\omega) = \begin{bmatrix} \boldsymbol{I}^{11} & \boldsymbol{I}^{12} \\ \boldsymbol{I}^{21} & \boldsymbol{I}^{22} \end{bmatrix}.$$

However, it was shown in Bhapkar (1989) that $\boldsymbol{I}(\theta;\omega) \geq \boldsymbol{I}_G(\theta;\omega)$; thus the relation (4.4) provides an inequality which is sharper than the inequality (4.5).

The inequality (4.4) is also seen to lead to a sharper lower bound to the variance (or covariance matrix) of unbiased estimators of any d_1-vector of functionally independent estimable functions $\boldsymbol{f}(\theta)$ of the parameter of interest θ, in the presence of nuisance parameters ϕ, than the Cramer-Rao inequality. If $\boldsymbol{T}$ is unbiased for $\boldsymbol{f}(\theta)$, we have (see, e.g., Rao (1965), p. 265)

$$V_\omega(\boldsymbol{T}) \geq \boldsymbol{F}(\omega)\boldsymbol{I}^{-1}(\omega)\boldsymbol{F}'(\omega) = \boldsymbol{F}_1(\theta)\boldsymbol{I}^{11}(\omega)\boldsymbol{F}_1'(\theta), \tag{4.6}$$

where $\boldsymbol{F}(\omega)$ is the $d_1 \times d$ matrix $[\partial \boldsymbol{f}/\partial\omega]$, while $\boldsymbol{F}_1(\theta)$ is the $d_1 \times d_1$ matrix $[\partial \boldsymbol{f}/\partial\theta]$. Now for the estimating function $\boldsymbol{g} = \boldsymbol{T} - \boldsymbol{f}(\theta)$, we have

$$\boldsymbol{I}_g = \boldsymbol{F}_1'(\theta)\boldsymbol{\Sigma}_g^{-1}(\omega)\boldsymbol{F}_1(\theta) = \boldsymbol{F}_1'(\theta)\boldsymbol{V}^{-1}\boldsymbol{F}_1(\theta).$$

In view of (4.4) we have $\boldsymbol{V}^{-1} \leq (\boldsymbol{F}_1')^{-1}\boldsymbol{I}_G\boldsymbol{F}_1^{-1}$ i.e.

$$V_\omega(\boldsymbol{T}) \geq \boldsymbol{F}_1(\theta)\boldsymbol{I}_G^{-1}(\theta;\omega)\boldsymbol{F}_1'(\theta), \tag{4.7}$$

provided $\boldsymbol{I}_G$ is nonsingular. However, $\boldsymbol{I}_G(\theta;\omega) \leq \boldsymbol{I}(\theta;\omega) = (\boldsymbol{I}^{11})^{-1}$ implies $\boldsymbol{I}_G^{-1} \geq \boldsymbol{I}^{11}$. Thus the inequality (4.7) is seen to be sharper than the inequality in (4.6).

5. Information concerning θ in statistic S

Consider now the statistic $S = S(X)$ where X has distribution P_ω which satisfies regularity conditions $\boldsymbol{R}$ in Section 2. Hereafter we assume the extended regularity conditions $\boldsymbol{R}^*$ defined below.

$\boldsymbol{R}^*$: (a) The conditions (i) and (ii) in $\boldsymbol{R}$.

(b) There exists a minimal sufficient statistic (S, T) for $\mathcal{P} = \{P_\omega : w \in \Omega\}$ and measures ν, η_s such that

$$p(x;\omega) = f(s;\omega)h(t;\omega \mid s)k(x \mid s,t), \tag{5.1}$$

where f is the p.d.f. of S with respect to ν, h is the conditional p.d.f. of T with respect to η_s, given $s = S(x)$, for almost all $(\nu)s$.

(c) f and h satisfy condition (i) in $\boldsymbol{R}$ for integrals with respect to ν and η_s, respectively.

(d) The Fisher information matrix $\boldsymbol{I}^{(S)}(\omega) = E_\omega \boldsymbol{l}^{(m)}_{(\omega)}(S)\boldsymbol{l}'^{(m)}_\omega(S)$ exists for the marginal distribution of S, where $\boldsymbol{l}^{(m)}_\omega(S)$ is the *marginal score function* of S given by $[\partial \log f(S;\omega)/\partial\omega]'$; similarly the conditional Fisher information matrix (of T) given s, is

$$\boldsymbol{I}(\omega \mid s) = E_\omega[\boldsymbol{l}^{(c)}_\omega(T;\omega \mid s)\boldsymbol{l}^{(c)\prime}_\omega \mid s],$$

with $\boldsymbol{l}^{(c)}_\omega(T \mid s)$ the *conditional score function* given s, viz. $[\partial \log h(T;\omega \mid s)/\partial\omega]'$.

Some redundancy among the conditions $\boldsymbol{R}^*$ is ignored for convenience of reference.

The following proposition concerning the integrated generalized information is easy to establish; see Bhapkar (1991) for the proof of its *local* version.

PROPOSITION 5.1. *If (S,T) is minimal sufficient for the family $\mathcal{P} = \{P_\omega, \omega \in \Omega\}$ of distributions of X, then for any distribution π in II,*

$$\boldsymbol{J}^{(S,T)}_\pi = \boldsymbol{J}_\pi. \tag{5.2}$$

Here $\boldsymbol{J}^{(S,T)}_\pi$ is defined for (S,T) as $\boldsymbol{J}_\pi$ is defined in Section 3 for X. In view of this proposition, hereafter we work with the reduced model based on (S,T).

For the sake of simplicity we will assume hereafter that π is a one-point distribution at ω.

We need to define a space analogous to $\mathcal{C}$ (in Section 3) and subspaces thereof. Now let

$$\mathcal{C} = \{c(s,t;\omega) : E_\omega(c) = 0, E_\omega c^2 < \infty, \text{ for all } \omega \in \Omega\}. \tag{5.3}$$

It is easy to show that $\mathcal{C}$ is a Hilbert space with inner product $\langle c_1, c_2\rangle = E_\omega c_1 c_2$. Define $\mathcal{G}$ to be the subspace of $\mathcal{C}$ spanned by $\bar{\mathcal{G}} = \{g(s,t;\theta) : g \in \mathcal{C}\}$, the set of all *regular* elements of $\mathcal{C}$ depending on ω only through θ. Denote the orthogonal complement of $\mathcal{G}$ in $\mathcal{C}$ by $\mathcal{U}$. (The asterisks are deleted here for simplicity of notation.) Thus,

$$\mathcal{C} = \mathcal{G} \oplus \mathcal{U}. \tag{5.4}$$

We will also need to work with the subspaces $\mathcal{C}(S) = \{c(s, \omega) : c \in \mathcal{C}\}$, the collection of all elements of $\mathcal{C}$ depending only on s and ω, and $\mathcal{C}(T \mid S)$ given by

$$\begin{aligned} \mathcal{C}(T \mid S) = \{c(s,t,\omega) \in \mathcal{C} : E_\omega(c \mid s) = 0, E_\omega(c^2 \mid s) < \infty, \\ \text{almost all } (\nu)s, \text{all } \omega \in \Omega\}. \end{aligned} \tag{5.5}$$

Finally, let $\mathcal{G}(T \mid S)$ and $\mathcal{G}(S)$ denote subspaces spanned by $\bar{\mathcal{G}} \cap \mathcal{C}(T \mid S)$ and $\bar{\mathcal{G}} \cap \mathcal{C}(S)$, respectively. With all the necessary quantities defined, we have the following proposition.

PROPOSITION 5.2. *The orthogonal complement $\mathcal{C}^\perp(S)$ of $\mathcal{C}(S)$ coincides with $\mathcal{C}(T \mid S)$. That is,*

$$\mathcal{C}^\perp(S) = \mathcal{C}(T \mid S). \tag{5.6}$$

PROOF. Let $c_1 \in \mathcal{C}(T \mid S)$. Then for any $c_2 \in \mathcal{C}(S)$,

$$\langle c_1, c_2 \rangle = E_\omega c_1 c_2 = E_\omega\{E_\omega(c_1 c_2 \mid S)\} = E_\omega[c_2 E_\omega\{(c_1 \mid S)\}] = 0.$$

Hence, $c_1 \in \mathcal{C}^\perp(S)$. Since c_1 is arbitrary, it follows that $\mathcal{C}(T \mid S) \subset \mathcal{C}^\perp(S)$. To prove the converse, $\mathcal{C}^\perp(S) \subset \mathcal{C}(T \mid S)$, let $c_0 \in \mathcal{C}^\perp(S)$. Now suppose $c_0 \notin \mathcal{C}(T \mid S)$. Then the function $\tilde{c}_0(S, \omega) = E_\omega(c_0 \mid S)$ is non-zero with positive probability. But $\tilde{c}_0(S, \omega) \equiv 0$ a.e. because $\tilde{c}_0 \in \mathcal{C}(S)$ (and hence, c_0 is orthogonal to $\tilde{c}_0$), leading to a contradiction. Hence, $c_0 \in \mathcal{C}(T \mid S)$ and $\mathcal{C}^\perp(S) \subset \mathcal{C}(T \mid S)$. □

From this proposition we can write down the following decompositions. Let $\mathcal{V}(S) = \mathcal{G}^\perp(S)$ in $\mathcal{C}(S)$ and $\mathcal{Y}(T \mid S) = \mathcal{G}^\perp(T \mid S)$ in $\mathcal{C}(T \mid S)$. Then

$$\mathcal{C} = \mathcal{C}(S) \oplus \mathcal{C}(T \mid S), \tag{5.7}$$

$$\mathcal{C}(S) = \mathcal{G}(S) \oplus \mathcal{V}(S), \tag{5.8}$$

$$\mathcal{C}(T \mid S) = \mathcal{G}(T \mid S) \oplus \mathcal{Y}(T \mid S). \tag{5.9}$$

Thus we have

$$\mathcal{C} = \mathcal{G}(S) \oplus \mathcal{V}(S) \oplus \mathcal{G}(T \mid S) \oplus \mathcal{Y}(T \mid S). \tag{5.10}$$

Next we consider the generalized Fisher information for θ in the reduced model (5.1), viz.,

$$\begin{aligned} \boldsymbol{I}_G(\theta; \omega) &= \min_{\boldsymbol{u} \in \mathcal{U}} E_\omega(\boldsymbol{l}_\theta(S,T) - \boldsymbol{u})(\boldsymbol{l}_\theta(S,T) - \boldsymbol{u})' \\ &= E_\omega \boldsymbol{g}^* \boldsymbol{g}^{*\prime}, \end{aligned} \tag{5.11}$$

where $\boldsymbol{g}^* =$ projection of $\boldsymbol{l}_\theta(S,T)$ onto $\mathcal{G}$. The expression (5.11) follows from Theorem 3.2. Here, and hereafter, by $\boldsymbol{u} \in \mathcal{U}$ we mean a vector $\boldsymbol{u} = \boldsymbol{u}(S,T;\omega)$ with elements belonging to the space $\mathcal{U}$, orthogonal to $\mathcal{G}$, defined earlier in this section. Similarly we define the generalized information for θ in the marginal distribution of S, as

$$\begin{aligned} \boldsymbol{I}_G^{(S)}(\theta,\omega) &= \min_{\boldsymbol{v}\in\mathcal{V}(S)} E_\omega(\boldsymbol{l}_\theta(S)-\boldsymbol{v})(\boldsymbol{l}_\theta(S)-\boldsymbol{v})' \\ &= E_\omega \hat{\boldsymbol{g}}\hat{\boldsymbol{g}}', \end{aligned} \tag{5.12}$$

where $\hat{\boldsymbol{g}} =$ projection of $\boldsymbol{l}_\theta(S)$ onto $\mathcal{G}(S)$.

Again, the definition of $\boldsymbol{I}_G^{(S)}$ is similar to that of $\boldsymbol{I}_G$, for the marginal distribution of S. The expression (5.12) follows from Theorem 3.2.

We now define the generalized Fisher information concerning θ in the conditional distribution of T, given S, by

$$\boldsymbol{I}_G^{(T|S)}(\theta;\omega) = \min_{\boldsymbol{y}\in\mathcal{Y}(T|S)} E_\omega[\boldsymbol{l}_\theta(T\mid S)-\boldsymbol{y}][\boldsymbol{l}_\theta(T\mid S)-\boldsymbol{y}]'. \tag{5.13}$$

PROPOSITION 5.3. *Assume conditions $\boldsymbol{R}^*$ and suppose $\mathcal{G}(T\mid S)$ is the subspace of $\mathcal{C}(T\mid S)$ spanned by $\bar{\mathcal{G}}\cap\mathcal{C}(T\mid S)$, then $\boldsymbol{I}_G^{(T|S)}$ exists and we have*

$$\boldsymbol{I}_G^{(T|S)}(\theta;\omega) = E_\omega\tilde{\boldsymbol{g}}\tilde{\boldsymbol{g}}', \tag{5.14}$$

where $\tilde{\boldsymbol{g}}$ is the projection of $\boldsymbol{l}_\theta(T\mid S)$ onto $\mathcal{G}(T\mid S)$.

The proof is parallel to that of Theorem 3.2 and, hence, it is omitted.

We have then the following theorem.

THEOREM 5.1. *Assume conditions $\boldsymbol{R}^*$ and let $\boldsymbol{I}_G$, $\boldsymbol{I}_G^{(S)}$ and $\boldsymbol{I}_G^{(T|S)}$ denote the generalized informations for θ, defined in* (5.11)–(5.13), *then*

$$\boldsymbol{I}_G(\theta;\omega) \geq \boldsymbol{I}_G^{(S)}(\theta;\omega) + \boldsymbol{I}_G^{(T|S)}(\theta;\omega). \tag{5.15}$$

PROOF. Consider the decomposition given earlier, viz.,

$$\mathcal{C} = \mathcal{G}(S)\oplus\mathcal{V}(S)\oplus\mathcal{G}(T\mid S)\oplus\mathcal{Y}(T\mid S).$$

Let $P_{\mathcal{G}(S)}$ and $P_{\mathcal{G}(T|S)}$ denote the projection operators on to subspaces $\mathcal{G}(S)$ and $\mathcal{G}(T\mid S)$, respectively. Since $P_{\mathcal{G}(S)}P_{\mathcal{G}(T|S)} = P_{\mathcal{G}(T|S)}P_{\mathcal{G}(S)} = 0$, $P_{\mathcal{G}(S)} + P_{\mathcal{G}(T|S)}$ is also a projection operator and satisfies

$$P_{\mathcal{G}(S)} + P_{\mathcal{G}(T|S)} = P_{\mathcal{G}(S)\oplus\mathcal{G}(T|S)}.$$

But,

$$\mathcal{G}(S)\oplus\mathcal{G}(T\mid S)\subset\mathcal{G}.$$

Hence, letting $\mathcal{G}_3 = \mathcal{G} \ominus (\mathcal{G}(S) \oplus \mathcal{G}(T \mid S))$

$$P_{\mathcal{G}} = P_{\mathcal{G}(S)} + P_{\mathcal{G}(T|S)} + P_{\mathcal{G}_3}.$$

Now, since $\boldsymbol{l}_\theta(S,T) = \boldsymbol{l}_\theta(S) + \boldsymbol{l}_\theta(T \mid S)$, we have

$$\begin{aligned} P_{\mathcal{G}}(\boldsymbol{l}_\theta(S,T)) &= P_{\mathcal{G}}(\boldsymbol{l}_\theta(S)) + P_{\mathcal{G}}(\boldsymbol{l}_\theta(T \mid S)) \\ &= P_{\mathcal{G}(S)}(\boldsymbol{l}_\theta(S)) + P_{\mathcal{G}(T|S)}(\boldsymbol{l}_\theta(T \mid S)) \\ &\quad + P_{\mathcal{G}_3}(\boldsymbol{l}_\theta(S,T)) \end{aligned}$$

because $P_{\mathcal{G}(S)}(c_1) = 0 = P_{\mathcal{G}(T|S)}(c_2)$ for any $c_1 \in \mathcal{C}(T \mid S)$ and $c_2 \in \mathcal{C}(S)$. Thus we have shown that

$$\boldsymbol{g}^* = \hat{\boldsymbol{g}} + \tilde{\boldsymbol{g}} + \tilde{\tilde{\boldsymbol{g}}}$$

where $\boldsymbol{g}^*$, $\hat{\boldsymbol{g}}$, $\tilde{\boldsymbol{g}}$ are given by (5.11), (5.12), (5.14) and $\tilde{\tilde{\boldsymbol{g}}} = P_{\mathcal{G}_3}(\boldsymbol{l}_\theta(S,T))$. The theorem now follows from the observation that $\hat{\boldsymbol{g}}$, $\tilde{\boldsymbol{g}}$ and $\tilde{\tilde{\boldsymbol{g}}}$ are orthogonal. □

COROLLARY 5.1. *Assume conditions $\boldsymbol{R}^*$ and suppose S and T are independent. Then*

$$\boldsymbol{I}_G(\theta;\omega) \geq \boldsymbol{I}_G^{(S)}(\theta;\omega) + \boldsymbol{I}_G^{(T)}(\theta;\omega). \tag{5.16}$$

The possible inequality sign, rather than a strict equality, might seem somewhat surprising in the relations (5.15), and especially in (5.16). The following example illustrates the need for a possible inequality.

Example 5.1. Let $X_1, \ldots, X_n$ be i.i.d. normal variables with mean μ and variance σ^2. For $\theta = \sigma^2$, it has been shown in Bhapkar (1989), that $I(\sigma^2;\omega) = n/2\sigma^4$, while $I_G(\sigma^2;\omega) = (n-1)/2\sigma^4$.

Now for $S = (X_1, \ldots, X_m)$, $T = (X_{m+1}, \ldots, X_n)$, $m < n$, we have $I_G^{(S)}(\sigma^2;\omega) = (m-1)/2\sigma^4$ and $I_G^{(T)}(\sigma^2;\omega) = (n-m-1)/2\sigma^4$; this does illustrate the need for a possible inequality sign in the statement (5.16).

This example also brings out an interesting difference between $I(\sigma^2;\omega)$ and $I_G(\sigma^2;\omega)$ as measures of information for σ^2, when μ is unknown, viz. in any *single* observation X_i, $I(\sigma^2;\omega) = 1/2\sigma^4$, while $I_G(\sigma^2;\omega) = 0$.

Example 5.2. As a generalization of Example 5.1, let now $X_1, \ldots, X_n$ be independent normal variables with variance σ^2 and means given by $E(\boldsymbol{X}) = \boldsymbol{A\beta}$, where $\boldsymbol{A}$ is a known $n \times p$ matrix of rank p, and $\boldsymbol{\beta}$ is an unknown regression vector. Then for $\theta = \sigma^2$, and $\omega = (\sigma^2, \boldsymbol{\beta})$, $I(\sigma^2;\omega) = n/2\sigma^4$, while $I_G(\sigma^2;\omega) = (n-p)/2\sigma^4$.

If we let here $S = \boldsymbol{X}'[\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}']\boldsymbol{X}$ and $\boldsymbol{T} = \boldsymbol{A}'\boldsymbol{X}$, then $(S, \boldsymbol{T})$ is minimal sufficient for ω, $\boldsymbol{T}$ is *p-ancillary* for σ^2 in the complete sense, while S is p-sufficient for σ^2 in the complete sense using the terminology of Bhapkar (1990, 1991). In view of the results established in these papers, we have $I_G(\sigma^2;\omega) = E_\omega I(\sigma^2 \mid T) = I^{(S)}(\sigma^2)$.

Example 5.3. For the distribution in Example 5.2, if we have $\theta = \boldsymbol{\beta}$, we do not have either p-ancillary or p-sufficient statistic for $\boldsymbol{\beta}$. However with parametrization $\theta = \boldsymbol{\beta}/\sigma^2$ we do have a p-ancillary statistic $T = \boldsymbol{X}'\boldsymbol{X}$ in the complete sense, in the terminology of Bhapkar (1989). Hence $\boldsymbol{I}_G(\theta; \omega) = E_\omega[\boldsymbol{I}(\theta \mid T)]$ for $\theta = \boldsymbol{\beta}/\sigma^2$. We note here that although $(\boldsymbol{\beta}, \sigma^2)$ are *orthogonal parameters* (see, e.g., Cox and Reid (1987)), (θ, σ^2) are no longer orthogonal if we take $\theta = \boldsymbol{\beta}/\sigma^2$.

Thus, in this example, conditioning on the p-ancillary statistic $T = \boldsymbol{X}'\boldsymbol{X}$ is indicated, for inference concerning $\theta = \boldsymbol{\beta}/\sigma^2$, on the basis of generalized information measure, $\boldsymbol{I}_G(\theta; \omega)$, but *not* on the basis of measure $\boldsymbol{I}(\theta; \omega)$, and $\boldsymbol{I}(\theta; \omega) \neq E_\omega[\boldsymbol{I}(\theta \mid T)]$.

For the special case $p = 1$ with common mean μ for all X_i, and $\theta = \mu/\sigma^2$, the conditional distribution of $S = \bar{X}$, given $T = t$, has the p.d.f.

$$k(s; \theta, t) = (t - ns^2)^{(n-3)/2} \exp\{ns\theta + c(t, \theta)\}$$

for $|s| \leq (t/n)^{1/2}$. This distribution does belong to one-parameter exponential family and, at $\theta = 0$, this distribution is equivalent to Student's distribution with $n - 1$ degrees of freedom for $S/\{[T - nS^2]/[n(n-1)]\}^{1/2}$, both conditionally and unconditionally.

Acknowledgement

The authors thank the referee for constructive suggestions that have led to the improved format of the paper.

References

Bhapkar, V. P. (1972). On a measure of efficiency of an estimating equation, *Sankhyā Ser. A*, **34**, 467–472.

Bhapkar, V. P. (1989). Conditioning on ancillary statistics and loss of information in the presence of nuisance parameters, *J. Statist. Plann. Inference*, **21**, 139–160.

Bhapkar, V. P. (1990). Conditioning, marginalization and Fisher information functions, *Probability Statistics and Design of Experiments* (ed. R. R. Bahadur), Wiley Eastern Limited, New Delhi.

Bhapkar, V. P. (1991). Loss of information in the presence of nuisance parameters and partial sufficiency, *J. Statist. Plann. Inference*, **28**, 185–203.

Chandrasekar, B. and Kale, B. K. (1984). Unbiased statistical estimation functions for parameters in presence of nuisance parameters, *J. Statist. Plann. Inference*, **9**, 45–54.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference, *J. Roy. Statist. Soc. Ser. B*, **49**, 1–39.

Ferreira, P. E. (1982). Multiparametric estimating equations, *Ann. Inst. Statist. Math.*, **34**, 423–431.

Friedman, A. (1982). *Foundations of Modern Analysis*, Dover, New York.

Godambe, V. P. (1960). An optimum property of a regular maximum likelihood estimation, *Ann. Math. Statist.*, **31**, 1208–1211.

Godambe, V. P. (1984). On ancillarity and Fisher information in the presence of nuisance parameter, *Biometrika*, **71**, 626–629.

Kale, B. K. (1962). An extension of the Cramér-Rao inequality for statistical estimation functions, *Skand. Aktuarietidskr.*, **45**, 80–89.

Rao, C. R. (1965). *Linear Statistical Inference*, Wiley, New York.