Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

Estimating Functions, Partial Sufficiency and Q-Sufficiency in the Presence of Nuisance Parameters

Vasant P. Bhapkar
University of Kentucky

Abstract

When there exists a statistic which has its distribution free of nuisance parameters, the optimality of the marginal score function can be investigated in the context of generalized Fisher information for parameters of interest. In the case of a partially sufficient statistic, i.e. a statistic sufficient for parameters of interest, the marginal score function is the optimal estimating function. With the new concept of q-sufficiency for parameters of interest, the marginal score function is operationally equivalent to the optimal estimating function.

AMS Subject Classification: 62A20, 62F99

Key Words: Generalized Fisher information matrix, completeness of a family, partial sufficiency.

1 INTRODUCTION

The optimality of the conditional score function as an estimating function for the parameter of interest, in the presence of unknown nuisance parameters, was established by Godambe (1976) in the situation where there exists a statistic which is ancillary for the parameter of interest and which is also complete for nuisance parameters. Such a statistic has been termed a complete p-ancillary statistic in an earlier paper (Bhapkar, 1989).

Refer to Liang and Zeger (1995) for a review of estimating functions theory, some discussion of optimality of the conditional score function under the conditions assumed by Godambe, and references to further work (for example, Lindsay 1982) to find approximately optimal estimating functions in more general situations.

The analogous question concerning optimality of the marginal score function in the *complete case*, has not yet been satisfactorily resolved. Lloyd (1987) considered this question; however, his assertion of optimality of the marginal score function has been shown recently by Bhapkar (1995) to be invalid.

In a special case where the two appropriate components of the sufficient statistic happen to be independent, the marginal score function does turn out to be optimal i.e., the property (6.2) holds. In such a special case, the appropriate component possesses the property of *p-sufficiency*, symbolized by the relation (6.1); this p-sufficiency then implies the optimality of the marginal score. However, in the general case where the two components of the sufficient statistic are not necessarily independent, the property of p-sufficiency does not necessarily hold (as shown by the counter-example by Bhapkar, 1995). This paper now establishes a certain weaker property, referred in this paper as q-sufficiency; this property is symbolized by the relation (7.7).

For example, in the case of a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ from a normal population with mean μ and variance $\sigma^2, T = \overline{X}$ is a complete p-ancillary statistic for σ^2 and $S = \sum_i (X_i - \overline{X})^2$ is p-sufficient for σ^2. Here S satisfies the property (6.1). However, in examples (4.1) and (5.1) S satisfies only the weaker relation (7.6), but not (6.1). Thus, here the marginal score function of S satisfies only the weaker optimality relation (7.7), but not (6.2), which is satisfied in the first example. In both these examples, the family of distributions of T, given $S = s$ and θ, is complete for the nuisance parameter ϕ.

The optimality of marginal score function, in its weaker or stronger forms, is shown to be related to certain generalized Fisher information functions. Section 2 introduces the basic terminology, Section 3 considers information in a statistic (or its marginal distribution) as well as the information in the conditional distribution given the statistic. Section 4 discusses the special case where the marginal distribution of a statistic is free of the nuisance parameters, while Section 5 deals with the optimality of estimating functions. The case where the statistic happens to be sufficient for parameter of interest (i.e. p-sufficient) is described in Section 6 while the last section establishes a somewhat weaker property (viz. q-sufficiency) that holds in general in the *complete* case, and in another specific situation.

2 INFORMATION IN ESTIMATING FUNCTION

Suppose that the random variable X has probability distribution P_ω and the probability density function (*pdf*) $p(x;\omega)$ with respect to a σ-finite measure μ over the measurable space $(\chi, \mathcal{A})$. The parameter $\omega \in \Omega$ and assume that $\omega = (\theta, \phi)$, where θ is the parameter of interest, ϕ is the nuisance parameter and θ, ϕ are variation independent in the sense that $\Omega = \Theta \times \Phi$, where Ω is an open interval in the d-dimensional space R^d, and d_1 is the dimensionality of θ.

Let $\boldsymbol{l}'_\omega(X) = [\partial \log p(x;\omega)/\partial\omega]$, be the row-vector of ω-scores, and its components $[\boldsymbol{l}'_\theta(X), \boldsymbol{l}'_\phi(X)]$ are referred to as θ-scores and ϕ-scores of X, respectively.

We assume the standard Cramér-Rao type regularity conditions:

$\boldsymbol{R}$: i. $\int p(x;\omega)d\mu(x) = 1$ can be differentiated twice under the integral sign with respect to elements of ω;

ii. The Fisher information matrix $\boldsymbol{I}(\omega) = E_\omega[\boldsymbol{l}_\omega(X)\boldsymbol{l}'_\omega(X)]$ is positive definite (pd).

If θ is real-valued, $g = g(x;\theta)$ is said to be a *regular unbiased estimating function* (RUEF) for θ if it satisfies the regularity conditions:

$\boldsymbol{R}_G$: i. $E_\omega g(X;\theta) = 0\,, E_\omega g^2(X;\theta) < \infty$;

ii. $\int g(x;\theta)p(x;\omega)d\mu(x) = 0$ can be differentiated under the integral sign with respect to elements of ω.

Operationally, $g(x;\theta) = 0$ is to be solved to produce an estimate $\hat{\theta} = \hat{\theta}(x)$.

In the d_1-dimensional case, we shall still use the term estimating function for a real-valued function $g = g(x;\theta)$ if it satisfies conditions $\boldsymbol{R}_G$; however to solve for θ we need an estimating equation $\boldsymbol{g}(x;\theta) = \boldsymbol{0}$, all elements of which satisfy conditions $\boldsymbol{R}_G$. Furthermore, in order to ensure that the equation leads to a solution $\hat{\theta} = \hat{\theta}(x)$, we also need the conditions:

(a) the covariance matrix $\boldsymbol{\sigma}_{\boldsymbol{g}}(\omega) \equiv E_\omega[\boldsymbol{g}\boldsymbol{g}']$ is pd.

(b) $\boldsymbol{G}(\omega) \equiv E_\omega[\partial\boldsymbol{g}(X;\theta)/\partial\theta]$ is nonsingular.

In any case, we denote by G the space of real-valued functions satisfying the regularity conditions $\boldsymbol{R}_G$.

The generalized Fisher information for θ, when $\omega = (\theta,\phi)$ is the *full* parameter in the distribution P_ω of X is defined (Godambe, 1984) in the one-dimensional case as

$$I_G(\theta;\omega) = \min_{u\in\mathcal{U}} E_\omega\left[l_\theta(X) - u(X;\omega)\right]^2 . \tag{2.1}$$

Here $\mathcal{U}$ is the space of real-valued functions $u = u(x;\omega)$ such that

$$\begin{aligned} (i)\quad E_\omega u(X;\omega) &= 0,\ E_\omega u^2(X;\omega) < \infty \\ (ii)\quad E_\omega[u(X;\omega)g(X;\theta)] &= 0\,,\ \text{for all } g \in G\,. \end{aligned} \tag{2.2}$$

On the other hand, the information concerning θ in the RUEF $g = g(x;\theta)$ is defined as

$$I_g(\theta;\omega) = \frac{E^2_\omega[\partial g(X;\theta)/\partial\theta]}{E_\omega g^2(X;\theta)}\,; \tag{2.3}$$

g_1 is then said to be more efficient than g if $I_{g_1}(\theta;\omega) \geq I_g(\theta;\omega)$ for all ω with strict inequality for at least one ω. We have, then, the *generalized information inequality* (Godambe, 1984)

$$I_g(\theta;\omega) \leq I_G(\theta;\omega)\,, \tag{2.4}$$

for all $g \in G$.

The multi-dimensional analogs of (2.1), (2.3) and (2.4) are (see, Bhapkar, 1989):

$$\begin{aligned} \boldsymbol{I}_G(\theta;\omega) &= \min_{\boldsymbol{u}\in\mathcal{U}} E_\omega\left[\boldsymbol{l}_\theta(X) - \boldsymbol{u}\right]\left[\boldsymbol{l}_\theta(X) - \boldsymbol{u}\right]'\,, && (2.5)\\ \boldsymbol{I}_{\boldsymbol{g}}(\theta;\omega) &= \boldsymbol{G}'(\omega)\boldsymbol{\sigma}_{\boldsymbol{g}}^{-1}(\omega)\boldsymbol{G}(\omega)\,, && (2.6)\\ \boldsymbol{I}_{\boldsymbol{g}}(\theta;\omega) &\leq \boldsymbol{I}_G(\theta;\omega)\,. && (2.7) \end{aligned}$$

Here, of course, $A \leq B$ for non-negative-definite (nnd) matrices means $B - A$ is nnd, min denotes the minimal matrix $\boldsymbol{M}^* = \boldsymbol{M}(\boldsymbol{u}^*)$ in the class of nnd matrices $\{\boldsymbol{M}(\boldsymbol{u})\}$ such that $\boldsymbol{M}(\boldsymbol{u}) \geq \boldsymbol{M}^*$ for all $\boldsymbol{u}$, and $\boldsymbol{u}$ denotes a vector with all elements $u \in \mathcal{U}$, which satisfy (2.2).

It has been proved (Bhapkar and Srinivasan, 1994) that $\boldsymbol{I}_G$ exists; in fact

$$\boldsymbol{I}_G(\theta;\omega) = E_\omega[\boldsymbol{g}^*\boldsymbol{g}^{*\prime}]\,, \tag{2.8}$$

where $\boldsymbol{g}^*$ is the *projection* of $\boldsymbol{l}_\theta(X)$ onto the space spanned by G. This G-form indeed generalizes the usual Fisher information matrix in the sense that

$$\boldsymbol{I}_G(\theta;\theta) = \boldsymbol{I}(\theta) = E_\theta[\boldsymbol{l}_\theta(X)\boldsymbol{l}'_\theta(X)]\,. \tag{2.9}$$

In order to prove (2.8) and the results in this paper we find the Hilbert space technique quite useful. Another motivation for this is the need to tackle the identifiability problem posed by the distinction between a function involving parameter θ; e.g. $g = g(x;\theta)$ and a function involving full parameter ω, e.g. $u = u(x;\omega)$, for the given ω.

To treat ω as a variable, along with x, Bhapkar and Srinivasan (1994) introduced an arbitrary probability distribution π over the measurable space (Ω, β) for "random variable" ω, and consider the joint distribution $P_\omega \times \pi$ of (X, ω) over $(\mathcal{X} \times \Omega)$.

Thus in the Appendix the space $\mathcal{C}$ of real-valued functions $c = c(x;\omega)$ is defined such that

1. $Ec(X;\omega) \equiv \int c(x;\omega)p(x;\omega)d\mu(x)d\pi(\omega) = 0$

2. $Ec^2(X;\omega) < \infty$.

Here we write E for joint expectation, while we write E_ω for expectation, given ω.

The technical discussion in this context, along with proofs where necessary, is given in the Appendix, in order to provide validity to statements in sections 4, 5 and 7 of the text of the paper, which specialize the general results for arbitrary π to the one-point distribution π at ω.

Thus $\mathcal{C}_\omega$ denotes the space of real-valued functions $c = c(x;\omega)$ such that

1. $E_\omega c(X;\omega) \equiv \int c(x;\omega)p(x;\omega)d\mu(x) = 0$
2. $E_\omega c^2(X;\omega) < \infty$.

Similarly, $\mathcal{G}$ denotes the closure $\bar{G}$ of the subspace G of functions $g = g(x;\theta)$, which depend on ω only through θ. Then $\mathcal{G}_\omega$ denotes the space with π a one-point distribution at ω. The orthogonal complement of $\mathcal{G}_\omega$ in $\mathcal{C}_\omega$ is then $\mathcal{U}_\omega$.

3 INFORMATION IN A STATISTIC

Suppose now (S,T) is a minimal sufficient statistic for the family $\{P_\omega^{(X)} : \omega \in \Omega\}$ of probability distributions $P_\omega^{(X)}$ of X. Without loss of generality, we replace hereafter X by (S,T) in view of the anticipated result stated here as a lemma (Bhapkar, 1991):

Lemma 3.1 *Under regularity conditions* $\boldsymbol{R}$ *for the distribution of* X, *and conditions* $\boldsymbol{R}_G$,

$$I_G^{(X)}(\theta;\omega) = I_G^{(S,T)}(\theta;\omega). \tag{3.1}$$

Now we want to investigate the information contained in S alone, in relation to (S,T); similarly we want to find out the maximum information, concerning θ, for estimating function based on S alone, in relation to such maximum information in the functions based on both S and T.

Let then $f(s;\omega)$ denote the marginal *pdf* of S with respect to measure ν, and $h(t;\omega|s)$ the *pdf* of conditional distribution of T given s, the value of S, with respect to measure η_s, when (S,T) has *pdf* $p(s,t;\omega)$ with respect to μ. In view of regularity conditions $\boldsymbol{R}$ for p, we assume the following conditions for f and h:

$\boldsymbol{R}^*$:
1. $\int f(s;\omega)d\nu(s) = 1$ can be differentiated, twice under the integral sign, with respect to elements of ω;
2. $\int h(t;\omega|s)d\eta_s(t) = 1$ can be differentiated twice with respect to elements of ω under the integral sign for almost all $(\nu)s$;

3. The Fisher information matrix $\boldsymbol{I}^{(S)}(\omega) = E_\omega[\boldsymbol{l}_\omega(S)\boldsymbol{l}'_\omega(S)]$ exists where $\boldsymbol{l}'_\omega(S) = [\partial \log f(s;\omega)/\partial\omega]$ is the marginal ω-score of S;
4. For almost all $(\nu)s$, the information matrix of T, given s, viz $\boldsymbol{I}^{(T|s)}(\omega) = E_\omega[\boldsymbol{l}_\omega(T|s)\boldsymbol{l}'_\omega(T|s)]$ exists where

$$\boldsymbol{l}_\omega(T|s) = [\partial \log h(T;\omega|s)/\partial\omega]$$

is the conditional ω-score of T.

Let $\mathcal{C}_\omega$ denote the Hilbert space of real functions $c = c(s,t;\omega)$ such that $E_\omega c = 0$ and $E_\omega c^2 < \infty$ (see Lemma A.2), and $\mathcal{G}_\omega = \bar{G}_\omega$ the closed linear subspace, where G_ω is the set of functions $g = g(s,t;\theta)$ in $\mathcal{C}_\omega$, which satisfy regularity conditions as in $\boldsymbol{R}_G$. Hereafter we drop the subscript ω from $\mathcal{C}_\omega, \mathcal{G}_\omega, \mathcal{U}_\omega, \ldots$ for simplicity of notation. Thus, we have

$$\mathcal{C} = \mathcal{G} \oplus \mathcal{U} \tag{3.2}$$

where $\mathcal{U}$ is the orthogonal complement of $\mathcal{G}$ in $\mathcal{C}$; the elements $u = u(s,t;\omega)$ are orthogonal to $g \in \mathcal{G}$ i.e. $E_\omega[gu] = 0$ for all $g \in \mathcal{G}$.

Now we denote by $\mathcal{C}(S)$ the subset of functions $c = c(s;\omega)$ in $\mathcal{C}$, which depend on (s,t) only through s. Similarly $\mathcal{G}(S) = \bar{G}(S)$ denotes the closure of $G(S)$, the subset of G depending on (s,t) only through s. Thus, $\mathcal{G}(S) = \mathcal{G} \cap \mathcal{C}(S)$.

Note that $\mathcal{C}(S)$ is itself a Hilbert space, which is a subspace of $\mathcal{C}$, and the inner-products (or norms) for both spaces coincide for $c_1, c_2 \in \mathcal{C}(S)$, in view of the fact that

$$\int c_1(s;\omega)c_2(s;\omega)dP_\omega^{(S,T)}(s,t) = \int c_1c_2dP_\omega^{(S)}(s)\,.$$

$\mathcal{G}(S)$ is a closed linear subspace of $\mathcal{C}(S)$ and we denote the orthogonal complement of $\mathcal{G}(S)$ by $\mathcal{V}(S)$; thus

$$\mathcal{C}(S) = \mathcal{G}(S) \oplus \mathcal{V}(S)\,. \tag{3.3}$$

Observe that $\mathcal{V}(S)$ contains both functions $u = u(s;\omega)$ in $\mathcal{U}$, not depending on t, and functions $\nu = \nu(s;\omega) = E_\omega[u(s,T;\omega|s]$ for the remaining u in $\mathcal{U}$, since $E_\omega[gu] = 0$ for all $g \in \mathcal{G}(S)$ for any $u \in \mathcal{U}$.

We can then define the generalized Fisher information matrix for S, with respect to parameter of interest θ when ω is the full parameter, as

$$\boldsymbol{I}_G^{(S)}(\theta;\omega) = \min_{\boldsymbol{\nu}\in\mathcal{V}} E_\omega[\boldsymbol{l}_\theta(S) - \boldsymbol{\nu}][\boldsymbol{l}_\theta(S) - \boldsymbol{\nu}]'\,, \tag{3.4}$$

in analogy with (2.5).

Similarly, for each value s of S, we consider the conditional distribution of T, given s, and define the Hilbert space $\mathcal{C}_s$ of real functions $c_s = c_s(t;\omega)$ satisfying

$$E_\omega(c_s|s) = 0,\ E_\omega(c_s^2|s) < \infty\,, \tag{3.5}$$

in view of Proposition A.2. The closed linear subspace $\mathcal{G}_s = \bar{G}_s$, and its orthogonal complement $\mathcal{Y}_s$ in $\mathcal{C}_s$, are defined like $\mathcal{C}_s$ above; thus we have

$$\mathcal{C}_s = \mathcal{G}_s \oplus \mathcal{Y}_s\,, \tag{3.6}$$

for each value s of S.

The generalized information in the conditional distribution $P_\omega^{(T|s)}$, given s, is defined now as

$$\boldsymbol{I}_G(\theta;\omega|s) = \min_{\boldsymbol{y}_s\in\mathcal{Y}_s} E_\omega\left\{[\boldsymbol{l}_\theta(T\,|s) - \boldsymbol{y}_s]\,[\boldsymbol{l}_\theta(T\,|s) - \boldsymbol{y}_s]'\,|s\right\}, \tag{3.7}$$

in analogy with (2.5).

Finally we define the subspace $\mathcal{C}(T\,|S)$ in $\mathcal{C}$ of functions $c(s,t;\omega) = c_s(t;\omega)$, $G(T\,|S)$ as the subspace in G of functions $(g(s,t;\theta) = g_s(t;\theta)$, $\mathcal{G}(T\,|S) = \bar{G}(T\,|S)$, and $\mathcal{Y}(T\,|S)$ as the subspace in $\mathcal{C}(T\,|S)$ of functions $y(s,t;\omega) = y_s(t;\omega)$. Thus these functions satisfy (3.5) a.e. $(P_\omega^{(S)})$ and, thus, $\mathcal{G}\ (T\,|S) \subset \mathcal{C}(T\,|S)$ and $(\mathcal{Y}(T\,|S) \subset \mathcal{C}(T\,|S)\,)$. We also note that $(\mathcal{G}(T\,|S) = \mathcal{G} \cap \mathcal{C}\ (T\,|S))$.

We now have the following lemma from Bhapkar and Srinivasan (1994), which is the special version of Lemma A.3.

Lemma 3.2

$$\mathcal{C} = \mathcal{C}(S) \oplus \mathcal{C}(T\,|S)\ . \tag{3.8}$$

PROOF It is easy to see that if $c \in \mathcal{C}(T\,|S)$, then $c \perp \mathcal{C}(S)$. We want to prove the converse that if $c \perp \mathcal{C}(S)$, then $c \in \mathcal{C}(T\,|S)$.

Let then $c \perp \mathcal{C}(S)$, and $E_\omega\,(c(s,T;\omega)\,|s) = c^*(s;\omega)$. Since $c^* \in \mathcal{C}(S)$, and $c \perp \mathcal{C}(S)$, we have $E_\omega(cc^*) = 0$. But $E_\omega(cc^*) = E_\omega\,[c^*(S;\omega)E_\omega(c|S)] = E_\omega[c^{*2}(S;\omega)]$. Hence $E_\omega(c^{*2}) = 0$, which implies $c^*(s;\omega) = 0$, a.e. $(P_\omega^{(S)})$. Thus, $c \in \mathcal{C}(T\,|S)$ and the lemma is established.

Note that the decomposition of the space $\mathcal{C}$ into $\mathcal{C}(S)$ and $\mathcal{C}(T\,|S)$ corresponds to the fact that, for any $c \in \mathcal{C}, E_\omega(c|s) \in \mathcal{C}(S)$, while $c - E_\omega(c|s) \in \mathcal{C}(T\,|S)$.

In view of Lemma A.4 we have $\mathcal{C}(T\,|S) = \mathcal{G}(T\,|S) \oplus \mathcal{Y}(T\,|S)$. The generalized information for θ in the conditional distribution of T, given S, is defined as

$$\boldsymbol{I}_G^{(T|S)}(\theta;\omega) = \min_{\boldsymbol{y}\in\mathcal{Y}(T|S)} E_\omega\,[\boldsymbol{l}_\theta(T\,|S) - \boldsymbol{y}]\,[\boldsymbol{l}_\theta(T\,|S) - \boldsymbol{y}]'\ . \tag{3.9}$$

Lemma 3.3 *In view of (3.7) and (3.9),*

$$\boldsymbol{I}_G^{(T|S)}(\theta;\omega) \geq E_\omega\left[\boldsymbol{I}_G(\theta;\omega\,|S)\right] \tag{3.10}$$

PROOF. The proof is immediately obtained by noting

$$\begin{aligned}\min_{\boldsymbol{y}\in\mathcal{Y}(T|S)} E_\omega[\boldsymbol{a}-\boldsymbol{y}][\boldsymbol{a}-\boldsymbol{y}]' &= \min_{\boldsymbol{y}\in\mathcal{Y}(T|S)} E_\omega\left\{E_\omega\left[(\boldsymbol{a}-\boldsymbol{y})(\boldsymbol{a}-\boldsymbol{y})'\,|S\right]\right\} \\ &\geq E_\omega \min_{\boldsymbol{y}_S\in\mathcal{Y}_S} E_\omega\left\{[\boldsymbol{a}_S-\boldsymbol{y}_S]\,[\boldsymbol{a}_S-\boldsymbol{y}_S]'\,|S\right\}\end{aligned}$$

with $\boldsymbol{a}(s,\ t;\omega)=\boldsymbol{l}_\theta(T\,|S)$ and $\boldsymbol{a}_s=\boldsymbol{l}_\theta(T\,|s)$.

We are now in a position to prove the following theorem;

Theorem 3.1 *Assume the regularity conditions $\boldsymbol{R}^*$ for the marginal distribution of S and the conditional distribution of T, given S. Then*

$$\text{(i)}\ \boldsymbol{I}_G^{(S,T)}(\theta;\omega) \geq \boldsymbol{I}_G^{(S)}(\theta;\omega)+\boldsymbol{I}_G^{(T|S)}(\theta;\omega)\,; \tag{3.11}$$

and

$$\text{(ii)}\ \boldsymbol{I}_G^{(S,T)}(\theta;\omega) \geq \boldsymbol{I}_G^{(S)}(\theta;\omega)+E_\omega\left[\boldsymbol{I}_G(\theta;\omega|S)\right]. \tag{3.12}$$

PROOF. Let $\boldsymbol{n}(\boldsymbol{u})=\boldsymbol{l}_\theta(S,T)-\boldsymbol{u}$, $\boldsymbol{m}(\boldsymbol{\nu})=\boldsymbol{l}_\theta(S)-\boldsymbol{\nu}$ and $\boldsymbol{r}(\boldsymbol{w})=\boldsymbol{l}_\theta(T\,|S)-\boldsymbol{w}$, where $\boldsymbol{\nu}(s;\omega)=E_\omega[\boldsymbol{u}\,|s]$ and $\boldsymbol{w}=\boldsymbol{u}-\boldsymbol{\nu}$ for any $\boldsymbol{u}\in\mathcal{U}$ (i.e. every element of $\boldsymbol{u}$ belongs to $\mathcal{U}$).

Then $\boldsymbol{\nu}\in\mathcal{V}(S)$, since $\boldsymbol{u}\perp\mathcal{G}(S)$ implies $E_\omega[\boldsymbol{u}g]=E_\omega[g\boldsymbol{\nu}]=0$ for all $g\in\mathcal{G}(S)$. On the other hand, if $g=0$ is the only g in $\mathcal{G}(S)$, then $\boldsymbol{\nu}\in\mathcal{V}(S)$ in view of (3.3). Also, for all $g\in\mathcal{G}(T\,|S)$, $E_\omega[g\boldsymbol{w}]=E_\omega[g(\boldsymbol{u}-\boldsymbol{v})]=E_\omega[g\boldsymbol{u}]-E_\omega[g\boldsymbol{\nu}]=\boldsymbol{0}$, and then $\boldsymbol{w}\in\mathcal{Y}(T\,|S)$ in view of Lemma A.4. It is also true that $\boldsymbol{w}\in\mathcal{Y}(T|S)$ if $g=0$ is the only g in $\mathcal{G}(T|S)$.

Now $\boldsymbol{n}(\boldsymbol{u})=\boldsymbol{m}(\boldsymbol{\nu})+\boldsymbol{r}(\boldsymbol{w})$ and we have

$$E_\omega[\boldsymbol{n}(\boldsymbol{u})\boldsymbol{n}'(\boldsymbol{u})]=E_\omega[\boldsymbol{m}(\boldsymbol{\nu})\boldsymbol{m}'(\boldsymbol{\nu})]+E_\omega[\boldsymbol{r}(\boldsymbol{w})\boldsymbol{r}'(\boldsymbol{w})]\,.$$

Since we have $\boldsymbol{\nu}\in\mathcal{V}(S)$ and $\boldsymbol{w}\in\mathcal{Y}(T\,|S)$ for every $\boldsymbol{u}\in\mathcal{U}$, it follows that

$$\min_{\boldsymbol{u}\in\mathcal{U}} E_\omega[\boldsymbol{n}(\boldsymbol{u})\boldsymbol{n}'(\boldsymbol{u})] \geq \min_{\boldsymbol{v}\in\mathcal{V}(S)} E_\omega[\boldsymbol{m}(\boldsymbol{u})\boldsymbol{m}'(\boldsymbol{u})]+\min_{\boldsymbol{y}\in\mathcal{Y}(T|S)} E_\omega[\boldsymbol{r}(\boldsymbol{y})\boldsymbol{r}'(\boldsymbol{y})]$$

i.e. inequality (i) holds. The inequality (ii) follows immediately from Lemma (3.3). Thus the theorem is established.

Remarks. (a) The relation (i) would be a strict equality if there exists $\boldsymbol{u}=\boldsymbol{u}^*$ for which $E_\omega[\boldsymbol{u}^*\,|s]\equiv\boldsymbol{\nu}^*$ minimizes $E_\omega[\boldsymbol{m}(\boldsymbol{v})\boldsymbol{m}'(\boldsymbol{v})]$ for $\boldsymbol{v}\in\mathcal{V}(S)$ and *simultaneously* $\boldsymbol{y}^*=\boldsymbol{u}^*-\boldsymbol{v}^*$ minimizes $E_\omega[\boldsymbol{r}(\boldsymbol{y})\boldsymbol{r}'(\boldsymbol{y})]$ for $\boldsymbol{y}\in\mathcal{Y}(T\,|S)$. The relation (ii) would be a strict equality if, furthermore, $\boldsymbol{y}^*=\boldsymbol{y}_s^*(t;w)$ where $\boldsymbol{y}_s^*$ minimizes $E_\omega[\boldsymbol{r}(\boldsymbol{y}_s)\boldsymbol{r}'(\boldsymbol{y}_s)\,|s]$ for $\boldsymbol{y}_s\in\mathcal{Y}_s$ a.e. $(P_\omega^{(S)})$.

(b) The inequality (3.12) was earlier given as Theorem 4.2 in Bhapkar (1991), subject to the condition that $\boldsymbol{I}_G(\theta;\omega)$ exists. Since existence of such a generalized information matrix has been established by Bhapkar and Srinivasan (1994), such an existence qualification is no longer needed. Furthermore, the present proof is mathematically more rigorous.

(c) For the special case discussed in next section, when S has distribution depending on ω only through θ, $\boldsymbol{I}_G^{(S)}(\theta;\omega)$ reduces to $\boldsymbol{I}^{(S)}(\theta)$ in the inequality (3.12). Such an inequality was earlier established as Theorem 3.1 in Bhapkar (1991) subject to existence qualification. The previous comments (a) and (b) apply to this case as well.

(d) The fact that we need a possible inequality in (3.11), rather than a strict equality, is seen from the following example: Suppose $X = (X_1, X_2)$, where X_1, X_2 are independent $N(\phi, \theta)$ variables. Then $I_G(\theta;\omega) = 1/2\theta^2$, but for $S = X_1, T = X_2$, we have $I_G^{(S)}(\theta;\omega) = 0$, and $I_G^{(T|S)}(\theta;\omega) = I_G^{(T)}(\theta;\omega) = 0$ as seen from Example 4.2 Bhapkar (1991).

4 SPECIAL CASE

Now we deal with the special case where the marginal distribution of S depends on ω only through θ. Then every element of the marginal score function of S, viz $\boldsymbol{l}_\theta(s)$ is a RUEF in view of condition $\boldsymbol{R}^*$(i).

We can now characterize all the RUEF in G in view of Lemma A.6 in the Appendix.

Theorem 4.1 *Assume that the joint distribution of (S,T), given ω, satisfies the regularity conditions $\boldsymbol{R}^*$ in Section 3, the marginal distribution of S depends only on θ, and the estimating functions $g = g(s,t;\theta)$, for parameter of interest θ, satisfy the regularity assumptions $\boldsymbol{R}_G$. Then any RUEF in G can be expressed in the orthogonal decomposition:*

$$g(s,t;\theta) = \boldsymbol{b}'(s,t;\theta)\boldsymbol{l}_\theta(s) + g_0(s;\theta) \tag{4.1}$$

where $E_\omega[\boldsymbol{b}(s,T;\theta)\,|s] = \boldsymbol{a}(\omega)$, and $g_0 \in G_0(S)$, the set of RUEF in $\mathcal{G}(S)$ uncorrelated with $\boldsymbol{l}_\theta(S)$.

Remark (a) Although we are assuming here that ω is fixed (i.e. π is a one-point distribution at the given ω), as in most of the text, the strict validity of the statements with $\omega = (\theta,\phi)$ as variables follows from the proofs given in Appendix (e.g. Lemma A.6 for Theorem 4.1) where $(S,T;\omega)$ are variables.

(b) Observe from Lemma A.6 that $\mathcal{G} = \mathcal{G}(S,T) \oplus \mathcal{G}_0(S)$, where $\mathcal{G}(S,T)$ is defined by (A.10), and $\mathcal{G}_1(S) \subset \mathcal{G}(S,T)$. In particular, if $G(T|S)$ is empty, G is not necessarily equal to $G(S)$. See Example 4.1 below in this regard.

Example 4.1. Let $X = (X_1, \ldots, X_n)$ be independent pairs $X_i = (Y_i, Z_i)$, where Y_i, Z_i are independent exponentially distributed variables with means ϕ and $\phi\theta$ respectively. Letting $Y = \sum_i Y_i$ and $Z = \sum_i Z_i$, (Y, Z) is minimally sufficient for $\omega = (\theta, \phi)$. Equivalently, (S, T) is a minimal sufficient statistic with $S = Z/Y$ and $T = Z$.

The marginal distribution of S is

$$f(s;\theta) = \frac{\Gamma(2n)}{\Gamma^2(n)} \frac{s^{n-1}\theta^n}{(\theta+s)^{2n}}, \; 0 < s < \infty$$

free of ϕ, and the conditional distribution of T, given s, has *pdf*

$$h(t;\omega\,|s) = \frac{t^{2n-1}}{\Gamma(2n)} \frac{e^{-t/\delta(s)}}{[\delta(s)]^{2n}}, \; 0 < t < \infty$$

where $\delta(s) = s\theta\,\phi/(s+\theta)$.

Here $G(T\,|S)$ is empty (in view of completeness of T, given s, for ϕ with fixed θ. We have $l_\theta(s) = n(s-\theta)/\theta(s+\theta)$. But there exists $g = g(s,t;\theta) = t(s-\theta)/\theta^2 s = b(s,t;\theta)l_\theta(s)$ where $b(s,t;\theta) = t(s+\theta)/n\theta s$ so that $E_\omega b|s = 2\phi$.

5 OPTIMALITY OF ESTIMATING FUNCTIONS

A RUEF $g^* = g^*(s,t;\theta)$ in G is said to be optimal (Godambe and Thompson, 1974) for estimating the one-dimensional parameter θ, in the presence of unknown ϕ, if $I_{g^*}(\theta;\omega) \geq I_g(\theta;\omega)$ for all $g \in G$, where I_g is the *information* concerning θ, defined by (2.3). The matrix analog of (2.3) is given by (2.6) for the case of multi-dimensional θ.

Now we wish to consider the marginal score function of S, when the distribution of S is free of the nuisance parameter ϕ. The general form of $g \in G$ has been shown to be (4.1) in this case.

For simplicity of proofs here we consider the case of one-dimensional θ.

We now show that $g_0(s;\theta)$ is the *non-informative* part of g in (4.1) in the sense that the information $I_{g_0}(\theta;\omega)$ is zero; also if $g(s,t;\theta) = g^*(s,t;\theta) + g_0(s;\theta)$ in (4.1), then $I_g(\theta;\omega) \leq I_{g^*}(\theta;\omega)$ with strict inequality unless $g_0(s;\theta) \equiv 0$.

In view of Theorem 4.1, g_0 is a RUEF belonging to $G(S)$, and it is uncorrelated with $l_\theta(S)$. Differentiating $\int g_0(s;\theta)f(s;\theta)d\nu(s) = 0$ with respect to θ under the integral sign we have

$$E_\theta\left[\frac{\partial g_0(S;\theta)}{\partial\theta}\right] = -E_\theta\left[g_0(S;\theta)l_\theta(S)\right] = 0\,.$$

Hence $I_{g_0}(\theta;\omega) = 0$ in view of (2.3).

More generally, for $g = g^* + g_0$, where $E_\omega[g^* g_0] = 0$, we have $E_\omega[g^2] = E_\omega[g^{*2}] + E_\omega[g_0^2]$. Also

$$E_\omega[\partial g/\partial\theta] = E_\omega[\partial g^*/\partial\theta] \quad + \quad E_\omega[\partial g_0/\partial\theta] = E_\omega[\partial g^*/\partial\theta] .$$

Hence

$$I_{g^*}(\theta;\omega) = \frac{[E_\omega(\partial g^*/\partial\theta)]^2}{E_\omega[g^{*2}]} \quad \geq \quad \frac{[E_\omega(\partial g^*/\partial\theta)]^2}{E_\omega[g^{*2}] + E_\omega[g_0^2]} = I_g(\theta;\omega) .$$

Thus we have proved

Proposition 5.1 *If $g(s,t;\theta) = g^*(s,t;\theta) + g_0(s;\theta)$, where $g_0(s;\theta)$ is uncorrelated with $l_\theta(S)$, and also with g^*, then $I_g(\theta;\omega) \leq I_{g^*}(\theta;\omega)$, with strict inequality unless $g_0(s;\theta) \equiv 0$.*

Now we consider the case where $G(T|S)$ is empty.

First suppose that S and T are independent. We plan then to show that $I_g(\theta;\omega) \leq I_{l_\theta(s)}(\theta;\omega)$ for any $g \in G$.

If $G(T|S)$ is empty, then we may assume $g = b(s,t;\theta)l_\theta(s)$ where $E_\omega[b|s] = a(\omega) \neq 0$, in view of (4.1) and Proposition 5.1. But $b(s,t;\theta) = b^*(t;\theta)$ for some b^* in view of independence of S and T. Thus, $g = b^*(T;\theta)l_\theta(S)$, so that $E_\omega(g^2) = E_\omega(b^{*2})E_\theta(l_\theta^2(S)) = E_\omega(b^{*2})I^{(S)}(\theta)$, where $I^{(S)}(\theta)$ is the usual Fisher information in S. Also, $\partial g/\partial\theta = (\partial b^*/\partial\theta)l_\theta(s) + b^*(\partial l_\theta/\partial\theta)$ so that

$$\begin{aligned} E_\omega\left[\frac{\partial g}{\partial\theta}\right] &= E_\omega\left[b^*(T;\theta)\right] E_\theta\left[\frac{\partial^2 \log f(S;\theta)}{\partial\theta^2}\right] \\ &= -I^{(S)}(\theta)E_\omega(b^*) . \end{aligned}$$

Hence, from (2.3),

$$I_g(\theta;\omega) = \frac{E_\omega^2(b^*(T;\theta))}{E_\omega(b^{*2}(T;\theta))} \cdot I^{(S)}(\theta) .$$

Since $E_\omega^2(b^*) \leq E_\omega(b^{*2})$, we have

Proposition 5.2 *If S and T are independent, the distribution of S depends only on θ and $G(T|S)$ is empty, then $I_g(\theta;\omega) \leq I_{l_\theta(S)}(\theta;\omega)$.*

The matrix analogs of Propositions 5.1 and 5.2 can be developed for the general case of vector θ in terms of the information matrix $\boldsymbol{I_g}(\theta;\omega)$, given by (2.6).

Thus the marginal score function of S, viz. $l_\theta(s)$, is the optimal RUEF for θ (uniquely except for multiplication by functions of θ alone), when $G(T|S)$ is empty, if S and T are independent.

If the family of conditional distributions $\{P_\omega^{(T|s)}; \phi \in \Phi\}$ of T, given s for $\phi \in \Phi$ is *complete* for every *fixed* θ, a.e. $(P_\omega^{(S)})$, then $G(T\,|S)$ is empty. Is then $l_\theta(s)$ an optimal RUEF if the distribution of S depends only on θ? Proposition 5.2 requires the additional assumption of independence of S and T to prove optimality of $l_\theta(s)$.

Lloyd (1987) asserted optimality of $l_\theta(s)$, without requiring independence, in the case of one-dimensional θ; a similar assertion was made by Bhapkar and Srinivasan (1993) for the general case. Both these assertions are now seen to be invalid in view of the following counter-example (Bhapkar, 1995).

Example 5.1. (continuation of Example 4.1) In Example 4.1, for given θ and s, T is complete for ϕ and hence $G(T\,|S)$ is empty. However $I_{l_\theta(s)}(\theta;\omega) = n^2/\theta^2(2n+1)$, while $I_g(\theta;\omega) = n/2\theta^2 > I^{(S)}(\theta) = I_{l_\theta(s)}(\theta;\omega)$, for $g = g(s,t;\theta) = t(s-\theta)/\theta^2 s$.

Note that although, here, both $l_\theta(s)$ and g produce the same estimate, viz. $\hat{\theta} = s, l_\theta$ and g are distinct functions, i.e., g is not of the type $k(\theta)l_\theta(s)$, and these distinct functions lead to different information functions.

6 PARTIAL SUFFICIENCY OF S FOR θ

If the marginal distribution of S depends on ω only through θ, S has been termed partially sufficient (p-sufficient) for θ if

$$I_G^{(X)}(\theta;\omega) = I^{(S)}(\theta)\,. \tag{6.1}$$

Thus, it has been shown by Bhapkar (Theorem 3.3, 1991) that S is p-sufficient for θ if the following condition holds:

(i) The conditional *pdf* of T, given s, depends on $\omega = (\theta,\phi)$ only through a parametric function $\delta = \delta(\omega)$, which is differentiable and is such that ω is a one-to-one function of $\eta = (\theta,\delta)$, for almost all $(P_\theta^{(S)})s$.

Earlier the term partial sufficiency had been used by some authors to describe the property of statistic T when $\delta(\omega)$ in (i) is ϕ itself. The property of partial sufficiency of T was, then, also referred to as S-sufficiency by Basu (1977), among others.

There is another situation where the property (6.1) holds; in this situation, condition (i) is replaced by the following condition:

(ii) The family of conditional probability distributions $\{P_\omega^{(T|s)} : \phi \in \Phi\}$ of T, given s, for $\phi \in \Phi$ is complete for each θ a.e. $(P_\theta^{(S)})$, and (S,T) are independent.

The proof that the equality (6.1) holds, under the regularity conditions $\boldsymbol{R}^*, \boldsymbol{R}_G^*$ and the additional condition (ii), when S has *pdf* depending only on θ, follows essentially from Proposition 5.2, in view of inequality (2.4). A direct proof has been given by Bhapkar (1990) under a condition (ii′), which is equivalent to (ii). In condition (ii′) T has been termed a complete p-ancillary statistic for θ.

The independence of (S, T) in the completeness condition (ii) is crucial, as shown by Example 5.1, and thus the assertions in Lloyd (1987) and Bhapkar and Srinivasan (1993) appear to be invalid.

If T is complete for ϕ, given s and θ, then the second terms on the right hand sides of both (3.11) and (3.12) are seen to vanish in view of (3.7) and (3.9). However, it *does not* follow that $\boldsymbol{I}_G(\theta;\omega) = \boldsymbol{I}_G^{(S)}(\theta;\omega)$, in view of comment (d) after Theorem (3.1). Thus the equality (6.1) is not necessarily true when S-distribution depends on ω only through θ, and T is complete for ϕ, given s and θ.

When the relation (6.1) holds, i.e. when S is p-sufficient for θ, then the marginal score function of S, viz. $\boldsymbol{l}_\theta(s)$, is the optimal RUEF for θ, in the sense that

$$\boldsymbol{I}_{\boldsymbol{g}}(\theta;\ \omega) \leq \boldsymbol{I}_{\boldsymbol{l}_\theta(s)}(\theta;\ \omega) \tag{6.2}$$

for all RUEF $\boldsymbol{g}$, in view of relations (2.4), or its matrix version (2.7), and (6.1).

Incidentally, example 5.1 provides a counter-example to the assertion in Corollary 3.2 in Bhapkar (1991). Although the statement of Theorem 3.1 in the 1991-paper is correct, as verified by the proof of Theorem 3.1 in the present paper, Corollary 3.2 was based on the relation (3.14) in the 1991-paper, which is in error. Thus, Example 5.1 has served as a counter-example to both Theorems 3.1 and 3.2 in Bhapkar and Srinivasan (1993). S in Example 5.1, thus, does not satisfy the p-sufficiency property (6.1), but rather the q-sufficiency property in the next section.

7 Q-SUFFICIENCY OF S FOR θ

If T is complete for ϕ, given s and θ, then $G(T|S)$ is empty. Lemma A.7 shows that all RUEF's for θ can be represented in the form $g(s,t;\theta) = \boldsymbol{b}'(s,t;\theta)\boldsymbol{l}_\theta(s) + g_0(s)$, where $E_\omega[\boldsymbol{b}(s,T;\theta)|s] = \boldsymbol{a}(\omega)$. Operationally, a d_1-vector $\boldsymbol{g} = \boldsymbol{g}(s,t;\theta)$ can be considered an estimating function for d_1-dimensional θ if the equation $\boldsymbol{g}(s,t;\theta) = \boldsymbol{0}$ can be solved to produce an estimate $\hat{\theta} = \hat{\theta}(s,t)$. Thus, if T is complete for ϕ, given s and θ.

$$\boldsymbol{g}(s,t;\theta) = \boldsymbol{B}(s,t;\theta)\boldsymbol{l}_\theta(s) + \boldsymbol{g}_0(s)\,, \tag{7.1}$$

where all the elements of $\boldsymbol{g}_0$ belong to $\mathcal{G}_0(S)$ and $E_\omega[\boldsymbol{B}(s,T;\theta|s] = \boldsymbol{A}(\omega)$. In order to find the optimal $\boldsymbol{g}$, which maximizes the information matrix $\boldsymbol{I}_g(\theta;\omega)$

given by (2.6), one may assume without any loss of generality that $\boldsymbol{g}_0 = \mathbf{0}$, in view of Proposition 5.2.

Now solving the estimating equation

$$\boldsymbol{B}(s,t;\theta)\boldsymbol{l}_\theta(s) = \mathbf{0} \tag{7.2}$$

for producing an estimator $\hat{\theta} = \hat{\theta}(s,t)$ is operationally equivalent, in fact, to solving the regular unbiased estimating equation $\boldsymbol{l}_\theta(s) = \mathbf{0}$. For instance, in the case $d_1 = 1, b(s,t;\theta)$ is *not* an unbiased estimating function since $E_\omega[b(s,T;\theta)|s] = a(\omega)$ implies $E_\omega b(S,T;\theta) = a(\omega)$ and $a(\omega) \neq 0$ in view of the assumption that $G(T|S)$ is empty.

One can then define a weaker concept (or property) of sufficiency for the parameters of interest, viz, q-sufficiency, if there exists statistic S with distribution depending on ω only through θ such that the inequality

$$I_{\boldsymbol{g}}(\theta;\omega) \leq I_{\boldsymbol{g}^*}(\theta;\omega) \tag{7.3}$$

for all RUEF $\boldsymbol{g}$ holds for some $\boldsymbol{g}^*$ of the form (7.2), which are operationally equivalent to $\boldsymbol{l}_\theta(s)$, in the sense that the estimator $\hat{\theta} = \hat{\theta}(s,t)$ is produced by the $\boldsymbol{l}_\theta(s)$ part. Note that in (7.2) the part $b(s,t;\theta)$, in the case $d_1 = 1$, is not a RUEF itself in the sense that $E_\omega b \neq 0$. In the d_1-dimensional case no row of $\boldsymbol{B}(s,t;\theta)$ in (7.2) can be a RUEF itself.

Thus, if T is complete for ϕ, given s and θ, then S is at least q-sufficient. If S and T happen to be independent, then S enjoys the stronger property of p-sufficiency.

If S happens to be q-sufficient, one could then define the *q-information* matrix

$$\boldsymbol{I}_Q(\theta;\omega) = \min_{\mathcal{U}^*} E_\omega[\boldsymbol{l}_\theta(S,T) - \boldsymbol{u}][\boldsymbol{l}_\theta(S,T) - \boldsymbol{u}]'] \,, \tag{7.4}$$

where now $\boldsymbol{u} = \boldsymbol{u}(S,T;\omega)$ is permitted to have elements in $\mathcal{U}^*$ which is the space of functions in $\mathcal{C}$ orthogonal to $\mathcal{G}^*$, which is obtained from $\mathcal{G}$ by excluding $g(s,t;\theta) = \boldsymbol{b}'(s,t;\theta)\ \boldsymbol{l}_\theta(s)$ such that $E_\omega[\boldsymbol{b}(s,T;\theta)|s] = \boldsymbol{a}(\boldsymbol{\omega})$, which is not of the form $\boldsymbol{a}^*(\theta)$. Thus $\mathcal{G}^*$ is the sub-space of functions $g^* = \boldsymbol{a}'(\theta)\boldsymbol{l}_\theta(s) + g_0(s;\theta)$, where $g_0 \in \mathcal{G}_0(S)$; hence $\mathcal{G}^* = \mathcal{G}(S)$. Since $\mathcal{G}^* \subset \mathcal{G}$, we have $\mathcal{U}^* \supset \mathcal{U}$, so that

$$\boldsymbol{I}_Q(\theta;\omega) \leq \boldsymbol{I}_G(\theta;\omega) \tag{7.5}$$

The property (6.1) is now replaced by the weaker property

$$\boldsymbol{I}_Q^{(S,T)}(\theta;\omega) = \boldsymbol{I}^{(S)}(\theta)\,; \tag{7.6}$$

we then have

$$\boldsymbol{I}_{\boldsymbol{g}}(\theta;\omega) \leq \boldsymbol{I}_{\boldsymbol{l}_\theta(s)}(\theta;\omega) = \boldsymbol{I}^{(S)}(\theta) = \boldsymbol{I}_Q(\theta;\omega) \tag{7.7}$$

for all $\boldsymbol{g}$ with elements in $\mathcal{G}^*$.

Another situation where S happens to be q-sufficient, but not necessarily p-sufficient, is somewhat similar to the one covered by condition (i) in Section 6 with the difference that the parametric function $\delta = \delta(\omega)$ does depend on s as well. Specifically, for the statistic (S, T), which is sufficient for ω, we have S-distribution depending on ω only through θ, and

(iii) the conditional pdf of T, given s, depends on $\omega = (\theta, \phi)$ only through $\delta_s = \delta(\omega; s)$, which is differentiable such that ω is a one-to-one function of (θ, δ_s), given s, for almost all s.

Under condition (iii), the conditional pdf $h(t; \omega|s) = h^*(t; \delta_s|s)$. Then for the case $d_1 = 1$ we have

$$\frac{\partial \log h(t; \omega|s)}{\partial \theta} = \left[\frac{\partial \log h(t; \omega|s)}{\partial \phi}\right] \boldsymbol{d}(\omega; s)\,, \tag{7.8}$$

for a suitable function $\boldsymbol{d}(\cdot; s)$, arguing as in Theorem 2.2 of Bhapkar (1991).

In view of regularity assumption $\boldsymbol{R}^*$ (ii) every element of $\partial \log h/\partial \phi$ belongs to the space $\mathcal{Y}_s$, given s. The same is true of $\partial \log h/\partial \theta$, i.e. $l_\theta(T|s)$, because of the relation (7.8).

Since $\mathcal{G}_s \oplus \mathcal{Y}_S$ is the orthogonal decomposition of $\mathcal{C}_s$, for every $g_s(t; \theta) \in \mathcal{G}_s$ we have

$$E_\theta\left[l_\theta(T|s) g_s(T; \theta)|s\right] = 0. \tag{7.9}$$

We have, thus,

$$E_\omega\left[\partial g_s(T; \theta)/\partial \theta|s\right] = 0\,, \tag{7.10}$$

again in view of $\boldsymbol{R}^*$ (ii); thus

$$E_\omega\left[\partial g(s, T; \theta)/\partial \theta|s\right] = 0$$

for every $g \in \mathcal{G}(T|S)$. It then follows that

$$E_\omega\left[\partial g(S, T; \theta)/\partial \theta\right] = 0 \tag{7.11}$$

for every $g \in \mathcal{G}(T|S)$.

Therefore, under assumption (iii), $I_g(\theta; \omega) = 0$ for $g \in \mathcal{G}(T|S)$. Hence g is optimal only if g has the form $g(s, t; \theta) = b(s, t; \theta) l_\theta(s)$, in view of Proposition 5.1, where $E_\omega b(S, T; \theta) = a(\omega) \neq 0$.

Thus, arguing as in the complete case, S happens to be q-sufficient for θ in the sense that the optimal $g^*(s, t; \theta)$ is a RUEF operationally equivalent to $l_\theta(s)$, and the property (7.7) holds, when assumption (iii) is satisfied.

References

Basu, D. (1977). On the elimination of nuisance parameters. *Journal of the American Statistical Association*, **72**, 355-366.

Bhapkar, V.P. (1989). Conditioning on ancillary statistics and loss of information in the presence of nuisance parameters. *Journal of Statistical Planning and Inference*, **21**, 139-160.

Bhapkar, V.P. (1990). Conditioning, marginalization and Fisher information functions. *Proceedings of the R.C. Bose Symposium*, R.R. Bahadur (editor), Wiley Eastern Ltd., 123-136.

Bhapkar, V.P. (1991). Loss of information in the presence of nuisance parameters and partial sufficiency. *Journal of Statistical Planning and Inference,* **28**, 185-203.

Bhapkar, V.P. (1995). Completeness and optimality of marginal likelihood estimating equations. *Communications in Statistics, Theory and Methods*, **24**, 945-952.

Bhapkar, V.P. and Srinivasan, C. (1993). Estimating functions: Fisher information and optimality. *Probability and Statistics*, Editors: S.K. Basu and B.K. Sinha, Narosa Publishing House, New Delhi, India, 165-172.

Bhapkar, V.P. and Srinivasan, C. (1994). On Fisher Information inequalities in the presence of nuisance parameters. *Annals of the Institute of Statistical Mathematics*, **46**, 593-604.

Godambe, V.P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika*, **63**, 277-284.

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

Godambe, V.P. and Thompson, M.E. (1974). Estimating equations in the presence of a nuisance parameter. *Annals of Statistics*, **2**, 568-571.

Liang, K.Y. and Zeger, S.L. (1995). Inference based on estimating functions in the presence of nuisance parameters. *Statistical Science*, **10**, 158-173.

Lindsay, B. (1982). Conditional score functions: some optimality results. *Biometrika*, **69**, 503-512.

Lloyd, C.J. (1987). Optimality of marginal likelihood estimating equations. *Communications in Statistics, Theory and Methods*, **16**, 1733-1741.

Rudin, Walter (1974). *Real and Complex Analysis*, McGraw-Hill Book Company, New York.

Appendix

Proposition A.1 *Let* $(\mathcal{X}, \mathcal{A}, \psi)$ *be a* σ*-finite measure space and* $L_2(\mathcal{X}, \psi)$ *the set of all real-valued functions* f *for which* $\int f^2(x)d\psi(x) < \infty$, *then* L_2 *is a Hilbert space.*

The inner product $< f_1, f_2 >$ is then given by $\int f_1 f_2 d\psi$ and the norm $\|f\|$ is defined by $\|f\|^2 = \int f^2 d\psi$.
The proof is given in mathematical texts (see, e.g., p. 81, Rudin, 1974). Such a proof can be modified, as appropriate, to establish

Lemma A.1 *Let $L(\mathcal{X}, \psi)$ be the set of all real-valued functions f for which $\int f(x)d\psi(x) = 0$ and $\int f^2(x)d\psi(x) < \infty$. Then $L(\mathcal{X}, \psi)$ is a Hilbert space with the inner product and norm defined as before.*

Lemma A.2 *Let $\mathcal{C} = \mathcal{C}(\mathcal{X} \times \Omega, \mu \times \pi)$ be the space of real-valued functions $c = c(x;\omega)$ satisfying $E(c) \equiv \int c(x;\omega)p(x;\omega)d\mu(x)d\pi(\omega) = 0$ and $E(c^2) \equiv \int c^2 p d\mu d\pi < \infty$, where $p(x;\omega)$ is the conditional pdf of X, given $\omega \in \Omega$, with respect to a σ-finite measure μ, and π an arbitrary probability measure over Ω. Then $\mathcal{C}$ is a Hilbert space with the inner product $< c_1, c_2 >$ defined as $E(c_1 c_2)$.*
Proof. The set $\mathcal{C}$ forms a linear space over the field R of real numbers. It remains to prove that $\mathcal{C}$ is complete, i.e. to show that every Cauchy sequence $\{c_n\}$ in $\mathcal{C}$ converges to an element c in $\mathcal{C}$ with respect to the norm $\|c\|$ defined by

$$\|c\|^2 = E(c^2) = \int c^2 p d\mu d\pi = \int d^2(x;\omega)d\psi(x;\omega)\,, \tag{A.1}$$

where $d(x;\omega) = c(x;\omega)[p(x;\omega)]^{1/2}$ and $\psi = \mu \times \pi$.
If $\{c_n\}$ is Cauchy, there exists a subsequence $\{c_{n_k}\}, n_1 < n_2 < \ldots$ such that

$$\|c_{n_{k+1}} - c_{n_k}\| < 2^{-k}, k = 1, 2, \ldots \tag{A.2}$$

Let

$$a_k = \sum_{i=1}^{k} |c_{n_{i+1}} - c_{n_i}|\,,\; a = \sum_{i=1}^{\infty} |c_{n_{i+1}} - c_{n_i}|\,.$$

In view of the triangle inequality, and (A.2), we have

$$\|a_k\| \leq \sum_{i=1}^{k} \|c_{n_{i+1}} - c_{n_i}\| < 1\,, k = 1, 2, \ldots$$

Applying Fatou's lemma to $\{a_k^2\}$,

$$\int \left(\liminf_{k \to \infty} a_k^2 \right) p d\psi \leq \liminf_{k \to \infty} \int a_k^2 p d\psi \leq 1\,,$$

i.e. $\|a\| \leq 1$. Thus $a(x;\omega) < \infty$ a.e. (ψ), and the series

$$c_{n_1}(x;\omega) + \sum_{i=1}^{\infty} \{c_{n_{i+1}}(x;\omega) - c_{n_i}(x;\omega)\}$$

converges absolutely, say to $c(x;\omega)$ a.e(ψ). Thus for $c_{n_k}(x;\omega) = c_{n_1}(x;\omega) + \sum_{i=1}^{k-1}\{c_{n_{i+1}} - c_{n_i}\}$

$$\lim_{k\to\infty} c_{n_k}(x;\omega) = c(x;\omega)\,, a.e.(\psi)\,. \tag{A.3}$$

We now show that $c \in \mathcal{C}$, and $\|c_n - c\| \to 0$ as $n \to \infty$.
For any $\epsilon > 0$ there exists N such that $\|c_n - c_m\| < \epsilon$ for m, n greater than N. Applying Fatou's lemma, for $m > N$ we have

$$\begin{aligned}\int (c - c_m)^2 p d\psi &= \int \liminf_{k\to\infty} (c_{n_k} - c_m)^2\, p d\psi \\ &\leq \liminf_{k\to\infty} \int (c_{n_k} - c_m)^2\, p d\psi \leq \epsilon^2\,. \end{aligned} \tag{A.4}$$

It follows that $E(c^2) = \int c^2 p d\psi < \infty$. Also

$$\begin{aligned}|E(c)| &= |E(c - c_m + c_m| = |E(c - c_m)| = |\int (c - c_m) p d\psi| \\ &\leq \left[\int (c - c_m)^2 p d\psi\right]^{1/2} \left[\int p d\psi\right]^{1/2},\end{aligned}$$

in view of Cauchy-Schwartz inequality. Since $\int p d\psi = \int p d\mu d\pi = \int d\pi = 1$, we have $|E(c)| \leq \epsilon$, in view of (A.4). Since ϵ is arbitrary, this proves that $E(c) = 0$. Thus $c \in \mathcal{C}$, and $\|c_m - c\| \to 0$ as $m \to \infty$ in view of (A.4). Thus the lemma is established.
Assume the regularity conditions $\boldsymbol{R}^*$, in Section 2, for the joint distribution of (S, T), given ω. Consider now the probability distribution $\pi = \pi_1 \times \pi_2$ of $\omega = (\theta, \phi)$ over $\Theta \times \Phi$.
$\mathcal{C}$ is the Hilbert space of real-valued functions $c = c(s, t; \omega)$, which satisfy

$$\begin{aligned} E(c) &\equiv \int c(s,t;\omega) p(s,t;\omega) d\mu(s,t) d\pi(\omega) = 0 \\ E(c^2) &\equiv \int c^2 p d\mu d\pi < \infty\,. \end{aligned}$$

See Lemma A.2 for the proof of assertion that $\mathcal{C}$ is a Hilbert space.
The subspaces (and their closures in $\mathcal{C}$, as needed) considered later on are defined below:

$$\begin{aligned} G &= \{g : g = g(s,t;\theta)\,, g \in \mathcal{C} \text{ and} \\ &\qquad g \text{ is differentiable with respect to elements of } \theta\,\} \\ \mathcal{G} &= \bar{G} \\ \mathcal{U} &= \mathcal{G}^{\perp} \text{ so that } \mathcal{C} = \mathcal{G} \oplus \mathcal{U} \\ \mathcal{C}(S) &= \{c : c = c(s;\omega) \text{ and } c \in \mathcal{C}\} \\ G(S) &= \{g : g = g(s;\theta) \text{ and } g \in G\} \\ \mathcal{G}(S) &= \bar{G}(S) \\ \mathcal{V}(S) &= \mathcal{G}(S)^{\perp} \text{ in } \mathcal{C}(S), \text{i.e. } \mathcal{C}(S) = \mathcal{G}(S) \oplus \mathcal{V}(S)\,. \end{aligned}$$

For a given value s of S, define

$$\begin{aligned}\mathcal{C}_s &= \{c_s : c_s = c_s(t;\omega)\,, E_\omega(c_s\,|s) \\ &\equiv \int c_s h(t;\omega\,|s)d\eta_s(t) = 0\,, \text{a.e. } (\pi)\,; \\ & E(c_s^2\,|s) = \int c_s^2 h d\eta_s d\pi(\omega) < \infty\}\,. \end{aligned} \tag{A.5}$$

The proof that $\mathcal{C}_s$ is itself a Hilbert space, given s, is very similar to that of Lemma A.2; first we have

Proposition A.2 *Every Cauchy sequence* $\{c_{s,n_k}\}, n_1 < n_2 < \ldots$ *in* $\mathcal{C}_s$ *converges to an element in* $\mathcal{C}_s$*; hence* $\mathcal{C}_s$ *is a Hilbert space.*
Proof. Arguing as in the proof of Lemma A.2, we have

$$\lim_{k\to\infty} c_{s,n_k}(t;\omega) = c_s(t;\omega)\,, \text{a.e.}(\eta_s \times \pi) \tag{A.6}$$

in view of (A.3). It remains to verify that $E_\omega[c_s(T;\omega)\,|s] = 0$, a.e.$(\pi)$; the remainder of the proof goes along the lines of Lemma A.2 proof. Now

$$\begin{aligned} E_\omega[c_s\,|s] &= \int (c_s - c_{s,m} + c_{s,m})dP_\omega^{(T|s)} = \int (c_s - c_{s,m})dP_\omega^{(T|s)} \\ &= \int \liminf_{k\to\infty} (c_{s,n_k} - c_{s,m})dP_\omega^{(T|s)} \le \liminf_k \int (c_{s,n_k} - c_{s,m})dP_\omega^{(T|s)}\,, \end{aligned}$$

by applying Fatou's lemma to $\{c_{s,n_k}\}$ as $k \to \infty$. Hence $E_\omega[c_s|s] \le 0$, in view of (A.6), noting that (A.6) holds a.e.(π). A similar argument applied to $-c_s$ gives the inequality $-E_\omega[c_s|s] \le 0$. Thus, $E_\omega[c_s|s] = 0$, a.e.(π). The proposition is established by arguing as in the proof of Lemma A.2.
In $\mathcal{C}_s$, we define the subspaces $\mathcal{G}_s$ and $\mathcal{Y}_s$ as given below.

$$\begin{aligned} G_s &= \{g_s : g_s = g_s(t;\theta), g_s \in \mathcal{C}_s \text{ and} \\ & \quad g_s \text{ is differentiable with respect to } \theta\} \\ \mathcal{G}_s &= \bar{G}_s \\ \mathcal{Y}_s &= \mathcal{G}_s^\perp \text{ in } \mathcal{C}_s, i.e. \mathcal{C}_s = \mathcal{G}_s \oplus \mathcal{Y}_s \end{aligned}$$

Finally we define in $\mathcal{C}$

$$\begin{aligned} \mathcal{C}(T|S) &= \{c : c = c(s,t;\omega), c \in \mathcal{C} \text{ and} \\ & \quad c(s,t;\omega) = c_s(t;\omega), c_s \in \mathcal{C}_s \text{ a.e.}(\nu \times \pi)\} \\ G(T|S) &= \{g : g = g(s,t;\theta), g \in G \text{ and} \\ & \quad g(s,t;\theta) = g_s(t;\theta), g_s \in G_s \text{ a.e.}(\nu \times \pi)\} \\ \mathcal{G}(T|S) &= \bar{G}(T|S) \\ \mathcal{Y}(T|S) &= \{y : y = y(s,t;\omega), y \in \mathcal{C} \text{ and} \\ & \quad y(s,t;\omega) = y_s(t;\omega), y_s \in \mathcal{Y}_s \text{ a.e.}(\nu \times \pi)\ \}\,. \end{aligned}$$

Lemma A.3 $\mathcal{C} = \mathcal{C}(S) \oplus \mathcal{C}(T|S)$.

PROOF Let $c^* = c^*(s;\omega) \in \mathcal{C}(S)$ and $c = c(s,t;\omega) \in \mathcal{C}(T|S)$. Then

$$E(cc^*) = \int c^*(s;\omega)\left[\int c(s,t;\omega)dP_\omega^{(T|s)}\right] dP_\omega^{(S)} d\pi(\omega) = 0$$

and thus $\mathcal{C}(S) \perp \mathcal{C}(T|S)$. Now we show that $\mathcal{C}(S)^\perp$ is $\mathcal{C}(T|S)$.
Suppose now $c \perp \mathcal{C}(S)$, and let $c^*(s;\omega) = E_\omega(c|s)$. Since $c^* \in \mathcal{C}(S)$, $E(cc^*) = 0$. But

$$\begin{aligned} E(cc^*) &= \int c^*(s;\omega)\left[\int c(s,t;\omega)dP_\omega^{(T|s)}\right] dP_\omega^{(S)} d\pi(\omega) \\ &= \int c^{*2}(s;\omega)dP_\omega^{(S)} d\pi(\omega)\,. \end{aligned}$$

Hence $c^*(s;\omega) = 0$, a.e.$(\nu \times \pi)$. Thus $c \in \mathcal{C}(T|S)$, and the lemma is proved.

Lemma A.4 $\mathcal{C}(T|S) = \mathcal{G}(T|S) \oplus \mathcal{Y}(T|S)$.
PROOF. If $y \in \mathcal{Y}(T|S)$ and $g \in \mathcal{G}(T|S)$, then

$$E(gy) = \int gy dP_\omega^{(T|s)} dP_\omega^{(S)} d\pi(\omega) = \int \left[\int g_s y_s dP_\omega^{(T|s)}\right] dP_\omega^{(S)} d\pi(\omega) = 0$$

Thus $\mathcal{G}(T|S) \perp \mathcal{Y}(T|S)$.
Suppose now $c \in \mathcal{C}(T|S)$ and $c \perp \mathcal{G}(T|S)$. Since $c \in \mathcal{C}(T|S)$, $c = c(s,t;\omega) = c_s(t;\omega)$, where $c_s \in \mathcal{C}_s$, a.e.$(\nu \times \pi)$. Consider the orthogonal decomposition of c_s in $\mathcal{C}_s$, viz

$$c_s(t;\omega) = g_s^*(t;\theta) + y_s^*(t;\omega)\,.$$

Consider now $g^*(s,t;\theta) = g_s^*(t;\theta)$; then $g^* \in \mathcal{G}(T|S)$ and, hence, $c \perp g^*$, i.e. $E(cg^*) = 0$. But

$$\begin{aligned} E(cg^*) &= \int cg^* dP_\omega^{(T|s)} dP_\omega^{(S)} d\pi(\omega) = \int \left[\int c_s(t;\omega) g_s^*(t;\theta) dP_\omega^{(T|s)}\right] dP_\omega^{(S)} d\pi(\omega) \\ &= \int \left[\int g_s^{*2}(t;\theta) dP_\omega^{(T|s)}\right] dP_\omega^{(S)} d\pi(\omega) \\ &= \int g^{*2}(s,t;\theta) dP_\omega^{(T|s)} dP_\omega^{(S)} d\pi(\omega)\,. \end{aligned}$$

Hence $g^*(s,t;\theta) = 0$, a.e.$(\mu \times \pi_1)$. Then $g_s^*(t;\theta) = 0$ a.e.$(\nu \times \pi_1)$, which implies $c_s = y_s^*$ a.e.$(\nu \times \pi)$ so that $c(s,t;\omega) = y^*(s,t;\omega)$, where $y^*(s,t;\omega) = y_s^*(t;\omega)$. Thus $c \in \mathcal{Y}(T|S)$, and the lemma is proved.

Lemma A.5 Suppose that the distribution of S, given ω, depends on ω only through θ,

$$\begin{aligned} \mathcal{G}_1(S) = \{ g_1 : g_1 = g_1(s;\theta) = \boldsymbol{a}'(\theta)\boldsymbol{l}_\theta(s) \in \mathcal{G}(S) \\ \text{for some } \boldsymbol{a}(\theta)\}\,, \end{aligned} \tag{A.7}$$

and $\mathcal{G}_0(S)$ is the sub-space in $\mathcal{G}(S)$ orthogonal to $\mathcal{G}_1(S)$. Then

$$\mathcal{G} = \mathcal{G}_0(S) \oplus \mathcal{G}^*(S,T)\,, \tag{A.8}$$

where

$$\begin{aligned} \mathcal{G}^*(S,T) &= \{g^* : g^* = g^*(s,t;\theta) \in \mathcal{G} \text{ and} \\ k^*(s;\theta) &\equiv \int g^*(s,t;\theta)dP_\omega^{(T|s)}d\pi_2(\phi) \in \mathcal{G}_1(S)\,,\ \text{a.e. } (\nu \times \pi_1)\}\,. \end{aligned} \tag{A.9}$$

PROOF. When the distribution of S, given ω, depends only on θ, then each component of $\boldsymbol{l}_\theta(s)$ belongs to $\mathcal{G}(S)$ in view of regularity condition $\mathbf{R}^*$ (i). Since $\mathcal{G}(S)$ is complete, it is a Hilbert space with decomposition $\mathcal{G}(S) = \mathcal{G}_0(S) \oplus \mathcal{G}_1(S)$.
We also note that $\mathcal{G}$ is a complete subspace of $\mathcal{C}$ and, hence, $\mathcal{G}$ is a Hilbert space. $\mathcal{G}^*(S,T)$ is seen to be orthogonal to $\mathcal{G}_0(S)$. It remains to show that $\mathcal{G}^*(S,T)$ is the orthogonal complement of $\mathcal{G}_0(S)$ in $\mathcal{G}$.
Let then $g \in \mathcal{G}$ and suppose $g \perp \mathcal{G}_0(S)$. Then for all $g_0 \in \mathcal{G}_0(S)$

$$\begin{aligned} 0 &= E[g_0 g] = \int g_0(s;\theta)\left[\int g(s,t;\theta)dP_\omega^{(T|s)}d\pi_2(\phi)\right]dP_\theta^{(S)}d\pi_1(\theta) \\ &= E[g_0 k^*]\,. \end{aligned}$$

Thus $k^* \perp \mathcal{G}_0(S)$. Since $k^* \in \mathcal{G}(S)$, it follows that $k^* \in \mathcal{G}_1(S)$. Thus the lemma is established.

Lemma A.6 Under the assumptions of Lemma A.5, let

$$\begin{aligned} \mathcal{G}(S,T) &= \{g : g = g(s,t;\theta) = \boldsymbol{b}'(s,t;\theta)\boldsymbol{l}_\theta(s) \text{ such} \\ \text{that } E_\omega[\boldsymbol{b}(s,T;\theta)|s] &= \boldsymbol{a}(\omega), \text{ and } g \in \mathcal{G}\}\,. \end{aligned} \tag{A.10}$$

(i) If there is a $g^* \in \mathcal{G}^*(S,T)$ orthogonal to $\mathcal{G}(S,T)$, then $k^*(s,\theta) = 0$, a.e. $(\nu\times\pi_1)$, and (ii) for every $g \in \mathcal{G}(S,T)$, we have the orthogonal decomposition

$$g = \boldsymbol{a}^{*\prime}(\theta)\boldsymbol{l}_\theta(s) + [\boldsymbol{b}'(s\,,t;\theta) - \boldsymbol{a}^{*\prime}(\theta)]\boldsymbol{l}_\theta(s)\,, \tag{A.11}$$

where $E_\omega[\boldsymbol{b}(s\,,T;\theta|s)] = \boldsymbol{a}(\omega)$, and $\int \boldsymbol{a}(\omega)d\pi_2(\phi) = \boldsymbol{a}^*(\theta)$.
Furthermore, (iii) if π is a one-point distribution at ω, then $\mathcal{G}$ has the orthogonal decomposition

$$\mathcal{G} = \mathcal{G}_0(S) \oplus \mathcal{G}(S,T)\,.$$

PROOF. Note first that $\mathcal{G}(S,T) \subset \mathcal{G}^*(S,T)$. Also observe that $\mathcal{G}_1(S) \subset \mathcal{G}(S,T)$. Hence if there is a $g^* \in \mathcal{G}^*(S,T)$ orthogonal to $\mathcal{G}(S,T)$, then we have for all $g_1 \in \mathcal{G}_1(S)$

$$0 = E[g_1 g^*] = \int g_1(s,\theta)k^*(s;\theta)dP_\theta^{(S)}d\pi_1(\theta)$$

in view of (A.9). However $k^* \in \mathcal{G}_1(S)$ in view of lemma A.5 and (i) follows.
To prove (ii), note that the two components of g in (A.11) are orthogonal. We need to show that the subspace spanned by the second component is

the orthogonal complement of $\mathcal{G}_1(S)$ in $\mathcal{G}(S,T)$. Let then $g \in \mathcal{G}(S,T)$ and suppose $g \perp \mathcal{G}_1(S)$. Then for all $g_1 \in \mathcal{G}_1(S)$

$$\begin{aligned} 0 = E[g_1 g] &= \int g_1(s;\theta)\boldsymbol{a}'(\omega)\boldsymbol{l}_\theta(s)dP_\theta^{(S)}d\pi(\omega) \\ &= \int \boldsymbol{a}_1'(\theta)\boldsymbol{l}_\theta(s)\boldsymbol{l}_\theta'(s)\boldsymbol{a}(\omega)dP_\theta^{(S)}d\pi(\omega) \\ &= \int \boldsymbol{a}_1'(\theta)\boldsymbol{I}^{(S)}(\theta)\boldsymbol{a}(\omega)d\pi(\omega)\,, \end{aligned}$$

in view of (A.7) and (A.10). Since this is true for all $\boldsymbol{a}_1(\theta)$, we have $\int \boldsymbol{a}(\omega)d\pi_2(\phi) = \mathbf{0}$. Thus (ii) follows.

Finally, if π is a one-point distribution at ω, then $E_\omega[g^*|s] = 0$ and, thus, $g^* \in \mathcal{G}(T|S)$, which is a subset of $\mathcal{G}(S,T)$. Since g^* is orthogonal to $\mathcal{G}(S,T)$, it follows that $g^* = 0$. Thus we have assertion (iii).

Lemma A.7. Under the assumptions of Lemma A.5 and A.6, suppose π is a one-point distribution at ω. If now $\mathcal{G}(T|S)$ is empty, every $g \in \mathcal{G}$ has a representation

$$g = \boldsymbol{b}'(s,t;\theta)\boldsymbol{l}_\theta(s) + g_0(s;\theta)\,, \tag{A.12}$$

where $g_0 \in \mathcal{G}_0(S)$ and $E_\omega[\boldsymbol{b}(s,T;\theta)|s] = \boldsymbol{a}(\omega) \neq \mathbf{0}$.

Remark. Although (A.12) gives the general representation of $\mathcal{G}$ for the case where $\mathcal{G}(T|S)$ is empty and π is a one-point distribution at ω, the decomposition (A.11) shows that g in $\mathcal{G}(S,T)$ belongs to $\mathcal{G}_1(S)$ only if $E_\omega[\boldsymbol{b}(s,T;\theta)|s] = \boldsymbol{a}^*(\theta)$ for some $\boldsymbol{a}^*$, i.e. $\boldsymbol{b}(s,t;\theta) = \boldsymbol{a}^*(\theta)$ in view of assumption that $\mathcal{G}(T|S)$ is empty. Then g is outside $\mathcal{G}(S)$ only if $E_\omega[\boldsymbol{b}(s,T;\theta)|s] = \boldsymbol{a}^*(\omega)$ for some function $\boldsymbol{a}^*$ which depends also on ϕ in a non-trivial manner. The orthogonal decomposition of $\mathcal{G}(S,T)$ into $\mathcal{G}_1(S)$ and its complement in $\mathcal{G}(S,T)$ as given in (A.11) is now possible only as

$$g(s,t;\theta) = \boldsymbol{a}'(\omega)\boldsymbol{l}_\theta(s) + [\boldsymbol{b}'(s,t;\theta) - \boldsymbol{a}'(\omega)]\boldsymbol{l}_\theta(s)$$

where $\boldsymbol{a}(\omega)$, for the *given* ω, depends non-trivially on co-ordinates ϕ of *given* ω.