Ann. Inst. Statist. Math.
39 (1987), Part A, 1-23

ANALYSIS OF MARGINAL AND CONDITIONAL DENSITY FUNCTIONS FOR SEPARATE INFERENCE

HISATAKA KUBOKI

(Received Jan. 16, 1985; revised June 23, 1986)

Summary

This paper discusses, with measure-theoretical rigor, some basic aspects of the theory of separate inference. To analyze densities of marginal and conditional submodels, certain operators are introduced. First a general concept of decomposition of a model is proposed, and the corresponding factorization of densities of the model is established. Next it is shown that the property of smoothness of a family of densities is retained in the operation of conditioning, and therefore it yields the differentiability of the conditional expectation of a real-valued statistic in a certain sense. On the basis of this result, two measures of the effectiveness of a submodel in separate inference are investigated.

1. Introduction

Separate inference is inference on parameters of interest from a part of the original model and data (Barndorff-Nielsen [1]). There are two key procedures for this. One is to decompose the model into several submodels with certain statistical structures. The other is to examine such submodels through a measure of the effectiveness of a submodel in the inference on parameters of interest. The main purpose of this paper is to discuss, with measure-theoretical rigor, the following basic aspects of the procedures: factorization of densities of a model; smoothness of a family of conditional densities; differentiability of the conditional expectation of a statistic.

To this end, we shall use certain operators for averaging, which were introduced by Pitman [11] and were developed by Kuboki [8]. The definitions and some further properties of them are described in Section 2. In particular, our interest is the case where they operate on

Key words and phrases: Efficacy matrix, factorization of densities, marginal and conditional densities, sensitivity matrix, separate inference, smoothness of a family of densities.

probability density functions. In Section 3, we shall investigate properties of the operators in this case, and give two propositions which are important to our discussions below.

First we shall deal with factorization of densities of a model. Let $\mathcal{P}_\Theta=\{P_\theta: \theta\in\Theta\}$ be a model for data X, i.e. a family of distributions of X, and let $r:\Theta\to\Gamma$ and $w:\Theta\to\Omega$ be two surjective parameter functions. We denote by $\mathcal{P}_\Gamma$ the partition of $\mathcal{P}_\Theta$ induced by r, i.e. $\mathcal{P}_\Gamma=\{\mathcal{P}_\gamma:\gamma\in\Gamma\}$, $\mathcal{P}_\gamma=\{P_\theta: \theta\in\Theta, r(\theta)=\gamma\}$; and denote by $\mathcal{P}_\Omega$ the partition of $\mathcal{P}_\Theta$ induced by w, i.e. $\mathcal{P}_\Omega=\{\mathcal{P}_\omega: \omega\in\Omega\}$, $\mathcal{P}_\omega=\{P_\theta: \theta\in\Theta, w(\theta)=\omega\}$. Consider any two statistics U and T such that U is a function of T. Then we shall introduce the following concept.

DEFINITION 1.1. We shall say that $(U, T|U)$ *induces* $(\mathcal{P}_\Gamma, \mathcal{P}_\Omega)$ if U and T satisfy the following two conditions.
(i) The marginal distributions of U depend on θ only through r.
(ii) For each event A of T and for each $\theta\in\Theta$, there exists a determination of the conditional probability of A given U that depends on θ only through w.

The condition (i) implies that the mapping $P_\theta\to P_\theta^U$, $\theta\in\Theta$, induces the same partition of $\mathcal{P}_\Theta$ as does r, where P_θ^U denote marginal distributions of U. Similarly, the condition (ii) implies that the mapping $P_\theta\to P_\theta^{T|U}$, $\theta\in\Theta$, induces the same partition of $\mathcal{P}_\Theta$ as does w, where $P_\theta^{T|U}$ denote the conditional distributions of T given U if they exist. The above concept includes various concepts of ancillarity and sufficiency found in the literature (Fraser [3], Sandved [13], Basu [2], Barndorff-Nielsen [1]). For example, take $T=X$. Then, U is ancillary for $\theta\in\Theta$ if r is a constant function, and U is sufficient for $\theta\in\Theta$ if w is a constant function.

In this paper, let us suppose that $\mathcal{P}_\Theta$ is dominated by a σ-finite measure m; and let us denote by p_θ a density of each $P_\theta\in\mathcal{P}_\Theta$. Then it will be useful to review Definition 1.1 in terms of the corresponding factorization of densities of T. Further, it will be useful to establish a factorization criterion of the densities for the concept in Definition 1.1. In Section 4, we shall consider these problems by using the operators introduced in Section 2.

Next we shall discuss Pitman's concept of smoothness of a family of densities. For ease of reference, we shall here describe the definition of the concept with some modification; see Pitman ([11], Chapter 3). Suppose that $\Theta=\Theta_1\times\cdots\times\Theta_n$, Θ_i being an open interval of R^1. We shall write $\theta={}^t(\theta_1,\cdots,\theta_n)$, where ${}^tM$ denotes the transpose of a matrix M. Further suppose that p_θ is differentiable a.e. m with respect to θ, i.e. there exists $p'_\theta={}^t(p'_{1\theta},\cdots,p'_{n\theta})$ such that for every fixed $l={}^t(l_1,\cdots,l_n)\in R^n$,

$$\lim_{\varepsilon\to 0}(p_{\theta+\varepsilon l}-p_\theta)/\varepsilon={}^t l p'_\theta \qquad \text{a.e. } m \text{ at every } \theta\in\Theta .$$

Define $\rho_p(\tau, \theta)$, the distance between P_τ and P_θ, $\tau, \theta\in\Theta$, by

$$\rho_p^2(\tau, \theta)=\int(\sqrt{p_\tau}-\sqrt{p_\theta})^2 dm .$$

Denote $\liminf_{\varepsilon\to 0} 2\rho_p(\theta+\varepsilon l, \theta)/|\varepsilon|$, which always exists, by $s_p(\theta|l)$; it is a measure of the sensitivity of the family of densities p_θ to small changes in θ in the direction l. We shall call $s_p^2(\theta|l)$ the *sensitivity* of the p_θ family (at θ in the direction l). Putting $\dot{p}_\theta=p'_\theta/\sqrt{p_\theta}$ or $=0$, according as $x\in S(p_\theta)=\{x: p_\theta(x)>0\}$ or $x\notin S(p_\theta)$, we can write the information matrix $I_p(\theta)$ of the p_θ family in the form

$$I_p(\theta)=\left[\int \dot{p}_\theta{}^t\dot{p}_\theta dm\right] .$$

Evidently, for every $\theta\in\Theta$ and every $l\in R^n$, $s_p^2(\theta|l)\geqq {}^t l I_p(\theta) l$.

DEFINITION 1.2. If for every fixed $l\in R^n$, $\lim_{\varepsilon\to 0} 4\rho_p^2(\theta+\varepsilon l, \theta)/\varepsilon^2$ exists, is finite, and is equal to ${}^t l I_p(\theta) l$ at every $\theta\in\Theta$, then we shall say that the p_θ family is *smooth* in Θ.

When the p_θ family is smooth in Θ, the value of ${}^t l I_p(\theta) l$ is the sensitivity of the family in the direction l at θ, so that then $I_p(\theta)$ is called the *sensitivity matrix* of the family.

It is proved by Pitman [11] that if the p_θ family is smooth in Θ, then so are the family of marginal densities of T and U. Further as pointed out by him, the property of smoothness is closely connected with the differentiability of the expectation of a real-valued statistic, and therefore with a certain property of the *efficacy matrix* of the statistic, a measure of the effectiveness of the statistic in investigating the value of θ. In Section 5, first we shall discuss the validity of the following two properties: smoothness of the family of conditional densities of T given U; and differentiability of the conditional expectation of a real-valued statistic given U in a certain sense. Next, on the basis of them, we shall introduce a measure of effectiveness of a statistic in investigating the value of θ conditionally on U. Here we remark that the conclusions are independent of the particular choice of that version of p_θ family which is smooth in Θ.

The last section discusses measures of the effectiveness of a submodel in inference on parameters of interest. In particular, let us consider the case where only θ_1 is the parameter of interest. Up to now, two measures are proposed to evaluate the effectiveness of a submodel in investigating only the value of θ_1. One is Liang's [9] measure of information about θ_1 and the other is Godambe's [5] measure of information about θ_1. We shall show that the former is related to sensitivity and the latter is related to efficacy. Furthermore, we shall discuss the relationship between the two measures when the conditional densities of T given U depend on θ only through θ_1. In connection with this, Godambe's [4] two concepts of ancillarity will also be discussed.

2. Notations and preliminaries

Let $(\mathcal{X}, \mathcal{F})$ be a measure space: a set $\mathcal{X}$ and a σ-algebra $\mathcal{F}$ of subsets of $\mathcal{X}$. T is a mapping from $\mathcal{X}$ into a space $\mathcal{T}$, and Π is a mapping from $\mathcal{T}$ into a space $\mathcal{U}$. Let $\mathcal{A}$ be a σ-algebra of subsets of $\mathcal{T}$, and $\mathcal{B}$ a σ-algebra of subsets of $\mathcal{U}$. We shall assume that T and Π are measurable. Define a measurable mapping U from $\mathcal{X}$ into $\mathcal{U}$ by $U=\Pi \circ T$.

Let m be a given σ-finite measure over $(\mathcal{X}, \mathcal{F})$. Then choose σ-finite measures μ over $(\mathcal{T}, \mathcal{A})$ and ν over $(\mathcal{U}, \mathcal{B})$ such that $m^T \ll \mu$ and $\mu^\Pi \ll \nu$, where $m^T, =mT^{-1}$, denotes the induced measure over $(\mathcal{T}, \mathcal{A})$ and $\mu^\Pi, = \mu\Pi^{-1}$, denotes the induced measure over $(\mathcal{U}, \mathcal{B})$. The σ-finiteness of m guarantees the existence of such measures μ and ν. It should be noted that m^T and μ^Π are sometimes not σ-finite even though m and μ are σ-finite. Evidently $m^U \ll \nu$, where $m^U, =mU^{-1}$, is the induced measure over $(\mathcal{U}, \mathcal{B})$.

Let us denote by $\mathcal{E}(\mathcal{X}, \mathcal{F}, m)$ the class of all extended real-valued $\mathcal{F}$-measurable functions whose m-integral exists, i.e.

$$\mathcal{E}(\mathcal{X}, \mathcal{F}, m)=\left\{\phi\colon \phi(\in)\mathcal{F},\ \int \phi^+ dm<\infty \text{ or } \int \phi^- dm<\infty\right\},$$

where the notation $\phi(\in)\mathcal{F}$ expresses the $\mathcal{F}$-measurability of ϕ. Of course, ϕ and ψ in $\mathcal{E}(\mathcal{X}, \mathcal{F}, m)$ are regarded as identical if $\phi=\psi$ a.e. m. Likewise, we shall define $\mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$ and $\mathcal{E}(\mathcal{U}, \mathcal{B}, \nu)$. For every $\phi \in \mathcal{E}(\mathcal{X}, \mathcal{F}, m)$, let us define $T^*\phi \in \mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$ by the following formula:

$$\int_{T^{-1}A} \phi dm=\int_A (T^*\phi)d\mu \qquad \text{for all } A \in \mathcal{A}. \tag{2.1}$$

This is justified as follows. For each $\phi \in \mathcal{E}(\mathcal{X}, \mathcal{F}, m)$, there exists the indefinite integral Φ over $(\mathcal{T}, \mathcal{A})$ such that $\Phi(A)=\int_{T^{-1}A} \phi dm$ for all $A \in \mathcal{A}$. Note that $\mu(A)=0 \Rightarrow m^T(A)=0 \Rightarrow \Phi(A)=0$. Hence $\mu \gg \Phi$, and so by the extended Radon-Nikodym theorem (Loève [10], p. 134) there exists a function $\psi \in \mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$, determined up to μ-equivalence, such that $\Phi(A)=\int_A \psi d\mu$ for all $A \in \mathcal{A}$. We shall write $\psi=T^*\phi$. This averaging operator T^* was introduced by Pitman [11]. Similarly, for every $\phi \in$

$\mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$ we define $\Pi^*\phi \in \mathcal{E}(\mathcal{U}, \mathcal{B}, \nu)$ by

(2.2) $$\int_{\Pi^{-1}B} \phi d\mu = \int_B (\Pi^*\phi) d\nu \qquad \text{for all } B \in \mathcal{B},$$

and for every $\phi \in \mathcal{E}(\mathcal{X}, \mathcal{F}, m)$ we define $U^*\phi \in \mathcal{E}(\mathcal{U}, \mathcal{B}, \nu)$ by

(2.3) $$\int_{U^{-1}B} \phi dm = \int_B (U^*\phi) d\nu \qquad \text{for all } B \in \mathcal{B}.$$

LEMMA 2.1. *We have* $U^* = \Pi^* \circ T^*$.

Here we shall list some properties of the operator Π^*. Lemmas 2.2 and 2.3 below are easy consequences of the defining relation (2.2). For Lemma 2.4 and other properties, see Pitman ([11], p. 101) and Kuboki [8]. We can also see that the operators T^* and U^* satisfy properties similar to them. From now on, denote the indicator function of a set E by χ_E, and for a function ϕ denote the set $\{\cdot : \phi(\cdot) \neq 0\}$ by $S(\phi)$.

LEMMA 2.2. (i) *Suppose that* $\phi(\in)\mathcal{A}$. *Then*

$$\phi \geqq 0 \text{ a.e. } \mu \Rightarrow \Pi^*\phi \geqq 0 \text{ a.e. } \nu \quad \text{and} \quad S(\phi) \subset \Pi^{-1}\{S(\Pi^*\phi)\} \text{ a.e. } \mu.$$

(ii) *Suppose that* $\phi, \psi(\in)\mathcal{A}$, *and that* $\phi, \psi \geqq 0$ *a.e.* μ. *Then*

$$S(\phi) \subset S(\psi) \text{ a.e. } \mu \Rightarrow S(\Pi^*\phi) \subset S(\Pi^*\psi) \text{ a.e. } \nu.$$

LEMMA 2.3. *Suppose that* $\phi, \psi \in \mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$, *and that* $B \in \mathcal{B}$. *Then*

$$\Pi^*(\chi_A\phi) = \Pi^*(\chi_A\psi) \text{ a.e. } \nu \text{ on } B \text{ for each } A \in \mathcal{A}$$
$$\Longleftrightarrow \phi = \psi \text{ a.e. } \mu \text{ on } \Pi^{-1}B.$$

LEMMA 2.4. (i) *Suppose that* $\phi, \psi \in \mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$. *If* $a\int \phi d\mu + b\int \psi d\mu$ *exists for constants* a *and* b, *then* $\Pi^*(a\phi + b\psi) = a\Pi^*\phi + b\Pi^*\psi$ *a.e.* ν.

(ii) *If* $\phi \in \mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$, *then* $|\Pi^*\phi| \leqq \Pi^*|\phi|$ *a.e.* ν.

(iii) *Suppose that* $\phi, \psi(\in)\mathcal{A}$. *If* ϕ^2 *and* ψ^2 *are* μ-*integrable, then* $\{\Pi^*(\phi\psi)\}^2 \leqq (\Pi^*\phi^2)(\Pi^*\psi^2)$ *a.e.* ν.

(iv) *Suppose that* $g(\in)\mathcal{B}$. *If* ϕ, $(g \circ \Pi)\phi \in \mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$, *then* $\Pi^*\{(g \circ \Pi)\phi\} = g\Pi^*\phi$ *a.e.* ν.

3. The operators on probability density functions

Let us consider the subclass $\mathcal{P}(\mathcal{X}, \mathcal{F}, m)$ consisting of all probability density functions in $\mathcal{E}(\mathcal{X}, \mathcal{F}, m)$. Put for each $p \in \mathcal{P}(\mathcal{X}, \mathcal{F}, m)$

(3.1) $$P(F) = \int_F p dm \qquad \text{for all } F \in \mathcal{F}.$$

As before, we shall denote by P^T the induced measure PT^{-1}, and by P^U the induced measure PU^{-1}. Then it follows from (2.1) and (2.3) that

$$(3.2) \qquad P^T(A)=\int_A (T^*p)d\mu \qquad \text{for all } A \in \mathcal{A}\,,$$

$$(3.3) \qquad P^U(B)=\int_B (U^*p)d\nu \qquad \text{for all } B \in \mathcal{B}\,.$$

Combining (3.1)–(3.3) with Lemma 2.2. (i), we see that if a random variable X over $(\mathcal{X}, \mathcal{F})$ is distributed with the density p relative to m, then T^*p is the density of the marginal distribution P^T of $T(X)$ relative to μ, and U^*p is the density of the marginal distribution P^U of $U(X)$ relative to ν.

Now for each $p \in \mathcal{P}(\mathcal{X}, \mathcal{F}, m)$, let us denote by $p^{T|U}$ any one of a whole family of nonnegative real-valued $\mathcal{A}$-measurable functions which agree with $T^*p/\{(U^*p)\circ\Pi\}$ a.e. μ on $\Pi^{-1}\{S(U^*p)\}$, that is

$$(3.4) \qquad p^{T|U}=T^*p/\{(U^*p)\circ\Pi\} \qquad \text{a.e. } \mu \text{ on } \Pi^{-1}\{S(U^*p)\}\,.$$

To stress this, we shall call $p^{T|U}$ a version of the conditional density of T given U. Note that any two versions are equal a.e. μ on Π^{-1} $\{S(U^*p)\}$. The following propositions show that $p^{T|U}$ has actually some properties of the conditional density. Here we shall denote by $E_P(\cdot|U)$ the conditional expectations taken with respect to the probability measure P given by (3.1) for each $p \in \mathcal{P}(\mathcal{X}, \mathcal{F}, m)$.

PROPOSITION 3.1. *For each $p \in \mathcal{P}(\mathcal{X}, \mathcal{F}, m)$, T^*p factorizes as follows.*

$$T^*p=\{(U^*p)\circ\Pi\}p^{T|U} \qquad a.e.\ \mu\,.$$

PROOF. From (3.4), it is obvious that the factorization holds on Π^{-1} $\{S(U^*p)\}$. Let us show that it also true on $\Pi^{-1}\{\mathcal{U}-S(U^*p)\}$. Since $T^*p\geqq 0$ a.e. μ, it follows from Lemmas 2.1 and 2.2. (i) that $\mathcal{T}-S(T^*p)$ $\supset \Pi^{-1}\{\mathcal{U}-S(U^*p)\}$ a.e. μ, and therefore $T^*p=0$ a.e. μ on $\Pi^{-1}\{\mathcal{U}-$ $S(U^*p)\}$. On the other hand, $\{(U^*p)\circ\Pi\}p^{T|U}=0$ on $\Pi^{-1}\{\mathcal{U}-S(U^*p)\}$. Thus the proof is completed.

PROPOSITION 3.2. *For each $p \in \mathcal{P}(\mathcal{X}, \mathcal{F}, m)$, let ϕ be a real-valued $\mathcal{A}$-measurable function whose expectation exists. Then*

$$E_P(\phi|U)=\Pi^*(\phi^+p^{T|U})-\Pi^*(\phi^-p^{T|U}) \qquad a.e.\ \nu \text{ on } S(U^*p)\,.$$

In particular, if $\phi p^{T|U} \in \mathcal{E}(\mathcal{T}, \mathcal{A}, \mu)$, then

$$E_P(\phi|U)=\Pi^*(\phi p^{T|U}) \qquad a.e.\ \nu \text{ on } S(U^*p)\,.$$

PROOF. Let us show that

$$E_P(\phi^+|U)=\Pi^*(\phi^+p^{T|U}) \qquad a.e.\ \nu \ on\ S(U^*p)\,.$$

This sufficies to prove the above assertions. First note that from Proposition 3.1, Lemma 2.4. (iv) and (3.2), it follows that for every $B\in\mathcal{B}$

$$\int_{\Pi^{-1}B}\phi^+dP^T=\int_{\Pi^{-1}B}\phi^+p^{T|U}\{(U^*p)\circ\Pi\}d\mu=\int_B\{\Pi^*(\phi^+p^{T|U})\}(U^*p)d\nu\,.$$

Next note that it follows from (3.3) that for every $B\in\mathcal{B}$

$$\int_{\Pi^{-1}B}\phi^+dP^T=\int_B\{E_P(\phi^+|U)\}(U^*p)d\nu\,.$$

Combining them, we have the desired result.

Now let us denote by $P^{T|U}$ the restriction of $E_P(\cdot|U)$ to the family $\{\chi_A: A\in\mathcal{A}\}$, and call it the *conditional probability function* of T given U; in other words, $P^{T|U}$ is a function on $\mathcal{A}$ whose values are $\mathcal{B}$-measurable functions $P^{T|U}(A)$ defined by $P^{T|U}(A)=E_P(\chi_A|U)$ a.e. ν. From Proposition 3.2, it follows that for each $A\in\mathcal{A}$,

$$P^{T|U}(A)=\Pi^*(\chi_A p^{T|U}) \qquad \text{a.e. } \nu \text{ on } S(U^*p)\,. \tag{3.5}$$

Thus we can regard $p^{T|U}$ as the density of $P^{T|U}$.

4. Factorization of densities of a model

In this section we shall use the same notations as in Definition 1.1. Let $\mathcal{P}_\Theta=\{P_\theta:\theta\in\Theta\}$ be a family of distributions of a random variables X over $(\mathcal{X},\mathcal{F})$, with densities p_θ relative to a σ-finite measure m. T is a statistic from $(\mathcal{X},\mathcal{F})$ into $(\mathcal{T},\mathcal{A})$, and U is a statistic from $(\mathcal{X},\mathcal{F})$ into $(\mathcal{U},\mathcal{B})$ such that $U=\Pi\circ T$. As before, let us denote by $\mathcal{P}_\Theta^T=\{P_\theta^T:\theta\in\Theta\}$ the family of marginal distributions of T, by $\mathcal{P}_\Theta^U=\{P_\theta^U:\theta\in\Theta\}$ the family of marginal distributions of U, and by $\mathcal{P}_\Theta^{T|U}=\{P_\theta^{T|U}:\theta\in\Theta\}$ the family of conditional probability functions of T given U. As shown in the previous section, $\mathcal{P}_\Theta^T$ has densities $f_\theta=T^*p_\theta$ relative to a σ-finite measure μ, $\mathcal{P}_\Theta^U$ has densities $g_\theta=U^*p_\theta$ relative to a σ-finite measure ν, and $\mathcal{P}_\Theta^{T|U}$ has densities $h_\theta=p_\theta^{T|U}$.

First we shall establish the following theorem.

THEOREM 4.1. *Suppose that $\mathcal{P}_\Theta\ll m$. Then $(U, T|U)$ induces $(\mathcal{P}_\Gamma, \mathcal{P}_\Omega)$ if and only if there exist nonnegative real-valued $\mathcal{B}$-measurable functions $\tilde{g}_{r(\theta)}$ and nonnegative real-valued $\mathcal{A}$-measurable functions $\tilde{h}_{w(\theta)}$ such that for every $\theta\in\Theta$*

$$f_\theta=(\tilde{g}_{r(\theta)}\circ\Pi)\tilde{h}_{w(\theta)} \qquad a.e.\ \mu\,, \tag{4.1}$$

(4.2) $$\tilde{g}_{r(\theta)}=g_\theta \qquad a.e.\ \nu\ ,$$

(4.3) $$\tilde{h}_{w(\theta)}=h_\theta \qquad a.e.\ \mu\ on\ \Pi^{-1}\{S(g_\theta)\}\ .$$

PROOF. It is obvious that the condition (i) of Definition 1.1 implies that g_θ depends on θ only through r. Hence g_θ is expressible as (4.2). Let us denote by $Q^{T|U}_{w(\theta)}$ the conditional probability functions of T given U satisfying the condition (ii) of Definition 1.1. Since $\mathcal{P}_\omega \ll m$ for every fixed $\omega \in \Omega$, it follows from Halmos and Savage ([7], Lemma 7) that there exists a probability measure over $(\mathcal{X}, \mathcal{F})$

(4.4) $$\Lambda_\omega=\sum_{i=1}^{\infty} c_i P_{\theta_i}\ , \qquad c_i>0 \text{ and } P_{\theta_i} \in \mathcal{P}_\omega\ ,$$

which is equivalent to $\mathcal{P}_\omega$ in the sense that for any $F \in \mathcal{F}$

$$\Lambda_\omega(F)=0 \Longleftrightarrow P_\theta(F)=0 \qquad \text{for every } P_\theta \in \mathcal{P}_\omega\ .$$

Putting $\lambda_\omega=d\Lambda_\omega/dm$, we obtain $d\Lambda^T_\omega/d\mu=T^*\lambda_\omega$ and $d\Lambda^U_\omega/d\nu=U^*\lambda_\omega$, where $\Lambda^T_\omega=\Lambda_\omega T^{-1}$ and $\Lambda^U_\omega=\Lambda_\omega U^{-1}$. Now we define $\tilde{h}_{w(\theta)}$ by

(4.5) $$\tilde{h}_{w(\theta)}=\begin{cases} T^*\lambda_{w(\theta)}/\{(U^*\lambda_{w(\theta)})\circ\Pi\} & \text{on } \Pi^{-1}\{S(U^*\lambda_{w(\theta)})\}\ , \\ 0 & \text{otherwise}\ . \end{cases}$$

We shall show that this function satisfies (4.3). Note that an argument similar to the proof of Proposition 3.1 yields $T^*\lambda_\omega=\{(U^*\lambda_\omega)\circ\Pi\}\tilde{h}_\omega$ a.e. μ. From this, we can easily see that for each $A \in \mathcal{A}$ and each $B \in \mathcal{B}$,

$$\int_{\Pi^{-1}B} \chi_A d\Lambda^T_\omega=\int_B \{\Pi^*(\chi_A\tilde{h}_\omega)\}(U^*\lambda_\omega)d\nu\ .$$

On the other hand, it follows from (4.4) that for each $A \in \mathcal{A}$ and each $B \in \mathcal{B}$,

$$\begin{aligned}\int_{\Pi^{-1}B} \chi_A d\Lambda^T_\omega &=\sum_{i=1}^{\infty} c_i \int_{\Pi^{-1}B} \chi_A dP^T_{\theta_i}=\sum_{i=1}^{\infty} c_i \int_B Q^{T|U}_\omega(A)dP^U_{\theta_i} \\ &=\int_B Q^{T|U}_\omega(A)d\Lambda^U_\omega=\int_B \{Q^{T|U}_\omega(A)\}(U^*\lambda_\omega)d\nu\ .\end{aligned}$$

From these two equalities, we have

$$Q^{T|U}_{w(\theta)}(A)=\Pi^*(\chi_A\tilde{h}_{w(\theta)}) \qquad \text{a.e. } \nu \text{ on } S(U^*\lambda_{w(\theta)})$$

for each $A \in \mathcal{A}$; on the other hand, from (3.5) we have

$$Q^{T|U}_{w(\theta)}(A)=\Pi^*(\chi_A h_\theta) \qquad \text{a.e. } \nu \text{ on } S(g_\theta)$$

for each $A \in \mathcal{A}$. Thus noting that $S(U^*\lambda_{w(\theta)})\supset S(g_\theta)$ a.e. ν because $P^U_\theta \ll$

$\Lambda^U_{w(\theta)} \ll \nu$, we get

$$\Pi^*(\chi_A \tilde{h}_{w(\theta)}) = \Pi^*(\chi_A h_\theta) \qquad \text{a.e. } \nu \text{ on } S(g_\theta)$$

for each $A \in \mathcal{A}$. From Lemma 2.3, it follows that $\tilde{h}_{w(\theta)}$ satisfies (4.3). The factorization (4.1) follows at once from (4.2), (4.3) and Proposition 3.1. Now it remains to establish the converse. However, it is trivially true.

Let us now suppose that the densities f_θ of $\mathcal{P}^T_\theta$ factorize as follows.

$$f_\theta = (\beta_{r(\theta)} \circ \Pi)\alpha_{w(\theta)} \qquad \text{a.e. } \mu \, , \tag{4.6}$$

where $\alpha_{w(\theta)}$ are nonnegative $\mathcal{A}$-measurable functions and $\beta_{r(\theta)}$ are nonnegative $\mathcal{B}$-measurable functions. Then the question arises whether $(U, T|U)$ induces $(\mathcal{P}_\Gamma, \mathcal{P}_\Omega)$ or not. Note that the Fisher-Neyman factorization criterion for sufficiency corresponds to the case where w is a constant function. Here we shall discuss this problem. Denote by $\mathcal{P}_{\Gamma\times\Omega}$ the partition of $\mathcal{P}_\Theta$ induced by r and w, i.e. $\mathcal{P}_{\Gamma\times\Omega} = \{\mathcal{P}_\gamma \cap \mathcal{P}_\omega : \gamma \in \Gamma, \ \omega \in \Omega\}$.

THEOREM 4.2. *Suppose that the densities f_θ of $\mathcal{P}^T_\theta$ factorize as* (4.6). *Then $(U, T|U)$ induces $(\mathcal{P}_{\Gamma\times\Omega}, \mathcal{P}_\Omega)$. If w is a function of r, then $(U, T|U)$ induces $(\mathcal{P}_\Gamma, \mathcal{P}_\Omega)$.*

PROOF. It follows from Lemmas 2.1 and 2.4. (iv) that

$$g_\theta = \beta_{r(\theta)} \Pi^* \alpha_{w(\theta)} \qquad \text{a.e. } \nu \, . \tag{4.7}$$

Thus the densities g_θ of P^U_θ depend on θ only through r and w; and therefore, they depend on θ only through r if w is a function of r. Next putting $D(\alpha_\omega) = \{t : 0 \leqq \alpha_\omega(t) < \infty\} \cap \Pi^{-1}\{u : 0 < \Pi^*\alpha_\omega(u) < \infty\}$, we define $\tilde{h}_{w(\theta)}$ by

$$\tilde{h}_{w(\theta)} = \begin{cases} \alpha_{w(\theta)} / \{(\Pi^*\alpha_{w(\theta)}) \circ \Pi\} & \text{on } D(\alpha_{w(\theta)}) \, , \\ 0 & \text{otherwise} \, . \end{cases}$$

Since $0 \leqq f_\theta < \infty$ a.e. μ and $0 \leqq g_\theta < \infty$ a.e. ν, it follows from (4.6) and (4.7) that

$$\Pi^{-1}\{S(g_\theta)\} \subset \Pi^{-1}\{u : 0 < \beta_{r(\theta)}(u) < \infty\} \subset \{t : 0 \leqq \alpha_{w(\theta)}(t) < \infty\} \qquad \text{a.e. } \mu \, ,$$

$$\Pi^{-1}\{S(g_\theta)\} \subset \Pi^{-1}\{u : 0 < \Pi^*\alpha_{w(\theta)}(u) < \infty\} \qquad \text{a.e. } \mu \, .$$

Hence $\Pi^{-1}\{S(g_\theta)\} \subset D(\alpha_{w(\theta)})$ a.e. μ. Thus we have

$$\tilde{h}_{w(\theta)} = f_\theta / (g_\theta \circ \Pi) \qquad \text{a.e. } \mu \text{ on } \Pi^{-1}\{S(g_\theta)\} \, .$$

This implies that for each $\theta \in \Theta$, $\tilde{h}_{w(\theta)}$ is a version of the density of $P^{T|U}_\theta$.

Thus the theorem is proved.

5. Smoothness of a family of conditional densities, differentiability of conditional expectation, and conditional efficacy

Let us further suppose that $\Theta=\Theta_1\times\cdots\times\Theta_n$, Θ_i being an open interval of R^1. Before discussing our problems, we shall introduce some notations.

Consider any two statistics U and T such that $U=\Pi\circ T$. As before, we write f_θ for T^*p_θ, and g_θ for U^*p_θ. For every $\tau, \theta\in\Theta$ define $\rho_f(\tau, \theta)$, the distance between P_τ^T and P_θ^T, and $\rho_g(\tau, \theta)$, the distance between P_τ^U and P_θ^U, by

$$\rho_f^2(\tau, \theta)=\int(\sqrt{f_\tau}-\sqrt{f_\theta})^2d\mu\ , \qquad \rho_g^2(\tau, \theta)=\int(\sqrt{g_\tau}-\sqrt{g_\theta})^2d\nu\ .$$

To fix ideas we assume that the densities h_θ of $\mathcal{P}_\theta^{T|U}$ are defined by

$$h_\theta=\begin{cases} f_\theta/(g_\theta\circ\Pi) & \text{on } \Pi^{-1}\{S(g_\theta)\}\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{5.1}$$

unless otherwise stated. Put for every $\tau, \theta\in\Theta$

$$\rho_h^{*2}(\tau, \theta)=\Pi^*(\sqrt{h_\tau}-\sqrt{h_\theta})^2\,.$$

Let us show that $\rho_h^*(\tau, \theta)$ is the distance between $P_\tau^{T|U}$ and $P_\theta^{T|U}$. It will suffice to discuss only the triangular inequality. First note that a nonnegative $\mathcal{B}$-measurable function ϕ exists such that $S(\phi)\supset S(g_\theta)$ a.e. ν for every $\theta\in\Theta$ and $\int\phi d\nu=1$; see Remark 5.1 below. Thus $\int(\phi\circ\Pi)h_\theta d\mu\leqq\int\phi d\nu=1$, and so $\int(\phi\circ\Pi)(\sqrt{h_\tau}-\sqrt{h_\theta})^2d\mu\leqq 2$, for every τ, $\theta\in\Theta$. Next using the Schwarz inequality with respect to Π^* (Lemma 2.4. (iii)), we have

$$\begin{aligned}\sqrt{\Pi^*\{(\phi\circ\Pi)(\sqrt{h_\tau}-\sqrt{h_\theta})^2\}}\leqq&\sqrt{\Pi^*\{(\phi\circ\Pi)(\sqrt{h_\tau}-\sqrt{h_\xi})^2\}}\\&+\sqrt{\Pi^*\{(\phi\circ\Pi)(\sqrt{h_\xi}-\sqrt{h_\theta})^2\}}\end{aligned}$$

a.e. ν for every $\theta, \tau, \xi\in\Theta$. Hence from Lemma 2.4. (iv) and (5.1)

$$\rho_h^*(\tau, \theta)\leqq\rho_h^*(\tau, \xi)+\rho_h^*(\xi, \theta) \qquad \text{a.e. } \nu \text{ for every } \theta, \tau, \xi\in\Theta\,.$$

Now define the mean difference $\rho_h(\tau, \theta\,;\theta)$ between $P_\tau^{T|U}$ and $P_\theta^{T|U}$ under P_θ^U by

$$\rho_h^2(\tau, \theta\,;\theta)=\int g_\theta\rho_h^{*2}(\tau, \theta)d\nu=\int(g_\theta\circ\Pi)(\sqrt{h_\tau}-\sqrt{h_\theta})^2d\mu$$

for every $\tau, \theta \in \Theta$. If $S(f_\tau)=S(f_\theta)$ a.e. μ, then we can easily see that $\rho_f^2(\tau, \theta)=E_\theta(\sqrt{f_\tau}/\sqrt{f_\theta}-1)^2$, $\rho_g^2(\tau, \theta)=E_\theta(\sqrt{g_\tau}/\sqrt{g_\theta}-1)^2$, and $\rho_h^2(\tau, \theta; \theta)= E_\theta(\sqrt{h_\tau}/\sqrt{h_\theta}-1)^2$, because $S(g_\tau)=S(g_\theta)$ a.e. ν; see Lemma 2.2. (ii).

Suppose for a moment that f_θ and g_θ are *loosely* differentiable with respect to θ, i.e. there exist functions $f_\theta'={}^t(f_{1\theta}', \cdots, f_{n\theta}')$ and $g_\theta'={}^t(g_{1\theta}', \cdots, g_{n\theta}')$ such that for every fixed $l={}^t(l_1, \cdots, l_n) \in R^n$

$$(5.2) \qquad (f_{\theta+\varepsilon l}-f_\theta)/\varepsilon \xrightarrow{l} {}^t l f_\theta' \quad \text{and} \quad (g_{\theta+\varepsilon l}-g_\theta)/\varepsilon \xrightarrow{l} {}^t l g_\theta' \qquad \text{as } \varepsilon \to 0\,,$$

at every $\theta \in \Theta$. Here we shall say that a sequence $\{\phi_n\}$ of real-valued functions converges *loosely* to ϕ, and write $\phi_n \xrightarrow{l} \phi$, if every subsequence of $\{\phi_n\}$ contains a subsequence which converges almost everywhere to ϕ; see Pitman ([11], p. 98). Now define $h_\theta'={}^t(h_{1\theta}', \cdots, h_{n\theta}')$ by

$$(5.3) \qquad h_{i\theta}'=\begin{cases} (f_{i\theta}' g_\theta \circ \Pi - f_\theta g_{i\theta}' \circ \Pi)/(g_\theta \circ \Pi)^2 & \text{on } \Pi^{-1}\{S(g_\theta)\}\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Then it follows from (3.4) and (5.2) that for every fixed $l \in R^n$

$$(h_{\theta+\varepsilon l}-h_\theta)/\varepsilon \xrightarrow{l} {}^t l h_\theta' \qquad \text{on } \Pi^{-1}\{S(g_\theta)\} \text{ as } \varepsilon \to 0\,,$$

at every $\theta \in \Theta$. Note that $h_\theta'=0$ a.e. μ on $\Pi^{-1}\{S(g_\theta)\}-S(f_\theta)$. Put

$$\dot{f}_\theta=\begin{cases} f_\theta'/\sqrt{f_\theta} & \text{on } S(f_\theta)\,, \\ 0 & \text{otherwise}\,; \end{cases}$$

$$\dot{g}_\theta=\begin{cases} g_\theta'/\sqrt{g_\theta} & \text{on } S(g_\theta)\,, \\ 0 & \text{otherwise}\,; \end{cases}$$

$$\dot{h}_\theta=\begin{cases} h_\theta'/\sqrt{h_\theta} & \text{on } S(f_\theta) \cap \Pi^{-1}\{S(g_\theta)\}\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Then the information matrices of the f_θ, g_θ and h_θ families are respectively given by

$$I_f(\theta)=\left[\int \dot{f}_\theta {}^t\dot{f}_\theta d\mu\right], \qquad I_g(\theta)=\left[\int \dot{g}_\theta {}^t\dot{g}_\theta d\nu\right],$$

$$I_h(\theta)=\left[\int (g_\theta \circ \Pi)(\dot{h}_\theta {}^t\dot{h}_\theta) d\mu\right].$$

Note that h_θ', $\dot{h}_\theta$ and $I_h(\theta)$ are independent of the particular densities h_θ given by (5.1); they are determined by f_θ and g_θ only. Denoting, as before, $\liminf_{\varepsilon\to 0} 2\rho_f(\theta+\varepsilon l, \theta)/|\varepsilon|$ by $s_f(\theta|l)$, $\liminf_{\varepsilon\to 0} 2\rho_g(\theta+\varepsilon l, \theta)/|\varepsilon|$ by $s_g(\theta|l)$, and $\liminf_{\varepsilon\to 0} 2\rho_h(\theta+\varepsilon l, \theta; \theta)/|\varepsilon|$ by $s_h(\theta|l)$, we shall briefly call $s_f^2(\theta|l)$ the sensitivity of the f_θ family, $s_g^2(\theta|l)$ the sensitivity of the g_θ family, and $s_h^2(\theta|l)$ the sensitivity of the h_θ family.

First we shall refer to Pitman's [11] results. If the p_θ family is smooth in Θ (Definition 1.2), then the family is differentiable in mean at every $\theta \in \Theta$, i.e. for every fixed $l \in R^n$,

$$\lim_{\varepsilon\to 0} \int |(p_{\theta+\varepsilon l}-p_\theta)/\varepsilon - {}^t l p'_\theta| dm = 0 ,$$

at every $\theta \in \Theta$. From this, it follows that for every fixed $l \in R^n$,

$$\text{(5.4)} \qquad \begin{aligned} &\lim_{\varepsilon\to 0} \int |(f_{\theta+\varepsilon l}-f_\theta)/\varepsilon - {}^t l T^* p'_\theta| d\mu = 0 , \\ &\lim_{\varepsilon\to 0} \int |(g_{\theta+\varepsilon l}-g_\theta)/\varepsilon - {}^t l U^* p'_\theta| d\nu = 0 , \end{aligned}$$

at every $\theta \in \Theta$. These imply that f_θ and g_θ are loosely differentiable with respect to θ, and that $f'_\theta = T^* p'_\theta$ and $g'_\theta = U^* p'_\theta$. Therefore when the p_θ family is smooth in Θ, $I_f(\theta)$, $I_g(\theta)$ and $I_h(\theta)$ are obtained from the substitution $f'_\theta = T^* p'_\theta$, $g'_\theta = U^* p'_\theta$. Further, then the following holds: for every fixed $l \in R^n$,

$$\text{(5.5)} \qquad \begin{aligned} &\lim_{\varepsilon\to 0} \int |4(\sqrt{f_{\theta+\varepsilon l}} - \sqrt{f_\theta})^2/\varepsilon^2 - ({}^t l \dot{f}_\theta)^2| d\mu = 0 , \\ &\lim_{\varepsilon\to 0} \int |4(\sqrt{g_{\theta+\varepsilon l}} - \sqrt{g_\theta})^2/\varepsilon^2 - ({}^t l \dot{g}_\theta)^2| d\nu = 0 , \end{aligned}$$

at every $\theta \in \Theta$. Thus noting that $\rho_p^2(\tau, \theta) \geqq \rho_f^2(\tau, \theta) \geqq \rho_g^2(\tau, \theta)$, we have the following theorem.

THEOREM 5.1 (Pitman [11], Theorem, pp. 19–20). *If the p_θ family is smooth in Θ, so are the f_θ family and the g_θ family, i.e. for every fixed $l \in R^n$,*

$$\lim_{\varepsilon\to 0} 4\rho_f^2(\theta+\varepsilon l, \theta)/\varepsilon^2 = {}^t l I_f(\theta) l , \qquad \lim_{\varepsilon\to 0} 4\rho_g^2(\theta+\varepsilon l, \theta)/\varepsilon^2 = {}^t l I_g(\theta) l ,$$

at every $\theta \in \Theta$; and the sensitivity matrices $I_p(\theta)$, $I_f(\theta)$ and $I_g(\theta)$ satisfy

$${}^t l I_p(\theta) l \geqq {}^t l I_f(\theta) l \geqq {}^t l I_g(\theta) l$$

for every $\theta \in \Theta$ and every $l \in R^n$.

Let φ be a real-valued $\mathcal{A}$-measurable statistic with a second moment for every $\theta \in \Theta$. The differntiability of the expectation $E_\theta(\varphi)$ with respect to θ is given by the following theorem.

THEOREM 5.2 (Pitman [11], Theorem I, p.31). *Suppose that the p_θ family is smooth in Θ. If φ has a second moment which is bounded* *in some neighbourhood of every fixed* $\theta \in \Theta$, *then for every fixed* $l \in R^n$,

$$\lim_{\varepsilon\to 0} \{E_{\theta+\varepsilon l}(\psi)-E_\theta(\psi)\}/\varepsilon = {}^t l \int \psi f'_\theta d\mu \qquad \textit{at every } \theta \in \Theta .$$

Let us turn to our problems. When the p_θ family is smooth in Θ, noting that $g_\theta = \Pi^* f_\theta$ and $g'_\theta = \Pi^* f'_\theta$, and that $f'_\theta = 0$ a.e. μ on $\Pi^{-1}\{S(g_\theta)\} - S(f_\theta)$, we have

$$I_h(\theta) = I_f(\theta) - I_g(\theta) \qquad \text{for every } \theta \in \Theta . \tag{5.6}$$

Then we shall prove the following theorem.

THEOREM 5.3. *Suppose that the* h_θ *family satisfies* (5.1). *If the* p_θ *family is smooth in* Θ, *then for every fixed* $l \in R^n$,

$$\lim_{\varepsilon\to 0} \int (g_\theta \circ \Pi) |(h_{\theta+\varepsilon l} - h_\theta)/\varepsilon - {}^t l h'_\theta| d\mu = 0 , \tag{5.7}$$

$$\lim_{\varepsilon\to 0} \int (g_\theta \circ \Pi) |4(\sqrt{h_{\theta+\varepsilon l}} - \sqrt{h_\theta})^2/\varepsilon^2 - ({}^t l \dot{h}_\theta)^2| d\mu = 0 , \tag{5.8}$$

at every $\theta \in \Theta$; *and therefore for every fixed* $l \in R^n$,

$$\lim_{\varepsilon\to 0} 4\rho_h^2(\theta+\varepsilon l, \theta; \theta)/\varepsilon^2 = {}^t l I_h(\theta) l \qquad \textit{at every } \theta \in \Theta .$$

PROOF. The above assertions follow at once from the argument given in the proof of Theorem 3.3 in Kuboki [8], though the h_θ family does not sometimes satisfy the assumption of the theorem, if we have only to show that

$$\lim_{\varepsilon\to 0} \int_{S(g_\theta)\cap S(g_{\theta+\varepsilon l})} |(g_{\theta+\varepsilon l} - g_\theta)/\varepsilon| d\nu = \int_{S(g_\theta)} |{}^t l g'_\theta| d\nu ,$$

$$\lim_{\varepsilon\to 0} \int_{S(g_\theta)\cap S(g_{\theta+\varepsilon l})} 4(\sqrt{g_{\theta+\varepsilon l}} - \sqrt{g_\theta})^2/\varepsilon^2 d\nu = \int_{S(g_\theta)} ({}^t l \dot{g}_\theta)^2 d\nu .$$

We shall prove the former; the latter is proved similarly from (5.5). Note that

$$\left| \int_{S(g_\theta)\cap S(g_{\theta+\varepsilon l})} |(g_{\theta+\varepsilon l} - g_\theta)/\varepsilon| d\nu - \int_{S(g_\theta)} |{}^t l g'_\theta| d\nu \right|$$
$$\leqq \int |(g_{\theta+\varepsilon l} - g_\theta)/\varepsilon - {}^t l g'_\theta| d\nu + \left| \int \chi_{S(g_\theta)\cap S(g_{\theta+\varepsilon l})} |{}^t l g'_\theta| d\nu - \int \chi_{S(g_\theta)} |{}^t l g'_\theta| d\nu \right| .$$

The first term on the right $\to 0$ as $\varepsilon \to 0$ by (5.4). Next note that for every fixed l, $\chi_{S(g_\theta)\cap S(g_{\theta+\varepsilon l})} \xrightarrow{l} \chi_{S(g_\theta)}$ as $\varepsilon \to 0$ because $g_{\theta+\varepsilon l} \xrightarrow{l} g_\theta$ as $\varepsilon \to 0$. Thus it follows from the dominated convergence theorem with respect to loose convergence (Pitman [11], Theorem, p. 99) that the second term $\to 0$ as $\varepsilon \to 0$, as was to be proved.

Remark 5.1. We shall show that for each $\theta\in\Theta$, there exist a common $\mathcal{B}$-measurable set S and a version h_θ of the density of $P_\theta^{T|U}$ such that

$$(5.9)\qquad \begin{gathered} S\supset S(g_\theta) \text{ a.e. } \nu\,, \\ h_\theta=0 \text{ a.e. } \mu \text{ on } \Pi^{-1}\{\mathcal{U}-S\} \quad\text{and}\quad \Pi^* h_\theta=1 \text{ a.e. } \nu \text{ on } S\,. \end{gathered}$$

This h_θ family satisfies the assumption of Theorem 3.3 in Kuboki [8], and therefore it satisfies (5.7) and (5.8). To see the above assertion, note that there exists a probability measure Λ over $(\mathcal{X}, \mathcal{F})$ such that for any $F\in\mathcal{F}$

$$\Lambda(F)=0 \Longleftrightarrow P_\theta(F)=0 \qquad \text{for every } \theta\in\Theta\,.$$

Putting $\lambda=d\Lambda/dm$, we get $d\Lambda^T/d\mu=T^*\lambda$ and $d\Lambda^U/d\nu=U^*\lambda$. Define $S=S(U^*\lambda)$. Since $P_\theta^U\ll\Lambda^U\ll\nu$, we can easily see that $S\supset S(g_\theta)$ a.e. ν. Further define

$$h_\theta=\begin{cases} f_\theta/(g_\theta\circ\Pi) & \text{on } \Pi^{-1}\{S(g_\theta)\}\,, \\ T^*\lambda/\{(U^*\lambda)\circ\Pi\} & \text{on } \Pi^{-1}\{S-S(g_\theta)\}\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

which satisfies (5.9).

Remark 5.2. Suppose that the p_θ family is smooth in Θ. Then it follows from (5.4) that

$$(5.10)\qquad \int f_\theta' d\mu=0 \quad\text{and}\quad \int g_\theta' d\nu=0\,, \qquad \text{at every } \theta\in\Theta\,.$$

Here we shall show that the following is also true:

$$(5.11)\qquad \Pi^*\{(g_\theta\circ\Pi)h_\theta'\}=0 \qquad \text{a.e. } \nu \text{ at every } \theta\in\Theta\,.$$

Let h_θ satisfy (5.9). Since $(g_\theta\circ\Pi)h_{\theta+\varepsilon l}$ and ${}^t l(g_\theta\circ\Pi)h_\theta'$ is μ-integrable for every $l\in R^n$, it follows from (5.9) and Lemma 2.4 that for every $l\in R^n$

$$\int (g_\theta\circ\Pi)|(h_{\theta+\varepsilon l}-h_\theta)/\varepsilon-{}^t l h_\theta'|d\mu \geqq \int |{}^t l\Pi^*\{(g_\theta\circ\Pi)h_\theta'\}|d\nu\,.$$

Letting $\varepsilon\to 0$, we see that (5.11) follows from (5.7).

Next let us discuss the differentiability of conditional expectations. From now on, we shall always assign 0 to the value of $E_\theta(\cdot|U)$ on a P_θ^U-null set $\mathcal{U}-S(g_\theta)$, i.e.

$$(5.12)\qquad E_\theta(\cdot|U)=0 \qquad \text{on } \mathcal{U}-S(g_\theta) \text{ for every } \theta\in\Theta\,.$$

Combining this with (5.1), we get from Proposition 3.2

$$E_\theta(\phi|U)=\Pi^*(\phi^+h_\theta)-\Pi^*(\phi^-h_\theta) \qquad \text{a.e. } \nu \text{ for every } \theta\in\Theta\,, \tag{5.13}$$

where ϕ is a real-valued $\mathcal{A}$-measurable function whose expectations exist.

Consider a real-valued $\mathcal{A}$-measurable statistic ψ such that $E_\theta\{E_\tau(\psi^2|U)\}<\infty$ for every τ in some neighbourhood of every fixed $\theta\in\Theta$. Then from (5.13)

$$E_\theta\{E_\tau(\psi^2|U)\}=\int(g_\theta\circ\Pi)\psi^2h_\tau d\mu\ .$$

Hence $(g_\theta\circ\Pi)\psi h_\tau$ is μ-integrable, and therefore from (5.13)

$$\int\Pi^*\{(g_\theta\circ\Pi)\psi h_\tau\}d\nu=\int g_\theta\{\Pi^*(\psi^+h_\tau)-\Pi^*(\psi^-h_\tau)\}d\nu=\int g_\theta E_\tau(\psi|U)d\nu\ .$$

THEOREM 5.4. *Suppose that the p_θ family is smooth in Θ, and that ψ is a real-valued $\mathcal{A}$-measurable statistic with a finite second moment for every $\theta\in\Theta$. Under the assumptions* (5.1) *and* (5.12), *if $E_\theta\{E_\tau(\psi^2|U)\}$ is bounded in some neighbourhood of every fixed $\theta\in\Theta$, then for every fixed $l\in R^n$*

$$\lim_{\varepsilon\to0}\int|g_\theta\{E_{\theta+\varepsilon l}(\psi|U)-E_\theta(\psi|U)\}/\varepsilon-{}^tl\Pi^*\{(g_\theta\circ\Pi)\psi h'_\theta\}|d\nu=0\ ,$$

at every $\theta\in\Theta$; and therefore for every $l\in R^n$

$$\{E_{\theta+\varepsilon l}(\psi|U)-E_\theta(\psi|U)\}/\varepsilon\xrightarrow{\ l\ }{}^tl\Pi^*\{(g_\theta\circ\Pi)\psi h'_\theta\}/g_\theta \qquad on\ S(g_\theta)\,,$$

at every $\theta\in\Theta$.

PROOF. For every fixed $\theta\in\Theta$, let $U(\theta)$ be a neighbourhood of θ, and suppose that $\int(g_\theta\circ\Pi)\psi^2h_{\theta+\varepsilon l}d\mu\leqq K_\theta$ provided $\theta+\varepsilon l\in U(\theta)$. Note that $(g_\theta\circ\Pi)\psi\,{}^tlh'_\theta$ is μ-integrable for every $l\in R^n$ because

$$\begin{aligned}(g_\theta\circ\Pi)|\psi\,{}^tlh'_\theta|&=(|\psi|\sqrt{f_\theta})\{\sqrt{g_\theta\circ\Pi}\,|{}^tl\dot{h}_\theta|\}\\&\leqq\psi^2f_\theta+(g_\theta\circ\Pi)({}^tl\dot{h}_\theta)^2 \qquad \text{a.e. } \nu\ .\end{aligned} \tag{5.14}$$

Thus from Lemma 2.4, it follows that for every $\theta+\varepsilon l\in U(\theta)$

$$\begin{aligned}&\int|g_\theta[\{E_{\theta+\varepsilon l}(\psi|U)-E_\theta(\psi|U)\}/\varepsilon]-{}^tl\Pi^*\{(g_\theta\circ\Pi)\psi h'_\theta\}|d\nu\\&\leqq\int(g_\theta\circ\Pi)|\psi(h_{\theta+\varepsilon l}-h_\theta)/\varepsilon-\psi\,{}^tlh'_\theta|d\mu\\&\leqq\int_{|\psi|\leqq c}(g_\theta\circ\Pi)|\psi(h_{\theta+\varepsilon l}-h_\theta)/\varepsilon-\psi\,{}^tlh'_\theta|d\mu\end{aligned}$$

$$+\int_{|\phi|>c}(g_\theta\circ\Pi)|\phi(h_{\theta+\varepsilon l}-h_\theta)/\varepsilon|d\mu+\int_{|\phi|>c}(g_\theta\circ\Pi)|\phi\,{}^t l h_\theta'|d\mu\ .$$

From (5.7), it follows that the first term on the right $\to 0$ as $\varepsilon\to 0$. Denoting the second term by $\alpha(\varepsilon, c)$, and using the Schwarz inequality, we have

$$\alpha^2(\varepsilon, c)=\left[\int_{|\phi|>c}|\phi|\sqrt{g_\theta\circ\Pi}\,(\sqrt{h_{\theta+\varepsilon l}}+\sqrt{h_\theta})\sqrt{g_\theta\circ\Pi}\,|(\sqrt{h_{\theta+\varepsilon l}}-\sqrt{h_\theta})/\varepsilon|d\mu\right]^2$$
$$\leqq 2\int\phi^2(g_\theta\circ\Pi)(h_{\theta+\varepsilon l}+h_\theta)d\mu\int_{|\phi|>c}(g_\theta\circ\Pi)(\sqrt{h_{\theta+\varepsilon l}}-\sqrt{h_\theta})^2/\varepsilon^2 d\mu$$
$$\leqq 4K_\theta\int_{|\phi|>c}(g_\theta\circ\Pi)(\sqrt{h_{\theta+\varepsilon l}}-\sqrt{h_\theta})^2/\varepsilon^2 d\mu\ ,$$

which $\to K_\theta\int_{|\phi|>c}(g_\theta\circ\Pi)({}^t l\dot{h}_\theta)^2 d\mu$ as $\varepsilon\to 0$ because of (5.8). Hence $\lim_{c\to\infty}\lim_{\varepsilon\to 0}\alpha(\varepsilon, c)=0$. Further it follows from (5.14) that the last term on the right $\to 0$ as $c\to\infty$. Thus the theorem is proved.

Finally, in connection with Theorems 5.3 and 5.4, we shall propose a measure of the effectiveness of a statistic in investigating the value of θ conditionally on U, and discuss its property. Before starting this, let us refer to Pitman's ([11], p. 30) definition of the efficacy (matrix) of a statistic.

Let ϕ be a real-valued $\mathcal{A}$-measurable statistic which has a finite second moment for every $\theta\in\Theta$. Suppose that $E_\theta(\phi)$ is differentiable with respect to θ, i.e. there exists $\eta(\theta;\phi)={}^t(\eta_1(\theta;\phi),\cdots,\eta_n(\theta;\phi))$ such that for every fixed $l\in R^n$, $\{E_{\theta+\varepsilon l}(\phi)-E_\theta(\phi)\}/\varepsilon\to{}^t l\eta(\theta;\phi)$ as $\varepsilon\to 0$ at every $\theta\in\Theta$. Define

$$J(\theta;\phi)=\eta(\theta;\phi)\,{}^t\eta(\theta;\phi)/V_\theta(\phi)\ ,$$

with the convention that $0/0=0$, where V denotes variance. We shall call it the *efficacy matrix* of ϕ at θ. It is a measure of the effectiveness of ϕ in estimating θ or in testing a hypothesis about θ. The value $\{{}^t l\eta(\theta;\phi)\}^2$ is an indication of the sensitivity of the distribution of ϕ to small changes in θ in the direction l. We want this to be large. The denominator $V_\theta(\phi)$ indicates the liability of ϕ to vary from observation to observation. We want this to be small. Thus high efficacy is a desirable characteristic of a statistic used in investigating the value of θ. Pitman said that ϕ is *regular* at $\theta\in\Theta$ if $E_\theta(\phi)$ has a derivative given by $\eta(\theta;\phi)=\int\phi f_\theta' d\mu$ at θ. If the p_θ family is smooth in Θ and if ϕ is regular at every $\theta\in\Theta$, then

$$\eta(\theta;\phi)=\int\{\phi-E_\theta(\phi)\}f_\theta' d\mu$$

because of (5.10). Applying the Schwarz inequality to this, we get

(5.15) $\quad {}^t lJ(\theta;\psi)l \leqq {}^t lI_f(\theta)l \qquad$ for every $\theta\in\Theta$ and every $l\in R^n$.

In particular, if $\psi=\phi\circ\Pi$, ϕ being $\mathcal{B}$-measurable, then

(5.16) $\quad {}^t lJ(\theta;\psi)l \leqq {}^t lI_g(\theta)l \qquad$ for every $\theta\in\Theta$ and every $l\in R^n$,

because $\eta(\theta;\psi)=\int\psi f_\theta' d\mu=\int\phi g_\theta' d\nu$. Theorem 5.2 gives a regularity condition on ψ.

Taking account of the above discussion, let us turn to our problem. Suppose that $E_\theta(\psi|U)$ is loosely differentiable with respect to θ on $S(g_\theta)$, i.e. there exists $\eta^*(\theta;\psi)={}^t(\eta_1^*(\theta;\psi),\cdots,\eta_n^*(\theta;\psi))$ such that for every fixed $l\in R^n$,

$$\{E_{\theta+\varepsilon l}(\psi|U)-E_\theta(\psi|U)\}/\varepsilon \xrightarrow{l} {}^t l\eta^*(\theta;\psi) \qquad \text{on } S(g_\theta),$$

as $\varepsilon\to 0$ at every $\theta\in\Theta$. Put $\eta^*(\theta;\psi)=0$ on $\mathcal{U}-S(g_\theta)$. Then we shall define the *conditional efficacy matrix* $J^*(\theta;\psi)$ of ψ given U at $\theta\in\Theta$ by

$$J^*(\theta;\psi)=\eta^*(\theta;\psi)\,{}^t\eta^*(\theta;\psi)/V_\theta(\psi|U),$$

with the convention that $0/0=0$, where $V(\cdot|U)$ denotes conditional variance given U. We shall say that ψ is *conditionally regular* at $\theta\in\Theta$ if $E_\theta(\psi|U)$ has a loose derivative given by $\eta^*(\theta;\psi)=\Pi^*\{(g_\theta\circ\Pi)\psi h_\theta'\}/g_\theta$ on $S(g_\theta)$ at θ. If the p_θ family is smooth in Θ, and if ψ is conditionally regular at every $\theta\in\Theta$, then

$$\eta^*(\theta;\psi)=\Pi^*[(g_\theta\circ\Pi)\{\psi-E_\theta(\psi|U)\circ\Pi\}h_\theta']/g_\theta \qquad \text{a.e. } \nu \text{ on } S(g_\theta).$$

To see this, note that as proved in (5.14), $(g_\theta\circ\Pi)\psi h_\theta'$ is μ-integrable, and therefore so is $(g_\theta\circ\Pi)\{E_\theta(\psi|U)\circ\Pi\}h_\theta'$ because $\{E_\theta(\psi|U)\}^2\leqq E_\theta(\psi^2|U)$ a.e. P_θ^U. Then the above equality follows at once from Lemma 2.4 and (5.11). Here using the Schwarz inequality with respect to Π^* (Lemma 2.4. (iii)), we have

$$\{{}^t l\eta^*(\theta;\psi)\}^2\leqq V_\theta(\psi|U)\Pi^*\{({}^t l\dot{h}_\theta)^2\} \qquad \text{a.e. } \nu \text{ on } S(g_\theta)$$

for every $\theta\in\Theta$ and every $l\in R^n$, where $V_\theta(\psi|U)=\Pi^*[\{\psi-E_\theta(\psi|U)\circ\Pi\}^2h_\theta]$ a.e. ν; recall (5.1) and (5.12). This inequality is also true on $\mathcal{U}-S(g_\theta)$. Thus for every $\theta\in\Theta$ and every $l\in R^n$,

(5.17) $\quad {}^t lJ^*(\theta;\psi)l\leqq\Pi^*\{({}^t l\dot{h}_\theta)^2\}$ a.e. ν, $\qquad {}^t lE_\theta\{J^*(\theta;\psi)\}l\leqq {}^t lI_h(\theta)l$.

If ψ satisfies the condition in Theorem 5.4, then it is conditionally regular at every $\theta\in\Theta$.

6. Discussion : sensitivity, efficacy, and other measures of information

Throughout this section, we use the same notations as in the previous section. Further we define the family of conditional densities by (5.1), and define conditional expectations by (5.12). Let us now suppose that the inference problem at hand relates only to $\theta^{(1)}={}^t(\theta_1,\cdots,\theta_k)$, $k<n$.

6.1. *Partial sensitivity and Liang's measure of information*

To simplify notation, we shall here represent p_θ, f_θ, g_θ and h_θ by a common symbol q_θ, and their derivatives or loose derivatives with respect to θ by a common symbol q'_θ; see (5.2) and (5.3) for loose derivatives. Further we shall denote by $s_q^2(\theta|l)$ the sensitivity of the q_θ family, and by $I_q(\theta)$ the information matrix of the q_θ family. In our problem, we may ignore that part of $s_q^2(\theta|l)$ which depends on small changes in $\theta^{(2)}={}^t(\theta_{k+1},\cdots,\theta_n)$, i.e. which depends on the direction $l^{(2)}={}^t(l_{k+1},\cdots,l_n)$, because information gained on $\theta^{(2)}$ is of no direct relevance to the problem. From this consideration we shall define the *partial sensitivity* $s_q^2(\theta;\theta^{(1)}|c)$ of the q_θ family with respect to $\theta^{(1)}$ (at $\theta\in\Theta$ in the direction $c\in R^k$) by

$$s_q^2(\theta;\theta^{(1)}|c)=\inf_{l\in L_c} s_q^2(\theta|l)\ ,$$

where $L_c=\{l\in R^n: l^{(1)}=c\}$ for every $c\in R^k$; $l^{(1)}={}^t(l_1,\cdots,l_k)$. In general, by using Fatou's lemma with respect to loose convergence (Pitman [11], Lemma, p. 98), it follows that

$$s_q^2(\theta;\theta^{(1)}|c)\geqq\inf_{l\in L_c} {}^tlI_q(\theta)l$$

for every $\theta\in\Theta$ and every $c\in R^k$. Note that the term on the right is a generalization of Liang's [9] measure of information about θ_1 in the q_θ family.

Suppose that all elements of $I_q(\theta)$ are finite for every $\theta\in\Theta$. Partition $I_q(\theta)$ as

$$I_q(\theta)=\begin{bmatrix} I_{11q}(\theta) & I_{12q}(\theta)\\ I_{21q}(\theta) & I_{22q}(\theta)\end{bmatrix}\ ,$$

where $I_{11q}(\theta)$ is of order $k\times k$, and put

$$I_q(\theta;\theta^{(1)})=I_{11q}(\theta)-I_{12q}(\theta)\{I_{22q}(\theta)\}^+I_{21q}(\theta)\ ,$$

where M^+ denotes the Moore-Penrose inverse of a matrix M. Then we have

$$\inf_{l\in L_c} {}^tlI_q(\theta)l={}^tcI_q(\theta;\theta^{(1)})c \tag{6.1}$$

for every $\theta \in \Theta$ and every $c \in R^k$; this is a direct consequence of a well known result on the extreme values of quadratic forms (Rao [12], p. 61). Further we can show that $I_q(\theta; \theta^{(1)})=0$ for every $\theta \in \Theta$ if and only if

$$(6.2) \qquad {}^t(q'_{1\theta}, \cdots, q'_{k\theta}) = A(\theta)\, {}^t(q'_{k+1\theta}, \cdots, q'_{n\theta}) \qquad \text{a.e. at every } \theta \in \Theta ,$$

where $A(\theta)$ is a matrix of order $k \times (n-k)$ depending on only θ.

Combining the above discussion with Theorems 5.1 and 5.3, we see that if the p_θ family is smooth in Θ, then for every $\theta \in \Theta$ and every $c \in R^k$

$$s_q^2(\theta; \theta^{(1)} | c) = {}^t c I_q(\theta; \theta^{(1)}) c .$$

For this reason, when the p_θ family is smooth in Θ, we shall call $I_q(\theta; \theta^{(1)})$ the *partial sensitivity matrix* of the q_θ family with respect to $\theta^{(1)}$ at $\theta \in \Theta$.

6.2. *Efficacy and Godambe's measure of information*

Here we shall discuss the case where $\theta^{(1)}=\theta_1$. Let T be sufficient for $\theta \in \Theta$. Consider a real-valued function $\psi = \psi_{\theta_1}(t)$ on $\Theta_1 \times \mathcal{T}$ such that $\psi_{\theta_1}(\in)\mathcal{A}$ for every fixed $\theta_1 \in \Theta_1$. An estimate of θ_1 is obtained by solving the equation $\psi_{\theta_1}(t)=0$ for θ_1. From this, ψ is called an estimating function. If $E_\theta(\psi_{\theta_1})=0$ for every $\theta \in \Theta$, then ψ is called unbiased (Godambe [4]). We shall say that ψ is conditionally unbiased if $E_\theta(\psi_{\theta_1}|U) =0$ a.e. P_θ^U for every $\theta \in \Theta$.

Throughout this subsection, let us suppose that the p_θ family is smooth in Θ, and that the supports $S(p_\theta)$ are independent of $\theta \in \Theta$. Therefore from Lemma 2.2. (ii), an $\mathcal{A}$-measurable set A exists such that for each $\theta \in \Theta$, $A=S(f_\theta)$ a.e. μ; and a $\mathcal{B}$-measurable set B exists such that for each $\theta \in \Theta$, $B=S(g_\theta)$ a.e. ν. To fix the idea, for each $\theta \in \Theta$, we take f_θ and g_θ so that they satisfy $A=S(f_\theta)$ and $B=S(g_\theta)$, respectively. For estimating functions ψ under consideration, we assume that $E_\tau(\psi_{\theta_1}^2)<\infty$ for every τ in some neighbourhood of every fixed $\theta \in \Theta$. We shall say that an estimating function ψ is *regular* or *conditionally regular* if for every fixed $\theta \in \Theta$, the statistic ψ_{θ_1} is regular or conditionally regular at θ, respectively. For a regular estimating function ψ, we shall define the *efficacy matrix* $J_\psi(\theta)$ of ψ at $\theta \in \Theta$ by $J(\theta; \psi_{\theta_1})$, the efficacy matrix of the statistic ψ_{θ_1} at θ; similarly, for a conditionally regular estimating function ψ, we shall define the *conditional efficacy matrix* $J_\psi^*(\theta)$ of ψ at $\theta \in \Theta$ by $J^*(\theta; \psi_{\theta_1})$. Naturally, one would prefer an estimating function with high efficacy in a certain sense.

Now let us denote by C the class of all regular and unbiased estimating functions; by C^* that subclass of C each member of which is conditionally regular and conditionally unbiased; and by C° the subclass consisting of all $\psi \in C$ such that $\psi_{\theta_1}=\phi_{\theta_1} \circ \Pi$, ϕ_{θ_1} being $\mathscr{B}$-measurable. Further, let us denote by $J_\psi(\theta;\theta_1)$ the (1, 1) entry of $J_\psi(\theta)$, and by $J_\psi^*(\theta;\theta_1)$ the (1, 1) entry of $J_\psi^*(\theta)$. Note that if $\psi \in C$, then all elements of $J_\psi(\theta)$ vanish except $J_\psi(\theta;\theta_1)$, because $E_{\theta+\epsilon l}(\psi_{\theta_1})-E_\theta(\psi_{\theta_1})=0$ for every l of the form $l={}^t(0, l_2, \cdots, l_n)$. Similarly, if $\psi \in C^*$, then all elements of $J_\psi^*(\theta)$ vanish except $J_\psi^*(\theta;\theta_1)$. Define

$$\mathscr{I}_f(\theta;\theta_1)=\sup_{\psi \in C} J_\psi(\theta;\theta_1)\ , \qquad \mathscr{I}_g(\theta;\theta_1)=\sup_{\psi \in C^\circ} J_\psi(\theta;\theta_1)\ ,$$
$$\mathscr{I}_h(\theta;\theta_1)=\sup_{\psi \in C^*} E_\theta\{J_\psi^*(\theta;\theta_1)\}\ . \tag{6.3}$$

Then combining (5.15)–(5.17) with (6.1), and taking account of the above observations, we obtain

$$\mathscr{I}_f(\theta;\theta_1)\leqq I_f(\theta;\theta_1)\ , \qquad \mathscr{I}_g(\theta;\theta_1)\leqq I_g(\theta;\theta_1)\ ,$$
$$\mathscr{I}_h(\theta;\theta_1)\leqq I_h(\theta;\theta_1)\ . \tag{6.4}$$

The quantities $\mathscr{I}_f(\theta;\theta_1)$, $\mathscr{I}_g(\theta;\theta_1)$ and $\mathscr{I}_h(\theta;\theta_1)$ are essentially the same as Godambe's [5] measures of information about θ_1 in the respective families, although his regularity conditions on estimating functions and on families of densities (Godambe [4]) are different from ours. For example, Godambe ([4], Theorem 2.2) has shown that the estimating function for which $J_\psi(\theta;\theta_1)$ attains its supremum in C can be often obtained from the efficient score. The following is a corresponding result of ours. Let us define $\bar{f}_\theta$, $\bar{g}_\theta$ and $\bar{h}_\theta$ by

$$\bar{f}_\theta=\begin{cases} f'_{1\theta}/f_\theta & \text{on } A\ , \\ (g'_{1\theta}\circ \Pi)/(g_\theta \circ \Pi) & \text{on } \Pi^{-1}B-A\ , \\ 0 & \text{otherwise}\ ; \end{cases}$$

$$\bar{g}_\theta=\begin{cases} g'_{1\theta}/g_\theta & \text{on } B\ , \\ 0 & \text{otherwise}\ ; \end{cases}$$

$$\bar{h}_\theta=\begin{cases} h'_{1\theta}/h_\theta & \text{on } A\cap \Pi^{-1}B\ , \\ 0 & \text{otherwise}\ . \end{cases}$$

These are essentially equal to $\partial \log f_\theta/\partial\theta_1$, $\partial \log g_\theta/\partial\theta_1$ and $\partial \log h_\theta/\partial\theta_1$, respectively, provided they exist. It follows from (5.3) that $\bar{f}_\theta=\bar{g}_\theta\circ \Pi +\bar{h}_\theta$ a.e. μ. Now put

$$\phi=\phi_\theta(t)=a(\theta)\bar{f}_\theta(t)+b_\theta(t)\ , \tag{6.5}$$

where $a(\theta)$ is a real-valued function on Θ, and $b_\theta(t)$ is a real-valued function on $\boldsymbol{\Theta}\times\mathcal{T}$ such that $E_\theta(b_\theta\phi_{\theta_1})=0$ for every $\phi\in C$ and for every $\theta\in\boldsymbol{\Theta}$. It is easy to prove that if $\phi\in C$, then $J_\phi(\theta;\theta_1)=\mathcal{J}_f(\theta;\theta_1)$. In fact, examining the proof of Theorem 2.2 in Godambe [4], we see that the assertion follows at once from the obvious relation:

$$E_\theta(\phi_{\theta_1}\phi_\theta)=a(\theta)\int\phi_{\theta_1}f'_{1\theta}d\mu=a(\theta)\eta_1(\theta;\phi_{\theta_1})\qquad\text{for every }\phi\in C\,.$$

Further suppose that the conditional densities h_θ depend on θ only through θ_1, say $h_\theta=h_{\theta_1}$, and suppose that the family $\mathcal{P}^U_{\theta_1}$ defined by

$$\mathcal{P}^U_{\theta_1}=\{P^U_\theta:\theta\in\{\theta_1\}\times\boldsymbol{\Theta}_2\times\cdots\times\boldsymbol{\Theta}_n\}\tag{6.6}$$

is complete for every fixed $\theta_1\in\boldsymbol{\Theta}_1$. Then, it can also be shown that $J_\phi(\theta;\theta_1)$ attains its supremum in C for $\phi=\bar{h}=\bar{h}_{\theta_1}(t)$, provided $\bar{h}\in C$. This result corresponds to Theorem 3.2 in Godambe [4]. Note that $\bar{h}$ is obtained from (6.5) by putting $a(\theta)=1$ and $-b_\theta=\bar{g}_\theta\circ\Pi$.

From now on, let us consider the case where the conditional densities h_θ depend on θ only through θ_1. To emphasize this, as before, we shall write h_{θ_1}, h'_{θ_1}, $\dot{h}_{\theta_1}$ and $\bar{h}_{\theta_1}$ for h_θ, h'_θ, $\dot{h}_\theta$ and $\bar{h}_\theta$, respectively. It follows from (5.11) that $E_\theta(\bar{h}_{\theta_1})=0$ for every $\theta\in\boldsymbol{\Theta}$, and therefore $\bar{h}=\bar{h}_{\theta_1}(t)$ is unbiased. Here we shall assume that $\bar{h}$ is regular, i.e. $\bar{h}\in C$. Moreover, we shall assume that $\bar{h}\in C^*$. Then for every $\theta\in\boldsymbol{\Theta}$,

$$E_\theta\{J^*_{\bar{h}}(\theta;\theta_1)\}=I_h(\theta;\theta_1)=J_{\bar{h}}(\theta;\theta_1)\,.\tag{6.7}$$

The left-hand equality follows from the easily proved,

$$\eta^*_1(\theta;\bar{h}_{\theta_1})=\Pi^*\{(\dot{h}_{\theta_1})^2\}=V_\theta(\bar{h}_{\theta_1}|U)\qquad\text{a.e. }\nu\,.$$

Next, noting that $h'_\theta=0$ a.e. μ on $\Pi^{-1}B-A$, we obtain from (5.3) and (5.11),

$$\begin{aligned}&\int\bar{h}_{\theta_1}f'_{1\theta}d\mu-\int(g_\theta\circ\Pi)(\dot{h}_{\theta_1})^2d\mu\\&\quad=\int(g'_{1\theta}\circ\Pi)\bar{h}_{\theta_1}h_{\theta_1}d\mu=\int(\bar{g}_\theta\circ\Pi)(g_\theta\circ\Pi)h'_{\theta_1}d\mu=\int\bar{g}_\theta\Pi^*\{(g_\theta\circ\Pi)h'_{\theta_1}\}d\nu=0\,.\end{aligned}$$

The right-hand equality is then obtained from the following observation:

$$\eta_1(\theta;\bar{h}_{\theta_1})=\int(g_\theta\circ\Pi)(\dot{h}_{\theta_1})^2d\mu=V_\theta(\bar{h}_{\theta_1})\,.$$

Taking account of (6.3) and (6.4), we see that (6.7) implies the relation,

$$I_h(\theta;\theta_1)=\mathcal{J}_h(\theta;\theta_1)\leqq\mathcal{J}_f(\theta;\theta_1)\leqq I_f(\theta;\theta_1)\tag{6.8}$$

for every $\theta\in\boldsymbol{\Theta}$. Further, it follows from (5.6) and (6.1) that for every

$\theta \in \Theta$,

$$I_f(\theta;\theta_1)=I_g(\theta;\theta_1)+I_h(\theta;\theta_1)\ . \tag{6.9}$$

Obviously, the property

$$I_h(\theta;\theta_1)=I_f(\theta;\theta_1) \qquad \text{for every } \theta\in\Theta \tag{6.10}$$

is one good quality in the h_{θ_1} family used in investigating the value of θ_1. Naturally, we can regard the property

$$\mathcal{I}_h(\theta;\theta_1)=\mathcal{I}_f(\theta;\theta_1) \qquad \text{for every } \theta\in\Theta \tag{6.11}$$

as another good quality in the h_{θ_1} family. Let us examine the relationship between (6.10) and (6.11).

Clearly, it follows from (6.8) that (6.10) entails (6.11). Furthermore, we see that (6.10) yields $\mathcal{I}_g(\theta;\theta_1)=0$ for every $\theta\in\Theta$ on account of (6.4) and (6.9). For example, suppose that the densities g_θ depend on θ only through r, say $g_\theta=\tilde{g}_{r(\theta)}$, where $r={}^t(r_1,\cdots,r_{n-1})$ is a continuously differentiable function of θ, and $\tilde{g}_r$ is differentiable a.e. ν with respect to r. Partition the Jacobian matrix $K_r(\theta)$ of r as $K_r(\theta)=[K_{1r}(\theta)\ K_{2r}(\theta)]$, $K_{1r}(\theta)$ being of order $(n-1)\times 1$ and $K_{2r}(\theta)$ being of order $(n-1)\times(n-1)$. Then, we shall say that the statistic U is ancillary with respect to θ_1 if rank $K_{2r}(\theta)=n-1$. Note that Godambe's ([4], Assumption 3.2) one concept of ancillarity with respect to θ_1 is a special case of this. If the rank condition is true, then

$$g'_{1\theta}={}^tK_{1r}(\theta)\{{}^tK_{2r}(\theta)\}^{-1}\,{}^t(g'_{2\theta},\cdots,g'_{n\theta}) \qquad \text{a.e. } \nu$$

at every $\theta\in\Theta$. Thus it follows from (6.2) that $I_g(\theta;\theta_1)=0$ for every $\theta\in\Theta$, which in turn implies (6.10) because of (6.9). Hence the above concept of ancillarity is justified by both property (6.10) and property (6.11).

Godambe's ([4], Assumption 3.4) another concept of ancillarity with respect to θ_1 is described as follows: the family $\mathcal{P}^U_{\theta_1}$ defined by (6.6) is complete for every fixed $\theta_1\in\Theta_1$. As mentioned above, in this case, (6.11) is true. However, (6.10) is generally not valid. To illustrate this, let us take X and Y to be independently normally distributed with variance θ_1 and mean θ_2; $\theta={}^t(\theta_1,\theta_2)\in(0,\infty)\times R^1$, and let $T={}^t(X,Y)$ and $U=X+Y$. This example was given by Godambe ([4], Example 4.3). Then we can easily show that $\mathcal{I}_h(\theta;\theta_1)=\mathcal{I}_f(\theta;\theta_1)=\theta_1^{-2}/2$ and $\mathcal{I}_g(\theta;\theta_1)=0$ for every θ. However, $I_g(\theta;\theta_1)=\theta_1^{-2}/2\neq 0$, and hence $I_h(\theta;\theta_1)<I_f(\theta;\theta_1)$.

Recently, Godambe [6] has shown that his two concepts of ancillarity can be unified with an extended concept of Fisher information. However, his new measure of information is also defined in a relation to the class C of all regular and unbiased estimating functions, as is the information $\mathcal{I}_f(\theta; \theta_1)$. On the other hand, the partial sensitivity or Liang's measure of information is related only to a given statistical model, i.e. a given family of distributions.

Acknowledgement

The author would like to acknowledge the referee for his useful comments and kind suggestions. Also the author would like to thank Professor M. Huzii for his helpful advice.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*, John Wiley and Sons, Chichester.

[2] Basu, D. (1977). On the elimination of nuisance parameters, *J. Amer. Statist. Ass.*, **72**, 355-366.

[3] Fraser, D. A. S. (1956). Sufficient statistics with nuisance parameters, *Ann. Math. Statist.*, **27**, 838-842.

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

[5] Godambe, V. P. (1980). On sufficiency and ancillarity in the presence of a nuisance parameter, *Biometrika*, **67**, 155-162.

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

[7] Halmos, P. R. and Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics, *Ann. Math. Statist.*, **20**, 225-241.

[8] Kuboki, H. (1984). A generalization of the relative conditional expectation—Further properties of Pitman's T^* and their applications to statistics, *Ann. Inst. Statist. Math.*, **36**, 181-197.

[9] Liang, K. Y. (1983). On information and ancillarity in the presence of a nuisance parameter, *Biometrika*, **70**, 607-612.

[10] Loève, M. (1977). *Probability Theory* I (4th ed.), Springer-Verlag, New York.

[11] Pitman, E. J. G. (1979). *Some Basic Theory for Statistical Inference*, Chapman and Hall, London.

[12] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (2nd ed.), John Wiley and Sons, New York.

[13] Sandved, E. (1972). Ancillary statistics in models without and with nuisance parameters, *Skand. AktuarTidskr.*, **55**, 81-91.