REMARKS ON SUFFICIENT STATISTICS

RYOICHI SHIMIZU

(Received Jan. 13, 1966)

1. Let Ω be an arbitrary set and let $\mathscr{P}=\{f(\boldsymbol{x};\theta):\theta\in\Omega\}$ be a family of probability density functions on an open subset $\mathfrak{X}$ of Euclidean n-space E^n such that

$$(1)\qquad f(\boldsymbol{x};\theta)>0,\quad \boldsymbol{x}=(x_1,\cdots,x_n)\in\mathfrak{X},\quad \theta\in\Omega,$$

and that

$$(2)\qquad \partial f(\boldsymbol{x};\theta)/\partial x_j,\quad j=1,\cdots,n$$

exist and are continuous in $\mathfrak{X}$ for all $\theta\in\Omega$. Barankin-Katz [2] treated sufficient statistics of minimal dimension. Barankin-Maitra [3] considered "Fisher-Darmois-Koopman-Pitman theorem" — it asserts that the existence of a sufficient statistic leads to an exponential family — for the case

$$f(\boldsymbol{x};\theta)=f_1(x_1;\theta)\ \cdots\ f_n(x_n;\theta),$$

where $f_j(x;\theta)$'s are not necessarily identical. These results were obtained under a regularity condition on θ, but as Barankin [1] recognized, and as we shall show below, they are obtainable without any assumption on Ω. Moreover, a global result is obtained for an extension of the above theorem without analyticity of $f_j(x;\theta)$'s.

2. Let $\mathscr{P}$ satisfy the conditions (1) and (2).

DEFINITION 1. Let N be a Borel subset of $\mathfrak{X}$. A statistic (i.e., a Borel measurable transformation on $\mathfrak{X}$) $S(\boldsymbol{x})$ is said to be sufficient in N (*for* $\mathscr{P}$) if $\boldsymbol{x},\boldsymbol{y}\in N$, and $S(\boldsymbol{x})=S(\boldsymbol{y})$ imply that $f(\boldsymbol{x};\theta)/f(\boldsymbol{y};\theta)$ is independent of θ. A statistic $T(\boldsymbol{x})$ is said to be necessary in N, if for any statistic $S(\boldsymbol{x})$ which is sufficient in N, $\boldsymbol{x},\boldsymbol{y}\in N$ and $S(\boldsymbol{x})=S(\boldsymbol{y})$ imply $T(\boldsymbol{x})=T(\boldsymbol{y})$.

DEFINITION 2. Let

(i) for a fixed θ_0, $g(\boldsymbol{x};\theta)=\log f(\boldsymbol{x};\theta)-\log f(\boldsymbol{x};\theta_0)$,

(ii) for any positive integer m,

$$M(\boldsymbol{x}; \theta_1, \cdots, \theta_m) = \text{rank of } \left\| \left(\frac{\partial g(\boldsymbol{x}; \theta_i)}{\partial x_j} \right)_{\boldsymbol{x}} \right\|_{\substack{i=1,\cdots,m \\ j=1,\cdots,n}},$$

(iii) $\sigma(\boldsymbol{x}) = \max\limits_{\substack{m \geqq 1 \\ \theta_1, \cdots, \theta_m}} M(\boldsymbol{x}; \theta_1, \cdots, \theta_m)$,

(iv) $R_r = \{\boldsymbol{x}_0 \mid \sigma(\boldsymbol{x}) = r$ in some neighbourhood of $\boldsymbol{x}_0\}$

$r = 0, 1, \cdots, n$,

(v) $\Omega_r = \{\boldsymbol{\theta} = (\theta_1, \cdots, \theta_r) \mid M(\boldsymbol{x}; \theta_1, \cdots, \theta_r) = r$ for some $\boldsymbol{x} \in R_r\}$,

$r = 1, \cdots, n$,

and

(vi) $A(\boldsymbol{\theta}) = \{\boldsymbol{x} \mid M(\boldsymbol{x}; \boldsymbol{\theta}) = r\} \cap R_r, \qquad \boldsymbol{\theta} \in \Omega_r$.

The following lemma 1 is a direct consequence of the definitions.

LEMMA 1. *$S(\boldsymbol{x})$ is sufficient (or necessary) in N, if for $\boldsymbol{x}, \boldsymbol{y} \in N$, $S(\boldsymbol{x}) = S(\boldsymbol{y})$ implies (or is implied by) $g(\boldsymbol{x}; \theta) = g(\boldsymbol{y}; \theta)$ for all $\theta \in \Omega$.*

LEMMA 2. *$A(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Omega_r$, $r = 1, \cdots, n$ are all open sets. If $R_r \neq \phi$ $(r \geqq 1)$, there exist $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots \in \Omega_r$ such that*

$$R_r = \bigcup_1^\infty A(\boldsymbol{\theta}_k). \tag{3}$$

$\mathfrak{X} - \bigcup\limits_0^n R_r$ is of Lebesgue measure zero.

PROOF. The openness of $A(\boldsymbol{\theta})$'s is clear. It is also clear that $R_r = \bigcup\limits_{\boldsymbol{\theta} \in \Omega_r} A(\boldsymbol{\theta})$. (3) follows then from the Lindelöf theorem. Let N be an arbitrary open subset of $\mathfrak{X} - R_0$ and let

$$r = \max_{\boldsymbol{x} \in N} \sigma(\boldsymbol{x}) = \sigma(\boldsymbol{x}_0) = M(\boldsymbol{x}_0; \theta_1, \cdots, \theta_r) \qquad \boldsymbol{x}_0 \in N.$$

Then

$$\boldsymbol{x}_0 \in R_r \subseteqq \bigcup_1^n R_r.$$

This means that open set $\bigcup\limits_1^n R_r$ is everywhere dense in $\mathfrak{X} - R_0$ and hence $\mathfrak{X} - \bigcup\limits_0^n R_r$ is of Lebesgue measure zero. q.e.d.

LEMMA 3. *If $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_r) \in \Omega_r$, then $S(\boldsymbol{x}) = (g(\boldsymbol{x}; \theta_1), \cdots, g(\boldsymbol{x}; \theta_r))$*

is necessary in $A(\boldsymbol{\theta})$ and sufficient in a suitable neighbourhood $A^{\gamma}(\boldsymbol{\theta})$ of every point $\boldsymbol{x}_{\gamma}$ of $A(\boldsymbol{\theta})$. Moreover, if $T(\boldsymbol{x})=(T_1(\boldsymbol{x}), \cdots, T_s(\boldsymbol{x}))$ is sufficient in some open subset N of $A(\boldsymbol{\theta})$, and if $T_i(\boldsymbol{x}), i=1, \cdots, s$ are real-valued and are continuously differentiable in N, then $s \geqq r$.

PROOF. Necessity is clear. From the definition of $A(\boldsymbol{\theta})$, we have

$$M(\boldsymbol{x}_{\gamma}; \theta_1, \cdots, \theta_r)=M(\boldsymbol{x}_{\gamma}; \theta_1, \cdots, \theta_r, \theta)=r, \ \boldsymbol{x}_{\gamma} \in A(\boldsymbol{\theta}), \ \theta \in \Omega .$$

Hence by the Implicit Function Theorem, we conclude that in some $A^{\gamma}(\boldsymbol{\theta}) \subseteqq A(\boldsymbol{\theta})$, we have $g(\boldsymbol{x}; \theta)=g(\boldsymbol{y}; \theta)$ for all $\theta \in \Omega$ if $g(\boldsymbol{x}; \theta_i)=g(\boldsymbol{y}; \theta_i)$, $i=1, \cdots, r$. This proves the first part of the lemma.

Write

$$y_j=T_j(\boldsymbol{x}) \qquad j=1, \cdots, s$$

$$\xi_i=g(\boldsymbol{x}; \theta_i) \qquad i=1, \cdots, r .$$

We suppose without loss of generality that

$$\max_{\boldsymbol{x} \in N} \text{ rank of } \left(\frac{\partial y_i}{\partial x_j}\right)=s .$$

Since $(\xi_1, \cdots, \xi_r)$ is necessary in $A(\boldsymbol{\theta})$, there exists a set of continuously differentiable functions $F_1, \cdots, F_r$ such that

$$\xi_i=F_i(y_1, \cdots, y_s) \qquad i=1, \cdots, r . \tag{4}$$

Differentiating (4), we obtain

$$\left(\frac{\partial \xi_i}{\partial x_j}\right)_{\substack{i=1, \cdots, r \\ j=1, \cdots, n}}=\left(\frac{\partial F_i}{\partial y_k}\right)_{\substack{i=1, \cdots, r \\ k=1, \cdots, s}} \cdot \left(\frac{\partial y_k}{\partial x_j}\right)_{\substack{k=1, \cdots, s \\ j=1, \cdots, n}} . \tag{5}$$

This implies $s \geqq r$. q.e.d.

Using the Lindelöf theorem and by reindexing, we have $R_r=\bigcup_{\alpha=1}^{\infty} A_{\alpha}$, where $A_{\alpha}=A^{\gamma}(\boldsymbol{\theta}_k) \subseteqq A(\boldsymbol{\theta}_k)$ for some γ and k.

Let

$$R_{r\alpha}=A_{\alpha}-\bigcup_{\beta=1}^{\alpha-1} A_{\beta}$$

$$D_{r\alpha}=\{S_{r\alpha}(\boldsymbol{x}) \mid \boldsymbol{x} \in R_{r\alpha}\} ,$$

where $S_{r\alpha}(\boldsymbol{x})=(g(\boldsymbol{x}; \theta_{1\alpha}^{r}), \cdots, g(\boldsymbol{x}; \theta_{r\alpha}^{r}))$ is necessary and sufficient in A_{α}.

Let

$$D_{r\alpha}=\bigcup_{i=1}^{\infty} D_{r\alpha i}$$

be a decomposition of $D_{r\alpha}$ into bounded Borel measurable sets in E^r such that $D_{r\alpha i}\cap D_{r\alpha j}=\phi$ $(i\neq j)$. Set

$$R_{r\alpha i}=S_{r\alpha}^{-1}(D_{r\alpha i})=\{\boldsymbol{x}\in R_{r\alpha}\mid S_{r\alpha}(\boldsymbol{x})\in D_{r\alpha i}\}.$$

Then

$$R_{r\alpha}=\bigcup_{i=1}^{\infty}R_{r\alpha i}\,,\qquad R_{r\alpha i}\cap B_{r\alpha j}=\phi\qquad(i\neq j)\,.$$

Since $D_{r\alpha i}$'s are bounded, there exist r-dimensional vectors $\eta_{r\alpha i}$ such that

$$D_{r\alpha i}^{*}=\{\eta+\eta_{r\alpha i}\mid\eta\in D_{r\alpha i}\}\,,\quad i=1,2,\cdots\quad(-\eta_{1\alpha i}\notin D_{1\alpha i})$$

are pairwise disjoint. Now we have

THEOREM 1. *The statistic defined by*

$$S(\boldsymbol{x})=\begin{cases}S_{r\alpha}(\boldsymbol{x})+\eta_{r\alpha i}, & \boldsymbol{x}\in R_{r\alpha i}\\ 0, & \boldsymbol{x}\in R_0\\ \boldsymbol{x}, & \boldsymbol{x}\in\mathfrak{X}-\bigcup_{0}^{n}R_r\end{cases}$$

is sufficient in $\bigcup_{0}^{n}R_r$ *and has minimal dimension in the sense of lemma 3 in the suitable neighbourhood of every point of* $\bigcup_{0}^{n}R_r$.

PROOF. It will be enough to show that $S(\boldsymbol{x})$ is sufficient. Suppose that

$$(6)\qquad S(\boldsymbol{x})=S(\boldsymbol{y})\,,\qquad \boldsymbol{x},\boldsymbol{y}\in\bigcup_{r=0}^{n}R_r\,.$$

Clearly $\boldsymbol{x},\boldsymbol{y}\in R_r$ for some r. If $r=0$, then (6) implies $g(\boldsymbol{x};\theta)=g(\boldsymbol{y};\theta)=0$ for all θ. If $r\geqq 1$, (6) holds if and only if $S_{r\alpha}(\boldsymbol{x})=S_{r\alpha}(\boldsymbol{y})$ and $\boldsymbol{x},\boldsymbol{y}\in R_{r\alpha}$ for some α. The desired result follows from lemma 3. q.e.d.

3. In this section we further assume that

$$f(\boldsymbol{x};\theta)=f_1(x_1;\theta)\cdots f_n(x_n;\theta)\,,$$

and

$$\mathfrak{X}=\mathfrak{X}_1\times\cdots\times\mathfrak{X}_n\,,$$

where $\mathfrak{X}_j$'s are open subsets of E^1, and $f_j(x;\theta)$'s are continuously differentiable in $\mathfrak{X}_j$ for all θ. Writing

$$g_j(x;\theta)=\log f_j(x;\theta)-\log f_j(x;\theta_0)\,,$$

we have

$$g(\boldsymbol{x};\theta)=\sum_{j=1}^{n} g_j(x_j;\theta)$$

and

$$\partial g(\boldsymbol{x};\theta)/\partial x_j=\partial g_j(x_j;\theta)/\partial x_j\,,\qquad j=1,\cdots,n\,.$$

THEOREM 2. *Let* $N=(a_1, b_1)\times\cdots\times(a_n, b_n)$ *be an open subset of* $\mathfrak{X}$, *and let* $r=\max_{x\in N}\sigma(\boldsymbol{x})<n$. *Then there are at least* $n-r$ *factors of* $f(\boldsymbol{x};\theta)$, $f_{r+1}(x_{r+1};\theta),\cdots,f_n(x_n;\theta)$, *say, which admit the representation,*

$$\log f_j(x;\theta)=c_{0j}(\theta)+\phi_j(x)+\sum_{k=1}^{r} c_k(\theta)\phi_{kj}(x) \tag{7}$$

$$x\in(a_j, b_j)\qquad j=r+1,\cdots,n\,,$$

where functions $\phi_j(x)$, $\phi_{kj}(x)$ *are continuously differentiable.*

PROOF. We have

$$A\equiv\det\left(g_j'(x_j^0;\theta_k)\right)_{j,\,k=1,\,\dots,\,r}\neq 0 \tag{8}$$

for some $\boldsymbol{x}_0=(x_1^0,\cdots,x_n^0)\in N$. Then we have for all $\theta\in\Omega$ and $x\in(a_j, b_j)$, $j=r+1,\cdots,n$,

$$\det\begin{bmatrix} g_1'(x_1^0;\theta_1) & \cdots & g_r'(x_r^0;\theta_1) & g_j'(x;\theta_1)\\ \cdot & & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ g_1'(x_1^0;\theta_r) & \cdots & g_r'(x_r^0;\theta_r) & g_j'(x;\theta_r)\\ g_1'(x_1^0;\theta) & \cdots & g_r'(x_r^0;\theta) & g_j'(x;\theta)\end{bmatrix}=0\,, \tag{9}$$

for otherwise, there would exist $\theta\in\Omega$ and $x_j^*\in(a_j, b_j)$ such that

$$\boldsymbol{x}^*=(x_1^0,\cdots,x_{j-1}^0, x_j^*, x_{j+1}^0,\cdots,x_n^0)\in N$$

and

$$r\geqq\sigma(\boldsymbol{x}^*)\geqq M(\boldsymbol{x}^*;\theta_1,\cdots,\theta_r,\theta)=r+1\,.$$

Expanding (9) by the $(r+1)$st column, we obtain

$$A_1(\theta)g_j'(x;\theta)+\cdots+A_r(\theta)g_j'(x;\theta_r)+Ag_j'(x;\theta)=0$$

$$\theta\in\Omega\,,\ x\in(a_j, b_j)\,.$$

Hence we have

$$\log f_j(x;\theta)=c_{0j}(\theta)+\phi_j(x)+\sum_{k=1}^{r} c_k(\theta)\phi_{kj}(x)$$

$$\theta\in\Omega,\ x\in(a_j, b_j)\,,\qquad j=r+1,\cdots,n$$

where

$$c_k(\theta) = -A_k(\theta)/A \quad \text{and} \quad \phi_j(x) = \log f_j(x;\, \theta_0). \qquad \text{q.e.d.}$$

Remark 1. We can easily verify the relations $A_i(\theta_j) = -\delta_{ij}A$, $i = 1, \cdots, r$, $j = 0, 1, \cdots, r$, which imply that $1, c_1(\theta), \cdots, c_r(\theta)$ are linearly independent. But this is not important when $f_j(x;\, \theta)$'s are not identical, since in this case $1, \phi_{1j}(x), \cdots, \phi_{rj}(x)$ are not necessarily linearly independent.

Remark 2. So far we have been concerned with $\mathfrak{X}$ or statistics. Exchanging the places of $\mathfrak{X}$ and Ω, we get similar results for Ω or parameters. Suppose that Ω is an open subset of E^ν, and that each $f(\boldsymbol{x};\, \theta)$ satisfies the following conditions:

$$(10) \qquad f(\boldsymbol{x};\, \theta) > 0, \qquad \boldsymbol{x} \in \mathfrak{X}, \qquad \boldsymbol{\theta} = (\theta^1, \cdots, \theta^\nu) \in \Omega\,,$$

and

$$(11) \qquad \partial f(\boldsymbol{x};\, \theta)/\partial\theta^i, \qquad i = 1, \cdots, \nu$$

exist and are continuous in Ω for all $\boldsymbol{x} \in \mathfrak{X}$.

No assumption is made on the sample space $\mathfrak{X}$ here.

DEFINITION 3. Let N be a Borel subset of Ω. A parameter (i.e., a Borel measurable transformation on Ω) $U(\theta)$ is said to be sufficient (or identifiable) in N if for $\theta, \tau \in N$, $U(\theta) = U(\tau)$ implies (or is implied by) $f(\boldsymbol{x};\, \theta) = f(\boldsymbol{x};\, \tau)$ for all $\boldsymbol{x}$.

Let $\boldsymbol{x}_0 \in \mathfrak{X}$ be fixed and let

$$k(\theta;\, \boldsymbol{x}) = \log f(\boldsymbol{x};\, \theta) - \log f(\boldsymbol{x}_0;\, \theta)\,.$$

Then we have the following

LEMMA 4. *A parameter $U(\theta)$ is sufficient (or identifiable) in N, if for $\theta, \tau \in N$, $U(\theta) = U(\tau)$ implies (or is implied by) $k(\theta;\, \boldsymbol{x}) = k(\tau;\, \boldsymbol{x})$ for all $\boldsymbol{x}$.*

PROOF. $f(\boldsymbol{x};\, \theta) = f(\boldsymbol{x};\, \tau)$ clearly implies $k(\theta;\, \boldsymbol{x}) = k(\tau;\, \boldsymbol{x})$. Conversely, suppose $k(\theta;\, \boldsymbol{x}) = k(\tau;\, \boldsymbol{x})$ for all $\boldsymbol{x}$. Then, we have

$$\frac{f(\boldsymbol{x};\, \theta)}{f(\boldsymbol{x}_0;\, \theta)} = \frac{f(\boldsymbol{x};\, \tau)}{f(\boldsymbol{x}_0;\, \tau)}\,.$$

Integrating over $\mathfrak{X}$, we get $f(\boldsymbol{x}_0;\, \theta_0) = f(\boldsymbol{x}_0;\, \tau_0)$. Hence we have

$$f(\boldsymbol{x};\, \theta) = f(\boldsymbol{x};\, \tau) \text{ for all } \boldsymbol{x}\,. \qquad \text{q.e.d.}$$

We can easily see that if we replace $\boldsymbol{x}$, θ, $g(\boldsymbol{x};\, \theta)$ and "necessary" by

θ, $\boldsymbol{x}$, $k(\theta; \boldsymbol{x})$ and "identifiable", respectively, all the results obtained in section 2 in terms of the statistics, hold for the parameters.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] E. W. Barankin, "Sufficient parameters: Solution of the minimal dimensionality problem," *Ann. Inst. Stat. Math.*, 12 (1960), 91-118.

[2] E. W. Barankin and M. Katz, Jr., "Sufficient statistics of minimal dimension," *Sankhyā*, 21 (1959), 217-246.

[3] E. W. Barankin and A. P. Maitra, "Generalization of the Fisher-Darmois-Koopman-Pitman-theorem on sufficient statistics," *Sankhyā*, 25 (1963), 217-244.

[4] E. B. Dynkin, "Necessary and sufficient statistics for a family of probability distributions," *Uspehi Matem. Nauk.*, 6 (1951), English translation which appeared in *Selected Translations in Math. Stat. and Prob.*, 1 (1961), 17-40.

[5] D. A. S. Fraser, "On sufficiency and the exponential family," *J. R. Statist. Soc. (B)*, 25 (1963), 115-123.