MATH. SCAND. 22 (1968), 197–202

SUFFICIENT DATA REDUCTION AND EXPONENTIAL FAMILIES

OLE BARNDORFF-NIELSEN and KARL PEDERSEN

Let $\mathscr{X}$ denote a region (an open connected set) in R^r, the r-dimensional Euclidean space, let $\mathscr{B}$ be the σ-algebra of Borel subsets of $\mathscr{X}$ and let $\mathscr{P}$ be a family of probability measures on $\mathscr{B}$. The elements of $\mathscr{P}$ are assumed absolutely continuous with respect to Lebesgue measure λ on $\mathscr{B}$ and f_P denotes the density of $P(\in \mathscr{P})$.

Take X as the random variable which is the identity map on $\mathscr{X}$ and let $X_1,\ldots,X_n$ be n independent observations of X.

In the present work we discuss two propositions of the Fisher–Darmois–Koopman–Pitman type. Propositions of this type state that, under certain regularity conditions on the densities f_P, if there exists a sufficient statistic $T=t(X_1,\ldots,X_n)$ which yields a reduction of the data $X_1,\ldots,X_n$ then the family $\mathscr{P}$ is exponential. The statistical interest of such results depends on what regularity conditions are imposed on f_P and on how the concept of reduction is formalized. The two propositions treated below seem to us particularly interesting in this respect. They represent generalizations to arbitrary dimensions r of results due to Brown (1964) and Dynkin (1951) for the case $r=1$.

For related work we refer to the papers in the bibliography, in particular to the excellent survey in the introductory section of Barankin and Maitra (1963).

Throughout the sequel it is assumed that f_P is strictly positive and continuous. Furthermore, k and n denote positive integers with $k<n$ and T, an arbitrary but fixed statistic, is assumed to be a continuous mapping of $\mathscr{X}^n$ into R^k. Let C be the space of real continuous functions on $\mathscr{X}$ and let C' be the space of real functions on $\mathscr{X}$ having first order continuous partial derivatives. Finally, let P_0 be an arbitrary but fixed element of $\mathscr{P}$. By the Fisher–Neyman factorization criterion T is sufficient provided that to each $P \in \mathscr{P}$ there exists a function h_P such that

$$(1) \qquad \frac{f_P(x_1)\ldots f_P(x_n)}{f_{P_0}(x_1)\ldots f_{P_0}(x_n)} = h_P(T(x^{(n)})), \qquad (x_1,\ldots,x_n) = x^{(n)} \in \mathscr{X}^n .$$

Received October 23, 1967.

Introducing the notations

$$\varphi_P = \log \frac{f_P}{f_{P_0}}, \quad \psi_P = \log h_P ,$$

(1) is equivalent to

$$(2) \qquad \varphi_P(x_1) + \ldots + \varphi_P(x_n) = \psi_P(T(x^{(n)})), \quad x^{(n)} \in \mathscr{X}^n .$$

Thus sufficiency of T may be described as follows.

Let S be the set of those $\varphi \in C$ for which there exists a function ψ such that

$$(3) \qquad \varphi(x_1) + \ldots + \varphi(x_n) = \psi(T(x^{(n)})), \quad x^{(n)} \in \mathscr{X}^n .$$

T is sufficient if $\varphi_P \in S$, $P \in \mathscr{P}$.

A family $\mathscr{P}$ is said to be *exponential* provided there exists a positive integer s, real functions $a|\mathscr{P}$, $\alpha_1|\mathscr{P},\ldots,\alpha_s|\mathscr{P}$ and real (measurable) functions $\tau_1|\mathscr{X},\ldots,\tau_s|\mathscr{X}$ and $b|\mathscr{X}$, $b \geqq 0$, such that (a.e.)

$$(4) \qquad f_P(x) = a(P)\, e^{\alpha(P)\tau^*(x)}\, b(x), \quad P \in \mathscr{P} ,$$

where $\alpha = (\alpha_1,\ldots,\alpha_s)$, $\tau = (\tau_1,\ldots,\tau_s)$, and τ^* denotes the transpose of τ. The smallest s which admits a representation of f_P of the form (4) is the *order* of $\mathscr{P}$. It is simple to see that $\mathscr{P}$ is exponential of order s if and only if $\dim V = s+1$ where dim denotes dimension and V stands for the linear subspace of C spanned by the constant functions and the functions φ_P, $P \in \mathscr{P}$.

We shall prove:

Suppose the probability densities f_P are strictly positive and continuous on the sample space $\mathscr{X}$, a region in R^r. Let k and n denote positive integers with $k < n$ and let T be a continuous, k-dimensional, sufficient statistic on $\mathscr{X}^n$.

(i) *If $k = 1$ then $\mathscr{P}$ is exponential of order* 1.

(ii) *If the densities f_P have continuous partial derivatives then $\mathscr{P}$ is exponential of order $\leqq k$.*

REMARKS. For $r = 1$ proposition (i) is due to Brown (1964) (cf. section 4 of that paper) while (ii) is a modified version of Theorem A, Brown (1964) which in turn was obtained by modification of results in Dynkin's (1951) paper.

Note that (ii) is subsumed under (i) when $k = 1$.

The assumption that T is sufficient in the strict sense that (3) holds for every $x^{(n)} \in \mathscr{X}^n$ and not just almost everywhere is indispensable. In fact, it imposes no restriction on the family $\mathscr{P}$ to require almost everywhere validity of (3) since to any $n>1$ there exists a continuous, real-valued function on $\mathscr{X}^n$ almost everywhere $1-1$ (see Denny (1964)).

PROOF. Let $S'=S\cap C'$. The first step in the proof of propositions (i) and (ii) is to show that they will follow from inequalities, subsequently proved, concerning the dimensions of the sets S and S' of solutions to the functional equation (3) with $n=k+1$.

S and S' are both linear spaces containing all constant functions on $\mathscr{X}$. The assumptions of (i) imply $V\subset S$ while those of (ii) imply $V\subset S'$. Thus, according to a previous remark, to verify (i) and (ii) it suffices to show

(i)′ $\dim S \leqq k+1$ when $k=1$

respectively,

(ii)′ $\dim S' \leqq k+1$.

Moreover, in verifying these inequalities it causes no loss of generality to assume $n=k+1$ as may be seen by fixing arbitrarily $x_{k+2},\ldots,x_n$, letting $\bar{T}$ be the section of T at $x_{k+2},\ldots,x_n$ and rewriting (3) in the form

$$\begin{aligned}\varphi(x_1)+\ldots+\varphi(x_{k+1}) &= \psi(\bar{T}(x^{k+1}))-\varphi(x_{k+2})-\ldots-\varphi(x_n)\\ &= \bar{\psi}(\bar{T}(x^{(k+1)})), \qquad x^{(k+1)}\in\mathscr{X}^{k+1}.\end{aligned}$$

(i)′ and (ii)′ will first be derived for $r=1$ by the methods of Brown and Dynkin and then extended to general r.

PROOF OF *(i)′* FOR $r=1$.

Here (3) with $n=k+1=2$ takes the form

$$\varphi(x_1)+\varphi(x_2) = \psi(T(x_1,x_2))$$

and we show $\dim S\leqq 2$.

LEMMA 1. *Let $\varphi\in S$ and let x_0, y_1, y_2 be points in $\mathscr{X}$. If $\varphi(y_1)=\varphi(y_2)$ and $T(x_0,y_1)<T(x_0,y_2)$ then φ is constant in a neighborhood of x_0.*

PROOF. Without loss of generality it can be assumed that $y_1<y_2$ and that

$$(5)\qquad t_1 = T(x_0,y_1) < T(x_0,y) < T(x_0,y_2) = t_2, \qquad y_1<y<y_2.$$

(If $y_1<y_2$ but (5) is not fulfilled, then let $y_1'=\sup\{y: y\leqq y_2 \text{ and } T(x_0,y)=t_1\}$ and $y_2'=\inf\{y: y\geqq y_1' \text{ and } T(x_0,y)=t_2\}$. Now y_1' and y_2' satisfy $y_1'<y_2'$ and (5), as well as the conditions of the Lemma.)

Since $\varphi(y_1)=\varphi(y_2)$ there exists a y_0 in the open interval (y_1,y_2) such that $\varphi(y_0)$ is either an absolute minimum or an absolute maximum for φ

on $[y_1,y_2]$. Suppose $\varphi(y_0)$ is an absolute minimum; the maximum case can be treated similarly. Then, for $t_0=T(x_0,y_0)$,

$$\psi(t_0) = \min\{\psi(t): t \in [t_1,t_2]\} .$$

By (5)

$$t_1 = T(x_0,y_1) < t_0 = T(x_0,y_0) < t_2 = T(x_0,y_2);$$

therefore, a neighborhood U of x_0 exists such that

$$T(x,y_1) < t_0 < T(x,y_2), \quad x \in U , \tag{6}$$

and

$$t_1 < T(x,y_0) < t_2, \quad x \in U . \tag{7}$$

From (6) and the continuity of T it follows that to every $x \in U$ there is an $\alpha(x)$ with $T(x,\alpha(x))=t_0$ and $y_1<\alpha(x)<y_2$. Hence, for $x \in U$

$$\begin{aligned}\psi(t_0) = \varphi(x)+\varphi(\alpha(x)) &\geqq \varphi(x)+\varphi(y_0) = \psi(T(x,y_0)) \\ &\geqq \psi(t_0) = \psi(T(x_0,y_0)) = \varphi(x_0)+\varphi(y_0) .\end{aligned}$$

None of the inequalities can be proper, and consequently $\varphi(x)=\varphi(x_0)$, $x \in U$.

Let φ_1 and φ_2 be arbitrary elements of S and suppose that φ_1 is not constant, i.e., there exist $y_1,y_2 \in \mathscr{X}$ with $\varphi_1(y_1) \neq \varphi_1(y_2)$. Then $T(x_0,y_1) \neq T(x_0,y_2)$ for every $x_0 \in \mathscr{X}$ and for some $a \in R$,

$$\varphi_2(y_1)-a\varphi_1(y_1) = \varphi_2(y_2)-a\varphi_1(y_2) .$$

The function $\varphi=\varphi_2-a\varphi_1$ is in S and Lemma 1 is applicable to φ for all $x_0 \in \mathscr{X}$. Hence, on account of the continuity of φ, there is a constant b such that $\varphi=b$, that is, $\varphi_2=a\varphi_1+b$. In other words, S is at most two-dimensional.

Proof of $(ii)'$ for $r=1$.

Lemma 2. *Let $\varphi_1,\ldots,\varphi_n$ be elements of C' and consider the mapping*

$$\Phi\colon x^{(n)} \to (\varphi_1(x_1)+\ldots+\varphi_1(x_n), \ldots, \varphi_n(x_1)+\ldots+\varphi_n(x_n)), \quad x^{(n)} \in \mathscr{X}^n ,$$

with Jacobian $J\colon x^{(n)} \to \{\varphi_i'(x_j)\}$. If $1,\varphi_1,\ldots,\varphi_n$ are linearly independent, then for some $x_0^{(n)} \in \mathscr{X}^n$, $\det J(x_0^{(n)}) \neq 0$.

Proof. The proof is by induction. The Lemma is clearly true for $n=1$. Suppose it holds for $n-1$ but $\det J \equiv 0$ for some $\varphi_1,\ldots,\varphi_n$ with $1,\varphi_1,\ldots,\varphi_n$ linearly independent. Expansion of the determinant by its last column yields

$$0 = a_1(x_1,\ldots,x_{n-1})\varphi_1'(x_n)+\ldots+a_n(x_1,\ldots,x_{n-1})\varphi_n'(x_n), \quad x^{(n)} \in \mathscr{X}^n . \tag{8}$$

Note that $a_n(x_1,\ldots,x_{n-1})$ is the determinant corresponding to $\varphi_1,\ldots,\varphi_{n-1}$. Thus, according to the induction assumption there exists an $x_0^{(n-1)} \in \mathscr{X}^{n-1}$ with $a_n(x_{01},\ldots,x_{0n-1}) \neq 0$. Insertion of $x_{01},\ldots,x_{0n-1}$ in (8) and integration with respect to x_n from x_0 to x yields

$$0 = a_1(x_{01},\ldots,x_{0n-1})\varphi_1(x) + \ldots + a_n(x_{01},\ldots,x_{0n-1})\varphi_n(x) - \\ - \sum_{i=1}^{n} a_i(x_{01},\ldots,x_{0n-1})\varphi_i(x_0) .$$

Since $a_n(x_{01},\ldots,x_{0n-1}) \neq 0$, this relation contradicts the linear independence of $1,\varphi_1,\ldots,\varphi_n$.

Suppose now that $\dim S' > k+1$. Then there exist functions $\varphi_1,\ldots,\varphi_n \in S'$ such that $1,\varphi_1,\ldots,\varphi_n$ are linearly independent $(n=k+1)$. By Lemma 2 there is a point $x_0^{(n)}$ with $\det J(x_0^{(n)}) \neq 0$ and hence a neighborhood of $x_0^{(n)}$ on which Φ is $1-1$. On the other hand $\Phi = \Psi(T)$ where

$$\Psi\colon\ (t_1,\ldots,t_k) \to (\psi_1(t_1,\ldots,t_k),\ldots,\psi_n(t_1,\ldots,t_k)),$$

whence follows that T must be 1-1 in that neighborhood. But this is impossible since T is a continuous function on $\mathscr{X}^{k+1}$ into R^k.

Generalizations to arbitrary r.

It will be convenient to stress the dependence of $\mathscr{X}$, S and S' on r by writing $\mathscr{X}_r$, S_r and S_r'. Let S_{0r} stand for either S_r or S_r'. The statement that $\dim S_{0r} \leq k+1$ is equivalent to the statement that for any set $\{\varphi_1,\ldots,\varphi_{k+1}\} \subset S_{0r}$ the range space of the mapping

$$x \to (\varphi_1(x),\ldots,\varphi_{k+1}(x)), \qquad x \in \mathscr{X}_r,$$

is contained in a hyperplane of R^{k+1}.

Thus, if $\dim S_{0r} > k+1$ for some $r > 1$, then there exist sets $\{\varphi_1,\ldots,\varphi_{k+1}\} \subset S_{0r}$ and $\{x_1,\ldots,x_{k+2}\} \subset \mathscr{X}_r$ such that the $k+2$ points

$$(\varphi_1(x_i),\ldots,\varphi_{k+1}(x_i)), \qquad i=1,\ldots,k+2,$$

do not lie in a hyperplane of R^{k+1}. Let γ be a continuously differentiable mapping on $\mathscr{X}_1$ into $\mathscr{X}_r$ whose range contains the points $x_1,\ldots,x_{k+2}$ (such a mapping clearly exists). Then the mapping

$$\tilde{T}\colon\ (\xi_1,\ldots,\xi_{k+1}) \to T(\gamma(\xi_1),\ldots,\gamma(\xi_{k+1})), \qquad \xi_1,\ldots,\xi_{k+1} \in \mathscr{X}_1,$$

is continuous and, letting $\tilde{\varphi}_i(\cdot) = \varphi_i(\gamma(\cdot))$,

$$\tilde{\varphi}_i(\xi_1) + \ldots + \tilde{\varphi}_i(\xi_{k+1}) = \psi(\tilde{T}(\xi_1,\ldots,\xi_{k+1})), \qquad \xi_1,\ldots,\xi_{k+1} \in \mathscr{X}_1,$$

Moreover, γ has been chosen so that $\tilde{\varphi}_i \in S_1$ if $\varphi_i \in S_r$ and $\tilde{\varphi}_i \in S_1'$ if $\varphi_i \in S_r'$ and since the range of the map

$$\xi \to (\tilde{\varphi}_1(\xi), \ldots, \tilde{\varphi}_{k+1}(\xi)), \qquad \xi \in \mathscr{X}_1,$$

is not contained in a hyperplane, a contradiction to the established validity of (i)′ and (ii)′ for $r = 1$ has been arrived at.

REFERENCES

1. E. W. Barankin and A. P. Maitra, *Generalizations of the Fisher–Darmois–Koopman–Pitman theorem on sufficient statistics*, Sankhyā Ser. A 25 (1963), 217–244.
2. L. Brown, *Sufficient statistics in the case of independent random variables*, Ann. Math. Stat. 35 (1964), 1456–1474.
3. G. Darmois, *Sur les lois de probabilité a estimation exhaustive*, C. R. Acad. Sci. Paris 260 (1935), 1265–1266.
4. J. L. Denny, *A continuous real-valued function on E^n almost everywhere* 1–1, Fund. Math. 55 (1964), 95–99.
5. E. B. Dynkin, *Necessary and sufficient statistics for a family of probability distributions*, Select. Transl. Math. Stat. and Prob. 1 (1961), 23–41. (Translation of the original paper in Russian from 1951.)
6. R. A. Fisher, *Two new properties of mathematical likelihood*, Proc. Roy. Soc. London 144 (A) (1934), 285–307.
7. L. H. Koopman, *On distributions admitting a sufficient statistic*, Trans. Amer. Math. Soc. 39 (1936), 399–409.
8. E. J. G. Pitman, *Sufficient statistics and intrinsic accuracy*, Proc. Cambridge Philos. Soc. 32 (1936), 567–579.
9. O. G. Zhuravlev, *Minimal sufficient statistics for a sequence of independent random variables*, Theor. Probability Appl. 7 (1963), 218-220.

UNIVERSITY OF ÅRHUS, DENMARK