35

Reprinted from Sankhyā : The Indian Journal of Statistics, Series A, Vol. 28, Parts 2&3, 1966

SUFFICIENCY FOR REGULAR MODELS

By D. A. S. FRASER

University of Toronto and *University of Copenhagen*

SUMMARY. For a general parameter space, and a fixed-carrier model, the sample-space gradient of the log-likelihood statistic is used to determine the local form of the minimal sufficient statistic. An extended form of the exponential model is defined and is characterized in terms of the likelihood gradient. Some global Koopman-Darmois-Pitman theorems are then developed that generalize those theorems of Dynkin, and Barankin and Maitra.

1. Introduction

Let x be a variable taking values in a closed region X of Euclidean space R^k (or, for most of the results, a finite dimensional manifold with a countable base); let θ be a parameter taking values in a general space Ω; let $f(x|\theta)$ be the probability density function and, for each x in an open set A whose closure $\bar{A} = X$, suppose that $f(x|\theta) > 0$ and the gradient *vector*

$$\frac{\partial}{\partial x} f(x|\theta)$$

exists. *Such a model will be called R-regular.* An open set A can be expressed as a countable union of disjoint connected open sets A_α.

$$A = \bigcup_{\alpha=1}^{\infty} A_\alpha.$$

In his analysis of sufficiency for a Koopman-Darmois-Pitman theorem, Dynkin (1951) makes similar assumptions for a real variable but requires in addition that $f(x|\theta)$ be continuous in x and the gradient vector be continuous in A. From his definitions it would seem that the set A in which the gradient vector exists could vary with θ but examination of the proof of his Theorem 2 excludes this apparent generality.

Barankin and Maitra (1963) make somewhat stronger assumptions. They require that A be connected, that Ω be a connected open Euclidean set, and that further derivatives exist and are continuous, and for their global theorem they assume that $f(x|\theta)$ is analytic on $A \times \Omega$.

Both Dynkin, and Barankin and Maitra restrict themselves to one-dimension variables for their principal results.

2. The extended exponential model

For an analysis of sufficiency when the open set A is disconnected it seems more appropriate to work with a generalization rather than with the ordinary exponential model.

Let $f(x|\theta)$ be an R-regular statistical model. *$f(x|\theta)$ is an extended exponential model if it has the form*

$$f(x|\theta) = \exp\{a(x) + \sum_{\alpha=1}^{\infty} \delta_\alpha(x)\varphi_\alpha(\theta) + \sum_{s=1}^{r} b_s(x)\psi_s(\theta)\}$$

where $a(x)$, $b_s(x)$ *are differentiable in* A *and* $\delta_\alpha(x)$ *is the characteristic function of the component domain* A_α.

As a simple example suppose that x is real, that $A_1 = (1, 2)$ and $A_2 = (3, 4)$ and that

$$\begin{aligned} f(x\,|\,\theta) &= c(\theta)\exp\{\theta^2+x\theta\} \qquad \text{on } A_1 \\ &= c(\theta)\exp\{\theta^3+x\theta\} \qquad \text{on } A_2 \\ &= \exp\{[\theta^2+d(\theta)]\delta_1(x)+[\theta^3+d(\theta)]\delta_2(x)+x\theta\}. \end{aligned}$$

This model has a basic exponential term $x\theta$ which is adjusted from A_1 to A_2 by an amount that depends on θ.

Ordinary exponential models can be standardized by removing linear dependencies among the functions of x and among the functions of θ. A similar simplification is possible for extended exponential models.

An extended exponential model

$$f(x\,|\,\theta) = \exp\{a(x)+\sum_{\alpha=1}^{\infty}\delta_\alpha(x)\varphi_\alpha(\theta)+\sum_{s=1}^{r}b_s(x)\psi_s(\theta)\}$$

is in canonical form with respect to references points x^0_α *in* A_α *and* θ_0 *in* Ω *if*

(i) $(\psi_1(\theta_0), \ldots, \psi_r(\theta_0)) = 0$.

(ii) $\psi_1(\theta), \ldots, \psi_r(\theta)$ *are linearly independent.*

(iii) $(b_1(x), \ldots, b_r(x)) = 0$ *at each* x^0_α.

(iv) $b_1(x), \ldots, b_r(x)$ *are linearly independent.*

(v) $(\varphi_1(\theta_0), \varphi_2(\theta_0), \ldots) = 0$.

The example can be written in the following canonical form.

$$\begin{aligned} f(x\,|\,\theta) = \exp\{&d(0)+(\theta^2+1.5\theta+d(\theta)-d(0))\delta_1(x) \\ &+(\theta^3+3.5\theta+d(\theta)-d(0))\delta_2(x)+(x-1.5\,\delta_1(x)-3.5\,\delta_2(x))\theta\}. \end{aligned}$$

Any extended exponential model can be put in canonical form with respect to given reference points. Condition (i) can be satisfied by writing $\psi_s(\theta)=\{\psi_s(\theta)-\psi_s(\theta_0)\}+\psi_s(\theta_0)$, and rearranging the exponent so as to incorporate the contribution from $\psi_s(\theta_0)$ into the term $a(x)$. If there is linear dependence among the new ψ functions, one of them can be expressed in terms of the others, and substitution and rearrangement leads to one less ψ function; this can be repeated until no linear dependency remains, thereby satisfying condition (ii). Condition (iii) can be satisfied by writing

$$b_s(x) = \Big\{b_s(x)-\sum_{1}^{\infty}b_s(x^0_\alpha)\,\delta_\alpha(x)\Big\}+\sum_{1}^{\infty}b_s(x^0_\alpha)\,\delta_\alpha(x)$$

substituting, and rearranging. Condition (iv) can be satisfied in the way that condition (ii) was satisfied, and condition (v) in the way that condition (i) was satisfied.

Some simple analysis shows that *the canonical form for an extended exponential model is essentially unique in the following way.*

(a) *The function $a(x)$ is unique.* It is equal to $\ln f(x \mid \theta)$ at θ_0.

(b) *The functions $\varphi_1(\theta), \varphi_2(\theta), \ldots$ are unique.* If likelihood is taken with respect to θ_0, these functions give the log-likelihood at the reference points : $l(\theta \mid x_1^0),\ l(\theta \mid x_2^0), \ldots$ . .

(c) *The linear space generated by the functions $\psi_1(\theta), \psi_2(\theta), \ldots$ is unique.* This is the smallest space containing the likelihood deviations, $l(\theta \mid x) - l(\theta \mid x_\alpha^0)$ with x in A_α.

(d) *For a given basis $\psi_1(\theta), \psi_2(\theta), \ldots$ the coefficients $b_1(x), b_2(x), \ldots$ are unique.*

These elements of the canonical form for an extended exponential model derive from the following separation of $\ln f(x|\theta)$.

$$\ln f(x \mid \theta) = \ln f(x \mid \theta_0) + \left\{ \sum_{\alpha=1}^{\infty} l(\theta \mid x_\alpha^0)\, \delta_\alpha(x) \right\} + \left\{ l(\theta \mid x) - \sum_{\alpha=1}^{\infty} l(\theta \mid x_\alpha^0)\, \delta_\alpha(x) \right\}.$$

The second bracket gives a likelihood contribution that comes from the *set* A_α in which the observation falls, the third gives the likelihood contribution that comes from the observation x relative to the reference point for the set A_α in which x falls.

For later reference it is convenient to have names for some elements of an *extended exponential model in canonical form.*

$$\psi_1(\theta), \ldots, \psi_r(\theta)$$

are *structural-basis parameters* or merely *basis parameters* when the context is explanatory. *The number r of structural-basis parameters is the order of the model. The extended exponential model is said to have structural basis* :

$$\psi_1(\theta), \ldots, \psi_r(\theta).$$

The following lemma derives simply from the preceding definitions :

Lemma : *$x_1, \ldots, x_n$ are independent variables. Each x_j has an extended exponential model with structural basis in a space over $\psi_1(\theta), \ldots, \psi_r(\theta)$ if and only if $(x_1, \ldots, x_n)$ is an extended exponential model with structural basis in a space over $\psi_1(\theta), \ldots, \psi_r(\theta)$.*

3. Sufficiency for x near x^0

Sufficiency in the neighbourhood of a sample space point was investigated by Barankin and Katz (1959) and provides the basis for the Koopman-Darmois-Pitman theorems in Barankin and Mairta (1963).

In this section a concept of local sufficiency is developed; it is, in essence, equivalent to the Barankin and Katz formulation but applies to the more general R-regular models and its development is based on the minimal sufficiency of the likelihood function (Dynkin, 1951 and Fraser, 1963).

Let $f(x \mid \theta)$ be an R-regular statistical model. The statistic

$$l(\theta \mid x) = \ln f(x \mid \theta) - \ln f(x \mid \theta_0)$$

is defined for all x in A and its values are functions on the parameter space Ω; it is the log-likelihood statistic and provides one mode of expression for the minimal sufficient statistic.

Consider the log-likelihood statistic in a neighbourhood of a point x^0. To a first derivative approximation it has the form

$$l(\theta \mid x) = l(\theta \mid x^0) + \sum_{i=1}^{k} (x_i - x_i^0)\, lg_i(\theta \mid x^0) + \ldots$$

where

$$lg_i(\theta \mid x^0) = \frac{\partial}{\partial x_i^0}\, l(\theta \mid x)$$

is the i-th coordinate of the *likelihood-gradient vector*. Thus, to a first derivative approximation at x^0 *the log-likelihood function has*

(i) a fixed contribution giving the form at x^0,

(ii) contributions from $lg_1(\theta \mid x^0), \ldots, lg_2(\theta \mid x^0)$ in the amounts $(x_1 - x_1^0), \ldots, (x_k - x_k^0)$.

Thus the local form of the minimal sufficient statistic depends essentially on the likelihood-gradient functions $lg_i(\theta \mid x^0)$.

For reference it is convenient to have some names : *The likelihoods gradients* $lg_1(\theta \mid x^0), \ldots, lg_k(\theta \mid x^0)$ *are structural parameters at the point* x^0; any linear combination of the parameters $lg_1(\theta \mid x^0), \ldots, lg_k(\theta \mid x^0)$ *is a structural parameter at the point* x^0. Let $V(x^0)$ designate the linear space over Ω formed by the structural parameters at x^0, it has dimension at most k. *The space* $V(x^0)$ *is the likelihood-gradient space or simply the gradient space at* x^0.

The log-likelihood function in a neighbourhood of x^0 has

(i) a fixed contribution $l(\theta \mid x^0)$, plus,

(ii) an element of the gradient space $V(x^0)$.

Thus the first derivative form of the minimal sufficient statistic takes values in a linear space of dimension at most k. If there is a differentiable sufficient statistic $t(x)$ with $l\ (\leqslant k)$ real coordinates, then the dimension of V is at most l. And if there is a minimal sufficient statistic with l differentiable coordinates in a neighbourhood of x^0 then l is the dimension of $V(x^0)$.

Some further names : *The dimension of the gradient space at* x^0, dim $V(x^0) = d(x^0)$, *is the local structural dimension of the model at* x^0.

The maximum value $d = \max_x d(x)$ *is the structural dimension of the model.*

Thus, *structural dimension is the dimension of a differentiable minimal sufficient statistic when one exists.*

4. Local sufficiency and independent variables

Consider independent variables x_1 and x_2 involving the same parameter θ. And let $l^{(j)}(\theta \mid x_j)$ be the likelihood function statistic, $V^{(j)}(x_j)$ be the gradient space, and let $d^{(j)}(x_j)$ and $d^{(j)}$ be the local and global structural dimensions of the j-th variable x_j.

For the composite variable (x_1, x_2) the likelihood-function statistic is

$$l(\theta \mid x_1, x_2) = l^{(1)}(\theta \mid x_1) + l^{(2)}(\theta \mid x_2).$$

In the neighbourhood of (x_1^0, x_2^0) the likelihood statistic has the form

$$l(\theta \mid x_1, x_2) = l^{(1)}(\theta \mid x_1^0) + l^{(2)}(\theta \mid x_2^0) + \sum_{i=1}^{k_1} (x_{1i} - x_{1i}^0)\, lg_i^{(1)}(\theta \mid x_1^0)$$

$$+ \sum_{i=1}^{k_2} (x_{2i} - x_{2i}^0)\, lg_i^{(2)}(\theta \mid x_2^0) + \ldots$$

Thus the structural parameters for the composite variable at (x_1^0, x_2^0) are given by the set-theory union of the structural parameters for x_1 at x_1^0 and the structural parameters for x_2 at x_2^0. The gradient space for the composite variable is obtained by addition.

$$V(x_1, x_2) = V^{(1)}(x_1) + V^{(2)}(x_2) = \{v^{(1)} + v^{(2)} \mid v^{(i)} \epsilon V^{(i)}(x_i)\}.$$

The dimensions satisfy inequalities

$$d(x_1\, x_2) \leqslant d^{(1)}(x_1) + d^{(2)}(x_2); \quad d \leqslant d^{(1)} + d^{(2)}.$$

Note that $d(x_o, x_2) = d^{(1)}(x_1) + d^{(2)}(x_2)$ if and only if the gradients $lg_i^{(1)}(\theta \mid x_1)$ are linearly independent of the gradients $lg_i^{(2)}(\theta \mid x_2)$.

5. Characterization of an extended exponential model

The analysis of local sufficiency in terms of gradient spaces provides a simple means for characterizing the extended exponential model.

Theorem 1 : *A necessary and sufficient condition that an R-regular statistical model $f(x \mid \theta)$ be an extended exponential model is that for all x in A*

$$V(x) \subset V$$

where V is a finite dimensional linear space of parameters.

Proof: (a) Let $f(x \mid \theta)$ be an extended exponential model in canonical form using basis parameters $\psi_1(\theta), \ldots, \psi_r(\theta)$.

$$l(\theta \mid x) = \ln f(x \mid \theta) - \ln f(x \mid \theta_0) = \sum_\alpha \delta_\alpha(x)\, \varphi_\alpha(\theta) + \sum_s b_s(x) \psi_s(\theta)$$

$$lg_i\,(\theta \mid x) = \sum_s \frac{\partial b_s(x)}{\partial x_i}\, \psi_s(\theta)$$

$$V(x) \subset \text{LS}\,\{\psi_1(\theta), \ldots, \psi_r(\theta)\}.$$

The right side of the last expression denotes the linear space generated by the parameters $\psi_1(\theta), \ldots, \psi_r(\theta)$.

(b) Let $f(x \mid \theta)$ be a model satisfying

$$V(x) \subset V$$

for all x in A, and let $\psi_1(\theta), \ldots, \psi_r(\theta)$ be a basis for the smallest space V satisfying this relationship. Then

$$lg_i(\theta \mid x) = c_{i1}(x)\psi_1(\theta)+\ldots+c_{ir}(x)\psi_r(\theta).$$

Since $lg_i(\theta \mid x)$ is integrable for each θ, so also are the coefficients $c_{sj}(x)$. Then

$$l(\theta \mid x) = l(\theta \mid x_\alpha^0)+C_1(x)\psi_1(\theta)+\ldots+C_r(x)\,\psi_r(\theta)$$

where the C's are integrals of c's along connected paths within a set A_α ; thus

$$l(\theta \mid x) = \sum_\alpha l(\theta \mid x_\alpha^0)\,\delta_\alpha(x)+\sum_s C_s(x)\psi_s(\theta)$$

and the model is an extended exponential model of order $\leqslant r$.

Corollary : *The order of an extended exponential model is the dimension of the smallest linear space V containing the structural parameters.*

$$\text{order} = \dim \{\sum_x V(x)\}.$$

Proof : Let $f(x \mid \theta)$ be an extended exponential model with basis parameters $\psi_1(\theta), \ldots, \psi_r(\theta)$. By the first part of the preceding proof,

$$V(x) \subset \text{LS}\,\{\psi_1(\theta), \ldots, \psi_r(\theta)\}.$$

Thus $\dim V \leqslant r$. By the second part of the proof the likelihood function can be expressed in terms of a basis for V, thus $r \leqslant \dim V$. Hence $r = \dim V$.

The structural dimension of an extended exponential model may be less than its order.

$$d = \max_x \dim V(x) \leqslant \dim \{\sum_x V(x)\}.$$

An extended exponential model on a connected open space A is an ordinary exponential model ; and its structural dimension must be at least one; otherwise it is a trivial model *not* dependent on θ.

Theorem 2 : *For a sample of r from an extended exponential model of order r, the structural dimension is necessarily r.*

Proof : Let V be the smallest linear space containing $\sum_x V(x)$; V has dimension r; hence some r (possibly fewer) of the $V(x)$ will generate V. Let $x_1', \ldots, x_r'$ be r points such that $V(x_1')+\ldots+V(x_r') = V$. For a sample of r the dimension at the point $(x_1', \ldots, x_r')$ is

$$d(x_1', \ldots, x_r') = \dim\ V(x_1', \ldots, x_r').$$
$$= r.$$

Thus the structural dimension of the sample model is r.

6. Sampling and sufficiency

The characterization of the extended exponential model permits simple and direct proofs for some Koopman-Darmois-Pitman theorems. These theorems are formulated in terms of structural dimension but *the concluding remarks in* Section 3 *allow a simple translation into dimensions for a sufficient statistic or minimal sufficient statistic.*

Theorem 3 : *If $x_1, \ldots, x_n$ are independent R-regular variables involving a parameter θ and if the structural dimension d for the variable $(x_1, \ldots, x_n)$ is achieved by at least two of the component variables, then all the variables are extended exponential with structural basis in a linear space over r parameters.*

Proof: Suppose that x_1 and x_2 achieve maximum structural dimension d at the points x_1^0 and x_2^0; let $V_1 = V_1(x_1^0)$, $V_2 = V_2(x_2^0)$. Then

$$V_1 \subset V_1(x_1^0)+V_2(x_2)+\ldots+V_n(x_n) = V(x_1^0, x_2, \ldots, x_n).$$

But $d = \dim V_1 \leqslant \dim V(x_1^0, x_2, \ldots, x_n) \leqslant d$, hence these must be equalities and $V_2(x_2), \ldots, V_n(x_n) \subset V_1$. By symmetry it follows that $V_1(x_1), V_3(x_3), \ldots, V_n(x_n) \subset V_2$. Hence $V_1 = V_2$, and all of $V(x_1), \ldots, V(x_n)$ are contained in $V_1 = V_2$. The proof is then completed by using Theorem 1 in Section 5.

For identically distributed variables a slight modification of this theorem seems appropriate.

Theorem 4 : *For an R-regular variable the structural dimension for a sample of n is equal to the structural dimension for a sample of $n+1$ if and only if the variable is an extended exponential variable.*

Proof: Let $(x_1^0, \ldots, x_n^0)$ be a point at which the maximum structural dimension d is attained for a sample of n. Then

$$d = \dim V(x_1^0, \ldots, x_n^0) \leqslant \dim V(x_1^0, \ldots, x_n^0, x_{n+1}) \leqslant d.$$

Hence $V(x_{n+1}) \subset V(x_1^0, \ldots, x_n^0)$ and by Theorem 1 in Section 5 it follows that the variable is extended exponential with d basis parameters.

The following theorem adds an extra conclusion to that found in Theorem 5.1 of Barankin and Maitra (1963).

Theorem 5 : *If the structural dimension of n independent R-regular variables with parameter θ is $s < n$, then $n-s$ of the variables are extended exponential with basis in an s-dimensional linear space V^0 which is generated by the likelihood gradient for the remaining s variables at some sample point. And if the structural dimension of these $n-s$ exponential variables is t, then an additional t variables are extended exponential with basis also in V^0.*

The second statement in the theorem says in effect that typically more than $n-s$ of the variables are extended exponential with a certain s structural parameters, and the only way that it can be just $n-s$ is for the structural dimension of the exponential variables to be zero, i.e., no variation with θ in the relative density within connected open sets. Then in particular for variables on connected open sample spaces, the number of extended exponential variables must be at least $n-s+1$, assuming that none of the variables is independent of θ.

Proof: (a) Let $(x_1^0, \ldots, x_n^0)$ be a point such that $V(x_1, \ldots, x_n)$ has maximum dimension s, and let

$$V^0 = V(x_1^0, \ldots, x_n^0) = V(x_1^0)+\ldots+V(x_n^0).$$

The s-dimensional space V^0 can be generated by some s (perhaps fewer) of the component spaces, say $V(x_1^0), \ldots, V(x_s^0)$. Consider then the relationship

$$V^0 = V(x_1^0)+\ldots+V(x_s^0) \subset V(x_1^0)+\ldots+V(x_s^0)+V(x_{s+1})+\ldots+V(x_n),$$

the left side has dimension s, the right side dimension at most s. Hence

$$V(x_{s+1}), \ldots, V(x_n) \subset V^0,$$

and by Theorem 1 in Section 5 the variables $(x_{s+1}, \ldots, x_n)$ are extended exponential with basis in the linear space V^0. By letting $\psi_1(\theta), \ldots, \psi_s(\theta)$ be a basis for V^0 this completes the first part of the theorem.

(b) Let t be the structural dimension of the exponential variables $(x_{s+1}, \ldots, x_n)$ with basis in V^0, let $(x^*_{s+1}, \ldots, x^*_n)$ be a point at which the maximum dimension t is attained, and let

$$V^* = V(x^*_{s+1}) + \ldots + V(x^*_n).$$

The t-dimensional space V^* can be generated by some t (perhaps fewer) of the component spaces, say $V(x^*_{n-t+1}), \ldots, V(x^*_n)$. Consider then

$$\begin{aligned} & V(x^0_1) + \ldots + V(x^0_{n-t}) + V(x^*_{n-t+1}) + \ldots + V(x^*_n) \\ & = V(x^0_1) + \ldots + V(x^0_{n-t}) + V^*. \end{aligned}$$

The first s of the component spaces generate V^0, and then with total dimension at most s it follows that the space is V^0. The space V^0 can then be generated from V^* by adding some $s-t$ (perhaps fewer) of the spaces $V(x^0_1), \ldots, V(x^0_{s-t})$, say the spaces $V(x^0_1), \ldots, V(x^0_{s-t})$. Consider then relationship :

$$\begin{aligned} V^0 &= V(x^0_1) + \ldots + V(x^0_{s-t}) + V(x^*_{n-t+1}) + \ldots + V(x^*_n) \\ &\subset V(x^0_1) + \ldots + V(x^0_{s-t}) + V(x_{s-t+1}) + \ldots + V(x_{n-t}) + V(x^*_{n-t+1}) + \ldots + V(x^*_n). \end{aligned}$$

The left side has dimension s; the last expression has dimension at most s. Hence

$$V(x_{s-t+1}), \ldots, V(x_{n-t}) \subset V^0$$

and by Theorem 1 in Section 5 the variables $x_{s-t+1}, \ldots, x_s$ are extended exponential with basis in V^0.

This theorem removes some of the restrictions found in the Barankin-Maitra theorem. It applies to vector-valued Euclidean variables generally rather than to real-valued variables, it does not have continuity and differentiability conditions with respect to the parameter, the open sample space need not be connected, the density function need not be analytic and it need not have a *continuous* gradient on the sample space as assumed in the Barankin-Maitra local analysis.

For identically distributed variables the theorem can be modified conveniently.

Theorem 6 : *If the structural dimension of a sample of n from an R-regular variable is $s < n$, then the variable is extended exponential of order s.*

REFERENCSE

BARANKIN, E. W. and KATZ, M. L., JR. (1959) : Sufficient statistics of minimal dimension. *Sankhyā*, **21**, 217-246.

BARANKIN, E. W. and MAITRA, A. P. (1963) : Generalization of the Fisher-Darmois-Koopman-Pitman theorem on sufficient statistics. *Sankhyā*, **25**, Series A, 217-244.

DYNKIN, E. B. (1951) : Necessary and sufficient statistics for a family of probability distributions. *Uspchi. Mat. Nauk* (N.S.), **6**, No. 1 (41), 68-90. (Also, *Selected Transl. Math. Stat. Prob.*, 17-40).

FRASER, D. A. S. (1963) : On sufficiency and the exponential family. *J. Roy. Stat. Soc.*, B, **25**, 115-123.

Paper received : November, 1964.