C.P. 31.3

NOTIONS OF SUFFICIENCY

by

S.L. Lauritzen

Contributed Paper
44th Session of the
International Statistical Institute

Madrid - España
September 12th - September 22nd 1983

NOTIONS OF SUFFICIENCY

Steffen L. Lauritzen

Aalborg University Centre, Denmark

INTRODUCTION

Fisher (1925) noted when discussing the notion of sufficiency that if $\hat{\theta}_m$ was a sufficient estimate of θ based on $x_1,\dots,x_m$ and $\hat{\theta}_n$ was a sufficient estimate of θ based on another sample $y_1,\dots,y_n$, then a sufficient estimate $\hat{\theta}_{mn}$ based on the combined sample would satisfy

$$\hat{\theta}_{mn} = f_{mn}(\hat{\theta}_m,\hat{\theta}_n)$$

provided the samples were independent. This recursive property of a sufficient statistic seems to have played a less prominent role in the discussion of sufficiency than deserved.

The property breaks down when samples are not independent. Notions of sufficiency developed to deal with dependent variables (e.g. stochastic processes) have however strong connections to this recursiveness.

RESULTS

Let $X_1,\dots,X_n,\dots$ be a sequence of random variables, whose distribution is governed by the law P_θ where $\theta\in\Theta$ is unknown. We shall consider sequences of statistics $t_n(x_1,\dots,x_n)$ and assume these to be sufficient in the classical sense, i.e. that the conditional distribution of $(X_1,\dots,X_n)$ given $Y_n = t_n(X_1,\dots,X_n) = y$ does not depend on θ. Suppose for technical reasons that all the random variables take values in discrete and at most countable sets.

We shall further say, that the sequence of statistics is

summarizing if for all n

$$p(x_1,\ldots,x_n|\theta) = f(t_n(x_1,\ldots,x_n)|\theta),$$

algebraically transitive if for all n

$$t_{n+1}(x_1,\ldots,x_n,x_{n+1}) = \phi_n(t_n(x_1,\ldots,x_n),x_{n+1})$$

transitive if for all n, $(X_1,\ldots,X_n)$ and Y_{n+1} are conditionally independent given Y_n

totally sufficient if for all n, $(X_1,\ldots,X_n)$ and $(X_{n+1},X_{n+2},\ldots)$ are conditionally independent given Y_n.

The notions "summarizing" and "algebraically transitive" were introduced by Freedman (1962), the latter under the name "S-structure". Transitivity was introduced by Bahadur (1954) and total sufficiency by Lauritzen (1974).

One can now show the following:

If (t_n) is totally sufficient and algebraically transitive, it is transitive.

If (t_n) is summarizing and algebraically transitive, it is totally sufficient and transitive.

If one further defines (t_n) to be minimal totally, sufficient, if it is a function of any other totally sufficient sequence (s_n), i.e. $t_n = \phi_n(s_n)$, we have:

It (t_n) is minimal totally sufficient, it is algebraically transitive and transitive.

Thus a minimal totally sufficient statistic has most of the nice properties one can hope for. Such a statistic can be shown to exist and essentially by uniquely defined by the conditional likelihood function which is obtained by considering

$$L(\theta, x_{n+1}, \ldots, x_{n+k} | x_1, \ldots, x_n) = p(x_1, \ldots, x_n | x_{n+1}, \ldots, x_{n+k}; \theta)$$

for all values of k. That is, the conditional likelihood function is obtained by conditioning on the future $(x_{n+1}, \ldots, x_{n+k})_{k \in \mathbb{N}}$, and treating this as a parameter.

For proofs of the results, see e.g. Lauritzen (1982) ch. II.

EXTENSIONS AND GENERALISATIONS

Most of the results can immediately be carried over to more general situations than the discrete case, since they in a sense are based on properties of the notion adequacy introduced by Skibinsky (1967), and on fundamental properties of conditional independence. The notion "summarizing" does of course have no meaning unless a kind of uniformity is defined, which e.g. could be given by group invariance. An interesting consequence of the results is to consider inference in stochastic processes based on the conditional likelihood, given the future.

BIBLIOGRAPHY

Bahadur R.R. (1954). Sufficiency and statistical decision functions. Ann. Math. Statist., 25, 423-462.

Fisher, R.A. (1925). Theory of statistical estimation. Proc. Camb. Philos Soc., 22, 700-725.

Freedman, D.A. (1962). Invariants under mixing which generalize de Finetti's theorem. Ann. Math. Statist., 34, 1194-1216.

Lauritzen, S.L. (1974). Sufficiency, prediction and extreme models. Scand. J. Statist., 1, 128-134.

Lauritzen, S.L. (1982). Statistical Models as Extremal Families. Aalborg University Press.

Skibinsky, M. (1967). Adequate subfields and sufficiency. Ann. Math. Statist., 38, 155-161.

SUMMARY

We discuss various extensions on notions related to sufficiency, with stochastic processes in mind.

RESUME

Considérons des extensions de la notion de sufficience en vue des processus stochastiques.