SUFFICIENT STATISTICS IN THE CASE OF INDEPENDENT RANDOM VARIABLES[1]

By L. Brown

Cornell University

1. Introduction. In many statistical situations the information obtained from the observation of n-independent identically distributed real random variables $X_1, \cdots, X_n$ can be condensed into one "sufficient statistic", $\phi(x_1, \cdots, x_n)$. In a well known sense the statistic ϕ contains as much information about the distribution of $X_1, \cdots, X_n$ as do the observations $x_1, \cdots, x_n$ themselves [1].

The Neyman factorization theorem [6], [9] gives one characterization of the situations in which a sufficient statistic can be employed. Suppose the distribution of each X_i is a priori known to be one of the distributions in the set $\{P_\theta(\cdot): \theta \varepsilon \Theta\}$ where each $P_\theta(x)$ has density $p_\theta(x)$ with respect to a fixed σ finite measure μ. Neyman's theorem tells how the densities $\{p_\theta(\cdot)\}$ must be related to each other through any statistic which is sufficient for the problem.

A more definitive characterization valid under certain additional assumptions of the densities $p_\theta(\cdot)$ in terms of the sufficient statistic is given by Koopman [7], and Darmois [3]. A further related result was proved by Dynkin [4]. This characterization states exactly what the functional form of the possible densities must be—specifically, that each density must be a member of a certain exponential family of densities (sometimes called a Koopman-Darmois family). This family is determined by the sufficient statistic.

The assumptions in the theorems of [3] and [7] include significant limitations on the form of the densities and on the form of the sufficient statistics. Dynkin [4] states a theorem in which a very minimal assumption is made on the form of the sufficient statistic, but the form of the densities involved is significantly restricted.

In the first main theorem of this paper—Theorem 2.1—a different approach is used. Almost the entire burden of the assumptions is on the form of the statistics involved. The second main theorem—Theorems 8.1 and 8.1′—makes one assumption on the form of ϕ which is generally satisfied. The remainder of its hypotheses are very weak. The conclusion is of a local nature, as opposed to the global nature of the conclusion of Theorem 2.1. These results are a fairly complete characterization of the situation when the conclusion is valid that each density is a member of a certain exponential family of densities.

Only the case of a real sufficient statistic is considered in detail in this paper. Some analogous results are clearly true for n-dimensional or even more general sufficient statistics. I hope to pursue these questions in a later paper.

Received 30 December 1963.

[1] This research was supported in part by the Office of Naval Research contract number Nonr-266(04) (NR 047-005).