SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Edited By : P. C. MAHALANOBIS

Vol. 10	Part 4	1950

COMPLETENESS, SIMILAR REGIONS, AND UNBIASED ESTIMATION—PART I

By E. L. LEHMANN

University of California

AND

HENRY SCHEFFÉ

Columbia University

1. Introduction

The aim of this paper is the study of two classical problems of mathematical statistics, the problems of similar regions and of unbiased estimation. The reason for studying these two problems together is that both are concerned with a family of measures and that essentially the same condition on this family insures a very simple solution of both.

The concepts of similar region and unbiased estimate were introduced at an early stage in the development of statistical theory, and both have proved extremely fruitful. On the other hand it seems rather difficult to justify either in a completely systematic development of statistics.

Similar regions were defined by Neyman and Pearson (1933) in connection with the problem of testing composite hypotheses. Suppose one wishes to test the hypothesis H that a random variable X, taking on values in a "sample space" of points x, is distributed according to some probability distribution P_θ^x of a family $\mathfrak{P}^x = \{P_\theta^x \mid \theta \epsilon \omega\}$. If the hypothesis is to be tested at level of significance α, the critical region A must satisfy the condition

$$P_\theta \ (A) \leqslant \alpha \qquad \text{for all } \theta \text{ in } \omega. \qquad \dots \ (1.1)$$

Neyman and Pearson replaced this by the much stronger condition of similarity

$$P_\theta^x \ (A) = \alpha \qquad \text{for all } \theta \text{ in } \omega. \qquad \dots \ (1.2)$$

In a number of important cases the problem reduces by this device to that of testing

J. Rojo (ed.), *Selected Works of E. L. Lehmann*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1412-4_23,

a simple hypothesis, and furthermore there exists among the similar regions one which is either uniformly most powerful or uniformly most powerful unbiased (Scheffé, 1942; Lehmann, 1947). On the other hand, in the same cases the most powerful test (among all those satisfying (1.1)) frequently depends very strongly on the specific alternative at which the power is maximized (Lehmann and Stein, 1948). In these situations therefore the simplifying value of the restriction (1.2) is considerable.

Somewhat analogous remarks hold for the problem of point estimation. Given again a family of measures $\mathfrak{P}^x = \{P_\theta^x \mid \theta \epsilon \omega\}$, it is desired to estimate some real-valued function $g(\theta)$. A reasonable estimate T would seem to be one which minimizes the risk for some given weight function; say, which minimizes

$$E_\theta[T - g(\theta)]^2 \qquad \dots \quad (1.3)$$

where E_θ denotes the expected value calculated with P_θ^x. Unfortunately, this estimate will depend on the value of θ for which (1.3) is minimized. If this value is θ_0 one clearly chooses $T = g(\theta_0)$. One way of avoiding this difficulty would be to replace (1.3) by

$$\sup_\theta \; E_\theta[T - g(\theta)]^2. \qquad \dots \quad (1.4)$$

Another approach consists in restricting the class of estimates by the intuitively rather appealing condition of unbiasedness

$$E_\theta(T) = g(\theta) \qquad \text{for all } \theta \text{ in } \omega. \qquad \dots \quad (1.5)$$

It appears that among the unbiased estimates there frequently exists one minimizing (1.3)—which now becomes the variance of T—uniformly in θ. (This clearly cannot happen if we omit the condition of unbiasedness unless $g(\theta)$ is constant).

In order to obtain the results alluded to above, we introduce the notion of completeness of a family of measures $\mathfrak{M}^x = \{M_\theta^x \mid \theta \epsilon \omega\}$. The family $\mathfrak{M}^x$ is said to be complete if

$$\int f(x) \, d \, M_\theta^x \, (x) = 0 \quad \text{for all } \theta \text{ in } \omega \qquad \dots \quad (1.6)$$

implies $f(x) = 0$ except on a set N with $M_\theta^x(N) = 0$ for all θ in ω. The family is said to be boundedly complete if this implication holds for all bounded functions f. Now let X be a random variable with distribution P_θ^x in the family $\mathfrak{P}^x$, let T be a sufficient statistic (not necessarily real-valued) for $\mathfrak{P}^x$, and denote by P_θ^t the distribution of T. We shall show that if $\mathfrak{P}^t$ is complete, then for any real valued estimable function $g(\theta)$ there exists an unbiased estimate with uniformly smallest variance, and that this is the only unbiased estimate of $g(\theta)$ which is a function of T only. This result is an immediate consequence of a theorem of Rao (1945) and Blackwell (1947) on unbiased estimation. We shall show also that whenever $\mathfrak{P}^t$ is boundedly complete, all similar regions A have a very simple structure (roughly speaking, that the conditional probability of X falling in A is independent of t), which was first described by Neyman

(1937), and that as a result the original composite hypothesis in these cases is reduced essentially to a simple one. The first to employ this method for finding the totality of similar regions was P. L. Hsu (1941).

The applicability of the above remarks hinges on the existence of a sufficient statistic T such that $\mathfrak{P}^t$ is complete (or boundedly complete). Now in general there are many different sufficient statistics, and the question arises, as to which has the best chance of satisfying the completeness condition. Actually, the question as to which is the appropriate sufficient statistic arises also in other statistical problems. Speaking intuitively, one makes use of sufficient statistics in order to reduce the complexity of a statistical problem without losing information of value. The latter condition being guaranteed by sufficiency, one is led to seek that sufficient statistic which reduces the statistical problem as far as possible, and hence to the definition (to be stated more precisely later with appropriate null set qualifications): A sufficient statistic T is said to be minimal if T is a function of all other sufficient statistics. Using this definition we prove that whenever a minimal sufficient statistic exists, a sufficient statistic can satisfy the condition of completeness only if it is equivalent to the minimal sufficient statistic in a certain sense. We also establish the existence of the minimal sufficient statistic when the sample space is Euclidean and the distribution of the sample possesses a generalized probablity density (defined in section 6), and we give a method of constructing it, which we show to be valid in this case. We remark that the result of our construction of the minimal sufficient statistic is essentially equivalent to the definition of sufficient statistic adopted by Koopman (1936) in a more special setting.

Some results for the problem of unbiased estimation are also found in the case where the minimal sufficient statistic is not complete: A formal theory is obtained which characterizes those estimable functions possessing unbiased estimates of uniformly minimum variance and these estimates. Finally a justification is given (at the end of section 4) for the restriction to sufficient statistics when testing hypotheses. (A justification in the case of point estimation was given by Rao and Blackwell). By permitting randomised decision functions, we show that given any test and any sufficient statistic there exists a test based only on the sufficient statistic and having identically the same power function as the given test.

In this paper (with the exception of section 6) the theory is developed in greater generality than is customary in statistical papers. The reason for not limiting ourselves to Euclidean spaces is that these are insufficient both for problems that have already arisen in sequential analysis and for statistical problems that may be expected to arise in connection with stochastic processes. Many of the difficulties encountered in the paper are associated with exceptional null sets. We believe that most of these difficulties are inherent in the nature of the problems treated, and that they would arise even if considerations were limited to Euclidean spaces.

It came to our attention while this paper was in proof that some of our results on minimum variance estimates were obtained earlier in papers by Rao (1947, 1949) and in an abstract by Seth (1949).

The material outlined above forms part I of the present study. In part II the general theory is applied to a number of more special problems. As the main application some theorems on similar regions (Neyman, 1941), type B_1 tests (Scheffé, 1942), and uniformly most powerful one-sided tests (Lehmann, 1947) are simplified and extended. These results are obtained by solving the differential equations introduced by Neyman and Scheffé. The theory of part I is then applied to the families of probability densities which are solutions of these equations, and also to some more general families of probability densities. Some of these results were summarized in a previous publication (Lehmann and Scheffé, 1947), and of these some were obtained independently and have been published since then by Ghosh (1948). Applications are also made to some non-parametric problems of estimation and testing. For example, a very simple proof is given of Scheffé's theorem (1943) concerning similar regions in the non-parametric case, this result is generalized, and the solution is given of a problem formulated by Halmos (1946) concerning point estimation.

2. Terminology and Notation

Unfortunately a considerable complexity of terminology and notation seems unavoidable; to minimize this we adopt the following conventions: Several spaces, to be denoted by W^x, W^t, etc., will have to be considered; here W^x is the whole space of points x, W^t is the space of t, etc. In each space there will be a fixed countably additive family of sets, $\mathfrak{F}^x$ in W^x, $\mathfrak{F}^t$ in W^t, etc. Here a family $\mathfrak{F}^x$ is said to be *countably additive* if it contains W^x, and with any set A in $\mathfrak{F}^x$ its complement $W^x - A$, and with any countable (i.e., finite or denumerable) number of sets in $\mathfrak{F}^x$ also their union. The sets in $\mathfrak{F}^x$ will be called *measurable* ($\mathfrak{F}^x$). We shall need to define *measurable functions* only for the case of real-valued functions: A function $f(x)$ defined on W^x is said to be measurable ($\mathfrak{F}^x$) if for every real c the set $\{x \mid f(x) < c\}$ is in $\mathfrak{F}^x$. A non-negative set function M^x defined for all A in $\mathfrak{F}^x$ is said to be a *measure* on $\mathfrak{F}^x$ if it is countably additive, that is, if for any disjoint sets $A_1, A_2, \ldots$ in $\mathfrak{F}^x$, $M^x(\cup_i A_i) = \Sigma_i M^x(A_i)$. A *probability measure* on $\mathfrak{F}^x$ is a measure M^x on $\mathfrak{F}^x$ for which $M^x(W^x) = 1$. Probability measures will usually be denoted by P^x, P^t, etc.

A family of measures M^x on $\mathfrak{F}^x$ will be denoted by $\mathfrak{M}^x$. It is convenient to index the members of the family by a subscript θ that takes on values in an abstract space ω, $\mathfrak{M}^x = \{M_\theta^x \mid \theta \epsilon \omega\}$. Similarly, we may write $\mathfrak{P}^x = \{P_\theta^x \mid \theta \epsilon \omega\}$ if the measures are probability measures. A set A in $\mathfrak{F}^x$ for which $M_\theta^x(A) = 0$ will be called a *null set for the measure* M_θ^x. A set will be called a *null set for the family* $\mathfrak{M}^x$ if it is a null set for every measure in the family. If a statement about the points of W^x is true for all x in $W^x - N$, we shall say it is valid *almost everywhere* (M_θ^x) if N is a null set for M_θ^x, we shall say it is valid *almost everywhere* ($\mathfrak{M}^x$) if N is a null set for $\mathfrak{M}^x$, and we shall abbreviate this by writing (*a.e.* M_θ^x) or (*a.e.*$\mathfrak{M}^x$) after the statement. An arbitrary function $t(x)$ from W^x to a space W^t generates a countably additive family $\mathfrak{F}^t$ of sets in W^t and a family $\mathfrak{M}^t$ of measures on W^t : the family $\mathfrak{F}^t$ consists of all those sets in W^t whose pre-images ("complete counter images") are in $\mathfrak{F}^x$, while the measure

M_θ^t on $\mathfrak{F}^t$ corresponding to M_θ^x on $\mathfrak{F}^x$ is defined for B in $\mathfrak{F}^t$ by $M_\theta^t(B) = M_\theta^x(t^{-1}(B))$, where $t^{-1}(B) = \{x \mid t(x) \epsilon B\}$ is the pre-image of B. It is easily seen that with this definition of $\mathfrak{F}^t$, a real-valued function $f(t)$ defined on W^t is measurable ($\mathfrak{F}^t$ if and only if the function $f(t(x))$ is measurable ($\mathfrak{F}^x$).

A family $\mathfrak{D}$ of sets D which are disjoint and cover W^x will be called a *deocmposition* of W^x; if all the elements D in $\mathfrak{D}$ are measurable ($\mathfrak{F}^x$) we will say that $\mathfrak{D}$ is a *measurable* ($\mathfrak{F}^x$) *decomposition.* If $t(x)$ is a function defined on W^x, and x^0 is any point of W^x, we shall say the set $\{x \mid t(x) = t(x^0)\}$ is a *contour* of $t(x)$ through x^0. Any function $t(x)$ defined on W^x determines a decomposition of W^x which may be denoted by $\mathfrak{D}_t$, the elements of $\mathfrak{D}_t$ being the contours of $t(x)$. Conversely, given any decomposition $\mathfrak{D}$ of W^x, a function $t(x)$ can be found (it is not unique, but the ranges of any two such functions are in 1:1 correspondence) such that $\mathfrak{D} = \mathfrak{D}_t$: For example, $t(x)$ may be taken to be the set-valued function whose value at x is the element of $\mathfrak{D}$ containing x. If for this function $t(x)$ we form according to the above description the countably additive family $\mathfrak{F}^t$, it turns out that it consists precisely of those sets of elements D of $\mathfrak{D}$ whose union is in $\mathfrak{F}^x$. This countably additive subfamily of $\mathfrak{F}^{x \mid \mathfrak{D}}$ generated by a decomposition $\mathfrak{D}$ or an associated function $t(x)$ will be denoted by $\mathfrak{F}^{x \mid \mathfrak{D}}$ or $\mathfrak{F}^{x \mid t}$. We say that a function $t(x)$ and decomposition $\mathfrak{D}$ are *associated* if $\mathfrak{D} = \mathfrak{D}_t$. For many purposes in probability and statistics only the decomposition associated with a function is of importance, it being immateral which of the different functions associated with the same decomposition is chosen.

We will encounter two kinds of equivalence for decompositions: Two decompositions $\mathfrak{D}$ and $\mathfrak{D}'$ of W^x are *equivalent* ($\mathfrak{P}^x$) *in the strict sense* if there exists a null set N for $\mathfrak{P}^x$ such that on W^x—N the decompositions $\mathfrak{D}$ and $\mathfrak{D}'$ coincide; $\mathfrak{D}$ and $\mathfrak{D}'$ are *equivalent* ($\mathfrak{P}^x$) *in the weak sense* if for every set A in $\mathfrak{F}^{x \mid \mathfrak{D}}$ there exists an A' in $\mathfrak{F}^{x \mid \mathfrak{D}'}$ and for every A$'$ in $\mathfrak{F}^{x \mid \mathfrak{D}'}$ there exists an A in $\mathfrak{F}^{x \mid \mathfrak{D}}$ such that A and A' differ by a null set for $\mathfrak{P}^x$, that is, $(A - A') \cup (A' - A)$ is a null set for $\mathfrak{P}^x$.

When one of the spaces being considered is the sample space of a statistical problem we shall always denote it by W^x. The family of measures on $\mathfrak{F}^x$ will then be a family $\mathfrak{P}^x$ of probability distributions. By saying that the random variable X (the "sample") is distributed according to P_θ^x we mean, as usual, that for any set A in $\mathfrak{F}^x$ the probability of X falling in A is P_θ^x (A). The term *statistic* will be used in this paper to mean any function of X, that is, if $t(x)$ is an arbitrary function of x defined on W^x, $T = t(X)$ will be called a statistic[1]. If $\mathfrak{F}^t$ and $\mathfrak{P}^t$ are defined as above then the probability of T falling in any set B in $\mathfrak{F}^t$ is $P_\theta^t(B)$. If $\mathfrak{D}_t$ is the decomposition of W^x associated with $t(x)$, it will be convenient to say also that it is associated with the statistic $T = t(X)$. We shall say that two statistics are *equivalent* ($\mathfrak{P}^x$) *in the strict sense*

[1] While for some purposes it may be convenient to restrict the definition of statistic, for example to functions whose associated decompositions are measurable ($\mathfrak{F}^x$), nothing is gained by such a restriction in this paper. We remark however that our construction of the minimal sufficient statistic in Section 6 does give a statistic satisfying this restriction.

or *in the weak sense* if their associated decompositions are. The reader may want to consider the statistical implications of these two kinds of equivalence of statistics.

Fundamental for this study is the notion of sufficient statistic. We base it on the definition given by Kolmogoroff (1933, p.41) of the conditional probability $P_\theta(A|t)$ of any set A in $\mathfrak{F}^x$ given that the statistic T has the value t: For each fixed θ and A, $P_\theta(A|t)$ is a real-valued point function of t, measurable ($\mathfrak{F}^t$), and defined implicitly by the equation

$$P_\theta^x(A \cap t^{-1}(B)) = \int_B P_\theta(A|t) dP_\theta^t, \qquad \dots \quad (2.1)$$

where B is any set in $\mathfrak{F}^t$, and (2.1) is regarded as an identity in B. For fixed θ and A, $P_\theta(A|t)$ is not defined uniquely, but if $f_{\theta,A}(t)$ and $g_{\theta,A}(t)$ are two determinations of $P_\theta(A/t)$, the set $N_{\theta,A}$ where they are unequal is a null set for P_θ^t. A statistic T is said to be a *sufficient statistic for* $\mathfrak{P}^x$ if there exists a determination of $P_\theta(A|t)$ indedent of θ, that is, if there exists a function $P(A|t)$ not depending on θ, measurable ($\mathfrak{F}^t$) for each fixed A, and such that for every θ in ω, every A in $\mathfrak{F}^x$, and any determination $P_\theta(A|t)$, $P(A|t) = P_\theta(A|t)$ except for t in a null set $B_{\theta,A}$ for P_θ^t. We note that two determinations of $P(A|t)$ must be equal except on a null set N_A for $\mathfrak{P}^t$. By putting $B = W^t$ in (2.1), we see that if T is a sufficient statistic for $\mathfrak{P}^x$, then for all A in $\mathfrak{F}^x$.

$$P_\theta(A) = \int_{W^t} P(A|t) dP_\theta^t. \qquad \dots \quad (2.2)$$

We shall have need also of the concept of conditional expectation as defined by Kolmogoroff (1933, p. 46). Suppose $\phi(x)$ is a real-valued measurable ($\mathfrak{F}^x$) function of x such that

$$E_\theta(\Phi) = \int_{W^x} \phi(x) d P_\theta, \qquad \dots \quad (2.3)$$

the expected value of the statistic $\Phi = \phi(X)$, calculated under the probability distribution P_θ^x, is finite. If $T = t(X)$ is a statistic (in general, not real-valued), then the conditional expected value of Φ, given $T = t$, calculated under P_θ^x, and to be denoted by $E_\theta(\Phi|t)$, is a point function of t, integrable ($\mathfrak{F}^t$, P_θ^t), defined implicitly by

$$\int_{t^{-1}(B)} \phi(x) d P_\theta^x = \int_B E_\theta(\Phi|t) d P_\theta^t, \qquad \dots \quad (2.4)$$

where B is in $\mathfrak{F}^t$, and (2.4) is regarded as an identity in B. For fixed θ, two determinations of $E_\theta(\Phi|t)$ are equal except on a null set N_θ of P_θ^t. By comparing the definitions (2.1) and (2.4) it is seen that if $\phi = \phi_A(x)$ is the characteristic function of a set A in $\mathfrak{F}^x$, then

$$P_\theta(A|t) = E_\theta(\Phi_A|t) \qquad (\text{a.e. } P_\theta^t). \qquad \dots \quad (2.5)$$

Returning to the general case of Φ with finite $E_\theta(\Phi)$, we remark that $E_\theta(\Phi|t)$ can also be calculated from the conditional probability $P_\theta(A|t)$ if this is known for all sets A in $\mathfrak{F}^x$, as was proved by Kolmogoroff (1933, p. 48). It then follows from the above definition of sufficient statistic that if T is a sufficient statistic for $\mathfrak{P}^x$, and if $E_\theta(\Phi)$ is finite for all θ in ω, there exists a measureable ($\mathfrak{F}^t$) point function of t, independent of θ, which we shall denote by $E(\Phi|t)$ and which has the property that for every θ in ω, $E(\Phi|t) = E_\theta(\Phi|t)$ except on a null set N_θ for $P_\theta{}^t$.

We shall have use for the following three formulae from the calculus of conditional probabilities: Suppose T is a sufficient statistic for $\mathfrak{P}^x$. (*i*) Then if $E_\theta(\Phi)$ is finite

$$\mathrm{E}_\theta(\Phi) = \int_{W^t} E(\Phi|t)\,d\,P_\theta{}^t. \qquad \ldots \quad (2.6)$$

This follows from (2.4) by taking $B = W^t$. (*ii*) If $f(t)$ is a real-valued measurable ($\mathfrak{F}^t$) function of t, and if $E_\theta(f(T)\Phi)$ and $E_\theta(\Phi)$ are finite for all θ in ω, then

$$E(f(T)\Phi|t) = f(t)E(\Phi|t) \qquad (\text{a.e. } \mathfrak{P}^t). \qquad \ldots \quad (2.7)$$

This formula is proved by Kolmogoroff (1933, p. 50), differently by Blackwell (1947, p. 105). (*iii*) It follows easily from (2.4) that if $c_1 \leqslant \phi(x) \leqslant c_2$ then

$$c_1 \leqslant E(\Phi|t) \leqslant c_2 \qquad (\text{a.e. } \mathfrak{P}^t). \qquad \ldots \quad (2.8)$$

We remark finally that the values of conditional probabilities and expectations, given a statistic, depend only on the decomposition associated with the statistic.

Of two statistics $T = t(X)$ and $T' = t'(X)$ we shall say *T is a function of T'* (a.e. $\mathfrak{P}^x$) if there exists a function $t = \psi(t')$ on $W^{t'}$ to W^t such that $t(x) = \psi(t'(x))$ (a.e. $\mathfrak{P}^x$). In terms of the decompositions $\mathfrak{D}_t$ and $\mathfrak{D}_{t'}$ associated with the functions $t(x)$ and $t'(x)$ this means there exists a null set N for $\mathfrak{P}^x$ such that on $W^x - N$ the decomposition $\mathfrak{D}_{t'}$ subdivides the decomposition $\mathfrak{D}_t$, that is every element of $\mathfrak{D}_{t'}$ is contained in an element of $\mathfrak{D}_t$. A sufficient statistic T for $\mathfrak{P}^x$ will be called a *minimal sufficient statistic* for $\mathfrak{P}^x$ if, for any other sufficient statistic T' for $\mathfrak{P}^x$, T is a function of T' (a.e. $\mathfrak{P}^x$). Sufficient conditions for the existence of minimal sufficient statistic, and a method of finding it, are given in section 6.

3. Completeness of a Family of Measures

Given a family $\mathfrak{M}^x = \{M_\theta{}^x|\theta\epsilon\omega\}$ of measures on the additive family $\mathfrak{F}^x$ of sets in W^x, consider integrals of the form

$$\int_{W^x} f(x)\,d\,M_\theta{}^x, \qquad \ldots \quad (3.1)$$

where $f(x)$ is real-valued and measurable ($\mathfrak{F}^x$). The value (if any) of this integral will in general depend on which measure $M_\theta{}^x$ of the family $\mathfrak{M}^x$ is used; (3.1) is a function of θ, which we may regard as a transform from a function f(x) defined on W^x to a function of θ defined on a part of ω. Under this transformation the function that is everywhere zero on W^x goes into the function that is everywhere zero on ω. Completeness means roughly that the zero function on W^x is the only function going into the zero function on ω; it is a unicity property of the transform. The exact definition is the following: The family $\mathfrak{M}^x$ of measures is *complete* if

$$\int_{W^x} f(x)\,d\,M_\theta{}^x = 0 \quad \text{for all } \theta \text{ in } \omega \qquad \ldots \quad (3.2)$$

implies $f(x) = 0$ (*a.e.* $\mathfrak{M}^x$). This definition of completeness is appropriate for the problem of unbiased estimation; for the problem of similar regions we require the weaker property of bounded completeness: The family $\mathfrak{M}^x$ is called *boundedly complete*[2] if the condition (3.2) and the condition that $f(x)$ is bounded jointly imply that $f(x) = 0$ (*a.e.* $\mathfrak{M}^x$).

We note that if $\mathfrak{M}^x$ is complete it is boundedly complete. The following simple example of a family of measures which is boundedly complete without being complete is a slight modification of an example constructed for a different purpose by Girshick, Mosteller, and Savage (1946):

Example 3.1: Suppose W^x is the real line, ω is the open interval $0<\theta<1$, and P_θ^x assigns the measure $(1-\theta)^2\theta^x$ to the points $x = 0, 1, 2...$, measure θ to the point $x = -1$, and measure zero to the complement of this set of points. The condition (3.2) then becomes

$$f(-1)\theta + \sum_{j=0}^{\infty} f(j)(1-\theta)^2\theta^j = 0.$$

We see that for $j = 0, 1, 2...$, $f(j)$ is the coefficient of θ^j in the Taylor series about the origin for the function $-f(-1)\theta(1-\theta)^{-2}$. Since for $|\theta|<1$,

$$\theta(1-\theta)^{-2} = \sum_{j=0}^{\infty} j\,\theta^j,$$

it follows that (3.2) is satisfied if and only if

$$f(j) = -jf(-1), \qquad j = 0, 1, 2,.... \qquad \dots \quad (3.3)$$

Hence if $f(x)$ satisfies (3.2) and is bounded, $f(j) = 0$ for $j = -1, 0, 1, 2,...$, that is, $f(x) = 0$ (*a.e.* $\mathfrak{P}^x$), and thus $\mathfrak{P}^x$ is boundedly complete. On the other hand, if $f(x)$ satisfies (3.3) and $f(-1) \neq 0$, then it satisfies (3.2) but not the condition $f(x) = 0$ (*a.e.* $\mathfrak{P}^x$), hence $\mathfrak{P}^x$ is not complete.

It is worth while to note that in general completeness or bounded completeness of a subset $\mathfrak{M}_1^x$ of $\mathfrak{M}^x$ does not imply the same for $\mathfrak{M}^x$, but that this implication does hold provided all null sets for $\mathfrak{M}_1^x$ are null sets for $\mathfrak{M}^x$.

We shall now give some simple examples of completeness. In each case the family $\mathfrak{P}^x$ whose completeness is discussed will be a family of probability measures that has been used extensively in statistics; it is of interest that in a number of these examples the question of completeness reduces to the problem of unicity for a transform that has been treated in the mathematical literature. In every case the space W^x will be the real line $-\infty<x<+\infty$, and the additive family $\mathfrak{F}^x$ on W^x may be taken as the class of Borel sets. The space ω of points θ may be taken as an appropriate set of real numbers. All the examples in which the family of measures is complete may be modified by replacing ω by certain proper subsets without destroying the completeness. We mention here that by means of some general theorems in Part II, one of which concerns completeness in product spaces, completeness may be proved for a large number of

[2] The property here called *bounded completeness* was called *completeness* by us in an earlier publication (1947).

families of measures in n-dimensional Euclidean spaces from these examples on the real line. Here and elsewhere in examples we use the "Nikodym derivative" notation

$$d\,P^x/d\mu^x = g(x)$$

to indicate that the measure P^x is absolutely continuous with respect to the measure μ^x and that

$$P^x(A) = \int_A g(x)d\mu^x$$

for all A in $\mathfrak{F}^x$.

Example 3.2: If $\mathfrak{P}^x$ is a family of normal probability distributions with mean θ and unit variance, then

$$d\,P_\theta^x/dx = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x-\theta)^2\right].$$

If the condition (3.2) is satisfied by $f(x)$ we find

$$\int_{-\infty}^{+\infty} f(x) \exp\left[-\frac{1}{2}x^2 + \theta x\right] dx = 0 \qquad \dots \quad (3.4)$$

for $-\infty < \theta < \infty$. Now (3.4) is the bilateral Laplace transform of the function $f(x)$ $\exp(-\frac{1}{2}x^2)$, and from the unicity theorem of this transform it follows that $f(x)$ $\exp(-\frac{1}{2}x^2) = 0 (a.e.\ L)$, where L denotes Lebesgue measure on $\mathfrak{F}^x$, and hence $f(x) = 0$ $(a.e.\ \mathfrak{P}^x)$. Thus $\mathfrak{P}^x$ is complete.

Example 3.3: For a family $\mathfrak{P}^x$ of normal distributions with zero mean and variance θ we have

$$d\,P_\theta^x/dx = (2\pi\theta)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\theta}x^2\right) \quad \text{for } 0 < \theta < \infty. \qquad \dots \quad (3.5)$$

This family is not complete or even boundedly complete because for every θ the density (3.5) is an even function of x, and hence (3.2) will be satisfied by any $f(x)$, which is an odd function of x and such that its product with (3.5) is integrable (L) for all $\theta > 0$. If we transform to a new set $\mathfrak{P}^t$ of measures P_θ^t by the transformation $t = x^2$ we find

$$dP_\theta^t/dt = \begin{cases} (2\pi t\theta)^{-\frac{1}{2}} \exp(-\frac{1}{2\theta}t) \text{ for } t > 0, \\ 0 \quad \text{for } t \leqslant 0. \end{cases}$$

Condition (3.2) written for $\mathfrak{P}^t$ instead of $\mathfrak{P}^x$ gives

$$\int_0^\infty f(t) t^{-\frac{1}{2}} \exp\left(-\frac{1}{2\theta}t\right) dt = 0 \qquad \dots \quad (3.6)$$

for $\theta > 0$. Letting $\tau = (2\theta)^{-1}$, and applying the unicity theorem for the unilateral Laplace transform we find $f(t) = 0$ $(a.e.\ \mathfrak{P}^t)$, that is, $\mathfrak{P}^t$ is complete.

Example 3.4: For a family $\mathfrak{P}^x$ of Cauchy distributions with zero median we have

$$dP_\theta^x/dx = [\pi\theta^{\frac{1}{2}}(1+x^2/\theta]^{-1} \qquad \text{for } \theta>0.$$

For symmetry reasons $\mathfrak{P}^x$ may again be seen not to be boundedly complete. If, as before, we let $t = x^2$, the completeness of the resulting $\mathfrak{P}^t$ follows from the unicity property of the Stieltjes transform

$$\int_0^\infty \frac{f(t)dt}{t+\theta}$$

Example 3.5: For a family of gamma distributions (chi-square distributions if θ is a half integer),

$$dP_\theta^x/dx = \begin{cases} 2^{-\theta}x^{\theta-1}\exp(-\frac{1}{2}x)/\Gamma(\theta) & \text{if } x>0, \\ 0 & \text{if } x \leqslant 0, \end{cases}$$

$\theta>0$, completeness follows from the unicity property of the Mellin transform

$$\int_0^\infty f(t)t^{\theta-1}\text{dt}.$$

Example 3.6: If P_θ^x is the uniform distribution on the interval $(0, \theta)$,

$$d\,P_\theta^x/dx = \begin{cases} 1/\theta \text{ if } 0<x<\theta, \\ 0 \text{ elsewhere,} \end{cases}$$

$\theta>0$, the completeness of $\mathfrak{P}^x$ follows from a theorem of Lebesgue : If

$$\int_0^\theta f(x)dx. = 0$$

for all θ in an interval, then $f(x) = 0$ (*a.e. L*) on the interval. The transform (3.1) of a function $f(x)$, which is involved in this case, is its indefinite integral.

Example 3.7: The family $\mathfrak{P}^x$ of uniform distributions on the intervals $(\theta, \theta+1)$,

$$dP_\theta{}^x/dx = \begin{cases} 1 \text{ if } \theta<x<\theta+1, \\ 0 \text{ elsewhere,} \end{cases}$$

$-\infty<\theta<\infty$, is not boundedly complete : If $f(x)$ is any periodic function with period 1 and

$$\int_0^1 f(x)dx = 0,$$

it is easily seen that

$$\int_{-\infty}^{+\infty} f(x)d\, P_\theta^x = 0$$

for all θ.

Example 3.8: Suppose $\mathfrak{P}^x$ is the family of Poisson distributions, so that P_θ^x assigns the measure $e^{-\theta}\theta^x/x!$ to the points $x = 0, 1, 2,\ldots$, and measure zero to the complement of this set. The condition (3.2) may be written

$$e^{-\theta} \sum_{j=0}^{\infty} f(j)\theta^j/j! = 0,$$

for $\theta > 0$. Since $f(j)$ is the coefficient of θ^j in the power series expansion of zero it follows that $f(x) = 0$ (*a.e.* $\mathfrak{P}^x$), and $\mathfrak{P}^x$ is complete.

Example 3.9: If $\mathfrak{P}^x$ is the family of binomial distributions corresponding to n independent trials with constant probability θ, $0<\theta<1$, P_θ^x assigns the probability

$$\binom{n}{x} \theta^x(1-\theta)^{n-x}$$

to the points $x = 0, 1,\ldots, n$ and zero probability elsewhere. Completeness of $\mathfrak{P}^x$ is implied by the theorem that if a polynomial of degree n vanishes for $n+1$ distinct values of the argument it is identically zero.

Example 3.10: Let $\mathfrak{P}^x$ be the family of hypergeometric distributions for fixed lot size N and fixed sample size n where θ is the number of defectives in the lot, $\theta = 0, 1,\ldots, N$; if P_θ^x is the probability of x defectives in the sample, P_θ^x assigns the discrete probabilites

$$\binom{\theta}{x} \binom{N-\theta}{n-x} \Big/ \binom{N}{n} \qquad \ldots (3.7)$$

to the points $x = 0, 1,\ldots, n$. In (3.7) we understand

$$\binom{m}{r} = 0 \text{ if } r<0 \text{ or } r>m.$$

Condition (3.2) then becomes

$$\sum_{j=0}^{n} f(j)\binom{\theta}{x} \binom{N-\theta}{n-x} = 0, \qquad \ldots (3.8)$$

and we find successively $f(0) = 0$, $f(1) = 0$, ..., $f(n) = 0$ by putting $\theta = 0, 1, \ldots, n$ in (3.8). Hence $\mathfrak{P}^x$ is complete.

Example 3.11: Let X_1, X_2, ... be a sequence of independent random variables each being capable of taking on the values 1 and 0 with probability p and $1-p$ respectively. Consider a sequential sampling scheme on these variables where the decision on whether or not to take an $N+1^{st}$ observation, when N observations have already been taken is made according to the value of $\sum_{i=1}^{N} X_i$. Let n be the total number of observations taken in one experiment—n is a random variable — and let $X = \sum_{i=1}^{n} X_i$.

The problem of bounded completeness for the family $\mathfrak{P}^x$ corresponding to a fixed stopping rule and the values $0<p<1$, was solved as follows in a series of papers by Girshick, Mosteller and Savage (1946), Wolfowitz (1946), and Savage (1947). For any N, consider $\left(\sum_{i=1}^{N} x_i,\ N - \sum_{i=1}^{N} x_i \right)$ as coordinates of a point in a plane, and define such a point to be accessible if it has positive probability under the sequential scheme being considered but is not a stopping point for this procedure so that, when this point is observed, another observation will be taken. As was shown in the above papers, a necessary and sufficient condition for bounded completeness is that, given any pair of accessible points P_1, P_2 corresponding to the same value of N (for any point P, N is the sum of the coordinates of P), all the points lying on the line segment connecting P_1 and P_2 and having integer valued coordinates are also accessible.

The concept of completeness developed in this section is related to that of minimality of a sufficient statistic, introduced at the end of secton 2, by

Theorem 3.1: *If T is a sufficient statistic for $\mathfrak{P}^x$ such that $\mathfrak{P}^t$ is boundedly complete, and if U is a minimal sufficient statistic for $\mathfrak{P}^x$, then T and U are equivalent ($\mathfrak{P}^x$) in the weak sense.*

The proof of this theorem can be based on the following

Lemma 3.1: *If $T = t(X)$ and $T' = t'(X)$ are two sufficient statistics for $\mathfrak{P}^x$, and if for every A in $\mathfrak{F}^x$, $P(A|t)$ and $P(A|t')$, considered as functions of x are equal (a.e. $\mathfrak{P}^x$), then T and T' are equivalent in the weak sense.*

To prove the lemma let $\mathfrak{F}^{x|t} = t^{-1}(\mathfrak{F}^t)$, and define $\mathfrak{F}^{x|t'}$ analogously. We shall prove that to any set A in $\mathfrak{F}^{x|t}$ there corresponds a set A' in $\mathfrak{F}^{x|t'}$ which differs from A by a null set for $\mathfrak{P}^x$. Let $\phi_A(x)$ be the characteristic function of the set A; then $\phi_A(x)$ depends on x through $t(x)$, say $\phi_A(x) = g(t(x))$. By (2.5) and (2.7),

$$P(A|t) = E(g(T)|t) = g(t) \qquad \text{(a.e. } \mathfrak{P}^t\text{)},$$

and thus

$$P(A|t(x)) = \phi_A(x) \qquad \text{(a.e. } \mathfrak{P}^x\text{)}.$$

But by hypothesis

$$P(A|t(x)) = P(A|t'(x)) \qquad \text{(a.e. } \mathfrak{P}^x\text{)}. \qquad \ldots \quad (3.9)$$

Define A' as the set in $\mathfrak{F}^{x|t'}$ where the right member of (3.9) equals 1. Now there exists a null set N for $\mathfrak{P}^x$ such that on $W^x - N$ the right member of (3.9) equals $\phi_A(x)$; hence the parts of A and A' in $W^x - N$ coincide, and the lemma is established.

Theorem 3.1 will thus be proved if we show that if there were an A_1 in $\mathfrak{F}^x$ for which $P(A_1/t(x))$ and $P(A_1|u(x))$, were not equal (a.e. $\mathfrak{P}^x$), then $\mathfrak{P}^t$ could not be boundedly complete. Consider the real-valued measurable ($\mathfrak{F}^x$) function $v(x)$ defined by

$$v(x) = P(A_1|t(x)) - P(A_1|u(x)).$$

We note that $|v(x)| \leqslant 1$ (a.e. $\mathfrak{P}^x$), and that the set in $\mathfrak{F}^x$ where $v(x) \neq 0$ is not a null set for $\mathfrak{P}^x$. If $V = v(X)$, it follows from (2.2) that

$$E_\theta(V) = 0 \qquad \ldots \quad (3.10)$$

for all θ in ω. Since U is a minimal sufficient statistic it is a function of T (a.e. $\mathfrak{P}^x$) and hence V is also a function of T (a.e. $\mathfrak{P}^t$). We can thus redefine $v(x)$ on a null set for $\mathfrak{P}^x$ so that the result is a measurable ($\mathfrak{F}^t$) function of t, say $f(t)$, with the properties that $|f(t)| \leqslant 1$, and the set in $\mathfrak{F}^t$ where $f(t) \neq 0$ is not a null set for $\mathfrak{P}^t$. This redefinition of $v(x)$ does not invalidate (3.10), and thus

$$\int_{W^t} f(t) dP_\theta^t = 0$$

for all θ in ω, and hence $\mathfrak{P}^t$ is not boundedly complete.

4. Similar Regions

A set A in $\mathfrak{F}^x$ is said to be a *similar region* of size α for the family $\mathfrak{P}^x$ of probability measures on $\mathfrak{F}^x$ if $P_\theta^x(A) = \alpha$ for all θ in ω. Neyman (1937) noted that if T is a sufficient statistic for $\mathfrak{P}^x$ and if the set A has the property

$$P(A|t) = \alpha \qquad \text{(a.e. } \mathfrak{P}^t\text{)}, \qquad \ldots \quad (4.1)$$

then A is a similar region of size α for $\mathfrak{P}^x$; this follows from (2.2). We whall say that a set A in $\mathfrak{F}^x$ has the *Neyman structure* with respect to the sufficient statistic T if it satisfies (4.1). Neyman did not investigate under what conditions, given a sufficient statistic T for $\mathfrak{P}^x$, all similar regions for $\mathfrak{P}^x$ have this structure with respect to T : this is of importance in the Neyman-Pearson theory of optimum tests, since there one wants to choose the "best" of *all* similar regions, and one therefore needs to know the totality of such regions. A partial answer is given by the following corollary to Theorem 4.1.

Corollary 4.1: *If T is a sufficient statistic for $\mathfrak{P}^x$, and if $\mathfrak{P}^t$ is boundedly complete, then a set A in $\mathfrak{F}^x$ is a similar region for $\mathfrak{P}^x$ if and only if it has the Neyman structure with respect to T.*

The situation in which the problem of similar regions arises is that sets A in $\mathfrak{F}^x$ are being considered as possible *critical regions* of statistical tests of a hypothesis : If the sample X falls in the critical region A the hypothesis is rejected by the test, while if it falls in $W^x - A$, the hypothesis is accepted. If A is a similar region for $\mathfrak{P}^x$, the probability of rejection is constant for all probability distribution in $\mathfrak{P}^x$. For many purposes it is convenient to employ a third category of points x such that when one of these points is observed one does not always reject or always accept, but rejects according to a chance method (say, with the help of a table of random mumbers), for which the probability of rejection is a predetermined number $\phi(x)$, $0<\phi(x)<1$. One may then extend the definition of the function $\phi(x)$ to the whole sample space W^x, by setting $\phi(x) = 1$ on the rejection set and $\phi(x) = 0$ on the acceptance set, and is thus led to the notion of critical function (Lehmann and Stein, 1948), which may be regarded as a special case of the randomized decision functions of Wald (1947): A *critical function* $\phi(x)$ is any measurable ($\mathfrak{F}^x$) function of x for which $0 \leqslant \phi(x) \leqslant 1$. Its use in testing hypotheses is that when $X=x$ one rejects the hypothesis with probability $\phi(x)$ according to a random process statistically independent of the random process governing X. The probability of rejection is then the expected value of $\Phi=\phi(X)$ calculated under P_θ^x, namely, $E_\theta(\Phi)$ as defined by (2.3). Critical regions are seen to correspond to the special case of critical functions which are characteristic functions of sets in W_x, that is, which take on only the values 0 and 1.

If $E_\theta(\Phi)$ is constant and equal to α for all θ in ω we shall say that $\phi(x)$ is a *similar critical function* of size α for $\mathfrak{P}^x$. Clearly, similar regions correspond to the special case of similar critical functions which are characteristic functions of sets: If A is a set in $\mathfrak{F}^x$ and ϕ_A is its characteristic function, then A is a similar region for $\mathfrak{P}^x$ if and only if ϕ_A is a similar critical function for $\mathfrak{P}^x$. We shall say that a critical function $\phi(x)$ has the *Neyman structure* with respect to a sufficient statistic T for $\mathfrak{P}^x$ if $E(\Phi|t) = \alpha$ (a.e. $\mathfrak{P}^t$). It is obvious from (2.5) that if a critical function ϕ is the characteristic function of a set A in $\mathfrak{F}^x$ then ϕ has the Neyman structure with respect to T if and only if A has the Neyman structure with respect to T. Therefore Corollary 4.1 follows from

Theorem 4.1: *If T is a sufficient statistic for $\mathfrak{P}^x$, a necessary and sufficient condition for all similar critical functions for $\mathfrak{P}^x$ to have the Neyman structure with respect to T is that $\mathfrak{P}^t$ be boundedly complete.*

We shall prove first sufficiency. Suppose $\phi(x)$ is a similar critical function of size α for $\mathfrak{P}^x$. Then from (2.6),

$$\int_{W^t} E(\Phi|t)d\, P_\theta^t = \alpha$$

for all θ in ω, so that

$$\int_{W^t} f(t)d\, P_\theta^t = 0 \qquad \dots \quad (4.2)$$

for all θ in ω, where

$$f(t) = E(\Phi|t)-\alpha. \qquad \dots \quad (4.3)$$

Since $0\leqslant\phi(x)\leqslant 1$, $0\leqslant E(\Phi)|t)\leqslant 1$ (a.e. $\mathfrak{P}^t$) by (2.8). Hence by taking a particular determination of $E(\Phi|t)$ in (4.3) and then redefining $f(t)$ on a null set for $\mathfrak{P}^t$, we may assume $f(t)$ to be bounded. Since $\mathfrak{P}^t$ is boundedly complete, (4.2) implies $f(t) = 0$ (a.e. $\mathfrak{P}^t$), that is, $\phi(x)$ has the Neyman structure with respect to T.

Next we shall prove necessity by showing that if $\mathfrak{P}^t$ were not boundedly complete there would exist similar critical functions (of every size α, $0<\alpha<1$) not having the Neyman structure with respect to T. If $\mathfrak{P}^t$ is not boundedly complete there exists a bounded measurable ($\mathfrak{F}^t$) function $f(t)$ such that

$$\int_{W^t} f(t)dP_\theta{}^t = 0$$

for all θ in ω, but

$$A_f = \{t|f(t) \neq 0\}$$

is not a null set for $\mathfrak{P}^t$. Suppose $|f(t)| < M$, and define

$$g(t) = c f(t)+\alpha,$$

where $0<c\leq M^{-1} \min \{\alpha, 1-\alpha\}$. Then $0\leqslant g(t)\leqslant 1$,

$$\int_{W^t} g(t)dP_\theta{}^t = \alpha \qquad \dots \quad (4.4)$$

for all θ in ω, and

$$g(t) \neq \alpha \text{ on } A_f. \qquad \dots \quad (4.5)$$

Now take as critical function $\phi(x) \equiv g(t(x))$. Then

$E(\Phi|t) = E(g(T)|t)$, and hence by (2.7)

$$E(\Phi|t) = g(t) \qquad (\text{a.e. } \mathfrak{P}^t). \qquad \dots \quad (4.6)$$

We see from (2.6), (4.6) and (4.4) that $\phi(x)$ is a similar critical function of size α for $\mathfrak{P}^x$, and from (4.6) and (4.5) that it does not have the Neyman structure wsth respect to T.

It may be appropriate to give here a simple example of a family $\mathfrak{P}^x$ of distributions for which there exist similar regions which do not have the Neyman structure with respect to a sufficient statistic T.

Example 4.1: Let $X = (X_1, X_2, \dots, X_n)$ be a random sample of size $n>1$ from the uniform distribution on the intervals $(\theta-\frac{1}{2}, \theta+\frac{1}{2})$. The sample space W^x may be taken as a Euclidean n-space, the additive family $\mathfrak{F}^x$ as the Borel sets in W^x, and ω as the real line $-\infty<\theta<+\infty$. The distribution $P_\theta{}^x$ of the sample is the uniform distribution on the n-dimensionsl cube $|x_i-\theta|<\frac{1}{2}$, $i = 1, 2, \dots, n$. Let $t_1(x) = \min_i x_i$, $t_2(x) = \max_i x_i$. The transformation $t = t(x) = (t_1(x), t_2(x))$ maps W^x into the two dimensional Euclidean space W^t, and the statistic $T = t(X) = (T_1, T_2)$ has the probability density

$$dP_\theta{}^t/d\mu^t = \begin{cases} c(t_2-t_1)^{n-2} & \text{if } \theta-\frac{1}{2}<t_1\leqslant t_2<\theta+\frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases} \qquad \dots \quad (4.7)$$

where $d\mu^t = dt_1dt_2$. It is known (it also follows from Example 6.4) that T is a sufficient statistic for $\mathfrak{P}^x$; it will be shown in Example 5.3 that $\mathfrak{P}^t$ is not boundedly complete. Similar regions not having the Neyman structure with respect to T can be based on the range $R = T_2-T_1$. Since R has a continuous distribution independent of θ, there exists for any $\alpha (0<\alpha<1)$ a constant r_α such that for all θ the probability is α that $R<r_\alpha$. The set

$$A_\alpha = \{x|t_2(x)-t_1(x)<r_\alpha)\}$$

is thus a similar region. Since the characteristic function of the set A_α can be expressed as a function of t, it follows from (2.5) and (2.7) that $P(A_\alpha|t)$ takes only the values 0 and 1 (a.e. $\mathfrak{P}^t$), that is, A_α does not have the Neyman structure with respect to T.

If in this example we take $n = 1$, an even simpler illustration of similar regions not possessing the Neyman structure with respect to a sufficient statistic can be given. On the interval (0,1) in W^x take any Borel set A_1 of Lebesgue measure α and then define A so that its characteristic function is periodic with period 1 and coincides with that of A_1 in (0,1). Then $X = X$ is a sufficient statistic for $\mathfrak{P}^x$ and $P(A|x) = 0$ or 1 (a.e. $\mathfrak{P}^x$).

We conclude this section by noting the following "justification" of sufficient statistics: Let Ω be a set in the θ-space containing ω, let $\bar{\mathfrak{P}}^x=\{P_\theta{}^x|\theta\epsilon\Omega\}$, and suppose the statistic $T=t(X)$ is sufficient for $\bar{\mathfrak{P}}^x$. Then given any critical function $\phi(x)$ for testing $H_0:\theta\epsilon\omega$, there exists a critical function $\phi_1(x)$ depending on x only through $t(x)$, and having the same power function as $\phi(x)$. The proof is similar to that of Rao and Blackwell's theorem (Section 5): Let $\psi(t) = E(\Phi|t)$ and let $\phi_1(x) = \psi(t(x))$; ψ is independent of θ in Ω since T is sufficient for $\bar{\mathfrak{P}}^x$. The power (probability of rejecting H_0) of $\phi(x)$ is

$$E_\theta(\Phi_1) = E_\theta(\psi(T)) = E_\theta(E(\Phi|T)),$$

and by (2.6) this equals $E_\theta(\Phi)$.

It should be remarked that an even stronger justification of sufficient statistics for all problems of statistical inference was given by Halmos and Savage (1949) who pointed out that by means of a sufficient statistic and the use of random variables with known distribution it is possible to construct a statistic having the same distribution as the original sample. Their argument however presupposes that there exists a determination of the conditional probability $P(A|t)$ that is a probability measure (a.e. $\mathfrak{P}^t$); the validity of this supposition has been established by Doob (1948, p.399) in the case that W^x is a Euclidean space. The justification for sufficient statistics given by us above for the problem of testing has the advantage of being valid without any restriction on the sample space.

5. Unbiased Estimation

In this section $\mathfrak{P}^x = \{P_\theta{}^x|\theta\epsilon\omega\}$ will denote the family of all distributions to which attention is restricted a priori in a particular problem of statistical inference. In considering real-valued statistics as possible solutions of problems of unbiased estimation we shall restrict ourselves to those with finite second moments. Let $\mathfrak{V}$ be the class of all real-valued statistics $V = v(X)$ for which $v(x)$ is measurable ($\mathfrak{F}^x$) and $E_\theta(V^2)$ is finite for all θ in ω. For any V in $\mathfrak{V}$, $E_\theta(V)$ is also finite for all θ in ω. We shall say that a real-valued function $g(\theta)$ is *estimable* if there exists a V in $\mathfrak{V}$ such that $E_\theta(V) = g(\theta)$ for all θ in ω. An unbiased estimate of $g(\theta)$ which is of uniformly minimum variance will be called a minimum variance estimate of $g(\theta)$, that is, a statistic V in $\mathfrak{V}$ is a *minimum variance estimate* of $g(\theta)$ if

(*i*) $E_\theta(V) = g(\theta)$ for all θ in ω, and

(*ii*) $\mathrm{Var}_\theta(V) \leqslant \mathrm{Var}_\theta(V')$ for all θ in ω and all V' in $\mathfrak{V}$ statisfying (*i*).

By applying the notion of completeness to a result of Rao and Blackwell, one obtains easily the following

Theorem 5.1: *If there exists a sufficient statistic T for $\mathfrak{P}^x$ such that $\mathfrak{P}^t$ is complete then every estimable function has a minimum variance estimate, and a statisic V in $\mathfrak{V}$ is a minimum variance estimate of its expected value if and only if it is a function of T (a.e. $\mathfrak{P}^x$).*

The result of Rao (1945) and Blackwell (1947) of which this theorem is a consequence states that if T is a sufficient statistic for $\mathfrak{P}^x$ and V in $\mathfrak{V}$ an unbiased estimate of $g(\theta)$, then if ψ is defined by $\psi(t) = E(V|t)$, $\Psi = \psi(T)$ is also an unbiased estimate of $g(\theta)$ with $\mathrm{Var}_\theta(\Psi) \leqslant \mathrm{Var}_\theta(V)$. equality holding for all θ if and only if V is a function of T (a.e. $\mathfrak{P}^x$).

To prove Theorem 5.1 suppose $g(\theta)$ is an estimable function; then there exists a V in $\mathfrak{V}$ for which $E_\theta(V) = g(\theta)$. Let Ψ be defined as above. To see that Ψ is a minimum variance estimate of $g(\theta)$, suppose V' in $\mathfrak{V}$ is any other unbiased estimate of $g(\theta)$, and define $\psi'(t) = E(V'|t)$. Then by the Rao-Blackwell theorem $\Psi' = \psi'(T)$ is an unbiased estimate of $g(\theta)$ and $\mathrm{Var}_\theta(\Psi') \leqslant \mathrm{Var}_\theta(V')$ for all θ. Since $E_\theta(\Psi' - \Psi) = 0$,

$$\int_{W^t} [\psi'(t) - \psi(t)] dP_\theta^t = 0$$

for all θ in ω, and hence it follows from the completeness of $\mathfrak{P}^t$ that $\psi'(t) = \psi(t)$ (a.e. $\mathfrak{P}^t$). Thus $\mathrm{Var}_\theta(\Psi) = \mathrm{Var}_\theta(\Psi') \leqslant \mathrm{Var}_\theta(V')$, and so Ψ is a minimum variance estimate of $g(\theta)$. The last part of the theorem is obtained from the condition for equality in the Rao-Blackwell theorem.

The application of Theorem 5.1 is illustrated by

Example 5.1: Suppose $\mathfrak{P}$ is the family of distributions of a random sample of size $n > 1$ from a normal population with mean θ_1 and variance θ_2. With $x = (x_1, x_2, \ldots, x_n)$ a point in the Euclidean space W^x, $d\mu^x = dx_1\, dx_2 \ldots dx_n$, and $\theta = (\theta_1, \theta_2)$ a point in the half plane $\omega = \{\theta \mid \theta_2 > 0\}$,

$$dP_\theta^x / d\mu^x = (2\pi\theta_2)^{-\frac{1}{2}n} \exp\left[-\frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2 \right]. \qquad \ldots \quad (5.1)$$

It is well-known that with

$$t_1(x) = \sum_{i=1}^{n} x_i / n,$$

$$t_2(x) = \sum_{i=1}^{n} \left[x_i - t_1(x) \right]^2,$$

$$t(x) = (t_1(x), t_2(x)),$$

$T = (T_1, T_2) = t(X)$ is a sufficient statistic for $\mathfrak{P}^x$ (in Example 6.2 it will be derived as a minimal sufficient statistic), and that T_1 and T_2 are independently distributed with joint probability density

$$dP_\theta^t/d\mu^t = \begin{cases} \dfrac{t_2^{\frac{n-3}{2}} e^{-\frac{1}{2\theta_2}[n(t_1-\theta_1)^2+t_2]}}{(2\theta_2)^{\frac{n}{2}}(\pi/n)^{1/2}\Gamma\left(\frac{n-1}{2}\right)} & \text{if } t_2 > 0, \\ 0 \text{ if } t_2 \leqslant 0. \end{cases}$$

In Part II of this study it will be proved that $\mathfrak{P}^t$ is complete. From the completeness of $\mathfrak{P}^t$ it follows at once that the following functions of the sample mean T_1 and sample variance T_2/n are minimum variance estimates of the indicated functions of θ_1 and θ_2: T_1 of θ_1; $T_2/(n-1)$ and $c_n\sqrt{T_2}$ of θ_2 and $\sqrt{\theta_2}$, respectively, where

$$c_n = 2^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}(n-1))\Gamma(\tfrac{1}{2}n);$$

$T_1 + a_p c_n\sqrt{T_2}$ of the (lower) 100 p percent point of the population, where a_p is determined from normal probability tables so that

$$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{a_p} \exp(-\tfrac{1}{2}X^2)dX = p.$$

These are all special cases of the statement following from Theorem 5.1 that any real-valued function of (T_1, T_2) with finite variance is a minimum variance estimate of its expected value.

Theorem 5.1 : can be extended easily to the case of simultaneous estimation of several real valued-functions of θ if, following Cramér (1946), we express the concentration of an unbiased estimate about its mean, in terms of its ellipsoid of concentration. Let $g_1(\theta), \ldots, g_m(\theta)$ be m real-valued functions of θ, and let $g(\theta) = (g_1(\theta), \ldots, g_m(\theta))$. We shall say that $V = (V_1, \ldots V_m)$ is an unbiased estimate of $g(\theta)$ if V_i is an unbiased estimate of $g_i(\theta)$ for $i = 1,\ldots m$. The statistic V will be said to be an estimate of $g(\theta)$ with maximum concentration if it is an unbiased estimate of $g(\theta)$, if $E_\theta(V_i^2) < \infty$ for $i = 1,\ldots,m$ and all θ in ω, and if for any other unbiased estimate U, the concentration elliposid of V is contained in that of U.

Theorem 5.2 : *If there exists a sufficient statistic* T *for* $\mathfrak{P}^x$ *such that* $\mathfrak{P}^t$ *is complete, then every estimable function* $g(\theta) = (g_1(\theta), \ldots, g_m(\theta))$ *for which* $g_1(\theta), \ldots, g_m(\theta), 1$ *are linearly independent, has an estimate with maximum concentration.*

The proof of this theorem follows easily from the following result[3] concerning quadratic forms: Let $\sum_{i,j=1}^{m} a_{ij}y_iy_j$, $\sum_{i,j=1}^{m} b_{ij}y_iy_j$ be two positive definite quadratic

[3] Pointed out to us by Professor E. W. Barankin.

forms, and let $\sum_{i,j=1}^{m} a^{ij}y_iy_j$, $\sum_{i,j=1}^{m} b^{ij}y_iy_j$ be the corresponding inverse forms. Then

$$\sum_{i,j=1}^{m} a_{ij}y_iy_j \leqslant \sum_{i,j=1}^{m} b_{ij}y_iy_j \text{ for all real } y\text{'s}$$

implies

$$\sum_{i,j=1}^{m} a^{ij}y_iy_j \geqslant \sum_{i,j=1}^{m} b^{ij}y_iy_j \text{ for all real } y\text{'s}.$$

Now let $V = (V_1,\ldots,V_m)$ be any unbiased estimate of $g(\theta)$, let $U_i = E(V_i|T)$, and let $U = (U_1,\ldots,U_m)$. Then for any real $y_1,\ldots y_m$, $\sum_{i=1}^{m} y_iV_i$ is an unbiased estimate of $\sum_{i=1}^{m} y_ig_i(\theta)$, and by Theorem 5.1, $\sum_{i=1}^{m} y_iU_i$ is a minimum variance estimate of $\sum_{i=1}^{m} y_ig_i(\theta)$. Comparing the variances of these two estimates we find

$$\sum_{i,j=1}^{m} \lambda^U{}_{ij}y_iy_j \leqslant \sum_{i,j=1}^{m} \lambda^V{}_{ij}y_iy_j$$

where $\lambda^U{}_{ij}$ and $\lambda^V{}_{ij}$ are the covariances of (Ui, U_j) and (V_i, V_j) respectively. Furthermore since $g_1(\theta),\ldots,g_m(\theta)$, 1 are linearly independent so are $V_1,\ldots,V_m,1$ and $U_1,\ldots,U_m,1$, and hence the above quadratic forms are positive definite. It follows that

$$\sum_{i,j=1}^{m} \lambda_U{}^{ij}y_iy_j \geqslant \sum_{i,j=1}^{m} \lambda_V{}^{ij}y_iy_j,$$

which proves the theorem since the ellipsoids of concentration of U and V are defined by the equations $\sum_{i,j=1}^{m} \lambda_U{}^{ij}y_iy_j = m+2$ and $\sum_{i,j=1}^{m} \lambda_V{}^{ij}y_iy_j = m+2$, respectively.

We shall now consider once more the problem of estimating a single real-valued function of θ, and we shall develop a formal theory of minimum variance estimates which answers certain questions about the class of estimable functions possessing minimum variance estimates, but which may be difficult to apply in specific problems. In addition to the class $\mathfrak{V}$ of statistics defined at the beginning of this section we define, for any sufficient statistic T, three classes $\mathfrak{V}_T$, $\mathfrak{V}_T{}^0$, and $\mathfrak{V}_T{}^1$. Here T will in general not be real-valued, and the application of the theory may be expected to be simplest if T is a minimal sufficient statistic for $\mathfrak{P}^x$. Let $\mathfrak{V}_T$ be the class of all statistics in $\mathfrak{V}$ which are functions of T (a.e. $\mathfrak{P}^x$). We define $\mathfrak{V}_T{}^0$ as the subclass of statistics V in $\mathfrak{V}_T$ for which $E_\theta(V) = 0$ for all θ in ω. If $\mathfrak{P}^t$ is complete, $\mathfrak{V}_T{}^0$ consists only of statistics $V = v(X)$ for which $v(x) = 0$ (a.e. $\mathfrak{P}^x$), and conversely. In statistical language we may say the $\mathfrak{V}_T{}^0$ consists of those functions of T which are estimates of zero (unbiased, with finite variance). Finally, $\mathfrak{V}_T{}^1$ is defined to be the subclass of $\mathfrak{V}_T$ whose members V satisfy the condition that VV^0 is in $\mathfrak{V}_T{}^0$ for every V^0 in $\mathfrak{V}_T{}^0$.

An equivalent condition is that $E_\theta(VV^0) = 0$, or $\text{Cov}_\theta(VV^0) = 0$ for every θ in ω and every V^0 in $\mathfrak{V}_T{}^0$. The class $\mathfrak{V}_T{}^1$ always contains all real constants.[4]

Theorem 5.3: *A statistic V in $\mathfrak{V}$ is a minimum variance estimate of its expected value if and only if it is a member of $\mathfrak{V}_T{}^1$. The class $\mathfrak{G}$ of all estimable functions $g(\theta)$ possessing minimum variance estimates is thus obtained by applying the operator E_θ to the members of $\mathfrak{V}_T{}^1$. If we identify all V in $\mathfrak{V}$ which are equivalent ($\mathfrak{P}^x$) in the strict sense, then the correspondence between $\mathfrak{G}$ and $\mathfrak{V}_T{}^1$ is 1:1.*

To prove the theorem suppose first that $V = v(X)$ in $\mathfrak{V}$ is a minimum variance estimate of its expected value. Define $\psi(t) = E(V|t)$. Then $v(x) = \psi(t(x))$ (a.e. $\mathfrak{P}^x$) by Rao and Blackwell's theorem, and thus V is in $\mathfrak{V}_T$. Let V^0 be any element of $\mathfrak{V}_T{}^0$ and let $U = V + \lambda V^0$, where λ is a constant. Clearly U is also an unbiased estimate of $g(\theta) = E_\theta(V)$, and

$$\text{Var}_\theta(U) - \text{Var}_\theta(V) = \lambda^2 \text{Var}_\theta(V^0) + 2\lambda \text{Cov}_\theta(VV^0).$$

This quadratic function of λ cannot be negative since V is a minimum variance estimate, and consequently it is easily found in either of the cases $\text{Var}_\theta(V^0) > 0$ or $= 0$, that $\text{Cov}_\theta(VV^0) = 0$ for all θ in ω and V^0 in $\mathfrak{V}_T{}^0$. Hence V is in $\mathfrak{V}_T{}^1$.

Next suppose that V is in $\mathfrak{V}_T{}^1$ and that U is another unbiased estimate of $g(\theta) = E_\theta(V)$. Then if $w(t) = E(U|t)$, $W = w(T)$ is also unbiased, and $\text{Var}_\theta(W) \leqslant \text{Var}_\theta(U)$ by Rao and Blackwell's theorem. Also $W - V$ is in $\mathfrak{V}_T$, $E_\theta(W - V) = 0$ for all θ in ω, and so $V^0 = W - V$ is in $\mathfrak{V}_T{}^0$. We now have

$$\text{Var}_\theta(U) \geqslant \text{Var}_\theta(W) = \text{Var}_\theta(V + V^0) = \text{Var}_\theta(V) + \text{Var}_\theta(V^0) \geqslant \text{Var}_\theta(V).$$

Hence V is a minimum variance estimate.

The last statement in Theorem 5.3 follows from the following

Lemma 5.1: *If V and V^1 are minimum variance estimates of $g(\theta)$ then $V = V^1$ with probability one for all θ in ω.*

To prove this let $h(\theta) = \text{Var}_\theta(V) = \text{Var}_\theta(V^1)$. If for some θ, $h(\theta) = 0$, then with probability one for this θ, $V = g(\theta) = V^1$. If $h(\theta) > 0$, let $\rho(\theta)$ be the correlation coefficient of V and V^1, form the unbiased estimate $U = \frac{1}{2}(V + V^1)$, and note that

$$h(\theta) \leqslant \text{Var}_\theta(U) = \text{Var}_\theta(\tfrac{1}{2}V + \tfrac{1}{2}V^1) = \tfrac{1}{4}h(\theta) + \tfrac{1}{4}h(\theta) + \tfrac{1}{2}h(\theta)\rho(\theta).$$

Thus $\rho(\theta) \geqslant 1$; hence $\rho(\theta) = 1$, and there exist constants A_θ, B_θ such that the probability P_θ is one that

[4] It seems to us in the light of Theorem 5.3 that the term "efficient estimate" should be reserved for the members of $\mathfrak{V}_T{}^1$, and that (absolute, as opposed to relative) "efficiency" of an estimate V of $g(\theta)$ should be defined only when V is unbiased and $g(\theta)$ is a member of the class $\mathfrak{G}$ of the theorem ; its efficiency (a function of θ) may then be defined as $\text{Var}_\theta(\Psi)/\text{Var}_\theta(V)$ if $\text{Var}_\theta(V) > 0$, 1 if $\text{Var}_\theta(V) = 0$, where Ψ is the minimum variance estimate of $g(\theta)$. We feel that efficiency should not be defined in such a way that the efficiency of the minimum variance estimate Ψ may be less than 1, as is the case for example if efficiency is defined relative to the lower bound for the variance of unbiased estimates given by the Cramér-Rao inequality.

$$V^1 = A_\theta + B_\theta V.$$

Hence

$$E_\theta(V^1) = A_\theta + B_\theta E_\theta(V); \qquad \dots \quad (5.2)$$

also, $\mathrm{Var}_\theta(V^1) = B_\theta^2 \mathrm{Var}_\theta(V)$, so that $B_\theta^2 = 1$. If $B_\theta = -1$, $A_\theta = 2g(\theta)$ from (5.2), and with probability P_θ equal to one, $V^1 = 2g(\theta) - V$ and $U = g(\theta)$. Hence $\mathrm{Var}_\theta(U) = 0$; but $\mathrm{Var}_\theta(U) \geqslant h(\theta) > 0$. From this contradiction we see that $B_\theta = 1$, hence from (5.2), $A_\theta = 0$, and thus $V = V^1$ with probability P_θ equal to one.

We shall now give two examples of the application of Theorem 5.3 in the case where T is a sufficient statistic (actually minimal) and $\mathfrak{P}^t$ is not complete.

Example 5.2 : We take $t(x) = x$ in Example 3.1. In defining statistics $V = v(X)$, the definition of the function $v(x)$ matters only at the points $x = -1$, 0, 1, 2,... $\mathfrak{V}_\Gamma = \mathfrak{V}$ is defined by the class of functions $v(x)$ for which the power series in θ

$$\sum_{j=0}^{\infty} [v(j)]^2 \theta^j$$

has a radius of convergence $\geqslant 1$. From the calculations of Example 3.1 we see that $\mathfrak{V}_T^0$ is defined by the class of functions $v(x)$ satisfying

$$v(j) = -jv(-1) \qquad (j = 0, 1, 2, \dots); \qquad \dots \quad (5.3)$$

for such functions the radius of convergence of the above series is equal to 1(∞, if $v(-1) = 0$). Finally $\mathfrak{V}_T^1$ is defined by the class of functions $v(x)$ satisfying the above series condition and such that $v(x)v^0(x)$ satisfies (5.3) for all $v^0(x)$ satisfying (5.3), that is,

$$v(j)v^0(j) = -jv(-1)\, v^0(-1) \qquad \dots \quad (5.4)$$

for all $v^0(x)$ such that

$$v^0(j) = -jv^0(-1). \qquad \dots \quad (5.5)$$

Combining (5.4) and (5.5), we find that $V = v(X)$ is in $\mathfrak{V}_T^1$ if and only if

$$v(j) = v(-1) \qquad (j = 1, 2, \dots),$$

with $v(0)$ arbitrary. Then for V in $\mathfrak{V}_T^1$,

$$E_\theta(V) = v(0)P_\theta\{X = 0\} + v(-1)P_\theta\{X \neq 0\}$$
$$= v(0)(1-\theta)^2 + v(-1)(2\theta - \theta^2).$$

The class $\mathfrak{G}$ of extimable functions $g(\theta)$ possessing minimum variance estimates is thus the two-parameter family of quadratic functions

$$g(\theta) = c_1 + c_2(1-\theta)^2,$$

and the minimum variance estimate of $g(\theta)$ is the statistic $V(X)$ taking on the value $c_1 + c_2$ if $X = 0$ and the value c_1 if $X \neq 0$.

Example 5.3: In this example we shall see that it is sometimes possible to prove that only constants have minimum variance estimates, without completely determining the class $\mathfrak{V}_T^0$ (although this would not be very difficult in the present example). As in Example 4.1, let $\mathfrak{P}^x$ be the family of distributions of a random sample of size $n>1$ from a rectangular population on the interval $(\theta-\frac{1}{2},\ \theta+\frac{1}{2})$. We define $T=(T_1, T_2)$ as in Example 4.1 and consider the class $\mathfrak{V}_T^0$. From (4.7) we see that a statistic of the type $V=f(T_1, T_2)$ (a.e. $\mathfrak{P}^x$) is in $\mathfrak{V}_T^0$ if and only if $E_\theta(V^2)<\infty$ and

$$\int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} \int_{\theta-\frac{1}{2}}^{t_2} f(t_1, t_2)(t_2-t_1)^{n-2}dt_1\, dt_2 = 0 \qquad \ldots \quad (5.6)$$

for all θ in $\omega=(-\infty, +\infty)$. We shall indicate how certain periodic solutions of (5.6) may be constructed; we remark in passing that there are also non-periodic solutions of a non-trivial kind.

Let

$$A_\theta = \{t|\theta-\tfrac{1}{2}<t_1 \leqslant t_2 \leqslant \theta+\tfrac{1}{2}\},$$

so that A_θ is a right triangle with hypotenuse on the 45^0 line, constituting the upper left half of a unit square S_θ centered at $(t_1, t_2)=(\theta, \theta)$, and with sides parallel to the axes. Hold $\theta=\theta_0$ fixed for the moment, and in A_{θ_0} define a Borel-measurable function $f=f(t_1, t_2)$ such that the condition (5.6) is satisfied for $\theta=\theta_0$, while

$$0 < \int\int_{A_{\theta_0}} f^2(t_2-t_1)^{n-2}dt_1dt_2<\infty.$$

We next extend the definition of f to the square S_{θ_0} by defining f symmetrically in the lower right half,

$$f(t_2,t_1) = f(t_1,t_2),$$

and we then extend the definition of f to the strip $0\leqslant t_2-t_1<1$ by the periodicity condition

$$f(t_1+\nu, t_2+\mu) = f(t_1, t_2) \quad (\nu, \mu = 0, \pm 1, \pm 2, \ldots).$$

With the aid of a figure showing the triangles A_{θ_0} and A_θ it is easy to see that (5.6) is satisfied for all θ, while

$$E_\theta(V^2) = E_{\theta_0}(V^2)<\infty.$$

We can now prove that the only estimable functions with minimum variance estimates are constants. Suppose $\psi(T_1, T_2)$ is in $\mathfrak{V}_T^1$ so that if $f(t_1, t_2)$ satisfies (5.6) and

$$E_\theta[f(T_1, T_2)]^2<\infty \ (\theta\epsilon\omega), \qquad \ldots \quad (5.7)$$

so does the product $\psi(t_1, t_2)f(t_1, t_2)$. It will suffice to show that in every A_θ, $\psi(t_1, t_2)$ is a constant (a.e. L), where L denotes Lebesgue measure; the constant must then be the same for all A_θ. Let θ_0 be any value of θ, henceforth held fixed, let A_+ be the part of A_{θ_0} where $\psi>0$, and let A_- be the part of A_{θ_0} where $\psi<0$. We shall show that it is impossible that $L(A_+)$ and $L(A_-)$ be both positive. If they were, we could define $f(t_1, t_2)$ in A_{θ_0} from

$$f(t_1,t_2)(t_2-t_1)^{n-2} = \begin{cases} 1/L(A_+) \text{ in } A_+, \\ -1/L(A_-) \text{ in } A_-, \\ 0 \text{ elsewhere in } A_{\theta_0}. \end{cases} \qquad \dots \quad (5.8)$$

Since f satisfies (5.6) for $\theta = \theta_0$, it can be extended to an f statisfying (5.6) for all θ by the method described above. If A_+ or A_- have points within every ϵ-distance of the 45^0 line, the f defined in this way is not bounded and the condition (5.7) may not be satisfied. However, if $L(A_+)$ and $L(A_-)$ are positive, and if $A^\varepsilon{}_+$ and $A^\varepsilon{}_-$ are the parts of A_+ and A_- outside the strip $0\leqslant t_2-t_1<\epsilon$, then for sufficiently small ϵ, $L(A_{\varepsilon+})$ and $L(A_{\varepsilon-})$ will also be positive. If the above definition of f is now modified by replacing A_+ and A_- in (5.8) by $A^\varepsilon{}_+$ and $A_{\varepsilon-}$, then f satisfies (5.6) and (5.7) for all θ. But then ψf cannot satisfy (5.6) for $\theta=\theta_0$ for in $A_{\varepsilon+}$ and $A^\varepsilon{}_-$, $\psi f>0$, while in the rest of A_{θ_0}, $\psi f = 0$, and thus

$$\int\int_{A_{\theta_0}} \psi f(t_2-t_1)^{n-2}dt_1dt_2>0.$$

We have now shown that if $V = \psi((T_1, T_2)$ (a.e. $\mathfrak{P}^x$) is in $\mathfrak{V}_T{}^1$, then the sets in A_{θ_0} where $\psi(t_1, t_2) > 0$, and where $\psi(t_1, t_2) < 0$, cannot both have positive Lebesgue measure. But if V is in $\mathfrak{V}_T{}^1$ so is $V-c$ for every constant c, and so it follows that for every c the sets in A_{θ_0} where $\psi>c$ and where $\psi<c$ cannot both have positive measure. From this it can be shown that for some constant c, $\psi=c$(a.e. L) in A_{θ_0}, and this proves that if V is in $\mathfrak{V}_T{}^1$ then $V = c$ (a.e. $\mathfrak{P}^x$). Thus the only estimable functions with minimum variance estimates are the constants.

6. CONSTRUCTION AND EXISTENCE OF MINIMAL SUFFICIENT STATISTICS

So far it has not been necessary to impose any restrictions on the family $\mathfrak{P}^x$ of probability measures. However, in this section we shall assume that the measures in $\mathfrak{P}^x$ are all absolutely continuous with respect to some measure μ^x on $\mathfrak{F}^x$ which is independent of θ and has the property that W^x is a countable union of sets in $\mathfrak{F}^x$ of finite measure μ^x. This is equivalent to assuming the existence for all θ in ω of a function $p_\theta(x)$ integrable $(\mathfrak{F}^x, \mu^x)$ such that for all A in $\mathfrak{F}^x$

$$P_\theta{}^x(A) = \int_A p_\theta(x)d\mu^x.$$

We shall refer to this situation by saying there exists a *generalized probability density* $p_\theta(x)$ *with respect to* μ^x (it is the "Nikodym derivative" $dP_\theta{}^x/d\mu^x$ mentioned in section 3). This situation includes of course the two important cases with fixed sample size known as the continuous case, where μ^x is Lebesgue measure on a Euclidean space W^x, and the discrete[5] case, where for all θ the possible positions of the sample point X are included in a fixed countable set $\{x^i\}$ $(i = 1,2,3, \ldots)$ independent of θ. In the latter case we may take for $\mu^x(A)$ the number of points x^i in A, and for $p_\theta(x^i)$, the probability that $X = x^i$ when the probability distribution of the sample is P_θ^x. The generalized density $p_\theta(x)$ is not uniquely determined by P_θ^x and μ^x; however, two determinations for the same θ are equal (a.e. μ^x and a.e. P_θ^x).

For most of the families $\mathfrak{P}^x$ ordinarily considered by statisticians there exists a generalized density $p_\theta(x)$ with respect to some μ^x, and if the "simplest" or "most natural" determination of $p_\theta(x)$ is used, a minimal sufficient statistic for $\mathfrak{P}^x$ can be found by applying to the family $\mathfrak{p} = \{p_\theta(x) | \theta \epsilon \omega\}$ an operation ϑ to be described below. Unfortunately the result of this operation does depend on which determination of the family $\mathfrak{p}$ is used. This introduces certain measure-theoretic difficulties, which can however be surmounted as we shall show later. We begin by defining the operation ϑ and by applying it to some examples of families of distributions of some statistical interest.

Let $\mathfrak{f}$ denote a family of real-valued functions $f(x)$ on W^x, and suppose these functions are indexed by a subscript θ taking on values in Λ, $\mathfrak{f} = \{f_\theta(x) | \theta \epsilon \Lambda\}$. The result of the operation ϑ on $\mathfrak{f}$ is a decomposition of W^x, to be denoted by $\vartheta(\mathfrak{f})$. For any point x^0 in W^x the element D of $\vartheta(\mathfrak{f})$ containing x^0, written $D(x^0)$, is defined as the set of all points x for which there exists a function $k(x, x^0) \neq 0$, not depending on θ, and such that $f_\theta(x) = k(x, x^0) f_\theta(x^0)$ for all θ in Λ. Roughly speaking, we may say that $D(x^0)$ consists of all x for which the ratio $f_\theta(x)/f_\theta(x^0)$ is independent of θ. We note that if x' is in $D(x^0)$, then x^0 is in $D(x')$; also, that

$$D^0 = \{x | f_\theta(x) = 0 \quad \text{for all } \theta \epsilon \Lambda\}$$

is an element of $\vartheta(\mathfrak{f})$. It may be shown (as in the proof of Theorem (6.1)) that if $\mathfrak{f}$ consists of a countable number of measurable ($\mathfrak{F}^x$) functions then $\vartheta(\mathfrak{f})$ is a measurable ($\mathfrak{F}^x$) decomposition.

We consider now five examples of applying the operation ϑ. In each case W^x is a Euclidean n-space, and $\mathfrak{F}^x$ may be taken as the family of Borel sets. In all but Example 6.1, μ^x will be taken as Lebsgue measure on $\mathfrak{F}^x$.

Example 6.1: Suppose $X = (X_1, X_2, \ldots, X_n)$ is a random sample from a binomial population with parameter θ, X_i taking on the values 1 and 0 with respective probabilities θ and $1-\theta$. Let μ^x assign measure 1 to each of the 2^n points in the set W_+^x consisting of the points $(x_1, x_2, \ldots, x_n)$ with $x_i = 0$ or 1, and measure zero to $W^x - W_+^x$. In this example we might take $\mathfrak{F}^x$ as the family of all subsets of W_+^x,

[5] In the general discrete case, while for each θ the possible values of X constitute a countable set A_θ, $A = \cup_\theta A_\theta$ can of course be non-denumerable. The case of non-denumerable A is not included in the present treatment.

or the family of all subsets of W^x, instead of the usual family of all Borel sets in W^x. For ω we take the open interval $0<\theta<1$, and for $p_\theta(x)$ the determination

$$p_\theta(x) = \begin{cases} \prod_{j=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \text{ if } x \epsilon W_+^x, \\ 0 \text{ otherwise.} \end{cases} \quad \ldots \quad (6.1)$$

For this specification, $D^0 = W^x - W_+^x$. For x^0 not in D^0, x is in $D(x^0)$ if and only if there exists $k(x, x^0)$ such that $p_\theta(x) = k(x, x^0) p_\theta(x^0)$, and hence

$$[\theta/(1-\theta)]^{\sum_i x_i - \sum_i x_i^0} = k(x, x^0). \quad \ldots \quad (6.2)$$

The left member of (6.2) is independent of θ if and only if

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i^0, \quad \ldots \quad (6.3)$$

and this conclusion would be valid even if (6.2) were required to hold only for two distinct values of θ instead of all values between 0 and 1. The decomposition resulting from the application of the operation ϑ to the family $\{p_\theta(x)\}$ of (6.1) may thus be described as follows: The decomposition is equivalent ($\mathfrak{P}^x$) in the strong sense to that associated with the statistic $\Sigma_i X_i$ or any 1:1 function of this. For any x^0 in W_+^x the element $D(x^0)$ consists of the $\binom{n}{\Sigma_i x_i^0}$ points in W_+^x satisfying (6.3). The decomposition contains $n+2$ elements D, namely D^0 and the $n+1$ sets D where $\sum_{i=1}^{n} x_i = \nu (\nu = 0, 1, \ldots, n)$. The same decomposition is obtained if the operation ϑ is applied to any subset of $\{p_\theta x)$ consisting of two or more elements.

Example 6.2: Let X be a random sample from a normal population with mean θ_1 and variance θ_2 as in Example 5.1. With the determination (5.1) of the probability density $p_\theta(x)$ we find D^0 is the empty set. Since the denominator of the fraction $p_\theta(x)/p_\theta(x^0)$ cannot vanish, $D(x^0)$ is the set where this ratio is independent of θ. This is seen to be the same as the set where

$$-\frac{1}{2\theta_2}\left(\sum_i x_i^2 - \sum_i x_i^{0^2}\right) + \frac{\theta_1}{\theta_2}\left(\sum_i x_i - \sum_i x_i^0\right)$$

is independent of θ, namely, the set where

$$\sum_i x_i^2 = \sum_i x_i^{0^2} \text{ and } \sum_i x_i = \sum_i x_i^0.$$

The decomposition induced by the operation ϑ is thus that associated with the statistic $(\sum_i X_i, \sum_i X_i^2)$; this is the same as that associated with the statistic (T_1, T_2) defined in Example 5.1. It may be verified in this case that one obtains the same result if the operation ϑ is applied to a set of any three of the members of $\{p_\theta(x)\}$, say for $(\theta_1, \theta_2) = (\theta_{1i}, \theta_{2i})$, $i = 1,2,3$, providing the three points $(\theta_{1i}, \theta_{2i})$ are not collinear.

Example 6.3: Suppose X is a random sample of n from a Cauchy distribution with median θ, and $\omega = \{\theta \mid -\infty < \theta < +\infty\}$. We take for $p_\theta(x)$ the usual determination

$$p_\theta(x) = \pi^{-n} \prod_{j=1}^{n} [1+(x_j-\theta)^2]^{-1}.$$

Again D^0 is empty and $D(x^0)$ is the set where the ratio $p_\theta(x)/p_\theta(x_0)$ is independent of θ. This ratio may be written

$$\frac{\prod_{j=1}^{n} [(\theta-x_j^0-i)(\theta-x_j^0+i]}{\prod_{j=1}^{n} [(\theta-x_j-i)(\theta-x_j+i)]},$$

where $i = \sqrt{-1}$. If this quotient is independent of θ, the polynomials of degree $2n$ in θ in the numerator and denominator must have equal roots, from which we conclude that $(x_1, x_2, \ldots, x_n)$ must be a permutation of $(x^0_1, x_2^0, \ldots, x_n^0)$. The decomposition induced by ϑ may be regarded as that associated with the set of "order statistics" of the sample, namely $(Z_1, Z_2, \ldots, Z_n)$, where $Z_1 \leqslant Z_2 \leqslant \ldots \leqslant Z_n$ is a rearrangement of $X_1, X_2, \ldots, X_n$. The same result would be found for any subset of $\{p_\theta(x)\}$ corresponding to $2n+1$ distinct values of θ.

Example 6.4: As in Examples 4.1 and 5.3 let X be a random sample of n from the uniform distribution on $(\theta-\frac{1}{2}, \theta+\frac{1}{2})$, $-\infty<\theta<\infty$. We determine $p_\theta(x)$ as 1 if $|x_i-\theta|<\frac{1}{2}$ for all i, as 0 otherwise. With $t_1(x) = \min_i x_i$, $t_2(x) = \max_i x_i$, we see that $p_\theta(x) = 1$ if

$$\theta-\tfrac{1}{2}<t_1(x)\leqslant t_2(x)<\theta+\tfrac{1}{2}, \quad \ldots \quad (6.4)$$

$p_\theta(x) = 0$ otherwise. Therefore

$$D^0 = \{x \mid t_2(x)-t_1(x) \geqslant 1\}.$$

Now for any x^0 and x, x is in $D(x^0)$ if and only if $p_\theta(x)$ and $p_\theta(x^0)$ vanish for the same set of θ in ω. It follows that for x^0 not in D^0, x is in $D(x^0)$ if and only if (6.4) is true for all θ in ω which satisfy

$$\theta-\tfrac{1}{2}<t_1(x^0)\leqslant t_2(x^0)<\theta+\tfrac{1}{2}.$$

It follows that for x^0 not in D^0, x is in $D(x^0)$ if and only if $t_j(x) = t_j(x^0)$, $j = 1, 2$. The decomposition imposed by the operation ϑ is thus equivalent ($\mathfrak{P}^x$) in the strong sense to that associated with the statistic (T_1, T_2) of Examples 4.1 and 5.3. The same decomposition is obtained if the operation ϑ is applied to a denumerable subset of $\{p_\theta(x)\}$ corresponding to a subset of ω everywhere dense in ω.

Example 6.5: Denote by $(X_1, X_2, \ldots, X_m)$ and $(X_{m+1}, \ldots, X_n)$ independent random samples from normal populations with means θ_1 and $\theta_1+\delta$, and variances θ_2 and θ_3, respectively. The null hypothesis of the Behrens-Fisher problem then specifies $\delta = \delta_0$, where δ_0 is a known constant, so that the ω for the family $\mathfrak{P}^x$ of probability distributions under the null hypothesis is the subset of a three-dimensional Euclidean space satisfying $-\infty<\theta_1<+\infty$, $\theta_2>0$, $\theta_3>0$. If we determine the density $p_\theta(x)$ as

$$p_\theta(x) = (2\pi)^{\frac{1}{2}n}\theta_2^{\frac{1}{2}m}\theta_3^{\frac{1}{2}(n-m)}e^{-\frac{1}{2\theta_2}\sum_{i=1}^{m}(x_i-\theta_1)^2-\frac{1}{2\theta_3}\sum_{j=m+1}^{n}(x_j-\theta_1-\delta_0)^2}$$

it is easy to show that the result of applying the operation ϑ to this three-parameter family of density functions is the decomposition associated with the statistic (T_1, T_2, T_3, T_4) with four real components consisting of the two sample means and the two sample variances.

The measure-theoretic difficulty in applying the operation ϑ to construct a minimal sufficient statistic is the following: If a minimal sufficient statistic for $\mathfrak{P}^x$ exists then it is unique up to equivalence ($\mathfrak{P}^x$) in the strong sense. If the family $\mathfrak{P}^x$ of probability measures possesses a generalized density $p_\theta(x)$ with respect to the measure μ^x, then $p_\theta(x)$ is not unique; two determinations of $p_\theta(x)$ for fixed θ differ on a set A_θ which is in $\mathfrak{F}^x$ and for which $\mu^x(A_\theta) = 0$. Now if ω is non-denumerably infinite, the union of A_θ for θ in ω may not be a null set for μ^x; it need not even be measurable ($\mathfrak{F}^x$). The decomposition $\vartheta(\mathfrak{p})$ resulting from the application of the operation ϑ to the family of generalized densities $\mathfrak{p}=\{p_\theta(x)|\theta\epsilon\omega\}$ depends on which determination is chosen for $\mathfrak{p}$. As long as ω is countable it can be shown (Theorem 6.1) that the decomposition $\vartheta(\mathfrak{p})$ is measurable ($\mathfrak{F}^x$), and is associated with a minimal sufficient statistic for $\mathfrak{P}^x$, and it is easily seen that the two decompositions $\vartheta(\mathfrak{p})$ resulting from two different determinations of $\mathfrak{p}$ are equivalent (μ^x and $\mathfrak{P}^x$) in the strong sense. However, if ω is not countable, a pathological choice of the determination of $\mathfrak{p}$ could lead to a decomposition $\vartheta(\mathfrak{p})$ that is not measurable ($\mathfrak{F}^x$), and the equivalence of two $\vartheta(\mathfrak{p})$ resulting from two different determinations of $\mathfrak{p}$ could not be proved. This shows that the application of the operation ϑ need not necessarily lead to the minimal sufficient statistic.

This difficulty is resolved as follows: We restrict ourselves to families $\mathfrak{P}^x$ for which the family $\{p_\theta(x)\}$ of generalized densities is separable in a certain sense to be defined below; this separability *always* holds when W^x is a Euclidean space. There is then a countable subset of $\{p_\theta(x)\}$ dense in the whole set, and we apply the operation ϑ to this countable subset. (Recall that in the above examples we noted that in every case the same result was obtained by applying the operation θ to a suitable countable subset of the densities). On the basis of the theorems below it follows that the result of applying the operation ϑ to the countable subset of the densities is a measurable ($\mathfrak{F}^x$) decomposition which is associated with a minimal sufficient statistic for the whole set $\mathfrak{P}^x$ of distributions. The result needed for countable subsets is

Theorem 6.1: *Suppose $\mathfrak{P}^x$ is a countable set of probability measures on $\mathfrak{F}^x$, possessing a generalized probability density function $p_\theta(x)$ with respect to μ^x. If the operation ϑ is applied to a particular determination of the family $\{p_\theta(x)\}$, the resulting* *decomposition is measurable* ($\mathfrak{F}^x$) *and is associated with a minimal sufficient statistic for* $\mathfrak{P}^x$. *The decompositions resulting from applying the operation* ϑ *to two different determinations of the family* $\{p_\theta(x)\}$ *are equivalent* (μ^x *and* $\mathfrak{P}^x$) *in the strong sense.*

The proof of this theorem rests on the following result[6] of Halmos and Savage (1949) generalizing a theorem of Neyman (1935):

Theorem 6.2: *A necessary and sufficient condition for the statistic* $T=t(X)$ *to be a sufficient statistic for a set of probability measures* $\mathfrak{P}^x = \{P_\theta^x | \theta\epsilon\omega\}$ *possessing a generalized density with respect to* μ^x *is that there exist functions* $g_\theta(t)$ *and* $h(x)$ *such that*

$$p_\theta(x) = g_\theta(t(x))h(x) \quad \ldots \quad (6.5)$$

for a suitable determination $p_\theta(x)$ *of this generalized density, where* $g_\theta(t(x))$ *is measurable* ($\mathfrak{F}^x$) *and* $h(x)$ *is integrable* ($\mathfrak{F}^x$, μ^x).

To prove Theorem 6.1 let us denote by θ_i the elements of the countable set ω, and by $f_i(x)$ some particular determination of the generalized density $p_{\theta_i}(x)$, $i = 1,2, \ldots$. Write $\mathfrak{f} = \{f_i(x)\}$, and $\vartheta(\mathfrak{f})$ for the decomposition generated by the operation ϑ on $\mathfrak{f}$.

We shall show first that $\vartheta(\mathfrak{f})$ is measurable ($\mathfrak{F}^x$). If $D_i^0 = \{x|f_i(x)=0\}$, then D_i^0 is measurable ($\mathfrak{F}^x$) since $f_i(x)$ is measurable ($\mathfrak{F}^x$); hence $D^0=\bigcap_i D_i^0$ is measurable ($\mathfrak{F}^x$). Suppose now that x^0 is not in D^0. Then there exists a smallest i for which $f_i(x^0)\neq 0$, say $I=I(x^0)$. Define $D_i(x^0)$ as the set where $f_I(x)\neq 0$ and

$$f_i(x) \,|\, f_I(x) = f_i(x^0) \,|\, f_I(x^0).$$

Since $f_i(x)$ and $f_I(x)$ are measurable ($\mathfrak{F}^x$) so is their quotient in $W^x-D_I^0$, the part of W^x where $f_I(x)\neq 0$, and hence the part $D_i(x^0)$ of $W^x-D_I^0$ where the quotient has a constant value is measurable ($\mathfrak{F}^x$). Therefore $\bar{D}(x^0)=\bigcap_i D_i(x^0)$ is measurable ($\mathfrak{F}^x$). If for any point x^0 in W^x, $D(x^0)$ denotes the element of $\vartheta(\mathfrak{f})$ containing x^0, it is easily verified that for x^0 not in $\bar{D}^0$, $D(x^0) = D(x^0)$, and for x^0 in D^0, $D(x^0)=D^0$. This completes the proof that $\vartheta(\mathfrak{f})$ is measurable ($\mathfrak{F}^x$).

Let $\mathfrak{f}'=\{f'_i(x)\}$ be another determination of the generalized densities; then if N_i is the set where $f'_i(x)\neq f_i(x)$, N_i is in $\mathfrak{F}^x$ and $\mu^x(N_i)=0$. Hence $N=\mathrm{U}_i N_i$ is also a null set for μ^x. On W^x-N, $\mathfrak{f}$ and $\mathfrak{f}'$ are identical hence so are $\vartheta(\mathfrak{f})$ and $\vartheta(\mathfrak{f}')$. This proves the strong equivalence (μ^x and $\mathfrak{P}^x$) of the two decompositions.

Next we shall prove that, if $D(x)$ is the element of $\vartheta(\mathfrak{f})$ which contains x, and if $T=t(X)$ is a statistic associated with $v\,\vartheta(\mathfrak{f})$,—in particular we could take $t(x)=D(x)$ —then $f_i(x)$ may be factored in the form (6.5) of Theorem 6.2. To this end we choose in each element D of $\vartheta(\mathfrak{f})$ a single point $x^0=\chi(D)$. If the set of points x^0 thus chosen is denoted by W^0, this choice determines a function $x^0=\psi(x)$ from W^x into W^0,

[6] We are indebted to Professors Halmos and Savage for giving us a manuscript copy of their paper containing this result long before publication. The reader will not find the above Theorem 6.2 there in exactly the form stated here; the connection is discussed in the Appendix to the present paper.

since x in W^x determines $D=D(x)$ in ϑ ($\mathfrak{t}$), and D determines $x^0=\chi(D)=\chi(D(x))=\psi(x)$. Take any D in $\vartheta(\mathfrak{t})$ and hold it fixed. Then for all x in D and $x^0=\chi(D)$,

$$f_i(x)=k(x,x^0)f_i(x^0),$$

where $k(x,x^0)\neq 0$, and $i=1,2,\ldots$ This may be written

$$f_i(x)=k(x,\psi(x))f_i(\chi(D)).$$

For x in W^x-D^0 define $\bar{h}(x)=k(x, \psi(x))\neq 0$, and for x in D^0 define $\bar{h}(x)=0$. Define $\bar{g}_i(t)=f_i(\chi(D))$, where the elements D of ϑ ($\mathfrak{t}$) are the contours of $t(x)$. We now have

$$f_i(x)=\bar{g}_i(t(x))\bar{h}(x), \qquad \ldots \quad (6.6)$$

and

$$\{x\,|\,\bar{h}(x)=0\}=D^0.$$

Now (6.6) is of the factored form (6.5) of Theorem 6.2 except that we do not know yet whether the factors can be determined so as to satisfy the measurability and integrability conditions of the theorem. For brevity in proving this let us denote by $\mathfrak{G}_t$ the family of sets in W^x each of which is a union of elements D of ϑ ($\mathfrak{t}$); the members of $\mathfrak{G}_t$ need not be in $\mathfrak{F}^x$. For $i=1,2,\ldots$, define

$$G_i=\{\, x\,|\,f_1(x)=f_2(x)=\ldots=f_{i-1}(x)=0,\ f_i(x)\neq 0\}.$$

It is clear that $\{G_i\}$ is a sequence of disjoint sets in $\mathfrak{F}^x$ with $\cup_i G_i = W^x-D^0$, and we shall now show that all G_i are in $\mathfrak{G}_t$: Let $H_i = \{x\,|\,\bar{g}_i(t(x)) = 0\}$, so H_i is in $\mathfrak{G}_t$. Let $K_i = \{x\,|\,f_i(x) = 0\}=\{x\,|\,\bar{g}_i(t(x))\bar{h}(x) = 0\} = H_i \cup D^0$, so K_i is also in $\mathfrak{G}_t$. Finally,

$$G_i=K_1 \cap K_2 \cap \ldots \cap K_{i-1} \cap (W^x-K_i)$$

is also in $\mathfrak{G}_t$. For x in $G_k (k = 1,2,\ldots)$ we have from (6.6)

$$f_k(x) = \bar{g}_k(t(x))\bar{h}(x)\neq 0. \qquad \ldots \quad (6.7)$$

Combining this with (6.6), we get for x in G_k and $i=1, 2,\ldots$,

$$f_i(x) = [\bar{g}_i(t(x))\,|\,\bar{g}_k(t(x))]f_k(x). \qquad \ldots \quad (6.8)$$

In G_k $(k = 1,2,\ldots)$ define

$$h(x) = 2^{-k}f_k(x), \qquad \ldots \quad (6.9)$$

and in $t(G_k)$ define

$$g_i(t) = 2^k\bar{g}_i(t)\,|\,\bar{g}_k(t),$$

noting from (6.7) that $\bar{g}_k(t)\neq 0$ in $t(G_k)$. This defines $h(x)$ in W^x-D^0 and $g_i(t)$ in $t(W^x-D^0)$ for $i=1,2,\ldots$; in D^0 define

$$h(x)=0, \qquad \ldots \quad (6.10)$$

and in $t(D^0)$, $g_i(t)=0$. To see that there is no inconsistency in our definition of $g_i(t)$ it is important to note that the sets $t(D^0)$ and $t(G_k)(k=1,2,\ldots)$ are all disjoint: This is a consequence of D^0 and $G_1, G_2, \ldots$ being disjoint sets in $\mathfrak{G}_t$. From (6.8) we have

$$f_i(x) = g_i(t(x))h(x). \qquad \ldots \quad (6.11)$$

From (6.9) and (6.10) we see that $h(x)$ is measurable ($\mathfrak{F}^x$) in each of the sets G_k and D^0, and since these form a countable disjoint covering of W^x by sets in $\mathfrak{F}^x$, $h(x)$ is measurable ($\mathfrak{F}^x$) in W^x. That it is also integrable ($\mathfrak{F}^x, \mu^x$) follows from

$$\int_{W^x} |h(x)|\, d\mu^x = \int_{W^x-D^0} h(x)d\mu^x = \sum_{k=1}^{\infty} \int_{G_k} f_k(x)d\mu^x$$

$$\leqq \not\geqq \quad \sum_k 2^{-k} \int_{W^x} f_k(x)d\mu^x = \sum_k 2^{-k} < \infty .$$

Since $g_i(t(x))=0$ in D^0 while from (6.11) $g_i\,(t(x))$ is the quotient of measurable ($\mathfrak{F}^x$) functions with non-vanishing denominator in W^x-D^0, $g_i(t(x))$ is measurable ($\mathfrak{F}^x$) in W^x.

We have now shown all the hypotheses of Theorem 6.2 to be satisfied and hence $T=t(X)$ is a sufficient statistic for $\mathfrak{P}^x$; it remains only to prove that it is minimal. To do this we must show that if $T'=t'(X)$ is any other sufficient statistic for $\mathfrak{P}^x$, then it is possible to find a null set N'' for $\mathfrak{P}^x$ such that in W^x-N'' every contour of $t'(x)$ is contained in a contour of $t(x)$.

If T' is sufficient for $\mathfrak{P}^x$, then by Theorem 6.2 there exists a determination of the generalized density, say $f_i'(x)$, such that for $i=1, 2, \ldots,$

$$f_i'(x) = g_i'(t'(x))h'(x),$$

where $h'(x)$ is measurable ($\mathfrak{F}^x$). The determination $\mathfrak{f}' = \{f'_i(x)\}$ may differ from the previous determination $\mathfrak{f} = \{f_i(x)\}$ to the extent that if $N_i' = \{x \mid f_i'(x)\neq f_i(x)\}$, then N_i' is a null set for μ^x. Hence $N' = \bigcup_i N_i'$ is a null set for μ^x and $\mathfrak{P}^x$, and $f_i'=f_i$ on W^x-N'. Let $N_0=\{x|h'(x)=0\}$. Then for all i, $f_i'(x)=0$ on N_0,

$$P_{\theta_i}{}^x(N_0) = \int_{N_0} f_i'(x)d\mu^x = 0,$$

and so N_0 is null set for $\mathfrak{P}^x$. I.. $N''=N'\cup N_0$, N'' is a null set for $\mathfrak{P}^x$, and on W^x-N'', $f_i'(x)=f_i(x)$ and $h'(x)\neq 0$.

We shall now show that in W^x-N'' every contour of $t'(x)$ is contained in a contour of $t(x)$. Let x^0 be any point in W^x-N'', and let $A(x^0)$ be the part in W^x-N'' of the contour of $t'(x)$ containing x^0,

$$A(x^0) = \{x|x\epsilon W^x-N'',\ t'(x) = t'(x^0)\}.$$

For all x in $A(x^0)$

$$f_i(x) = f_i'(x) = g_i'(t'(x^0))h'(x);$$

in particular,

$$f_i(x^0) = g_i'(t'(x^0))h'(x^0).$$

Since $h'(x^0) \neq 0$,

$$f_i(x) = k(x,x^0)f_i(x^0) \quad \ldots \quad (6.12)$$

for all x in $A(x^0)$, where

$$k(x,x^0) = h'(x) | h'(x^0) \neq 0.$$

It follows from (6.12) that in $W^x - N''$, $A(x^0)$ is contained in $D(x^0)$, the contour of $t(x)$ containing x^0. This completes the proof of Theorem 6.1.

In proving an existence theorem for minimal sufficient statistics for families $\mathfrak{P}^x$ which may be non-denumerably infinite, we shall employ the following two lemmas whose proof is obvious.

Lemma 6.1: *If the statistic T is sufficient for the family $\mathfrak{P}^x$ of probability measures it is sufficient for any subfamily of $\mathfrak{P}^x$.*

Lemma 6.2: *If T is a sufficient statistic for the family $\mathfrak{P}^x$, if it is a minimal sufficient statistic for the subfamily $\mathfrak{P}_1{}^x$ of $\mathfrak{P}^x$, and if every null set for $\mathfrak{P}_1{}^x$ is a null set for $\mathfrak{P}^x$, then T is a minimal sufficient statistic for $\mathfrak{P}^x$.*

Our extension of the existence theorem from the countable to the not necessarily countable case involves the notion of separability of a function space with respect to convergence in the mean of order one. It $f = f(x)$ and $g=g(x)$ are two real-valued functions on W^x, integrable ($\mathfrak{F}^x$, μ^x), we define their distance as

$$\delta_\mu\ (f, g) = \int_{W^x} | f(x - g(x)| d\mu^x.$$

We shall say that a family of real-valued functions, $\mathfrak{f} = \{f_\theta(x)|\theta\epsilon\omega\}$, each integrable ($\mathfrak{F}^x,\mu^x$), is *separable* ($\mu^x$) if there exists a fixed countable subset $\mathfrak{f}_1$ of $\mathfrak{f}$ such that for every f_θ in $\mathfrak{f}$ there is a sequence $\{g_i | i=1,2\ldots\}$ of functions g_i in $\mathfrak{f}_1$ for which $\delta_\mu\ (f,g_i) \to 0$ as $i \to \infty$. It will be convenient to say that the countable subset $\mathfrak{f}_1$ of this definition is *dense* (μ^x) in $\mathfrak{f}$. We remark that if W^x is a Euclidean space and $\mathfrak{F}^x$ is the family of Borel sets in W^x then any family $\{p_\theta(x)\}$ of generalized densities with respect to μ^x is separable: This is a consequence of the separability[7] (μ^x) of the family of all functions integrable ($\mathfrak{F}^x$, μ^x), and the separability of any subfamily of a separable family in a metric space. We can now conclude that the result of applying the operation ϑ in each of Examples 6.1 to 6.5 was to give a minimal sufficient statistic: This follows from (*i*) the remark just made about Euclidean W^x, (*ii*) the remarks in the examples, about the same decomposition being obtained if the operation ϑ is applied to certain countable subsets of the generalized densities, and (*iii*) the following theorem.

[7] This may be proved from the Approximation Theorem on page 4 of *Ergodentheorie* by Hopf (1948, Chelsea, N.Y.).

Theorem 6.3: *If the family* $\mathfrak{P}^x$ *of probability measures on* W^x *possesses a generalized density* $p_\theta(x)$ *with respect to* μ^x, *and if the family* $\mathfrak{p} = \{p_\theta(x) \mid \theta \epsilon \omega\}$ *of densities is separable* (μ^x), *then there exists a minimal sufficient statistic for* $\mathfrak{P}^x$, *and it may be constructed by applying the operation* ϑ *to any countable set* $\mathfrak{p}_1$ *dense* (μ^x) *in* $\mathfrak{p}$.

To prove the theorem let $\mathfrak{f}_1 = \{f_i(x)\}(i = 1,2,\ldots)$ be a particular determination of a countable subfamily $\mathfrak{p}_1$ of densities dense (μ^x) in $\mathfrak{p}$, and let $\mathfrak{P}_1{}^x$ be the countable subfamily of $\mathfrak{P}^x$ corresponding to $\mathfrak{p}_1$. In order to apply Lemma 6.2 we must show that if N is a null set for $\mathfrak{P}_1{}^x$ it is a null set for $\mathfrak{P}^x$. Let $P_{\theta_0}{}^x$ be any member of $\mathfrak{P}^x$, let $f_0(x) = dP_{\theta_0}{}^x/d\mu^x$, and let $\{g_i\}$ be a sequence in $\mathfrak{f}_1$ for which $\delta_\mu(f_0, g_i) \to 0$ as $i \to \infty$. We have

$$\int_N g_i(x) d\mu^x = 0$$

for all i. Now

$$\int_N f_0(x) d\mu^x = \left| \int_N f_0 d\mu^x \right| = \left| \int_N f_0 d\mu^x - \int_N g_i d\mu^x \right|$$

$$= \left| \int_N (f_0 - g_i) d\mu^x \right| \leqslant \int_N |f_0 - g_i| d\mu^x$$

$$\leqslant \int_{W^x} |f_0 - g_i| d\mu^x = \delta_\mu(f_0, g_i) \to 0.$$

Hence

$$\int_N f_0(x) d\mu^x = 0$$

and N is a null set for $\mathfrak{P}^x$.

Suppose $T = t(X)$ is a minimal sufficient statistic for $\mathfrak{P}_1^x$, obtained by applying the operation ϑ to $\mathfrak{f}_1$. To prove T is a minimal sufficient statistic for $\mathfrak{P}^x$ it remains only to prove that T is sufficient for $\mathfrak{P}^x$. That T is sufficient for $\mathfrak{P}_1{}^x$ means that there exists a real-valued function $P(A|t)$, measurable $(\mathfrak{F}^t)$ for fixed A, independent of θ, and such that

$$P^{\theta_x}(A \cap t^{-1}(B)) = \int_B P(A|t) dP_\theta{}^t \quad \ldots \quad (6.13)$$

for all A in $\mathfrak{F}^x$, B in $\mathfrak{F}^t$, and θ in ω_1, where ω_1, is the subset of ω corresponding to $\mathfrak{P}_1{}^x$. To prove T sufficient for $\mathfrak{P}^x$ it thus suffices to show that if (6.13) is valid for all θ in ω_1, then it is valid for all θ in ω. Let us write B' as an abbreviation of $t^{-1}(B)$.

We may transform (6.13) to

$$\int_{A \cap B'} dP_\theta{}^x = \int_{B'} \pi(A, x) dP_\theta{}^x,$$

Where $\pi(A, x) = P(A|t(x))$, or

$$\int_{A \cap B'} p_\theta(x)\, d\mu^x = \int_{B'} \pi(A, x) p_\theta(x) d\mu^x. \quad \ldots \quad (6.14)$$

We note $0 \leqslant \pi(A,x) \leqslant 1$ (a.e. $\mathfrak{P}_1{}^x$) by (2.5) and (2.8), and hence (a.e. $\mathfrak{P}^x$). Let $\theta = \theta_0$

be any θ in ω, write $p_{\theta_0}(x) = g(x)$, and let $\{g_i\}$ be a sequence in $\mathfrak{f}_1$ for which $\delta_\mu(g, g_i) \to 0$ as $i \to \infty$. From

$$\int_{A\cap B'} g_i d\mu^x = \int_{B'} \pi g_i d\mu^x \qquad \dots \quad (6.15)$$

we shall prove

$$\int_{A\cap B'} g d\mu^x = \int_{B'} \pi g d\mu^x \qquad \dots \quad (6.16)$$

by letting $i \to \infty$ and showing that the limits of the right and left members of (6.15) are the corresponding members of (6.16).

$$\left| \int_{A\cap B'} g_i d\mu^x - \int_{A\cap B'} g d\mu^x \right| \leqslant \int_{A\cap B'} |g_i - g| d\mu^x$$

$$\leqslant \delta_\mu(g_i, g) \to 0.$$

$$\left| \int_{B'} \pi g_i d\mu^x - \int_{B'} \pi g d\mu^x \right| \leqslant \int_{B'} |\pi| \, |g_i - g| d\mu^x$$

$$\leqslant \int_{B'} |g_i - g| d\mu^x \leqslant \delta_\mu(g_i, g) \to 0.$$

This completes the proof of Theorem 6.3.

APPENDIX

With only notational changes, the result on sufficient statistics and the factorization of generalized probability densities stated by Halmos and Savage (1949), of which Theorem 6.2 above is a slightly different version, is their following

Corollary 1: *Let $\mathfrak{F}^x$ and $\mathfrak{G}^t$ be countably additive families of sets in the spaces W^x and W^t, respectively. Let $t = t(x)$ be a function from W^x onto W^t such that for every set B in $\mathfrak{G}^t$, $t^{-1}(B)$ is in $\mathfrak{F}^x$, and let $\mathfrak{G}^{x|t} = t^{-1}(\mathfrak{G}^t)$. Let $\mathfrak{M}^x = \{M_\theta^x | \theta \epsilon \omega\}$ be a family of finite measure on $\mathfrak{F}^x$ with all M_θ^x absolutely continuous with respect to the finite measure μ^x on $\mathfrak{F}^x$. Then a necessary and sufficient condition for $T = t(X)$ to be a sufficient statistic*[8] *for $\mathfrak{M}^x$ is that for every θ in ω, $f_\theta(x) = dM_\theta^x/d\mu^x$ be factorable in the form $f_\theta = g_\theta h$, where $0 \leqslant g_\theta$ is measurable ($\mathfrak{G}^{x|t}$), $0 \leqslant h$ and $0 \leqslant g_\theta h$ are integrable ($\mathfrak{F}^x$, μ^x), and $h(x) = 0$ (a.e. μ^x) on every null set for $\mathfrak{M}^x$.*

Some of the simplifications we have obtained arise from the restrictions that, first, we always take for $\mathfrak{G}^t$ the family $\mathfrak{F}^t$ of all sets in W^t whose pre-images are in

[8] If $\mathfrak{M}^x$ is not a family of probability measures, T is not a random variable. Halmos and Savage define statistic, conditional probability, and sufficient statistic for this case. We need not consider these definitions here except to remark that in the case that $\mathfrak{M}^x$ is a family of probability measures—the only case considered by us—the definitions agree with ours.

$\mathfrak{F}^x$, and that, secondly, we assume $\mathfrak{M}^x$ to be a family of probability measures. Theorem 6.2 follows from Corollary 1 by means of the following six steps:

1^0. The assumption that $t(x)$ satisfies the condition that pre-images of sets in $\mathfrak{G}^t$ are in $\mathfrak{F}^x$ is automatically fulfilled for an arbitrary function $t(x)$ because we always take $\mathfrak{G}^t = \mathfrak{F}^t$ as just mentioned.

2^0. It is pointed out by Halmos and Savage that in their work the assumption on μ^x that $\mu^x(W^x)$ is finite may be replaced by the assumption that W^x is a countable union of sets of finite measure (μ^x).

3^0. The assumption that $g_\theta h$ is non-negative and integrable ($\mathfrak{F}^x$, μ^x) may be dropped because of our assumption that $\mathfrak{M}^x$ is a family of probability measures, so that $g_\theta h = f_\theta$ may be assumed non-negative and

$$\int_{W^x} g_\theta h d\mu^x = 1.$$

4^0. The condition of non-negativeness on g_θ and h may be dropped, since we may assume f_θ non-negative and then redefine g_θ and h as $|g_\theta|$ and $|h|$ without affecting the product $g_\theta h$, the measurability of g_θ, or the integrability of h.

5^0. The assumption that $g_\theta(x)$ is measurable ($\mathfrak{G}^{x|t}$) is equivalent to the assumption that it is of the form $\bar{g}_\theta(t(x))$, where $\bar{g}_\theta(t)$ is measurable ($\mathfrak{G}^t$) by Lemma 2 in the same paper by Halmos and Savage. But with our choice of $\mathfrak{G}^t = \mathfrak{F}^t$, $\bar{g}_\theta(t)$ is measurable ($\mathfrak{G}^t$) if and only if $\bar{g}_\theta(t(x))$ is measurable ($\mathfrak{F}^x$).

6^0. The proof that the assumption $h(x)=0$ (a.e. μ^x) on every null set for $\mathfrak{M}^x$, may be dropped requires a little more effort. If h satisfies all the remaining conditions of Corollary 1 we shall show that $\bar{h}$ can be defined which also satisfies all the remaining conditions and furthermore $\bar{h}(x) = 0$ (a.e. μ^x) on every null set for $\mathfrak{M}^x$.

Let $f_\theta = g_\theta h$ denote a particular determination of $dM_\theta^x/d\mu^x$ for which $f_\theta \geqslant 0$. B Lemma 7 of Halmos and Savage, there exists a countable subset $\mathfrak{M}_1^x$ of $\mathfrak{M}^x$, say $\mathfrak{M}_1^x = \{M_\theta^x | \theta = \theta_1, \theta_2, \ldots\}$, such that all null sets for $\mathfrak{M}_1^x$ are null sets for $\mathfrak{M}^x$. Denote by A^0 the set in $\mathfrak{F}^x$ where $f_{\theta_i}(x) = 0$ for all i. Then A_0 is a null set for $\mathfrak{M}_1^x$ and hence for $\mathfrak{M}^x$. Define

$$\bar{h}(x) = \begin{cases} h(x) \text{ for } x \text{ in } W^x - A_0, \\ 0 \text{ for } x \text{ in } A_0. \end{cases}$$

For all A in $\mathfrak{F}^x$,

$$\int_A g_\theta h d\mu^x = M_\theta^x(A) = M_\theta^x(A - A_0) = \int_{A-A0} g_\theta h d\mu^x$$

$$= \int_{A-A0} g_\theta \bar{h} d\mu^x = \int_A g_\theta \bar{h} d\mu^x,$$

and hence $g_\theta h = g_\theta \bar{h}$ (a.e. μ^x). We may thus use $\bar{f}_\theta = g_\theta \bar{h} \geqslant 0$ instead of $f_\theta = g_\theta h$ as a determination of $dM_\theta^x/d\mu^x$, and $\bar{h}$ obviously satisfies the condition of integrability ($\mathfrak{F}^x$, μ^x) since h did.

It remains only to prove that if N is a null set for $\mathfrak{M}^x$ then $\bar{h}(x) = 0$ (a.e. μ^x) on N. Let N' be the part of N where $\bar{h} \neq 0$; we need to prove $\mu^x(N') = 0$. Now

$$0 = M_{\theta_i}{}^x(N') = \int_{N'} g_{\theta_i} \bar{h} d\mu^x.$$

Let N_i be the part of N' where $g_{\theta_i}\bar{h} > 0$, so $\mu^x(N_i) = 0$. On $N' - N_i$, $g_{\theta_i}\bar{h} = 0$, $\bar{h} \neq 0$, and thus $g_{\theta_i} = 0$. If $N'' = \text{U}_i N_i$, then $\mu^x(N'') = 0$, and on $N' - N''$, $g_{\theta_i} = 0$ for all i. Hence $N' - N''$ is contained in A_0, and $\bar{h} = 0$ on $N' - N''$. But $\bar{h} \neq 0$ on N', and so $N' - N''$ is the empty set. Since N' contains N'', $N' = N''$, and therefore $\mu^x(N') = 0$.

References

Blackwell, David (1947): Conditional expectation and unbiased sequential estimation. *Annals of Math. Stat.*, **18**, 105-110.

Cramér, Harald (1946): *Mathematical Methods of Statistics*, Princeton.

Doob, J. (1948): Asymptotic properties of Markoff transition probabilities. *Trans. Amer. Math. Soc.*, **63**, 393-421.

Ghosh, M. N. (1948): On the problems of similar regions. *Sankhyā*, **8**, 329-338.

Girshick, M. A., Mosteller, F., and Savage, L.J. (1946): "Unbiased estimates for certain binomial sampling problems with applications," *Annals of Math. Stat.*, **17**, 13-23.

Halmos, Paul R. (1946): The theory of unbiased estimation. *Annals of Math. Stat.*, **17**, 34-43.

Halmos, P. R. and Savage, L. J. (1949): "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math. Stat.*, **20**, 225-241.

Hsu, P. L. (1941): "Analysis of variance from the power-function standpoint", *Biometrika*, **32**, 62-69.

Kolmogoroff, A. (1933): Grundbegriffe der Wahrscheinlichkeitsrechnung. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, **2**, no. 3 Reprinted by Chelsea, New York (1946).

Koopman, B. O. (1936): On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.*, **39**, 399-409.

Lehmann, E. L. (1947): On optimum tests of composite hypotheses with one constraint. *Annals of Math. Stat.*, **18**, 473-494.

Lehmann, E. L. and Scheffé, H. (1947): On the problem of similar regions. *Proc. Nat. Acad. Sci.*, **33**, 382-386.

Lehmann, E. L. and Stein, C. (1948): Most powerful tests of composite hypotheses, I. Normal Distributions. *Annals of Math. Stat.*, **19**, 495-516.

Neyman, J. (1937): Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans. Roy. Soc. London*, series A, **236**, 333-380.

Neyman, J. (1941): On a statistical problem arising in routine analyses and in sampling inspections of mass production. *Annals of Math. Stat.*, **12**, 46-76.

Neyman, J. and Pearson, E. S. (1933): On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. London*, series A, **231**, 289-337.

Rao, C. R. (1945): Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, **37**, no. 3. 81-91.

Rao, C. R. (1947): Minimum variance and the estimation of several parameters. *Proc. Camb. Phil. Soc.*, **43**, 280-283.

RAO, C. R. (1949): Sufficient statistics and minimum variance estimates. *Proc. Camb. Phil. Soc.*, **45**, 213-218.

SAVAGE, L. J. (1947): A uniqueness theorem for unbiased sequential bionomial estimation. *Annals of Math. Stat.* **18**, 295-297.

SCHEFFÉ, HENRY (1942): On the theory of testing composite hypotheses with one constraint. *Annals of Math. Stat.*, **13**, 280-293.

SCHEFFÉ, HENRY (1943): On a measure problem arising in the theory of non-parametric tests. *Annals of Math Stat.*, **14**, 227-233.

SETH, G. R. (1949) abstract: Certain properties of the multiparameter unbiased estimates. *Annals of Math. Stat.*, **20**, 142.

WALD, ABRAHAM (1947): Foundations of a general theory of sequential decision functions. *Econometrica* **15**, 279-313.

WOLFOWITZ, J. (1946): On sequential binomial estimation," *Annals of Math. Stat.* **17**, 489-493.