OPTIMAL ALLOCATION OF OBSERVATIONS[1)]

M. H. DeGroot

(Received Feb. 15, 1965)

1. Introduction and summary

The concept of sufficient experiments, as introduced by Blackwell [1], [2], [3], and recently treated by LeCam [6], provides a method of comparing the effectiveness of certain statistical experiments. Roughly speaking, an experiment involving the observation of some random variable (or random vector) X^* is sufficient for another experiment involving the observation of some other random variable X if it is possible, from an observation on X^* and an auxiliary randomization, to generate a random variable with the same distribution as X for all possible values of any unknown parameters. Precise definitions are given in the above references and, in the context to be considered here, in Section 2 below. In this paper we will explore the relevance of this concept in problems in which a fixed total number of observations must be allocated in some optimal fashion among various possible alternatives. The following example is a prototype of the problems studied.

Consider a population of coins. Associated with each coin is a probability Z of heads which may vary from coin to coin. Suppose that it is desired to make some inference about the distribution of Z over the population. For example, one may be interested in estimating $\Pr(Z>1/2)$ (i.e., the proportion of coins that are biased in favor of heads), or in investigating the hypothesis that $\Pr(Z=1/2)\leqq\alpha$ (i.e., that the proportion of fair coins is at most α). Suppose that one can select a random sample of k coins from the population, toss the ith coin n_i times ($i=1, \cdots, k$), and observe the results. Given that the total number $\sum_{i=1}^{k} n_i$ of tosses is fixed, the problem is to find an optimum choice of k and $n_1, \cdots, n_k$. An optimum allocation will, in general, depend on the particular parametric family of distributions to which the distribution of Z is assumed to belong as well as on the type of information about the distribution that is desired. However, for a given family, it may happen

1) This research was supported in part by the Office of Ordnance Research under contract DA-36-061-ORD 447 and in part by the National Science Foundation under grant NSF-G25087.

that one experiment X^* defined by the allocation $(k^*; n_1^*, \cdots, n_{k^*}^*)$ is sufficient for any other experiment X defined by any other allocation $(k; n_1, \cdots, n_k)$ with $\sum_{i=1}^k n_i = \sum_{i=1}^{k^*} n_i^*$. We then say that the allocation $(k^*; n_1^*, \cdots, n_{k^*}^*)$ is optimal, regardless of the type of inference to be made or the kind of information desired.

The problem can be formulated in a more general setting as follows. A random sample $Z_1, \cdots, Z_k$ is drawn from a certain unknown distribution. However, the sample values cannot be observed without error. Thus, for each $i (i=1, \cdots, k)$, instead of observing the value of Z_i we observe the values of a set of random variables $X_{i1}, \cdots, X_{in_i}$ which, conditionally on any given value of Z_i, are independent and identically distributed with a known common conditional distribution. Thus, the random variables $X_{i1}, \cdots, X_{in_i}$ represent n_i (conditionally) independent measurements made on the unknown randomly selected Z_i. The problem considered is that of finding an optimal choice of k and $n_1, \cdots, n_k$.

A familiar situation of this type is the random effects model in a one-factor analysis of variance. In this model the effects $Z_1, \cdots, Z_k$ are regarded as a random sample from a normal distribution and the observations $X_{ij} (j=1, \cdots, n_i)$ are, given $Z_i = z_i$, independent and normally distributed with mean $z_i (i=1, \cdots, k)$ and variance τ^2. In order to satisfy our assumption that the conditional distributions are known, we assume that τ^2 is a known number. A sufficient and, hence, optimal allocation for this situation is found in Section 3. The question of finding suitable allocations in the more complex situation where the variances are unknown is discussed in [8], p. 236, and in [9].

In Section 2 the definition of an optimal allocation is given and some results that are useful in determining whether one of the two extreme allocations, in which 1 observation is made on each of n randomly selected Z's or all n observations are made on a single randomly selected Z, is optimal.

In Sections 3 and 4 it is shown how the existence of complete, sufficient statistics under various allocations can be used to simplify the search for an optimal allocation. Several examples are given illustrating the results.

In Section 5 problems of the optimal allocation of Bernoulli observations, as exemplified by the coin-tossing problem given above, are studied from the point of view developed in the earlier sections.

2. Sufficient experiments and optimal allocation

Let Ω be the set of all possible values of some unknown parameter θ. An experiment X consists of the abstract random variable X, taking values in a space $\mathscr{X}$ on which there is defined a σ-field $\mathscr{A}$ of subsets, together with a family $\{P_\theta;\ \theta\in\Omega\}$ of probability distributions of X over $(\mathscr{X}, \mathscr{A})$. Thus, we speak of performing the experiment $X=\{\mathscr{X};\ \mathscr{A};\ P_\theta,\ \theta\in\Omega\}$ or, equivalently, of observing the random variable X.

Suppose $X=\{\mathscr{X};\ \mathscr{A};\ P_\theta,\ \theta\in\Omega\}$ and $Y=\{\mathscr{Y};\ \mathscr{B};\ Q_\theta,\ \theta\in\Omega\}$ are two experiments with the same parameter space Ω. The concept of X being sufficient for Y, as embodied in the following two definitions due to Blackwell, [1], [2], is fundamental to the present study.

DEFINITION 2.1. A *stochastic transformation* from X to Y is a non-negative function $\pi(\cdot\mid\cdot)$ defined on $\mathscr{B}\times\mathscr{X}$ such that (i) for each fixed $x\in\mathscr{X}$, $\pi(\cdot\mid x)$ is a probability measure on $\mathscr{B}$, and (ii) for each fixed $B\in\mathscr{B}$, $\pi(B\mid\cdot)$ is an $\mathscr{A}$-measurable function on $\mathscr{X}$.

DEFINITION 2.2. The experiment X is *sufficient* for the experiment Y if there exists a stochastic transformation $\pi(\cdot\mid\cdot)$ from X to Y such that

$$Q_\theta(B)=\int_{\mathscr{X}}\pi(B\mid x)dP_\theta(x) \tag{2.1}$$

for all $B\in\mathscr{B}$ and $\theta\in\Omega$.

The import of this definition is that if X is sufficient for Y then it is possible, by means of an observation on X and an auxiliary randomization as specified by the probability distribution $\pi(\cdot\mid x)$, to generate a random variable having the same distribution as Y for all values of θ. In other words, it is possible to perform an experiment equivalent to Y.

It is this concept of a sufficient experiment that will now be considered in the explicit context of the optimal allocation of observations.

Let Z be a random variable (or random vector) with distribution P_θ depending on the unknown real-valued parameter (or vector of parameters) θ, where $\theta\in\Omega$, the parameter space. Suppose that it is possible to draw a random sample $Z_1,\cdots,Z_k$ from the distribution P_θ, but that the values of $Z_1,\cdots,Z_k$ cannot be observed. Rather, for each value of i $(i=1,\cdots,k)$, what can be observed are the values of n_i random variables $X_{i1},\cdots,X_{in_i}$ which, conditionally on any given value $Z_i=z$, are independent, each with the known conditional distribution function $G(\cdot\mid z)$.

Since $Z_1,\cdots,Z_k$ are independent, the random vectors $(X_{11},\cdots,X_{1n_1})$, $\cdots,(X_{k1},\cdots,X_{kn_k})$ are also independent. Specifically, the joint distribution function $H_\theta(x_{11},\cdots,x_{1n_1},\cdots,x_{kn_k})$ of $\{X_{ij};\ j=1,\cdots,n_i,\ i=1,\cdots,k\}$ is, for each $\theta\in\Omega$,

$$H_\theta(x_{11},\cdots,x_{1n_1},\cdots,x_{k1},\cdots,x_{kn_k})=\prod_{i=1}^{k}\int_{\mathscr{Z}}\left[\prod_{j=1}^{n_i}G(x_{ij}\mid z_i)\right]dP_\theta(z_i)\,. \tag{2.2}$$

If R_n denotes n-dimensional Euclidean space and $\mathscr{B}_n$ denotes the class of n-dimensional Borel sets, then the observational procedure just described may be expressed as the experiment

$$X=\{R_n;\ \mathscr{B}_n;\ H_\theta,\ \theta\in\Omega\}\qquad (n=\sum_{i=1}^{k} n_i)\,. \tag{2.3}$$

DEFINITION 2.3. A specific choice of values for k and $n_1, \cdots, n_k$ is called an *allocation* and denoted by $A(\sum_{i=1}^{k} n_i;\ k;\ n_1, \cdots, n_k)$. Each allocation $A(\sum_{i=1}^{k} n_i;\ k;\ n_1, \cdots, n_k)$ results in an experiment X defined by (2.3) and (2.2). It represents an allocation of a total of $\sum_{i=1}^{k} n_i$ observations among k random variables $Z_1, \cdots, Z_k$ in such a way that n_i observations are made on Z_i $(i=1, \cdots, k)$. The random variables $Z_1, \cdots, Z_k$ are called *the individuals on which observations are made*; k is the number of individuals on which observations are made. Accordingly, the notation $A(n;\ k;\ n_1, \cdots, n_k)$ is used only when $n, k, n_1, \cdots, n_k$ are all positive integers with $\sum_{i=1}^{k} n_i=n$.

DEFINITION 2.4. For any given positive integer n, let $\mathscr{C}_n$ denote the class of all allocations $A(n;\ k;\ n_1, \cdots, n_k)$ or equivalently, the class of all experiments X for which the total number of observations is n.

DEFINITION 2.5. An experiment X in $\mathscr{C}_n$ (or equivalently, an allocation $A(n;\ k;\ n_1, \cdots, n_k)$) is *optimal in* $\mathscr{C}_n$ if it is sufficient for every other experiment in $\mathscr{C}_n$.

The central problem with which we will be concerned in the remainder of this paper is that of exploring the possibility that $\mathscr{C}_n$ contains an optimal allocation.

DEFINITION 2.6. Let $X=\{\mathscr{X};\ \mathscr{A};\ P_\theta,\ \theta\in\Omega\}$ and $Y=\{\mathscr{Y};\ \mathscr{B};\ Q_\theta,\ \theta\in\Omega\}$ be two experiments with the same parameter space Ω. We denote by $X\times Y=\{\mathscr{X}\times\mathscr{Y};\ \mathscr{A}\times\mathscr{B};\ P_\theta\times Q_\theta,\ \theta\in\Omega\}$ the composite experiment consisting of independent observations of X and Y. In particular, in terms of allocations, the composite experiment $A(m;\ k;\ m_1, \cdots, m_k)\times A(n;\ l;\ n_1, \cdots, n_l)$ is equivalent to the experiment $A(m+n;\ k+l;\ m_1, \cdots, m_k, n_1, \cdots, n_l)$ in the sense that the $m+n$ observations have the same distribution from both points of view.

In general, it is relatively difficult to establish the optimality of a particular allocation since it must be shown to be sufficient for every other allocation in $\mathscr{C}_n$. However, the next two theorems provide helpful reductions in studying the allocation $A(n;\ n;\ 1, \cdots, 1)$ (in which 1 observation is made on each of n individuals) and the allocation $A(n;\ 1;\ n)$ (in which all n observations are made on 1 individual). Before presenting the theorems we state the following elementary result ([3], p. 332).

LEMMA 2.7. *Let* $X_1, \cdots, X_r$ *and* $Y_1, \cdots, Y_r$ *be experiments such* *that X_i is sufficient for $Y_i (i=1, \cdots, r)$. Then $X_1 \times \cdots \times X_r$ is sufficient for $Y_1 \times \cdots \times Y_r$.*

THEOREM 2.8. *Let n be a fixed positive integer. A necessary and sufficient condition that $A(m; m; 1, \cdots, 1)$ be optimal in $\mathscr{C}_m$ for every value of m, $1 \leqq m \leqq n$, is that $A(m; m; 1, \cdots, 1)$ be sufficient for $A(m; 1; m)$ for every value of m, $1 \leqq m \leqq n$.*

PROOF. Suppose that $A(m; m; 1, \cdots, 1)$ is sufficient for $A(m; 1; m)$ for every value of m, $1 \leqq m \leqq n$. Let r be a positive integer, $1 \leqq r \leqq n$. It must be shown that $A(r; r; 1, \cdots, 1)$ is sufficient for every other allocation $A(r; k; r_1, \cdots, r_k)$ in $\mathscr{C}_r$. But $A(r; r; 1, \cdots, 1)$ can be regarded as the composite experiment

$$(2.4) \quad A(r; r; 1, \cdots, 1) = A(r_1; r_1; 1, \cdots, 1) \times \cdots \times A(r_k; r_k; 1, \cdots, 1)$$

and $A(r; k; r_1, \cdots, r_k)$ can be regarded as the composite experiment

$$(2.5) \quad A(r; k; r_1, \cdots, r_k) = A(r_1; 1; r_1) \times \cdots \times A(r_k; 1; r_k).$$

The desired result now follows from Lemma 2.7, since each component of (2.4) is sufficient for the corresponding component of (2.5).

The converse of the theorem is trivial. If $A(m; m; 1, \cdots, 1)$ is optimal in $\mathscr{C}_m$ then it is sufficient for every allocation in $\mathscr{C}_m$. In particular, it is sufficient for $A(m; 1; m)$.

THEOREM 2.9. *Let n be a fixed positive integer. A necessary and sufficient condition that $A(m; 1; m)$ be optimal in $\mathscr{C}_m$ for every value of m, $1 \leqq m \leqq n$, is that, for each value of m, $1 \leqq m \leqq n$, $A(m; 1; m)$ be sufficient for every allocation in $\mathscr{C}_m$ of the form $A(m; 2; m_1, m_2)$.*

PROOF. Suppose that for each value of m, $1 \leqq m \leqq n$, $A(m; 1; m)$ is sufficient for every allocation of the form $A(m; 2; m_1, m_2)$. Let r be a positive integer, $1 \leqq r \leqq n$. It must be shown that $A(r; 1; r)$ is sufficient for every other allocation $A(r; k; r_1, \cdots, r_k)$ in $\mathscr{C}_r$. It follows from the hypothesis that $A(r; 1; r)$ is sufficient for $A(r; 2; r_1, \sum_{i=2}^{k} r_i)$; that $A(\sum_{i=2}^{k} r_i; 1; \sum_{i=2}^{k} r_i)$ is sufficient for $A(\sum_{i=2}^{k} r_i; 2; r_2, \sum_{i=3}^{k} r_i)$; and, in general, that $A(\sum_{i=j}^{k} r_i; 1; \sum_{i=j}^{k} r_i)$ is sufficient for $A(\sum_{i=j}^{k} r_i; 2; r_j, \sum_{i=j+1}^{k} r_i)$ for $j=1, \cdots, k-1$. Since $A(\sum_{i=j}^{k} r_i; 2; r_j, \sum_{i=j+1}^{k} r_i)$ can be regarded as the composite experiment $A(r_j; 1; r_j) \times A(\sum_{i=j+1}^{k} r_i; 1; \sum_{i=j+1}^{k} r_i)$, it follows from Lemma 2.7 and the above results that each allocation in the following sequence is sufficient for its successor in the sequence:

$$(2.6) \qquad A(r;\, 1;\, r) = A(\sum_{i=1}^{k} r_i;\, 1;\, \sum_{i=1}^{k} r_i)\,,$$

$$A(r_1;\, 1;\, r_1) \times A(\sum_{i=2}^{k} r_i;\, 1;\, \sum_{i=2}^{k} r_i),\, \cdots,$$

$$A(r_1;\, 1;\, r_1) \times \cdots \times A(r_j;\, 1;\, r_j) \times A(\sum_{i=j+1}^{k} r_i;\, 1;\, \sum_{i=j+1}^{k} r_i),\, \cdots,$$

$$A(r_1;\, 1;\, r_1) \times \cdots \times A(r_k;\, 1;\, r_k).$$

Since the final allocation in (2.6) is equivalent to $A(r;\, k;\, r_1, \cdots, r_k)$, and since sufficiency is clearly a transitive relation, it follows that $A(r;\, 1;\, r)$ is sufficient for $A(r;\, k;\, r_1, \cdots, r_k)$. The converse of the theorem again is trivial. These results will be illustrated in subsequent sections of the paper.

It should be noted that if there exists a function φ of Z such that $P(X=\varphi(Z) \mid Z=z)=1$ for all values of z, then clearly it would be redundant to make more than one observation on any individual and, hence, the allocation $A(n;\, n;\, 1, \cdots, 1)$ is optimal.

3. Sufficient statistics

We will now show how the familiar concept of a sufficient statistic can often be used to simplify the search for an optimal allocation.

Let $X=\{\mathscr{X};\, \mathscr{A};\, P_\theta,\, \theta \in \Omega\}$ be a given experiment. Recall that $S=s(X)$ is a sufficient statistic for X if it is possible to generate, from S and an auxiliary randomization, a random variable having the same distribution as X for all $\theta \in \Omega$. It is clear that if $S=s(X)$ is a sufficient statistic for the experiment X and if $T=t(Y)$ is a sufficient statistic for another experiment Y with the same parameter space Ω, then X is sufficient for Y if and only if the experiment in which only S is observed is sufficient for the experiment in which only T is observed. Thus, in the search for an optimal allocation one can restrict himself to studying the distributions of sufficient statistics. This, of course, often has the value of greatly reducing the dimensions of the spaces between which a stochastic transformation must be constructed.

Now consider the allocation $A(n;\, 1;\, n)$ defined in the preceding section. Suppose that the conditional distribution of an observation X, given the individual $Z=z$, is $G(\cdot \mid z)$. Then the joint conditional distribution of n observations $X_1, \cdots, X_n$, all made on the same individual, is $G_n(\cdot \mid z)$, the n-dimensional product probability distribution each of whose components is $G(\cdot \mid z)$. The performance of the experiment $A(n;\, 1;\, n)$ can be thought of as follows. First, the experiment $Z=\{\mathscr{Z};\, \mathscr{A};\, P_\theta,\, \theta \in \Omega\}$, the selection of one individual, is carried out. Then without learning the value of z that was obtained, it is followed by the experiment $Y=\{R_n;\ \mathscr{B}_n;\ G_n(\cdot \mid z),\ z \in \mathscr{Z}\}$ in which the components of the vector $(X_1, \cdots, X_n)$ of observations are independent and z is the parameter. Together, the two stages of experimentation result in the overall experiment $Y^*Z=\{R_n;\ \mathscr{B}_n;\ H_\theta,\ \theta \in \Omega\}=A(n;\ 1;\ n)$, where H_θ, defined in Section 2, is the marginal joint distribution of $X_1, \cdots, X_n$.

The next result is helpful in identifying a sufficient statistic.

THEOREM 3.1. *If $T=t(X_1, \cdots, X_n)$ is a sufficient statistic for the experiment $Y=\{R_n;\ \mathscr{B}_n;\ G_n(\cdot \mid z),\ z \in \mathscr{Z}\}$ then T is a sufficient statistic for the allocation $A(n;\ 1;\ n)$.*

PROOF. Let $\mathscr{T}$ be the sample space of T and let $\mathscr{D}$ be the σ-field of all subsets D of $\mathscr{T}$ such that $t^{-1}(D) \in \mathscr{B}_n$. Furthermore, for each $z \in \mathscr{Z}$, let α_z be the probability distribution of T when the distribution of $(X_1, \cdots, X_n)$ is $G_n(\cdot \mid z)$, and for each $\theta \in \Omega$, let β_θ be the probability distribution of T when the distribution of $(X_1, \cdots, X_n)$ is H_θ. Recall that

$$(3.1) \qquad H_\theta(B)=\int_{\mathscr{Z}} G_n(B \mid z)dP_\theta(z),\ B \in \mathscr{B}_n, \text{ and hence, for each } D \in \mathscr{D},$$

$$(3.2) \qquad \beta_\theta(D)=H_\theta(t^{-1}(D))=\int_{\mathscr{Z}} G_n(t^{-1}(D) \mid z)dP_\theta(z)=\int_{\mathscr{Z}} \alpha_z(D)dP_\theta(z)\ .$$

Since T is a sufficient statistic for the experiment Y, there exists a stochastic transformation $\pi(\cdot \mid \cdot)$ from T to Y such that, for every $B \in \mathscr{B}_n$ and $z \in \mathscr{Z}$,

$$(3.3) \qquad G_n(B \mid z)=\int_{\mathscr{T}} \pi(B \mid t)d\alpha_z(t)\ .$$

But it now follows from (3.1), (3.3) and (3.2) that, for every $B \in \mathscr{B}_n$ and $\theta \in \Omega$,

$$(3.4) \qquad \begin{aligned} H_\theta(B)&=\int_{\mathscr{Z}} \int_{\mathscr{T}} \pi(B \mid t)d\alpha_z(t)dP_\theta(z) \\ &=\int_{\mathscr{T}} \pi(B \mid t)d\beta_\theta(t), \end{aligned}$$

since the interchange of the order of integration needed to attain the final expression in (3.4) is justified. Thus, it is seen from (3.4) that $\pi(\cdot \mid \cdot)$ satisfies the requirements of Definition 2.2 and, hence, that T is a sufficient statistic for the allocation $A(n;\ 1;\ n)$.

COROLLARY 3.2. *Suppose that for each value of $m (m=1, 2, \cdots)$ the statistic $t_m(X_1, \cdots, X_m)$ is a sufficient statistic for the experiment $Y=\{R_m;\ \mathscr{B}_m;\ G_m(\cdot \mid z),\ z \in \mathscr{Z}\}$. Then, for all values of k and $n_1, \cdots, n_k$, the* *random vector*

(3.5) $$\{t_{n_1}(X_{11}, \cdots, X_{1n_1}), \cdots, t_{n_k}(X_{k1}, \cdots, X_{kn_k})\}$$

is a joint sufficient statistic for the allocation $A(n; k; n_1, \cdots, n_k)$.

PROOF. The allocation $A(n; k; n_1, \cdots, n_k)$ can be regarded as the composition $A(n_1; 1; n_1)\times \cdots \times A(n_k; 1; n_k)$ of k independent experiments. The desired result follows from Theorem 3.1.

The following discussion illustrates some of the results that have been derived in the present and preceding sections. The first result deals with a situation where for each value of n the allocation $A(n; n; 1, \cdots, 1)$ is optimal.

THEOREM 3.3. *Suppose that Z is normally distributed with unknown mean θ and known variance σ^2. Suppose that an observation made on an individual Z is, conditionally on $Z=z$, normally distributed with mean z and known variance τ^2. Then for all values of $n(n=1, 2, \cdots)$, the allocation $A(n; n; 1, \cdots, 1)$ is optimal in $\mathscr{C}_n$.*

PROOF. It follows from the above conditions that the marginal distribution of each observation is normal with mean θ and variance $\sigma^2+\tau^2$. Under the allocation $A(n; n; 1, \cdots, 1)$, in which 1 observation is made on each of n individuals, we obtain n independent observations, each with this marginal distribution. It is well-known that $S=\frac{1}{n}\sum_{i=1}^{n} X_i$ is a sufficient statistic for this experiment and the distribution of S is normal with mean θ and variance $(\sigma^2+\tau^2)/n$.

Under the allocation $A(n; 1; n)$, in which n observations are made on 1 individual, we observe $Y_1, \cdots, Y_n$ which, conditionally on $Z=z$, are independent and normally distributed, each with mean z and variance τ^2. It is well known that $T=\frac{1}{n}\sum_{i=1}^{n} Y_i$ is a sufficient statistic for this (conditional) experiment and, hence, by Theorem 3.1, T is a sufficient statistic for the allocation $A(n; 1; n)$. The conditional distribution of T, given $Z=z$, is normal with mean z and variance τ^2/n and it follows that the marginal distribution of T is normal with mean θ and variance $\sigma^2+(\tau^2/n)$.

But the experiment in which S is observed is sufficient for the experiment in which T is observed. This is seen by noting that if we define an auxiliary random variable U that is normally distributed independently of S, with mean 0 and variance $\frac{n-1}{n}\sigma^2$, then $S+U$ has the same distribution as T for all values of θ. This in turn implies that the allocation $A(n; n; 1, \cdots, 1)$ is sufficient for the allocation $A(n; 1; n)$. Since this is true for all positive integers n, it follows from Theorem 2.8 that $A(n; n; 1, \cdots, 1)$ is optimal in $\mathscr{C}_n$ for every value of n.

Thus, in making an inference about the mean of a normal distribution with known variance when the observations are themselves subject to a normally distributed error, $A(n; n; 1, \cdots, 1)$ is the optimal allocation.

The next results concern situations where the allocation $A(n; 1; n)$, in which all n observations are made on 1 individual, is optimal.

THEOREM 3.4. *Suppose that for each $\theta \in \Omega$, the random variable Z takes only two values, a_θ and b_θ, with $P_\theta(Z=a_\theta)=P_\theta(Z=b_\theta)=1/2$. For each value z of Z, let $G(\cdot \mid z)$ be the conditional distribution function of an observation X given $Z=z$. Suppose that there exists a function $\psi(X)$ such that, for all $\theta \in \Omega$, if X has distribution function $G(\cdot \mid a_\theta)$ then $\psi(X)$ has distribution function $G(\cdot \mid b_\theta)$, and if X has distribution function $G(\cdot \mid b_\theta)$ then $\psi(X)$ has distribution function $G(\cdot \mid a_\theta)$. Under these conditions the allocation $A(n; 1; n)$ is optimal in $\mathscr{C}_n$ for all values of n $(n=1, 2, \cdots)$.*

PROOF. Under the allocation $A(n; 1; n)$ we obtain n observations whose joint conditional distribution function, given $Z=z$, is $\prod_{i=1}^{n} G(x_i \mid z)$. Thus, the marginal joint distribution function of $X_1, \cdots, X_n$ is, for each $\theta \in \Omega$,

$$\frac{1}{2}\left[\prod_{i=1}^{n} G(x_i \mid a_\theta)+\prod_{i=1}^{n} G(x_i \mid b_\theta)\right]. \tag{3.6}$$

Now consider any other allocation of the form $A(n; 2; n_1, n_2)$. Under this allocation we observe $Y_1, \cdots, Y_{n_1}, Y_{n_1+1}, \cdots, Y_n$ whose joint distribution function is, for each $\theta \in \Omega$,

$$\frac{1}{2}\left[\prod_{i=1}^{n_1} G(y_i \mid a_\theta)+\prod_{i=1}^{n_1} G(y_i \mid b_\theta)\right] \cdot \frac{1}{2}\left[\prod_{i=n_1+1}^{n} G(y_i \mid a_\theta)+\prod_{i=n_1+1}^{n} G(y_i \mid b_\theta)\right]. \tag{3.7}$$

To show that $A(n; 1; n)$ is sufficient for $A(n; 2; n_1, n_2)$ we exhibit the following stochastic transformation from $(X_1, \cdots, X_n)$ to $(Y_1, \cdots, Y_n)$. Let

$$(U_1, \cdots, U_n)=\begin{cases} (X_1, \cdots, X_n) \text{ with probability } 1/2, \\ (X_1, \cdots, X_{n_1}, \psi(X_{n_1+1}), \cdots, \psi(X_n)) \\ \qquad\qquad \text{with probability } 1/2. \end{cases} \tag{3.8}$$

It now follows from (3.6) and the assumed properties of the function $\psi(X)$ that the joint distribution function of $(U_1, \cdots, U_n)$, conditionally on each of the two possibilities listed in (3.8), is

$$
(3.9)\quad \begin{cases} \dfrac{1}{2}\left[\prod_{i=1}^{n} G(u_i \mid a_\theta)+\prod_{i=1}^{n} G(u_i \mid b_\theta)\right] \\ \qquad \text{if } (U_1, \cdots, U_n)=(X_1, \cdots, X_n); \\ \dfrac{1}{2}\left[\prod_{i=1}^{n_1} G(u_i \mid a_\theta)\prod_{i=n_1+1}^{n} G(u_i \mid b_\theta)+\prod_{i=1}^{n_1} G(u_i \mid b_\theta)\prod_{i=n_1+1}^{n} G(u_i \mid a_\theta)\right] \\ \qquad \text{if } (U_1, \cdots, U_n)=(X_1, \cdots, X_{n_1}, \psi(X_{n_1+1}), \cdots, \psi(X_n)) . \end{cases}
$$

Since each of the above possibilities occurs with probability 1/2, the joint distribution function of $(U_1, \cdots, U_n)$ is the average of the two expressions given in (3.9). When this result is compared with (3.7) it is seen that $(U_1, \cdots, U_n)$ and $(Y_1, \cdots, Y_n)$ have the same distribution for all $\theta \in \Omega$.

It follows that $A(n; 1; n)$ is sufficient for $A(n; 2; n_1, n_2)$ and hence, by Theorem 2.9, $A(n; 1; n)$ is optimal in $\mathscr{C}_n$ for all values of n.

COROLLARY 3.5. *Suppose that $P(Z=\theta)=P(Z=-\theta)=1/2$, where $\theta \geqq 0$ is an unknown parameter. Suppose that the conditional distribution of an observation, X, given $Z=z$, is normal with mean z and variance σ^2. Then the allocation $A(n; 1; n)$ is optimal in $\mathscr{C}_n$ for all values of $n(n=1, 2, \cdots)$.*

PROOF. The function $\psi(X)=-X$ satisfies the hypothesis of Theorem 3.4 and the desired result follows immediately.

Another example of this type will be given in Section 5.

4. Complete families of distributions

In problems in which a stochastic transformation from one experiment X to another Y cannot readily be found, some criterion is needed for deciding whether or not X is sufficient for Y. In this section we will derive a criterion of this type that is often useful and is expressed in terms of the standard statistical concepts of completeness and unbiasedness.

It is interesting to note that although we will be concerned below with the existence and construction of unbiased estimators of various functions of a parameter θ, we will not be concerned with their usefulness as estimators. They are merely devices used in the construction of sufficient experiments. Thus, the present section illustrates how some of the mathematical techniques and methods developed for the standard theory of unbiased estimation can be utilized in statistical studies (based, e.g., on the likelihood principle or the Bayesian mode of analysis) that are not explicitly concerned with unbiased estimation of the parameters of interest.

Recall that for a given experiment $X=\{\mathscr{X};\ \mathscr{A};\ P_\theta,\ \theta\in\Omega\}$, the family of distributions $\{P_\theta,\ \theta\in\Omega\}$ is said to be *complete* if every $\mathscr{A}$-measurable function $f(\cdot)$ satisfying

$$(4.1)\qquad \int_{\mathscr{X}} f(x)dP_\theta(x)=0 \qquad \text{for all } \theta\in\Omega\,,$$

is such that $f(x)=0$ a.e. (X). The condition $f(x)=0$ a.e. (X) means that $P_\theta\{x: f(x)\neq 0\}=0$ for every $\theta\in\Omega$.

A real-valued function $\varphi(\cdot)$ defined on Ω is said to be *non-negatively estimable* (X) if there exists a non-negative, $\mathscr{A}$-measurable function $f(\cdot)$ on $\mathscr{X}$ such that

$$(4.2)\qquad \int_{\mathscr{X}} f(x)dP_\theta(x)=\varphi(\theta) \qquad \text{for all } \theta\in\Omega\,.$$

The proof of the next theorem closely follows the usual proof of the existence of conditional distributions, as given, e.g., in [7], p. 44, or [5], p. 31, and, hence, many of the details are omitted. It should be emphasized, however, that the present context is different from the context in the above references.

THEOREM 4.1. *Let* $X=\{\mathscr{X};\ \mathscr{A};\ P_\theta,\ \theta\in\Omega\}$ *and* $Y=\{\mathscr{Y};\ \mathscr{B};\ Q_\theta,\ \theta\in\Omega\}$ *be two experiments. Suppose that the family* $\{P_\theta,\ \theta\in\Omega\}$ *is complete. Suppose also, that* $\mathscr{Y}$ *is a Euclidean space and that the* σ*-field* $\mathscr{B}$ *is the class of all Borel sets. Then* X *is sufficient for* Y *if and only if, for each fixed* $B\in\mathscr{B}$*, the function* $Q_\theta(B)$*, considered as a function on* Ω*, is non-negatively estimable* (X).

PROOF. Suppose that for each Borel set $B\in\mathscr{B}$, there exists a non-negative, $\mathscr{A}$-measurable function $f(B\mid\cdot)$ such that

$$(4.3)\qquad \int_{\mathscr{X}} f(B\mid x)dP_\theta(x)=Q_\theta(B) \qquad \text{for all } \theta\in\Omega.$$

To show that X is sufficient for Y we must show that there exists a stochastic transformation $\pi(\cdot\mid\cdot)$ from X to Y such that (4.3) holds for all $B\in\mathscr{B}$ and all $\theta\in\Omega$ when $f(\cdot\mid\cdot)$ is replaced by $\pi(\cdot\mid\cdot)$.

Since $\{P_\theta,\ \theta\in\Omega\}$ is complete and Q_θ is, for each $\theta\in\Omega$, a probability distribution, it can be shown that

$$(4.4)\qquad \begin{cases} \text{(i)}\ \ f(\mathscr{Y}\mid x)=1 \ \text{ a.e. } (X),\ f(\phi\mid x)=0 \ \text{ a.e. } (X); \\ \text{(ii)}\ \ \text{For each } B\in\mathscr{B},\ 0\leqq f(B\mid x)\leqq 1 \ \text{ a.e. } (X); \\ \text{(iii)}\ \ \text{For each sequence } \{B_j;\ j=1,2,\cdots\} \text{ of disjoint} \\ \qquad \text{events from } \mathscr{B}, \\ \qquad f(\bigcup_{j=1}^{\infty} B_j\mid x)=\sum_{j=1}^{\infty} f(B_j\mid x) \ \text{ a.e. } (X)\,. \end{cases}$$

It now follows, as in the standard proofs referred to above, that $f(\cdot \mid \cdot)$ can be replaced by a function $\pi(\cdot \mid \cdot)$ such that for *each* $x \in \mathscr{X}$, the above properties (4.4) hold and all integrals (4.3) remain unchanged. This completes the proof in one direction.

The proof of the converse is trivial since, if X is sufficient for Y, equation (2.1) immediately reveals that $Q_\theta(B)$ is non-negatively estimable (X).

Theorem 4.1 is particularly useful for dealing with situations where the distributions P_θ and Q_θ are discrete. The following corollary is a simple consequence of the theorem.

COROLLARY 4.2. *Let* $X=\{\mathscr{X};\ \mathscr{A};\ P_\theta,\ \theta \in \Omega\}$ *be an experiment such that the random variable* X *has, for each* $\theta \in \Omega$, *a discrete distribution specified by*

$$P_\theta(X=x)=p_\theta(x) \qquad \text{for } x=x_1, x_2, \cdots . \tag{4.5}$$

Similarly, let $Y=\{\mathscr{Y};\ \mathscr{B};\ Q_\theta,\ \theta \in \Omega\}$ *be an experiment such that* Y *has a discrete distribution for each* $\theta \in \Omega$ *given by*

$$Q_\theta(Y=y)=q_\theta(y) \qquad \text{for } y=y_1, y_2, \cdots . \tag{4.6}$$

Assume that the family $\{P_\theta,\ \theta \in \Omega\}$ *is complete; i.e., if* $\sum_{i=1}^{\infty} a_i p_\theta(x_i)=0$ *for all* $\theta \in \Omega$, *then* $a_i=0 (i=1, 2, \cdots)$. *Then* X *is sufficient for* Y *if and only if, for each value* $y_j (j=1, 2, \cdots)$ *there exist non-negative constants* $\{a_i;\ i=1, 2, \cdots\}$, *depending on* y_j, *such that*

$$q_\theta(y_j)=\sum_{i=1}^{\infty} a_i p_\theta(x_i) \qquad \text{for all } \theta \in \Omega . \tag{4.7}$$

It is shown in the next section (Example 5.2) that the explicit assumption that the functions $f(B \mid \cdot)$ in (4.3), and the constants a_i in (4.7), are non-negative is necessary.

The preceding results are valuable in the search for an optimal allocation since it is often true that both the allocations $A(n;\ 1;\ n)$ and $A(n;\ n;\ 1, \cdots, n)$ admit a complete, real-valued sufficient statistic. We now turn to a class of problems for which this is so: the allocation of Bernoulli observations.

5. Optimal allocation of Bernoulli observations

In this section we will present a few results for the special class of allocation problems in which the random variable Z takes values in the unit interval [0, 1] with a distribution belonging to a given family $\{P_\theta,\ \theta \in \Omega\}$, and for which the conditional distribution of an observation X, given $Z=z$, is specified by

$$\Pr(X=1 \mid Z=z)=z\,, \qquad \Pr(X=0 \mid Z=z)=1-z\,. \tag{5.1}$$

It follows from Theorem 3.1 that the sum T_n of the observations is a sufficient statistic for the allocation $A(n; 1; n)$ in which n observations are made on one individual. The distribution of T_n is, for each $\theta \in \Omega$, given by

$$q_{n,t}(\theta)=P_\theta(T_n=t)=\binom{n}{t}\int_{[0,1]} z^t(1-z)^{n-t}dP_\theta(z)\,, \qquad t=0, 1, \cdots, n\,. \tag{5.2}$$

It is seen from (5.2) that the distribution of T_n involves only the first n moments of the distribution P_θ. Thus, the probability distribution of the observations obtained from any allocation in which at most r observations are made on any individual involves at most the first r moments of P_θ. In particular, the probability distribution of the observations obtained from the allocation $A(n; n; 1, \cdots, 1)$ depends only on the mean $E_\theta(Z)$. If $E_\theta(Z)=c$, a constant, for all $\theta \in \Omega$, the allocation $A(n; n; 1, \cdots, 1)$ is useless.

It follows from (5.1) that the marginal distribution of an observation X is given by $P_\theta(X=1)=E_\theta(Z)$, $P_\theta(X=0)=1-E_\theta(Z)$. Since the n observations obtained under the allocation $A(n; n; 1, \cdots, 1)$ are independent, the sum S_n of the observations is a sufficient statistic for this allocation. The distribution of S_n is, for each $\theta \in \Omega$, given by

$$p_{n,s}(\theta)=P_\theta(S_n=s)=\binom{n}{s}[E_\theta(Z)]^s[1-E_\theta(Z)]^{n-s}\,, \qquad s=0, 1, \cdots, n\,. \tag{5.3}$$

The above discussion makes it clear that if there are two values of θ, say, θ_1 and θ_2, for which $E_{\theta_1}(Z)=E_{\theta_2}(Z)$ but $E_{\theta_1}(Z^r)\neq E_{\theta_2}(Z^r)$ for some value of r, $1<r\leqq n$, then the allocation $A(n; n; 1, \cdots, 1)$ cannot be optimal in $\mathscr{C}_n$.

For any positive constants α and β let $b(\cdot \mid \alpha, \beta)$ denote the beta density function defined by

$$b(z \mid \alpha, \beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}z^{\alpha-1}(1-z)^{\beta-1} \tag{5.4}$$

for $0<z<1$, and $b(z \mid \alpha, \beta)=0$ elsewhere. It follows from the preceding comments that if the family of distributions of Z is the family of all beta distributions, then the allocation $A(n; n; 1, \cdots, 1)$ cannot be optimal. It is shown in the next theorem, however, that if the distribution of Z is restricted to an appropriate subfamily, no two members of which have the same mean, the allocation $A(n; n; 1, \cdots, 1)$ is optimal.

THEOREM 5.1. *Suppose that the density function of Z is the beta* *density* $b(z \mid \theta, k-\theta)$, *as defined by (5.4), where* $k>0$ *is a known constant and* θ *is an unknown parameter,* $0<\theta<k$. *Then, for every value of* $n (n=1, 2, \cdots)$, *the allocation* $A(n; n; 1, \cdots, 1)$ *is optimal in* $\mathscr{C}_n$.

PROOF. The distribution of the sufficient statistic S_n for the allocation $A(n; n; 1, \cdots, 1)$ is given by (5.3) where, as is well-known, $E_\theta(Z)=\theta/k$. Since θ/k can take any value in the interval (0, 1), this family of distributions is complete. Furthermore, by (5.2), the distribution of the sufficient statistic T_n for the allocation $A(n; 1; n)$ is given by

$$q_{n,t}(\theta)=\binom{n}{t}\frac{\Gamma(k)}{\Gamma(\theta)\Gamma(k-\theta)}\int_{[0,1]} z^{t+\theta-1}(1-z)^{n-t+k-\theta-1}dz \tag{5.5}$$

$$=\binom{n}{t}\frac{\Gamma(k)}{\Gamma(k+n)}\cdot\frac{\Gamma(\theta+t)}{\Gamma(\theta)}\cdot\frac{\Gamma(k-\theta+n-t)}{\Gamma(k-\theta)}, \quad 0<\theta<k .$$

It now follows from Theorem 2.8 and Corollary 4.2 that, in order to show that $A(n; n; 1, \cdots, 1)$ is optimal in $\mathscr{C}_n$ for all values of n, it is sufficient to show that for each value of t, $t=0, 1, \cdots, n$, there exist non-negative constants $\{a_s; s=0, 1, \cdots, n\}$, depending on t and n such that

$$q_{n,t}(\theta)=\sum_{s=0}^{n} a_s\binom{n}{s}\left(\frac{\theta}{k}\right)^s\left(1-\frac{\theta}{k}\right)^{n-s}, \quad 0<\theta<k . \tag{5.6}$$

From the explicit expression (5.5) for $q_{n,t}(\theta)$ it is seen that there is an expansion of the form (5.6) if and only if

$$\frac{\Gamma(\theta+t)}{\Gamma(\theta)}\cdot\frac{\Gamma(k-\theta+n-t)}{\Gamma(k-\theta)}=\sum_{s=0}^{n} b_s\theta^s(k-\theta)^{n-s}, \quad 0<\theta<k , \tag{5.7}$$

for some non-negative constants $\{b_s; s=0, 1, \cdots, n\}$ depending on t and n. But

$$\frac{\Gamma(\theta+t)}{\Gamma(\theta)}\cdot\frac{\Gamma(k-\theta+n-t)}{\Gamma(k-\theta)}=\left[\prod_{i=0}^{t-1}(\theta+i)\right]\left[\prod_{j=0}^{n-t-1}(k-\theta+j)\right] \tag{5.8}$$

$$=\left[\sum_{i=0}^{t}\alpha_i\theta^i\right]\left[\sum_{j=0}^{n-t}\beta_j(k-\theta)^j\right] ,$$

where the coefficients α_i and β_j are all non-negative. It should be noted that the final expression in (5.8) is valid for all values of t, including $t=0$ and $t=n$, even though the middle expression of (5.8) does not hold at these values. Finally, recall that for $i\geqq 0$, $j\geqq 0$, and $i+j\leqq n$, there exist non-negative constants γ_{ijs}, $s=0, 1, \cdots, n$, such that

$$\theta^i(k-\theta)^j=\sum_{s=0}^{n}\gamma_{ijs}\theta^s(k-\theta)^{n-s}, \quad 0<\theta<k . \tag{5.9}$$

More specifically,

$$\theta^i(k-\theta)^j=\sum_{s=i}^{n-j}\binom{n-i-j}{s-i}\frac{1}{k^{n-i-j}}\theta^s(k-\theta)^{n-s}\,,\quad 0<\theta<k\,. \tag{5.10}$$

Together, (5.8) and (5.9) yield an expansion of the desired form (5.7). This completes the proof.

In connection with the remark made following the proof of Corollary 4.2 in the preceding section, the next example shows that there may exist constants a_i satisfying (4.7), some of which are negative. Thus, in applying Theorem 4.1 or Corollary 4.2 it is always important to verify the non-negativity of the required solutions.

Example 5.2. Suppose that Z can take only two given values z_1 and z_2, $0<z_1<z_2<1$, and $P_\theta(Z=z_1)=1-P_\theta(Z=z_2)=\theta$, where θ is an unknown parameter, $0\leqq\theta\leqq 1$. Under the allocation $A(n;\,1;\,n)$ the distribution of the sufficient statistic T_n is, as defined by (5.2),

$$q_{n,t}(\theta)=\theta\binom{n}{t}z_1^t(1-z_1)^{n-t}+(1-\theta)\binom{n}{t}z_2^t(1-z_2)^{n-t}\,,\qquad t=0,\,1,\,\cdots,\,n\,. \tag{5.11}$$

Under the allocation $A(n;\,n;\,1,\,\cdots,\,1)$ the distribution of the sufficient statistic S_n is, as defined by (5.3)

$$p_{n,s}(\theta)=\binom{n}{s}[\theta z_1+(1-\theta)z_2]^s[1-\theta z_1-(1-\theta)z_2]^{n-s}\,,\qquad s=0,\,1,\,\cdots,\,n\,. \tag{5.12}$$

The allocation $A(n;\,n;\,1,\,\cdots,\,1)$ is sufficient for the allocation $A(n;\,1;\,n)$ if and only if, for each value of $t(t=0,\,1,\,\cdots,\,n)$, there exist non-negative constants $a_s(s=0,\,1,\,\cdots,\,n)$, depending on t and n, such that

$$q_{n,t}(\theta)=\sum_{s=0}^{n}a_s p_{n,s}(\theta)\qquad(0\leqq\theta\leqq 1)\,. \tag{5.13}$$

It follows immediately from (5.12) that

$$(1/n)\sum_{s=0}^{n}s p_{n,s}(\theta)=\theta z_1+(1-\theta)z_2\qquad(0\leqq\theta\leqq 1)\,. \tag{5.14}$$

Hence, if the constants $a_s(s=0,\,1,\,\cdots,\,n)$ are defined, for a given value of t, to be

$$a_s=\binom{n}{t}z_1^t(1-z_1)^{n-t}\left(\frac{nz_2-s}{nz_2-nz_1}\right)+\binom{n}{t}z_2^t(1-z_2)^{n-t}\left(\frac{s-nz_1}{nz_2-nz_1}\right)\,, \tag{5.15}$$

then (5.13) is satisfied. Furthermore, since the family of distributions defined by (5.12) for $0 \leqq \theta \leqq 1$ is complete, the constants defined by (5.15) are, for each value of t, the only ones that satisfy (5.13). However, if $n \geqq 2$ then for some values of s and t (e.g., $s=0$, $t=n$), $a_s<0$. Hence, there is no set of non-negative constants satisfying (5.13) and $A(n; n; 1, \cdots, 1)$ is not sufficient for $A(n; 1; n)$.

The next result provides an example for Bernoulli observations in which, for each value of n, the allocation $A(n; 1; n)$ is optimal in $\mathscr{C}_n$.

THEOREM 5.3. *Suppose that the distribution of Z is given by $P_\theta(Z=\theta)=P_\theta(Z=1-\theta)=1/2$, where θ is an unknown parameter $0 \leqq \theta \leqq 1/2$. Then for every value of $n(n=1, 2, \cdots)$ the allocation $A(n; 1; n)$ is optimal in $\mathscr{C}_n$.*

PROOF. The function $\psi(X)=1-X$ satisfies the hypotheses of Theorem 3.4. The desired result follows immediately.

6. Acknowledgments

This study was suggested initially by some of the work of H. D. Block and J. Marschak in [4]. I am deeply indebted to Jack Nadler for many valuable discussions concerning the material presented here and to D. P. Gaver, J. Marschak, and M. M. Rao for their helpful comments.

CARNEGIE INSTITUTE OF TECHNOLOGY

REFERENCES

[1] D. Blackwell, "Comparison of experiments," *Proc. Second Berkeley Symp. Math. Statist. Prob.*, University of California Press, (1951), 93-102.

[2] D. Blackwell, "Equivalent comparisons of experiments," *Ann. Math. Statist.*, 24 (1953), 265-272.

[3] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, Wiley, New York, 1954.

[4] H. D. Block and J. Marschak, "Random orderings and stochastic theories of responses," *Contributions to Probability and Statistics*, ed. by Olkin et al., Stanford Univ. Press, (1960), 97-132.

[5] J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.

[6] L. LeCam, "Sufficiency and approximate sufficiency," *Ann. Math. Statist.*, 35 (1964), 1419-1455.

[7] E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, New York, 1959.

[8] H. Scheffé, *The Analysis of Variance*, Wiley, New York, 1959.

[9] A. Zinger, "A note on optimum allocation for a one-way layout," *Biometrika*, 49 (1962), 563-564.