On Sufficiency and Invariance

D. BASU*, *University of Chicago and Indian Statistical Institute*

1. SUMMARY

Let $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ be a given statistical model and let $\mathscr{G}$ be the class of all one-to-one, bimeasurable maps g of $(\mathscr{X}, \mathscr{A})$ onto itself such that g is measure-preserving for each $P\varepsilon\mathscr{P}$, i.e. $Pg^{-1} = P$ for all P. Let us suppose that there exists a least (minimal) sufficient sub-field $\mathscr{L}$. Then, for each $L\varepsilon\mathscr{L}$, it is true that $g^{-1}L$ is $\mathscr{P}$-equivalent to L for each $g\varepsilon\mathscr{G}$, i.e., the least sufficient sub-field is almost $\mathscr{G}$-invariant. It is demonstrated that, in many familiar statistical models, the least sufficient sub-field and the sub-field of all almost $\mathscr{G}$-invariant sets are indeed $\mathscr{P}$-equivalent. The problem of data reduction in the presence of nuisance parameters has been discussed very briefly. It is shown that in many situations the principle of invariance is strong enough to lead us to the standard reductions. For instance, given n independent observations on a normal variable with unknown mean

* This research was supported by Research Grant No. NSF GP-3707 from the Devision of Mathematical, Physical and Engineering Sciences of the National Science Foundation, and by the Statistics Branch, Office of Naval Research.

A. DasGupta (ed.), *Selected Works of Debabrata Basu*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4419-5825-9_23, © Springer Science+Business Media, LLC 2011

(the nuisance parameter) and unknown variance, it is shown how the principle of invariance alone can reduce the data to the sample variance.

2. DEFINITIONS AND PRELIMINARIES

(a) The basic probability *model* is denoted by $(\mathscr{X}, \mathscr{A}, \mathscr{P})$, where $\mathscr{X} = \{x\}$ is the sample space, $\mathscr{A} = \{A\}$ the σ-field of events and $\mathscr{P} = \{P\}$ the family of probability measures.

(b) By *set* we mean a typical member of $\mathscr{A}$. By *function* (usually denoted by f) we mean a measurable mapping of $(\mathscr{A}, \mathscr{A})$ into the real line.

(c) A set A is *$\mathscr{P}$-null* if $P(A) = 0$ for all $P \epsilon \mathscr{P}$. Two sets A_1 and A_2 are *$\mathscr{P}$-equivalent* if their symmetric difference is $\mathscr{P}$-null. Two functions f_1 and f_2 are $\mathscr{P}$-equivalent if the set of points where they differ is $\mathscr{P}$-null. The relation symbol $\sim$ stands for $\mathscr{P}$-equivalence.

(d) By *sub-field* we mean a sub-σ-field of $\mathscr{A}$. A *statistic* is a measurable mapping of $(\mathscr{X}, \mathscr{A})$ into any measurable space. We identify a statistic with the sub-field it induces (see pp. 36–39 of [6]).

(e) By the *$\mathscr{P}$-completion* $\bar{\mathscr{A}}_0$ of a sub-field $\mathscr{A}_0$ we mean the least sub-field that contains $\mathscr{A}_0$ and all $\mathscr{P}$-null sets. Observe that $\bar{\mathscr{A}}_0$ may also be characterized as the class of all sets that are $\mathscr{P}$-equivalent to some member of $\mathscr{A}_0$. Two sub-fields are $\mathscr{P}$-equivalent if they have identical $\mathscr{P}$-completions.

(f) By *transformation* (usually denoted by g) we mean a one-to-one, bimeasurable mapping g of $(\mathscr{X}, \mathscr{A})$ onto itself such that the family

$$\mathscr{P} g^{-1} = \{Pg^{-1} \mid P \epsilon \mathscr{P}\}$$

of induced probability measures is the same as the family $\mathscr{P}$. A transformation g is called *model-preserving* if

$$Pg^{-1} \equiv P, \text{ for all } P \epsilon \mathscr{P}.$$

Observe that, if g is any transformation, then so also is g^n for each integral (positive or negative) n and that the identity map is always model-preserving. Also observe that any transformation carries $\mathscr{P}$-null sets into $\mathscr{P}$-null sets.

(g) Given a transformation g, the sub-field $\mathscr{A}(g)$ of *g-invariant sets* is defined as

$$\mathscr{A}(g) = \{A \mid g^{-1}A = A\}.$$

The $\mathscr{P}$-completion $\overline{\mathscr{A}}(g)$ of $\mathscr{A}(g)$ is then the class of all *essentially g-invariant* sets, i.e., sets that are $\mathscr{P}$-equivalent to some g-invariant set.

(h) The set A is *almost g-invariant* if $g^{-1}A \sim A$. It is easy to demonstrate that every almost g-invariant set is also essentially g-invariant and vice versa (see Lemma 1 for a sharper result). Thus, $\overline{\mathscr{A}}(g)$ is also the class of all almost g-invariant sets.

(i) Given a class $\mathscr{G}$ of transformations g, the three sub-fields of α) *$\mathscr{G}$-invariant*, β) *essentially $\mathscr{G}$-invariant* and γ) *almost $\mathscr{G}$-invariant* sets are defined as follows :

α) $\mathscr{A}(\mathscr{G}) = \bigcap \mathscr{A}(g)$, ($\mathscr{G}$-invariant)

β) $\overline{\mathscr{A}}(\mathscr{G}) = \mathscr{P}$-completion of $\mathscr{A}(\mathscr{G})$, (essentially $\mathscr{G}$-invariant)
and

γ) $\widetilde{\mathscr{A}}(\mathscr{G}) = \bigcap \overline{\mathscr{A}}(g)$, (almost $\mathscr{G}$-invariant).

Observe that $\mathscr{A}(\mathscr{G}) \subset \overline{\mathscr{A}}(\mathscr{G}) \subset \widetilde{\mathscr{A}}(\mathscr{G})$. With some assumptions on $\mathscr{G}$, one can prove (see Theorem 4 on p. 225 in [6]) the equality of $\overline{\mathscr{A}}(\mathscr{G})$ and $\widetilde{\mathscr{A}}(\mathscr{G})$. That $\overline{\mathscr{A}}(\mathscr{G})$ can be a very small sub-field compared to $\widetilde{\mathscr{A}}(\mathscr{G})$ is shown in example 1.

(j) A function f is α) *$\mathscr{G}$-invariant*, β) *essentially $\mathscr{G}$-invariant*, or γ) *almost $\mathscr{G}$-invariant*, according as

α) $f = f(g)$, for all $g \in \mathscr{G}$,

β) $f \sim$ some $\mathscr{G}$-invariant function,
or

γ) $f \sim f(g)$, for all $g \in \mathscr{G}$.

Observe that f satisfies the definitions α), β) or γ) above if and only if f is measurable with respect to the corresponding sub-field defined in (i).

(k) When the class $\mathscr{G}$ of transformations g happens to be a group (with respect to the operation of composition of transformations), the sub-field $\mathscr{A}(\mathscr{G})$ of $\mathscr{G}$-invariant sets is easily recognized as follows. For each $x \in \mathscr{X}$, define the *orbit* O_x as

$$O_x = \{x' \mid x' = gx \text{ for some } g \in \mathscr{G}\}.$$

The orbits define a partition of $\mathscr{X}$, and the sub-field $\mathscr{A}(\mathscr{G})$ is the class of all (measurable) sets that are the unions of orbits. The sub-field of essentially $\mathscr{G}$-invariant sets is then the $\mathscr{P}$-completion of $\mathscr{A}(\mathscr{G})$. Our main concern, in this paper, is the sub-field $\tilde{\mathscr{A}}(\mathscr{G})$ and this is not so easily understood in terms of the orbits—unless $\mathscr{G}$ has a structure simple enough to ensure the equality of $\bar{\mathscr{A}}(\mathscr{G})$ and $\tilde{\mathscr{A}}(\mathscr{G})$ (see lemma 1 and example 1).

3. A MATHEMATICAL INTRODUCTION

Let $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ be a given probability model and let g be a fixed model-preserving transformation [definition 2(f)]. We remark that the identity map is trivially model-preserving. In many instances of statistical interest there exist fairly wide classes of such transformations. However, it is not difficult to construct examples where no non-trivial transformation is model-preserving. See for instance example 5.

For each bounded function f [definition 2(b)], define the associated sequence $\{f_n\}$ of (uniformly bounded) functions as follows :

$$f_n(x) = [f(x)+f(gx)+\ldots+f(g^n x)]/(n+1), \qquad n = 1, 2, 3, \ldots$$

Since g is measure-preserving for each $P \in \mathscr{P}$, the pointwise ergodic theorem tells us that the set N_f where $\{f_n(x)\}$ fails to converge is $\mathscr{P}$-null [definition 2(c)].

If we define f^* as

$$f^*(x) = \begin{cases} \lim f_n(x) & \text{when } x \notin N_f, \\ 0 & \text{otherwise} \end{cases}$$

then, it is easily seen that

i) f^* is $\mathscr{A}(g)$-measurable [definition 2(g)], i.e., it is g-invariant [definition 2(i)], and that

ii) for all $B \epsilon \mathscr{A}(g)$ and $P \epsilon \mathscr{P}$

$$\int_B f dP = \int_B f^* dP.$$

This is another way of saying that f^* is the conditional expectation of f, given the sub-field $\mathscr{A}(g)$ of g-invariant sets. Since the definition of f^* does not involve P, we have the following theorem (lemma 2 in [4]).

Theorem 1. *If the transformation g preserves the model $(\mathscr{X}, \mathscr{A}, \mathscr{P})$, then the sub-field $\mathscr{A}(g)$ is sufficient.*

[*Remark* : Note that in the proof of theorem 1 we have not used the assumptions that g is a one-to-one map and that it is bimeasurable. The proof remains valid for any measurable mapping of $(\mathscr{X}, \mathscr{A})$ into itself that is measure-preserving for each $P \epsilon \mathscr{P}$. A similar remark will hold true for a number of other results to be stated later. However, in a study of the statistical theory of invariance (see [6]) it seems appropriate to restrict our attention to one-to-one, bimeasurable maps of $(\mathscr{X}, \mathscr{A})$ onto itself that preserve the model either wholly or partially.]

Now, given a class $\mathscr{G}$ of model-preserving transformations g, what can we say about the sufficiency of the sub-field

$$\mathscr{A}(\mathscr{G}) = \bigcap \mathscr{A}(g)$$

of $\mathscr{G}$-invariant [definition 2(i)] sets? The intersection of two sufficient sub-fields is not necessarily sufficient. However, it is known (see Theorem 4 and Corollary 2 of [3]) that the intersection of the $\mathscr{P}$-completions [definition 2(e)] of a countable number of sufficient sub-fields is sufficient. Using this result, we have the following theorem (theorem 2 in [4]).

Theorem 2. *If the class $\mathscr{G}$ of model-preserving transformations g is countable, then the sub-field*

$$\tilde{\mathscr{A}}(\mathscr{G}) = \bigcap \tilde{\mathscr{A}}(g)$$

of almost $\mathcal{G}$-invariant sets [definition 2(i)(γ)] *is sufficient.*

Given a countable class $\mathcal{G}$ of transformations, consider the larger class $\mathcal{G}^*$ of transformations of the type $\alpha_1\alpha_2 \ldots \alpha_n$, where each α_i is such that either α_i or α_i^{-1} belongs to $\mathcal{G}$, and n is an arbitrary positive integer. The following properties of $\mathcal{G}^*$ are easy to check :

a) $\mathcal{G}^*$ is a group (the group operation being composition of transformations),

b) $\mathcal{G}^*$ is countable,

c) $\mathcal{A}(\mathcal{G}^*) = \mathcal{A}(\mathcal{G})$ and $\tilde{\mathcal{A}}(\mathcal{G}^*) = \tilde{\mathcal{A}}(\mathcal{G})$.

Now, let A be an arbitrary almost $\mathcal{G}^*$-invariant set, i.e., $g^{-1}A \sim A$ (equivalently, $gA \sim A$) for all $g \in \mathcal{G}^*$. Consider the set

$$B = \bigcap gA$$

where the intersection is taken over all $g \in \mathcal{G}^*$. Since $\mathcal{G}^*$ is a group, the set B must be $\mathcal{G}^*$-invariant. Again, since $\mathcal{G}^*$ is countable, and each gA is $\mathcal{P}$-equivalent to A, we have $B \sim A$ We have thus established the following lemma.

Lemma 1. *For any countable class $\mathcal{G}$ of transformations [definition 2(f)], every almost $\mathcal{G}$-invariant set is essentially $\mathcal{G}$-invariant, i.e., $\tilde{\mathcal{A}}(\mathcal{G}) = \mathcal{A}(\mathcal{G})$.*

Theorem 2, together with lemma 1 and the observation that a sub-field that is $\mathcal{P}$-equivalent to a sufficient sub-field is itself sufficient, leads to the following theorem.

Theorem 3. *If $\mathcal{G}$ is a countable class of model-preserving transformations, then the sub-field $\mathcal{A}(\mathcal{G})$ of $\mathcal{G}$-invariant sets is sufficient.*

[*Remark* : Note that in Theorem 3 we have used the one-to-oneness and bimeasurability of our transformations.]

Before proceeding further, let us consider an example which shows that Theorem 3 is no longer true if we drop the condition that $\mathcal{G}$ is countable.

Example 1. Let $\mathscr{P}$ be a family of continuous distributions on the real line $\mathscr{X}$ where $\mathscr{A}$ is the σ-field of Borel-sets. Let $\mathscr{G} = \{g\}$ be the class of all one-to-one maps of $\mathscr{X}$ onto $\mathscr{X}$ such that $\{x | gx \neq x\}$ is finite. Clearly, $\mathscr{G}$ is a group and every member of $\mathscr{G}$ is model-preserving. It is easy to check that the sub-field $\mathscr{A}(\mathscr{G})$ of $\mathscr{G}$-invariant sets consists of ϕ and $\mathscr{X}$ only, and hence is not sufficient. The sub-field $\widetilde{\mathscr{A}}(\mathscr{G})$ of almost $\mathscr{G}$-invariant sets is the same as $\mathscr{A}$ and is sufficient.

Consider another example.

Example 2. Let $\mathscr{X}$ be the n-dimensional Euclidean space and $\mathscr{P}$ the class of all probability measures (on the σ-field $\mathscr{A}$ of all Borel-sets) that are symmetric (in the co-ordinates). If $\mathscr{G} = \{g\}$ is the group of all permutations (of the co-ordinates) then $\mathscr{A}(\mathscr{G})$ is the sub-field of all sets that are symmetric (in the co-ordinates) and is sufficient. Since the empty set is the only $\mathscr{P}$-null set, the two sub-fields $\mathscr{A}(\mathscr{G})$ and $\widetilde{\mathscr{A}}(\mathscr{G})$ are the same here.

Given a probability model $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ we ask ourselves the following questions.

1. How wide is the group $\mathscr{G}$ of all model-preserving transformations ?
2. Is the sub-field $\widetilde{\mathscr{A}}(\mathscr{G})$ of almost $\mathscr{G}$-invariant sets sufficient?
3. If a least (minimal) sufficient sub-field $\mathscr{L}$ exists, then is it true that $\mathscr{L} \sim \widetilde{\mathscr{A}}(\mathscr{G})$?
4. What is $\widetilde{\mathscr{A}}(\mathscr{G})$, when $\mathscr{G}$ is the class of all transformations that partially preserve the model in a given manner ?

4. STATISTICAL MOTIVATION

Suppose we have two different systems of measurement co-ordinates for the outcome of a statistical experiment, so that, if a typical outcome is recorded as x under the first system, the same outcome is recorded as gx under the second system. Let us suppose further that the two statistical variables x and gx

have the same domain (sample space) $\mathscr{X}$, the same family $\mathscr{A}$ of events, and the same class of probability measure $\mathscr{P}$. Furthermore, if $P\epsilon\mathscr{P}$ holds for x, then the same probability measure P also holds for gx. The second system of measurement co-ordinates may then be represented mathematically as a model-preserving transformation g of the statistical model $(\mathscr{X}, \mathscr{A}, \mathscr{P})$. The principle of invariance then stipulates that the decision rule should be invariant with respect to the transformation g—and this irrespective of the actual decision problem. If the choice of a new system of measurement co-ordinates leaves the problem (whatever it is) entirely unaffected, then so also should be every reasonable inference procedure.

Thus, if g is model-preserving, the principle of invariance leads to the invariance reduction of the model $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ to the simpler model $(\mathscr{X}, \mathscr{A}(g), \mathscr{P})$, where $\mathscr{A}(g)$ is the sub-field of g-invariant sets. To put it differently, the principle of invariance requires every decision function to be g-invariant or $\mathscr{A}(g)$-measureable. Consider now the class $\mathscr{G}$ of all model-preserving transformations g. Must we, following the principle of invariance, insist that every decision rule be $\mathscr{G}$-invariant (i.e. g-invariant for every $g\epsilon\mathscr{G}$)? We have already noted in example 1 that, when $\mathscr{P}$ consists of a family of non-atomic measures, the class $\mathscr{G}$ is large enough to reduce the sub-field $\mathscr{A}(\mathscr{G})$ of $\mathscr{G}$-invariant sets to the trivial one (consisting of only the empty set and the whole space). Obviously, we cannot (must not) reduce $\mathscr{A}$ all the way down to $\mathscr{A}(\mathscr{G})$ or even to $\overline{\mathscr{A}}(\mathscr{G})$—the sub-field of essentially $\mathscr{G}$-invariant sets. A logical compromise (with the principle) would be to reduce $\mathscr{A}$ to the sub-field $\widetilde{\mathscr{A}}(\mathscr{G})$ of all almost $\mathscr{G}$-invariant sets—that is to insist upon the decision function to be almost $\mathscr{G}$-invariant.

The principle of sufficiency is another reduction principle of the omnibus type. If $\mathscr{S}$ be a sufficient sub-field, then this principle tells us not to use a decision function that is not $\mathscr{S}$-measurable. Suppose there exists a least (minimal) sufficient subfield $\mathscr{L}$. Following the principle of sufficiency, we reduce the model $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ to the model $(\mathscr{X}, \mathscr{L}, \mathscr{P})$.

Which of the two reductions (invariance or sufficiency) is more extensive ? In other words, what is the relation between $\widetilde{\mathscr{A}}(\mathscr{G})$ and $\mathscr{L}$?

Theorem 1 tells us that $\mathscr{A}(g)$ is sufficient (for each g in $\mathscr{G}$). Since $\mathscr{L}$ is the least sufficient sub-field we have (from the definition of $\mathscr{L}$)

$$\bar{\mathscr{L}} \subset \bar{\mathscr{A}}(g) \text{ for all } g \in \mathscr{G}.$$

Theorem 4 follows at once.

Theorem 4. $$\bar{\mathscr{L}} \subset \bigcap \bar{\mathscr{A}}(g) = \widetilde{\mathscr{A}}(\mathscr{G}).$$

[*Remark*: Theorem 4 does not establish the sufficiency of $\widetilde{\mathscr{A}}(\mathscr{G})$. [See example 1 in [3].]

We thus observe that the invariance reduction (in terms of the group $\mathscr{G}$ of all model-preserving transformations) of a model can never be more extensive than its maximal sufficiency reduction (if one such reduction is available). The principal question raised in this paper is, "When is $\widetilde{\mathscr{A}}(\mathscr{G})$ equal to $\bar{\mathscr{L}}$?" We shall show later that in many familiar situations the sub-fields $\widetilde{\mathscr{A}}(\mathscr{G})$ and $\mathscr{L}$ are essentially equal. This raises the question about the nature of $\widetilde{\mathscr{A}}(\mathscr{G})$, where $\underline{\mathscr{G}}$ is the class of all transformations that preserve the model partially in some well-defined manner. This question is discussed in a later section. In the next two sections we give two alternative approaches to Theorem 4.

5. WHEN A BOUNDEDLY COMPLETE SUFFICIENT SUB-FIELD EXISTS

Let us suppose that the sub-field $\mathscr{L}$ is sufficient and boundedly complete. We need the following lemma.

Lemma 2. *If z is a bounded $\mathscr{A}$-measurable function such that*

$$E(z \mid P) \equiv 0 \text{ for all } P \in \mathscr{P},$$

then, for all bounded $\mathscr{L}$-measurable functions f, it is true that

$$E(zf \mid P) \equiv 0 \quad \text{for all } P \in \mathscr{P}.$$

The proof of this well-known result is omitted. Now, let S be an arbitrary member of $\mathscr{L}$, and let g be an arbitrary model-preserving transformation. Let $S_0 = g^{-1}S$. Since g is model-preserving we have

$$P(S) \equiv Pg^{-1}S \equiv P(S_0) \quad \text{for all} \quad P\epsilon\mathscr{P}.$$

Writing I_S for the indicator of S, and noting that $I_S - I_{S_0}$ and I_S satisfy the conditions for z and f in Lemma 2, we at once have

$$E[(I_S - I_{S_0})I_S \,|\, P] \equiv 0, \quad \text{for all} \quad P\epsilon\mathscr{P},$$

or
$$P(S) \equiv P(SS_0), \quad \text{for all} \quad P\epsilon\mathscr{P}.$$

Hence,
$$\begin{aligned} P(S\Delta S_0) &\equiv P(S) + P(S_0) - 2P(SS_0) \\ &\equiv 2[P(S) - P(SS_0)] \\ &\equiv 0, \quad \text{for all } P\epsilon\mathscr{P}. \end{aligned}$$

That is, for all $S\epsilon\mathscr{L}$, the two sets S and $g^{-1}S$ are $\mathscr{P}$-equivalent. In other words,

$$\mathscr{L} \subset \bar{\mathscr{A}}(g),$$

and since g is an arbitrary element of $\mathscr{G}$ (the class of all model-preserving transformations) we have the following theorem.

Theorem 4 (a). *If $\mathscr{L}$ is a boundedly complete sufficient sub-field then* $\mathscr{L} \subset \tilde{\mathscr{A}}(\mathscr{G})$.

[*Remark*: Since bounded completeness of $\mathscr{L}$ implies that $\mathscr{L}$ is the least sufficient sub-field, Theorem 4(a) is nothing but a special case of Theorem 4. The proof of Theorem 4(a) is simple and amenable to a generalization to be discussed later.]

6. THE DOMINATED CASE

We make a slight digression to state a useful lemma. Let T be a measurable map of $(\mathscr{X}, \mathscr{A})$ into $(\mathscr{Z}, \mathscr{B})$. Let P and Q be two probability measures on $\mathscr{A}$ and let PT^{-1} and QT^{-1} be the corresponding measures on $\mathscr{B}$. Suppose that Q dominates P and let $f = dP/dQ$ be the Radon-Nikodym derivative defined on $\mathscr{X}$. It is then clear that QT^{-1} dominates PT^{-1}. Let $h = (dPT^{-1})/(dQT^{-1})$. The function hT on $\mathscr{X}$ (defined as $hT(x) = h(Tx)$) is $T^{-1}(\mathscr{B})$-measurable and satisfies the following relation.

Lemma 3. $$hT = E(f|T^{-1}(\mathscr{B}), Q),$$

i.e., hT is the conditional expectation of f, given $T^{-1}(\mathscr{B})$ and Q.

The proof of this well-known lemma consists of checking the identity

$$\int_B f dQ = \int_B hT dQ, \quad \text{for all } B \epsilon T^{-1}(\mathscr{B}).$$

Corollary. *If $T^{-1}(\mathscr{B})$ is Q-equivalent to $\mathscr{A}$, then f and hT are Q-equivalent.*

Now, returning to our problem, let $\mathscr{G}$ be the class of all transformations g that preserves the model $(\mathscr{X}, \mathscr{A}, \mathscr{P})$. Let us suppose that $\mathscr{P}$ is dominated by some σ-finite measure. It follows that there exists a countable collection $P_1, P_2, \ldots,$ of elements in $\mathscr{P}$ such that the convex combination

$$Q = \Sigma c_i P_i, \quad c_i > 0, \quad \Sigma c_i = 1,$$

dominates the family $\mathscr{P}$. Let $f_P = dP/dQ$ be a fixed version of the Radon-Nikodym derivative of P with respect to Q. The factorization theorem for sufficient statistics asserts that a subfield $\mathscr{A}_0$ is sufficient if and only if f_P is $\overline{\mathscr{A}_0}$-measurable for every $P \epsilon \mathscr{P}$, where $\overline{\mathscr{A}_0}$ is the $\mathscr{P}$-completion of $\mathscr{A}_0$. We now prove that f_P is $\widetilde{\mathscr{A}}(\mathscr{G})$-measurable for every P. (The $\mathscr{P}$-completion of $\widetilde{\mathscr{A}}(\mathscr{G})$ is itself.)

Since $Pg^{-1} = P$ for all $P \epsilon \mathscr{P}$, it follows that $Qg^{-1} = Q$. From Lemma 3 we have

$$f_P g = E(f_P | g^{-1}(\mathscr{A}), Q).$$

The assumption that g is one-to-one and bimeasurable implies that $g^{-1}(\mathscr{A}) = \mathscr{A}$. Hence $f_P g = f_P$ a.e.w. [Q]. Since Q dominates $\mathscr{P}$, it follows that f_P is almost g-invariant, i.e., is $\widetilde{\mathscr{A}}(g)$ measurable. Since, in the above argument, g is an arbitrary element of $\mathcal{G}$, we have the following theorem. (See problem 19 on p. 253 in [6].)

Theorem 4(b). *If $\mathcal{P}$ is dominated, then $\tilde{\mathscr{A}}(\mathcal{G})$ is sufficient.*

[*Remark* : Since, in the dominated set-up, the least sufficient sub-field $\mathscr{L}$ always exists and since, in this set-up, any sub-field containing $\mathscr{L}$ is necessarily sufficient, it is clear that Theorem 4(b) is nothing but an immediate corollary to Theorem 4. Also note that in the present proof (of Theorem 4(b)) we had to draw upon our supposition that g is one-to-one and bimeasurable. This section has been written only with the object of drawing attention to some aspects of the problem.]

7. EXAMPLES

Example 3 : Let y be a real random variable having a uniform distribution over the unit interval. For each c in $[0, 1)$ define the transformation g_c as

$$g_c y = y + c \pmod 1$$

In this example, $\mathcal{P}$ consists of a single measure and each g_c is model preserving (measure-preserving). If $\mathcal{G}_0 = \{g_c \mid 0 < c < 1\}$, then the only $\mathcal{G}_0$-invariant sets are the empty set and the whole of the unit interval. Here, $\mathscr{A}(\mathcal{G}_0)$ is sufficient.

[*Remark* : In this case, there are a very large number of measure-preserving transformations that are not one-to-one maps of [0, 1] onto itself. For example, let $a_n(y)$ be the nth digit in the decimal representation of y and let

$$gy = \sum_{k=1}^{\infty} \frac{a_{n_k}(y)}{10^k}$$

where $\{n_k\}$ is a fixed increasing sequence of natural numbers.]

Again, if x has a fixed continuous distribution on the real line with cumulative distribution function F, then the class $\mathcal{G}_0$ of transformations g_c defined as

$$g_c x = F^{-1}[F(x) + c \pmod 1)], \quad 0 < c < 1$$

are all model-preserving for x. [In case F is not a strictly increasing function of x, we define $F^{-1}(y) = \inf \{x \mid F(x) = y\}$.]

Thus, for any fixed continuous distribution on the real line, there always exists a large class of measure-preserving transformations.

Example 4. Let $\mathscr{X} = [0, \infty)$ and let x have a uniform distribution over the interval $[0, \theta)$, where θ is an unknown positive integer. Here $\mathscr{P}$ consists of a countable infinity of probability measures. For each c in $[0, 1)$, define the transformation g_c as

$$g_c x = [x] + \{x - [x] + c \pmod{1}\},$$

where $[x]$ is the integer-part of x.

Here, each g_c is model-preserving. The minimal (least) sufficient statistic $[x]$ is also the maximal-invariant with respect to the group $\mathscr{G}_0 = \{g_c\}$ of model-preserving transformations defined above. Thus, if $\mathscr{L}$ is the sub-field (least sufficient) generated by $[x]$ and $\mathscr{A}(\mathscr{G}_0)$ is the sub-field of $\mathscr{G}_0$-invariant sets, we have

$$\mathscr{L} = \mathscr{A}(\mathscr{G}_0). \tag{a}$$

Now, if $\mathscr{G}$ is the class of all model-preserving transformations, then (as we have seen in Example 1) the sub-field $\mathscr{A}(\mathscr{G})$ will reduce to the level of triviality. (It will consist of only the empty set and the whole set $\mathscr{X}$.) However, from Theorem 4 we have

$$\bar{\mathscr{L}} \subset \tilde{\mathscr{A}}(\mathscr{G}). \tag{b}$$

Since the group $\mathscr{G}_0$ has a *decent* structure, we can apply Stein's theorem (theorem 4 on p. 225 in [6]) to prove that

$$\mathscr{A}(\mathscr{G}_0) \sim \tilde{\mathscr{A}}(\mathscr{G}_0). \tag{c}$$

Since $\mathscr{G}_0 \subset \mathscr{G}$, we at once have

$$\tilde{\mathscr{A}}(\mathscr{G}) \subset \tilde{\mathscr{A}}(\mathscr{G}_0). \tag{d}$$

Putting the relations (a), (c), and (d) together we have

$$\widetilde{\mathscr{A}}(\mathscr{G}) \subset \overline{\mathscr{L}} \tag{e}$$

Putting (b) and (e) together we finally have

$$\overline{\mathscr{L}} = \widetilde{\mathscr{A}}(\mathscr{G}),$$

i.e., the least-sufficient sub-field (rather, the $\mathscr{P}$-completion of any version of the least sufficient sub-field) and the sub-field of almost $\mathscr{G}$-invariant sets are identical.

The chain of arguments, detailed as above, is of a general nature and will be used repeatedly in the sequel.

Example 5. Let x be a normal variable with unit variance and mean equal to either μ_1 or μ_2. Does there exist a non-trivial transformation of $\mathscr{X}$ (the real line) into itself that preserves each of the two measures ? That the answer is "no" is seen as follows. Let $\mathscr{G}$ be the class of all model-preserving transformations. In view of theorem 4, $\widetilde{\mathscr{A}}(\mathscr{G})$ contains the least sufficient sub-field $\mathscr{L}$. But, in this example, the likelihood ratio (which is the least sufficient statistic) is

$$\exp\left[(\mu_1-\mu_2)x - \frac{1}{2}(\mu_1^2-\mu_2^2)\right],$$

and this is a one-to-one measurable function of x. Thus, every set is almost $\mathscr{G}$-invariant. And this implies that every g in $\mathscr{G}$ must be equivalent to the identity map.

Example 6. Let $x_1, x_2, \ldots, x_n$ be n independent observations on a normal variable with known mean and unknown standard deviation σ. Without loss of generality we may assume the mean to be zero. Let $\mathscr{G}_0$ be the group of all linear orthogonal transformations of the n-dimentional Euclidean space onto itself. Clearly, every member of $\mathscr{G}_0$ is model-preserving. That the class $\mathscr{G}$ of all model-preserving transformations is much wider than $\mathscr{G}_0$ is seen as follows. Let

$$y_i = \phi_i(x_i)\,|x_i|,\ i = 1, 2, \ldots, n,$$

where $\phi_1, \phi_2, \ldots, \phi_n$ are arbitrary skew-symmetric (i.e., $\phi(x) = -\phi(-x)$ for all x) functions on the real line that take only the two values -1 and $+1$. It is easily checked that, whatever the value of σ, the two vectors $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are identically distributed, i.e., the above transformation (though non-linear) is model-preserving. However, the sub-group $\mathscr{G}_0$ is large enough to lead us to Σx_i^2—which is the least sufficient statistic—as the maximal invariant. In view of the decent structure of the sub-group $\mathscr{G}_0$, the arguments given for example 4 are again available to prove the equality of $\bar{\mathscr{L}}$ and $\tilde{\mathscr{A}}(\mathscr{G})$.

8. TRANSFORMATIONS OF A SET OF NORMAL VARIABLES

This section is devoted to a study of the special case (model) of n independent normal variables $x_1, x_2, \ldots, x_n$ with equal unknown means μ and equal unknown standard deviations σ. Hence Σx_i and Σx_i^2 jointly constitute the least sufficient statistic. If $\bar{\mathscr{L}}$ is the $\mathscr{P}$-completion of the sub-field $\mathscr{L}$ induced by $(\Sigma x_i, \Sigma x_i^2)$, then we know from Theorem 4, that

$$\bar{\mathscr{L}} \subset \tilde{\mathscr{A}}(\mathscr{G})$$

where $\mathscr{G}$ is the class of all the model-preserving transformations of $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ to $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$. For any model-preserving transformation from $\boldsymbol{x}$ to $\boldsymbol{y}$, we, therefore, have

$$\Sigma x_i \sim \Sigma y_i$$

and

$$\Sigma x_i^2 \sim \Sigma y_i^2$$

i.e., the statistic $(\Sigma x_i, \Sigma x_i^2)$ is almost $\mathscr{G}$-invariant. If we can demonstrate the existence of a *'decent'* sub-group $\mathscr{G}_0$ of $\mathscr{G}$ for which the statistic $(\Sigma x_i, \Sigma x_i^2)$ is the maximal invariant, then (following the method of proof indicated in examples 4 and 6) we can show that $\bar{\mathscr{L}}$ is indeed equal to $\tilde{\mathscr{A}}(\mathscr{G})$.

Let $\mathscr{G}_0$ be the sub-group of all linear model-preserving transformations. Do there exist non-trivial linear model-preserving transformations (i.e., linear transformations that are not a permutation of the co-ordinates)? That the answer is "yes" is seen as follows. Let $\mathscr{M} = \{M\}$ be the family of all orthogonal $n \times n$ matrices with the initial row as

$$\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right).$$

Now, if $x = (x_1, x_2, \ldots, x_n)$ are independent $N(\mu, \sigma)$'s and we define

$$y' = Mx'$$

(where x' is the corresponding column vector), then the y_i's are independent normal variables with equal standard deviation σ and with means as follows :

$$E(y_1) = \sqrt{n}\,\mu, \quad E(y_i) = 0, \qquad i = 2, 3, \ldots, n.$$

Thus, for an arbitrary pair of members M_1 and M_2 in $\mathscr{M}$, we note that M_1x' and M_2x' are identically distributed (whatever the values of μ and σ). Therefore, the two vectors x' and $M_2^{-1}M_1x$ are identically distributed for each μ and σ. In other words the linear transformation defined by the matrix

$$M_2^{-1}\,M_1 \qquad (=M_2'M_1,$$

since M_2 is orthogonal) is model-preserving for each pair M_1, M_2 of members from $\mathscr{M}$.

[*Remark* : Later on we shall have some use for this way of generating members of $\mathscr{G}_0$.]

For example, the 4×4 matrix

$$\begin{pmatrix} 1/2 & 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & 1/2 \\ -1/2 & 1/2 & 1/2 & 1/2 \end{pmatrix}$$

defines a member of $\mathscr{G}_0$ for $n = 4$.

Let H be a typical $n \times n$ matrix that defines a model-preserving linear transformation. We are going to make a brief digression about the nature of H. From the requirement that each co-ordinate of

$$y' = Hx'$$

has mean μ and standard deviation σ, we at once have that the elements in each row of H must add up to unity and that their squares also must add up to unity. From the mutual independence of the y_i's, it follows that the row vectors of H must be mutually orthogonal. Thus, H must be an orthogonal matrix with unit row sums. Now, for each model-preserving H, its inverse

$$H^{-1} \ (= H', \text{ since } H \text{ is orthogonal})$$

is necessarily also model-preserving and, thus, the columns of H must also add up to unity, etc.

It is easily checked that the sub-group $\mathscr{G}_0 = \{H\}$ of linear model-preserving transformations of the n-space onto itself may also be characterized as the class of all linear transformations that preserve both the sum and the sum of squares of the co-ordinates. The author came to learn that G. W. Haggastrom of the University of Chicago had come upon these matrices from this point of view and had a brief discussion on them in an unpublished work of his. We call such matrices by the name Haggstrom-matrix.]

Going back to our problem, we have to demonstrate that the statistic $T = (\Sigma x_i, \Sigma x_i^2)$ is a maximal invariant with respect to the sub-group $\mathscr{G}_0 = \{H\}$ of model-preserving linear transformations. For this we have only to prove that if $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$ are any two points in n-space such that $T(\mathrm{a}) = T(\mathrm{b})$ then there exists a Haggstrom-matrix H such that

$$\boldsymbol{b}' = H\boldsymbol{a}'.$$

Let M_1 and M_2 be two arbitrary members of $\mathscr{M}$—the class of all orthogonal $n \times n$ matrices with the leading row as $(1/\sqrt{n}, \ldots,$

$1/\sqrt{n}$). If $\boldsymbol{\alpha} = M_1\boldsymbol{a}'$ and $\boldsymbol{\beta} = M_2\boldsymbol{b}'$, then we have, from $T(\boldsymbol{a}) = T(\boldsymbol{b})$, that $\alpha_1 = \beta_1$ that and that

$$\sum_2^n \alpha_i^2 = \sum_2^n \beta_i^2.$$

It follows that there exists an $(n-1)\times(n-1)$ orthogonal matrix that will transform $(\alpha_2, \alpha_3, \ldots, \alpha_n)$ to $(\beta_2, \beta_2, \ldots, \beta_n)$. And this, in turn, implies that there exists an $n\times n$ orthogonal matrix K, with the first row as $(1, 0, 0, \ldots, 0)$, such that

$$\boldsymbol{\beta}' = K\boldsymbol{\alpha}'.$$

Thus, the transformation $M_2^{-1}KM_1$ takes $\boldsymbol{a}$ into $\boldsymbol{b}$. Now, note that, since K is an orthogonal matrix with the initial row as $(1, 0, 0, \ldots, 0)$, the matrix KM_1 is orthogonal with its initial row as $(1/\sqrt{n}, 1/\sqrt{n}, \ldots, 1/\sqrt{n})$, i.e. $KM_1 \epsilon \mathscr{M}$. In other words, there exists a matrix of the form $M_2^{-1}M$ (with M_2 and M belonging to $\mathscr{M}$) which transforms $\boldsymbol{a}$ into $\boldsymbol{b}$. We have already noted that all such matrices belong to $\mathscr{G}_0$, and this completes our proof that $(\Sigma x_i, \Sigma x_i^2)$ is a maximal invariant with respect to $\mathscr{G}_0$.

In example 6 of the previous section, we considered the particular case of the foregoing problem where μ is known. Consider now the other particular case where σ is known and μ is the only unknown parameter. In this case Σx_i is the least sufficient statistic. Now, it is no longer possible to produce a sub-group $\mathscr{G}_0$ of linear model-preserving transformations such that Σx_i is the maximal invariant with respect to $\mathscr{G}_0$. This is because every linear model-preserving transformation must necessarily be orthogonal and would thus preserve Σx_i^2 also. We proceed as follows.

Suppose (without loss of generality) that $\sigma = 1$. Let $\boldsymbol{y}' = M\boldsymbol{x}'$ where M is a fixed member of the class $\mathscr{M}$ of orthogonal $n\times n$ matrices with the initial row as $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$. Observe that the y_i's are mutually independent and that $y_2, y_3, \ldots, y_n$ are standard normal variables. Let F stand for the cumulative distribution function of a standard normal variable. Let c_2,

$c_2, \ldots, c_n$ be arbitrary constants in $[0, 1)$. If we define $z_1 = y_1$ and

$$z_i = F^{-1}[F(y_i)+c_i \pmod{1}], \qquad i = 2, 3, \ldots, n,$$

then $z = (z_1, z_2, \ldots, z_n)$ has the same distribution as that of y and so it follows that x' and $M^{-1}z'$ are identically distributed. Observe that we have described above a model-preserving transformation corresponding to each $(n-1)$-vector $(c_2, c_3, \ldots, c_n)$ with the c_i's in $[0, 1)$. It is easly checked that Σx_i is a maximal invariant with respect to the above class of model-preserving transformations.

9. PARAMETER-PRESERVING TRANSFORMATIONS

Let $\gamma = \gamma(P)$ be the parameter of interest. That is, the experimenter is interested only in the characteristic $\gamma(P)$ of the measure P (that actually holds) and considers all other details about P to be irrelevent (nuisance parameters). We define a γ-preserving transformation as follows.

Definition. The transformation [see Definition 2(f)] g is γ-preserving if $\gamma(Pg^{-1}) = \gamma(P)$, for all $P \in \mathscr{P}$.

If g is γ-preserving then so also is g^{-1}. The composition of any two γ-preserving transformations is also γ-preserving. Let $\mathscr{G}_\gamma$ be the group of all γ-preserving transformations.

In the particular case where $\gamma(P) = P$, the γ-preserving transformations are what we have so far been calling model-preserving. If $\mathscr{G}$ is the class of all model-preserving transformations, then note that $\mathscr{G} \subset \mathscr{G}_\gamma$ for every γ.

If γ is the parameter of interest, then the principle of invariance leads to the reduction of $\mathscr{A}$ to the sub-field $\widetilde{\mathscr{A}}(\mathscr{G}_\gamma)$. This section is devoted to a study of the sub-field $\widetilde{\mathscr{A}}(\mathscr{G}_\gamma)$.

Since $\mathscr{G} \subset \mathscr{G}_\gamma$ we have

$$\widetilde{\mathscr{A}}(\mathscr{G}_\gamma) \subset \widetilde{\mathscr{A}}(\mathscr{G}).$$

Let us suppose that the least sufficient sub-field $\mathscr{L}$ exists. We have discussed a number of examples where $\widetilde{\mathscr{A}}(\mathscr{G})$ and $\widetilde{\mathscr{L}}$ are identical. [The author believes that the above identification is true under very general conditions.] In all such situations we therefore have

$$\tilde{\mathscr{A}}(\mathscr{G}_\gamma) \subset \tilde{\mathscr{L}}.$$

That is, when the interest of the experimenter is concentrated on some particular characteristic $\gamma(P)$ or P, the principle of invariance will usually reduce the data more extensively than the principle of sufficiency.

Theorems 4, 4(a) and 4(b) tell us that

$$\tilde{\mathscr{L}} \subset \tilde{\mathscr{A}}(\mathscr{G}).$$

The following theorem gives us a similar lower bound for $\tilde{\mathscr{A}}(\mathscr{G}_\gamma)$.

A set A is called γ-oriented if, for every pair $P_1, P_2 \epsilon \mathscr{P}$ such that $\gamma(P_1) = \gamma(P_2)$, it is true that $P_1(A) = P_2(A)$. In other words, a γ-oriented set is one whose probability depends on P through $\gamma(P)$. A sub-field is γ-oriented if every member of the sub-field is. The following theorem* is a direct generalization of Theorem 4(a).

Theorem 5. *Let $\mathscr{G}_\gamma$ be the class of all γ-preserving transformations and let $\mathscr{B}$ be a sub-field that is*

i) *γ-oriented and*

ii) *contained in the boundedly complete sufficient sub-field $\mathscr{L}$ (which exists).*

Then,

$$\mathscr{B} \subset \tilde{\mathscr{A}}(\mathscr{G}_\gamma).$$

Proof : Let $g \epsilon \mathscr{G}_\gamma$ and $B \epsilon \mathscr{B}$ and let $B_0 = g^{-1} B$

Then

$$P(B_0) = P(g^{-1}B) \qquad (\because \quad B_0 = g^{-1}B)$$

$$= Pg^{-1}(B).$$

* The author wishes to thank Professor W. J. Hall of the University of North Carolina for certain comments that eventually led to this theorem.

Since g is γ-preserving, we have $\gamma(Pg^{-1}) = \gamma(P)$. And since B is γ-oriented we have

$$Pg^{-1}(B) = P(B).$$

Therefore,

$$P(B_0) = P(B) \quad \text{or} \quad R(I_B - I_{B_0} | P) \equiv 0.$$

The rest of the proof is the same as in Theorem 4(a).

In the next section we repeatedly use the above theorem.

10. SOME TYPICAL INVARIANCE REDUCTIONS

Let $x_1, x_2, \ldots, x_n$ be n independent and identical normal variables with unknown μ and σ. Suppose the parameter of interest is σ. Let $\mathcal{G}_\sigma$ be the class of all σ-preserving transformations. If $\bar{x}$ is the mean and s, the standard deviation of the n observations, then the pair $(\bar{x}, s)$ is a complete sufficient statistic. Also s is σ-oriented. From Theorem 5, we then have that s is almost $\mathcal{G}_\sigma$-invariant. Thus, the principle of invariance cannot reduce the data beyond s. That s is indeed the exact (upto $\mathcal{P}$-equivalence) attainable limit of invariance reduction is shown as follows.

Let $\{H\}$ be the class of all $n \times n$ Haggstrom matrices (see section 8), i.e., each H is an orthogonal matrix with unit row and column sums. Let c be an arbitrary real number and let $\boldsymbol{i}$ stand for the n-vector $(1, 1, \ldots, 1)$. Consider all linear transformations of the type

$$\boldsymbol{y}' = H\boldsymbol{x}' + (c\boldsymbol{i})'.$$

If $\mathcal{G}_\sigma^*$ is this class of transformations, then it is easily verified that $\mathcal{G}_\sigma^*$ is a sub-group of $\mathcal{G}_\sigma$. That $\Sigma(x_i - \bar{x})^2$ is a maximal invariant with respect to the sub-group $\mathcal{G}_\sigma^*$ of σ-preserving transformations is seen as follows.

Let $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$ be two arbitrary n-vectors such that

$$\Sigma(a_i - \bar{a})^2 = \Sigma(b_i - \bar{b})^2.$$

Let $c = \bar{b} - \bar{a}$. Then the two vectors $\boldsymbol{a} + c\boldsymbol{i}$ and $\boldsymbol{b}$ have equal sums and sums of squares (of co-ordinates). Hence there exists a

Haggstrom matrix H that transforms $\boldsymbol{a}+c\boldsymbol{i}$ into $\boldsymbol{b}$. In other words, the transformation

$$\boldsymbol{y}' = H\boldsymbol{x}' + (c\boldsymbol{i})'$$

maps $\boldsymbol{a}$ into $\boldsymbol{b}$.

If $\mathscr{B}(s)$ is the sub-field generated by s, then we have just proved that $\mathscr{B}(s) = \mathscr{A}(\mathscr{G}_\sigma^*)$. The proof of the $\mathscr{P}$-equivalence of

$$\mathscr{B}(s) \quad \text{and} \quad \widetilde{\mathscr{A}}(\mathscr{G}_\sigma)$$

will follow the familiar pattern set up earlier for example 4 in section 7.

Now, suppose that our parameter of interest is $\gamma = \mu/\sigma$. The statistic $\bar{x}/s$ is γ-oriented and hence (Theorem 5) is almost $\mathscr{G}_\gamma$-invariant, where $\mathscr{G}_\gamma$ is the group of all γ-preserving transformations.

Consider now the sub-group $\mathscr{G}_\gamma^*$ of all linear transformations of the type

$$\boldsymbol{y}' = cH\boldsymbol{x}'$$

where $c > 0$ and H is a Haggstrom matrix.

It is easily checked that the maximal invariant with respect to the sub-group $\mathscr{G}_\gamma^*$ is the statistic $\bar{x}/s$ and hence, the sub-field $\mathscr{B}(\bar{x}/s)$ generated by $\bar{x}/s$ is $\mathscr{P}$-equivalent to the sub-field $\widetilde{\mathscr{A}}(\mathscr{G}_\gamma)$ of all almost $\mathscr{G}_\gamma$-invariant sets.

The case where our parameter of interest is $|\mu|/\sigma$ is very similar. Once again we observe that the statistic $|\bar{x}|/s$ is oriented towards $|\mu|/\sigma$. Hence, making use of Theorem 5 and observing that $|\bar{x}|/s$ is the maximal invariant with respect to the sub-group of linear parameter-preserving transformations of the form

$$\boldsymbol{y}' = cH\boldsymbol{x}' \ (c \neq 0,\ H \text{ a Haggstrom matrix}),$$

we are able to show that the invariance reduction of the data is to the statistic $|\bar{x}|/s$.

11. SOME FINAL REMARKS

a) In statistical literature, the principle of invariance has been used in a rather half-hearted manner (see, for example, [5] and [6]). We do not find any consideration given to the present project of reducing the data with the help of the whole class $\mathcal{G}_\gamma$ of γ-preserving transformations. In fact, the question of how extensive the class $\mathcal{G}_\gamma$ can be has escaped general attention. One usually works in the framework of a relatively small and simple sub-group of $\mathcal{G}_\gamma$ and invokes simultaneously the two different principles of sufficiency and invariance for the purpose of arriving at a *satisfactory* data reduction. The main object of the present article is to investigate how far the principle of invariance by itself can take us.

b) The main limitation of the principle of sufficiency is that it does not recognize nuisance parameters. Several attempts have been made to generalize the idea of sufficiency so that one gets an effective data reduction in the presence of nuisance parameters. Not much success has, however, been achieved in this direction.

c) On the other hand, the invariance principle usually falls to pieces when faced with a discrete model. The Bernoulli experimental set-up is one of the rare discrete models that the invariance principle can tackle. If $x_1, x_2, \ldots, x_n$ are n independent zero-one variables with probabilities θ and $1-\theta$, then all permutations (of co-ordinates) are model-preserving. And they reduce the data directly to the least sufficient statistic $r = x_1 + \ldots + x_n$. Take however, the following simple example. Let x and y be independent zero-one variables, where

$$P(x = 0) = \theta, \quad 0 < \theta < 1,$$

and

$$P(y = 0) = 1/3.$$

Now the identity-map is the only available model-preserving transformation. The principle of sufficiency reduces the data immediately to x.

d) The object of this article was not to make a critical evaluation of the twin principles of data reduction. Yet, the author finds it hard to refrain from observing that both the principles of sufficiency and invariance are extremely sensitive to changes in the model. For example, the spectacular data reductions we have achieved in the many examples considered here become totally unavailable if the basic normality assumption is changed ever so slightly.

References

[1] Bahadur, R. R. "Sufficiency and Statistical Decision Functions," *Ann. Math. Statist.*, 25 (1954), 423–462.

[2] Basu, D. "Sufficiency and Model-Preserving Transformations," Inst. of Statistics, Mimeo Series No. 420, Univ. of North Carolina, Chapel Hill, 1965.

[3] Burkholder, D. L. "Sufficiency in the Undominated Case," *Ann. Math. Statist.*, 32 (1961), 1191–1200.

[4] Farrell, R. H. "Representation of Invariant Measures," *Illinois J. of Math.*, 6 (1962), 447–467.

[5] Hall, W. J., R. A. Wijsman, and J. K. Gosh. "The Relationship between Sufficiency and Invariance," *Ann. Math. Statist.*, 36 (1965), 575–614.

[6] Lehmann, E. L. *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1959.

(*Received Sept. 17, 1965.*)