IMPARTIAL DECISION RULES AND SUFFICIENT STATISTICS[1]

By Raghu Raj Bahadur and Leo A. Goodman

The University of Chicago

Summary. A class of decision problems concerning k populations was considered in [1] and it was shown that a particular decision rule is the uniformly best 'impartial' decision rule for many problems of this class. The present paper provides certain improvements of this result. The authors define impartiality in terms of permutations of the k samples rather than in terms of the k ordered values of an arbitrarily chosen real-valued statistic as in the earlier paper. They point out that (under conditions which are satisfied in the standard cases of k independent samples of equal size) if the same function is a sufficient statistic for each of the k samples then the conditional expectation of an impartial decision rule given the k sufficient statistics is also an impartial decision rule. A characterization of impartial decision rules is given which relates the present definition of impartiality with the one adopted in [1]. These results, together with Theorem 1 of [1], yield the desired improvements. The argument indicated here is illustrated by application to a special case.

1. Introduction. Let $\pi_1, \pi_2, \cdots, \pi_k$ be a given set of populations and let the distribution function of a single observation x_i from π_i be

$$\Pr(x_i \leqq z) = G(z, \theta_i) \qquad (-\infty < z < \infty), \tag{1}$$

where θ_i is an unknown parameter (not necessarily real-valued), $i = 1, 2, \cdots, k$. Write $\omega = (\theta_1, \theta_2, \cdots, \theta_k)$ and let Ω be the set of all points ω which are regarded as being possible in the given case. Suppose that n independent observations are drawn from π_i, giving the sample $(x_{i1}, x_{i2}, \cdots, x_{in}) = u_i$ say, $i = 1, 2, \cdots, k$, and let the combined sample point $(u_1, u_2, \cdots, u_k)$ be denoted by v. Let $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ be an ordered set of functions p_i of the combined sample point v such that

$$0 \leqq p_i(v) \leqq 1, \qquad \sum_{i=1}^{k} p_i(v) = 1 \tag{2}$$

for all v. Then $d(v)$ is said to be a *decision rule*. The statistical problems which motivate this definition may be described as follows.

Suppose that it is desired to determine appropriate sampling rates $p_1, p_2, \cdots, p_k$ for $\pi_1, \pi_2, \cdots, \pi_k$, respectively, p_i being the relative proportion of x's which will be drawn in future from π_i, $(0 \leqq p_i \leqq 1, \sum_1^k p_i = 1)$. For example, the given populations may be k varieties of grain, x_i the yield (bushels per acre or dollars profit per acre) from π_i, and p_i the proportion of the available land on which π_i is to be grown, $i = 1, 2, \cdots, k$. Again, $\pi_1, \pi_2, \cdots, \pi_k$ may be

[1] This paper was prepared in connection with research sponsored by the Office of Naval Research.

sources of a manufactured article, x_i the relevant quality characteristic (e.g., number of hours of service) of an article supplied by π_i, and $p_1, p_2, \cdots, p_k$ the relative proportions in which a consumer obtains the articles he needs from $\pi_1, \pi_2, \cdots, \pi_k$, respectively. The mixed population obtained by using a given set $d^0 = (p_1^0, p_2^0, \cdots, p_k^0)$ of sampling rates is characterized by the distribution function

$$G(z \mid \omega, d^0) = \sum_{i=1}^{k} p_i^0 G(z, \theta_i) \qquad (-\infty < z < \infty), \tag{3}$$

where the component distribution functions $G(z, \theta_i)$ are given by (1), and the object is to determine d^0 in such a way that $G(z \mid \omega, d^0)$ has properties which are desirable in the given case. (For instance, it may be desirable to minimize $G(a \mid \omega, d^0)$, where a is a given constant, or to maximize $G(b + \epsilon \mid \omega, d^0) - G(b - \epsilon \mid \omega, d^0)$ where b and $\epsilon > 0$ are given constants.) If the parameters θ_i were known, an appropriate d^0 could (presumably) be determined, but otherwise the statistician must resort to sampling the populations and take d^0 to be a function of the sample values. If samples of fixed size n are drawn from each population, we see that the statistician will be using a decision rule, say $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$. The expected distribution function of the mixed population obtained by using the rule $d(v)$ is (cf. (3))

$$\begin{aligned} H(z \mid \omega, d) &= E[G(z \mid \omega, d(v)) \mid \omega] \\ &= \sum_{i=1}^{k} G(z, \theta_i) E[p_i(v) \mid \omega] \qquad (-\infty < z < \infty), \end{aligned} \tag{4}$$

where $E[p_i(v) \mid \omega]$ denotes the expected value of $p_i(v)$ when the true parameter point is ω. The statistician's problem is to construct a decision rule $d^*(v)$ such that $H(z \mid \omega, d^*)$ has properties which are desirable in the given case.

A special version of applications of this type is the following. For brevity, write $G_i(z) = G(z, \theta_i)$ and let $\lambda(G)$ be a real-valued functional on the distribution functions $G_i(z)$, for example, $\lambda(G) = \int_{-\infty}^{\infty} z \, dG(z)$ or $\lambda(G) = G(b + \epsilon) - G(b - \epsilon)$, where b and $\epsilon > 0$ are given constants. Writing $\lambda_i = \lambda(G_i)$, suppose that it is desired to select a population, π_r say, from the given set $\pi_1, \pi_2, \cdots, \pi_k$ such that $\lambda_r = \max\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$. Since the G_i's are unknown, it will in general be impossible to effect a (or the) correct selection with certainty, but if it is agreed to make the selection depend on the outcome of drawing samples of size n from each population, the most general selection procedure is to use a decision rule, $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ say, in the following manner: given v, the statistician performs a random experiment whose outcome ρ takes on only the values $1, 2, \cdots, k$ with

$$\Pr(\rho = i) = p_i(v) \qquad (i = 1, 2, \cdots, k)$$

and selects π_ρ. The probabilities of selecting $\pi_1, \pi_2, \cdots, \pi_k$ are then

$$E[p_1(v) \mid \omega], E[p_2(v) \mid \omega], \cdots, E[p_k(v) \mid \omega], \tag{5}$$

respectively. The problem in such a case might be to construct a decision rule $d^*(v)$ such that the probability of a correct selection in using $d^*(v)$ is "as large as possible."

In view of the applications (cf. (4), (5)) we shall say that two decision rules $d(v) = (p_1(v), \cdots, p_k(v))$ and $d'(v) = (p_1'(v), \cdots, p_k'(v))$ are *equivalent* if $E[p_i(v) \mid \omega] = E[p_i'(v) \mid \omega]$ for $i = 1, 2, \cdots, k$ and all ω in Ω.

We shall concern ourselves primarily with a class of decision rules which seems to be of interest on intuitive grounds. This is the class of impartial decision rules (see [1], [2]). Let us consider the case $k = 2$. Then a decision rule $d(v)$ is said to be impartial if $d(u_1, u_2) = (\alpha, \beta)$ implies $d(u_2, u_1) = (\beta, \alpha)$. In other words, $d(v) = (p_1(v), p_2(v))$ is an impartial decision rule if $p_1(u_2, u_1) = p_2(u_1, u_2)$, $p_2(u_2, u_1) = p_1(u_1, u_2)$ for all $v = (u_1, u_2)$. In the general case, a decision rule $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ is said to be *impartial* if for any $v = (u_1, u_2, \cdots, u_k)$ and any permutation $i_1\, i_2 \cdots i_k$ of $1\ 2\ 3 \cdots k$ we have $p_j\,(u_{i_1}, u_{i_2}, \cdots, u_{i_k}) = p_{i_j}\,(u_1, u_2, \cdots, u_k)$ for $j = 1, 2, \cdots, k$.

The main result of this paper (Theorem 2) applies to cases whose essential feature is the existence of a function $s(u)$, not necessarily real-valued, on n dimensional sample space such that $s_i = s(u_i)$ is a sufficient statistic for θ_i $(i = 1, 2, \cdots, k)$, and such that the conditional distribution of u_i with s_i fixed equal to c is the same for each i. (The necessary conditions are always satisfied if, for example, the k populations are all (i) normal, or (ii) rectangular, or (iii) exponential, or (iv) binomial, or (v) of Poisson type.) Then $t(v) = (s_1, s_2, \cdots, s_k)$ is a sufficient statistic for ω, and it is clear, upon taking conditional expectations, that corresponding to any decision rule $d(v)$ there exists a decision rule $d^*(t(v))$ which is equivalent to $d(v)$ (Theorem 1). It is not immediately obvious, however, that if $d(v)$ is an impartial decision rule then this equivalent rule $d^*(t(v))$ will also be impartial. We show, in proving Theorem 2, that this is indeed the case. The question raised on page 374 of [1] with reference to Example 2 of that paper is thus answered in the affirmative. Our final result (Theorem 3) gives a characterization of impartial decision rules which relates the present definition to the one adopted in [1]. It might be pointed out that impartiality is a special case of invariance (cf. [6]), so that this is a special case of the following proposition: the conditional expectation of an invariant decision rule is also an invariant decision rule. A discussion of the general proposition will appear elsewhere.

Now, it is known (Theorem 1 of [1]) that there exist two impartial decision rules, called $d_1^*(t(v))$ and $d_k^*(t(v))$, such that in many applications $d_1^*(t(v))$ is the worst one and $d_k^*(t(v))$ the best one in the class of all impartial decision rules of the form $d^*(t(v))$ whatever the unknown parameter point ω may be; that is, $d_1^*(t(v))$ and $d_k^*(t(v))$ are the uniformly worst and uniformly best decision rules in the class of all impartial decision rules which are based on the sufficient statistics $s_1, s_2, \cdots, s_k$ alone. Theorem 2 shows that in these applications $d_1^*(t(v))$ and $d_k^*(t(v))$ are in fact uniformly worst and uniformly best in the class of *all* impartial decision rules. (Theorem 1 of [1] is stated and proved only in the "continuous" case, but can be extended to cover the discrete case as well; the necessary modifications become evident upon comparing Theorem 3 of the present paper with the development in [1].) By way of illustration of the argument indicated here, in the final section of the paper we consider certain problems connected with the case when $\pi_1, \pi_2, \cdots, \pi_k$ are normal populations having unknown means m_i and a common variance σ^2 (which may or may not be known) and prove a result (Theorem 4) which generalizes Example 1 of [1] as also a result due to Simon [2] for the case $k = 2$.

2. Theorems. The reader is referred to [3] for an account of such measure-theoretic terms and results as we use without explanation in what follows. Throughout this section we write '(sub)set' for 'Borel-measurable (sub)set' and 'function' for 'Borel-measurable function.' Functions whose range is not specified are understood to be real-valued. R^q denotes a *fixed subset* of the set of all points $z = (x_1, x_2, \cdots, x_q)$ with real coordinates x_i. (In our discussion, some of the spaces R^q will be given at the outset, and all the others will be defined explicitly in terms of them.) For any subset A of R^q, $\chi_A(z)$ denotes the characteristic function of A; that is, $\chi_A(z) = 1$ for z in A and $= 0$ for z in $R^q - A$. Let f be a nonnegative function on R^q and let λ be a measure on R^q (more precisely, a measure on the subsets of R^q) such that

$$\int_{R^q} f(z)\, d\lambda = 1.$$

Let Z be a random variable taking values in R^q such that the probability of event $\{Z \varepsilon A\}$ is

$$\int_A f(z)\, d\lambda$$

for all sets A. We then say that Z is distributed (on R^q) according to $f(z)\, d\lambda$.

Let $U_1, U_2, \cdots$, and U_k be independent random variables whose joint distribution is governed by a parameter ω taking values in a space Ω. Each U_i takes values in a set R^n of points u. Let s be a function on R^n onto a set R^m of points y. Let $h(u)$ be a nonnegative function of u, and let μ be a σ-finite measure on R^n. Corresponding to each ω in Ω and each $i = 1, 2, \cdots, k$ let $g_i(y:\omega)$ be a nonnegative function of y such that

$$\int_{R^n} h(u) g_i(s(u):\omega)\, d\mu = 1.$$

It is assumed that U_i is distributed according to $h(u)g_i(s(u):\omega)\, d\mu$ $(i = 1, 2, \cdots k)$.

Let R^{nk} $(= R^n \times R^n \times \cdots \times R^n)$ be the set of all points $v = (u_1, u_2, \cdots, u_k)$ with u_i in R^n $(i = 1, 2, \cdots, k)$, and write

$$\alpha(v) = \prod_{i=1}^{k} h(u_i),$$

$$t(v) = (s(u_1), s(u_2), \cdots, s(u_k)),$$

$$\beta(t(v):\omega) = \prod_{1=1}^{k} g_i(s(u_i):\omega).$$

Let $\mu^{(k)}$ be the product measure $\mu \times \mu \times \cdots \times \mu$ on R^{nk}. Then $V = (U_1, U_2, \cdots, U_k)$ is distributed according to $\alpha(v)\ \beta(t(v):\omega)\ d\mu^{(k)}$ If ϕ is a function on R^{nk}, we shall denote the expected value (if it exists) of $\phi(V)$, that is, the integral

$$\int_{R^{nk}} \phi(v)\alpha(v)\beta(t(v):\omega)\ d\mu^{(k)},$$

by $E[\phi(v) \mid \omega]$.

Let R^{mk} $(= R^m \times R^m \times \cdots \times R^m)$ be the set of all values of t as v ranges over R^{nk}, and let the generic point of R^{mk} be denoted by w or by $(y_1, y_2, \cdots, y_k)$. It can be shown that the preceding assumptions and definitions imply that t is a sufficient statistic for ω when the sample space is R^{nk}; that is, corresponding to each subset A of R^{nk} there exists a function ϕ_A, $0 \leqq \phi_A \leqq 1$, on R^{mk} such that

$$E[\chi_B(t(v))\ \chi_A(v) \mid \omega] = E[\chi_B(t(v))\ \phi_A(t(v)) \mid \omega]$$

for all subsets B of R^{mk} and all ω in Ω. ($\phi_A(w)$ is called the conditional probability of the event $\{V \varepsilon A\}$ given $t(V) = w$ and any ω in Ω). This property of t does not, however, suffice for our purpose; we require in addition the following result concerning the structure of the functions $\phi_A(w)$.

LEMMA. *Corresponding to each y in R^m there exists a probability measure λ_y on R^n such that for each A and w we may take $\phi_A(w) = \nu_w(A)$, where, for fixed $w = (y_1, y_2, \cdots, y_k)$, ν_w is the product measure $\lambda_{y_1} \times \lambda_{y_2} \times \cdots \times \lambda_{y_k}$ on R^{nk}.*

A proof of the lemma can be constructed along the following lines. (i) There exist functions $g_i(s:\omega)$ and a fixed probability measure μ^* on R^n such that U_i is distributed according to $g_i(s(u):\omega)\ d\mu^*$ $(i = 1, 2, \cdots, k; \omega \varepsilon \Omega)$. (ii) There exist functions $\lambda_y(C)$ such that, for each y, λ_y is a probability measure on R^n, and for each subset C of R^n, $\lambda_y(C)$ is the conditional measure of C given $s(u) = y$ when μ^* is the unconditional measure on R^n (cf. Exercise (5) on page 210 of [3]). (iii) For each $i = 1, 2, \cdots, k$ and any set C, $\lambda_y(C)$ is the conditional probability of the event $\{U_i \varepsilon C\}$ given $s(U_i) = y$ and any ω in Ω. Finally, (iv) "the conditional joint distribution of $U_1, U_2, \cdots$, and U_k given $s(U_i) = y_i$, $i = 1, 2, \cdots, k$, and ω is the product of the individual conditional distributions under the corresponding individual conditions", and the lemma follows. We omit the detailed verifications.

The reader should satisfy himself that (with a suitable definition of the sufficient statistic t in each case) the lemma applies to all standard cases of k independent samples of equal size. It is therefore likely to prove useful also in contexts other than the present one.

Now let $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ be a decision rule. Write

$$p_i^*(w) = \int_{R^{nk}} p_i(v)\ d\nu_w.$$

Since ν_w is a probability measure, it is clear from (2) that $0 \leqq p_i^*(w) \leqq 1$, $\sum_{i=1}^k p_i^*(w) \equiv 1$, so that $d^*(t(v)) = (p_1^*(t(v)), p_2^*(t(v)), \cdots, p_k^*(t(v))$ is a decision rule. It follows from Exercise (6) on page 211 of [3] that $p_i^*(w)$ is the conditional expectation of $p_i(V)$ given $t(V) = w$ and any ω in Ω; that is,

$$E[\chi_B(t(v))p_i(v) \mid \omega] = E[\chi_B(t(v))\, p_i^*(t(v)) \mid \omega]$$

for all subsets B of R^{mk} and all ω in Ω ($i = 1, 2, \cdots, k$). Taking $B = R^{mk}$ we see that $d^*(t(v))$ is equivalent to $d(v)$. Thus we have

THEOREM 1. *Corresponding to any decision rule $d(v)$ there exists an equivalent decision rule $d^*(t(v))$.*

Suppose now that $d(v)$ is an impartial decision rule. It is easy to see that in that case $d^*(t(v))$ must also be impartial. Consider the case $k = 2$. Then for any point (y_1, y_2) of R^{m2} we have

$$\begin{aligned} p_2^*(y_1, y_2) &= \int_{R^{n2}} p_2(v)\, d\nu_{(y_1, y_2)}(v) \\ &= \int_{R^n} \left\{ \int_{R^n} p_2(u_1, u_2)\, d\lambda_{y_1}(u_1) \right\} d\lambda_{y_2}(u_2) \\ &= \int_{R^n} \left\{ \int_{R^n} p_1(u_2, u_1)\, d\lambda_{y_1}(u_1) \right\} d\lambda_{y_2}(u_2) \\ &= \int_{R^n} \left\{ \int_{R^n} p_1(u_1, u_2)\, d\lambda_{y_1}(u_2) \right\} d\lambda_{y_2}(u_1) \\ &= \int_{R^{n2}} p_1(v)\, d\nu_{(y_2, y_1)}(v) = p_1^*(y_2, y_1) \end{aligned}$$

and the impartiality of $d^*(t(v))$ is proved. A parallel argument applies to the general case. Hence

THEOREM 2. *Corresponding to any impartial decision rule $d(v)$ there exists an equivalent impartial decision rule $d^*(t(v))$.*

We remind the reader that the $d^*(t(v))$ which we have constructed in terms of the given $d(v)$ is equivalent to $d(v)$ in virtue of the fact that $d^*(t(v))$ is the conditional expectation of $d(V)$ given $t(V) = t(v)$ and any ω in Ω. In many statistical applications, the loss incurred in adopting a particular decision $d^0 = (p_1^0, p_2^0, \cdots, p_k^0)$ when ω is the parameter point is of the form

$$l(\omega, d^0) = \sum_{i=1}^{k} f_i(\omega) p_i^0, \quad f_i(\omega) \geqq 0.$$

For each ω let $c(\omega, x)$ be a bounded convex function of x defined for $\min_i \{f_i(\omega)\} \leqq x \leqq \max_i \{f_i(\omega)\}$ and write $\psi(\omega, d^0) = c(\omega, l(\omega, d^0))$. Then ψ is a convex function of d^0 for each fixed ω. It follows from Lemma 3.1 of [4] that, irrespective of the particular weights $f_i(\omega)$ and particular function $c(\omega, x)$, we have

$$E[\psi(\omega, d^*(t(v))) \mid \omega] \leqq E[\psi(\omega, d(v)) \mid \omega] \text{ for all } \omega.$$

Our immediate purpose in stating this consequence of the relation between $d(v)$ and $d^*(t(v))$ will be served by noting the following easy corollaries of the general result: (i) The expected loss when using $d^*(t(v))$ always equals the expected loss when using $d(v)$, and (ii) the *variance* of the loss when using $d^*(t(v))$ never exceeds the variance when using $d(v)$. Now, the equivalence of $d^*(t(v))$ and $d(v)$ also implies (i), but results such as (ii) do not follow from equivalence alone. There is, therefore, a somewhat stronger justification than the one given by Theorems 1 and 2 for using decision rules which depend on the outcome v only through t.

We shall now give a useful representation of impartial decision rules. Let $\phi(u)$ be a real valued function on R^n and for any $v = (u_1, u_2, \cdots, u_k)$ set $\phi_i = \phi(u_i)$, $i = 1, 2, \cdots, k$. Let $\mathfrak{D}(\phi)$ be the class of all impartial decision rules which are based on $\phi_1, \phi_2, \cdots, \phi_k$ alone, that is, all impartial decision rules of the form $d(v) = (p_1(\phi_1, \phi_2, \cdots, \phi_k), p_2(\phi_1, \cdots, \phi_k), \cdots, p_k(\phi_1, \cdots, \phi_k))$. Since $\phi(u)$ is a given function, $\mathfrak{D}(\phi)$ will, in general, be a subclass of the class of all impartial decision rules, but may coincide with it. In any case, for given v, let $\phi_{(1)}, \phi_{(2)}, \cdots, \phi_{(k)}$ be the k (not necessarily distinct) numbers ϕ_i arranged in ascending order of magnitude and write

$$a_{ij} = \begin{cases} 1 & \text{if } \phi_i = \phi_{(j)}, \\ 0 & \text{otherwise} \end{cases} \qquad (i, j = 1, 2, \cdots, k).$$

THEOREM 3. *A decision rule $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ is a member of $\mathfrak{D}(\phi)$ if and only if there exist functions $\lambda_j(z_1, z_2, \cdots, z_k)$, $j = 1, 2, \cdots, k$, such that*

$$0 \leqq \lambda_j \leqq 1, \qquad \sum_{j=1}^{k} \lambda_j \equiv 1$$

and such that for each $i = 1, 2, \cdots, k$

$$p_i(v) = \sum_{j=1}^{k} \frac{a_{ij}}{\sum_{i=1}^{k} a_{ij}} \lambda_j(\phi_{(1)}, \phi_{(2)}, \cdots, \phi_{(k)})$$

for all v.

The proof is by direct verification and is omitted.

3. An application. Let $\pi_1, \pi_2, \cdots, \pi_k$ be normal populations, π_i having an unknown mean m_i and variance σ^2. Write $\theta_i = (m_i, \sigma)$, $\omega = (\theta_1, \theta_2, \cdots, \theta_k)$, and let Ω be the set of all points ω which are regarded as being possible in the given case. Let $g_i(\omega)$, $i = 1, 2, \cdots, k$, be functions defined on Ω such that $m_i \leqq m_j$ implies $g_i \leqq g_j$, $i, j = 1, 2, \cdots, k$. Suppose that samples $u_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ of n independent observations are drawn from each of the populations π_i, giving the combined sample point $v = (u_1, u_2, \cdots, u_k)$. For any decision rule $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ and any ω in Ω let the expected loss, or risk, by given by

$$r(d \mid \omega) = \max_i \{g_i(\omega)\} - \sum_{i=1}^{k} g_i(\omega) \cdot E[p_i(v) \mid \omega]. \tag{6}$$

Regarded as a function of ω, $r(d \mid \omega)$ is called the *risk function* of $d(v)$. The problem is to construct (if possible) an impartial decision rule $d^*(v)$ such that $r(d^* \mid \omega)$ is as small as possible no matter what ω may be. We shall show that the problem has a solution which is independent of the functions g_i. We shall also describe two determinations of the functions g_i which seem to be of special interest.

For any v, set $\bar{x}_i = n^{-1} \sum_{j=1}^{n} x_{ij}$, $i = 1, 2, \cdots, k$, and $\bar{x}_{(1)} = \min \{\bar{x}_i\}$, $\bar{x}_{(k)} = \max \{\bar{x}_i\}$, and let $a(v)$ = number of $\bar{x}_i$'s which equal $\bar{x}_{(1)}$, $b(v)$ = number of $\bar{x}_i$'s which equal $\bar{x}_{(k)}$. (Of course, we have $\Pr(a(v) = b(v) = 1 \mid \omega) = 1$ for all ω.) Write

$$p_i^{(1)}(v) = \begin{cases} \dfrac{1}{a(v)} & \text{if } \bar{x}_i = \bar{x}_{(1)}, \\ 0 & \text{otherwise}, \end{cases} \qquad (i = 1, 2, \cdots, k),$$

$$p_i^{(k)}(v) = \begin{cases} \dfrac{1}{b(v)} & \text{if } \bar{x}_i = \bar{x}_{(k)}, \\ 0 & \text{otherwise} \end{cases} \qquad (i = 1, 2, \cdots, k).$$

It is then clear that $d_1(v) = (p_1^{(1)}(v), p_2^{(1)}(v), \cdots, p_k^{(1)}(v))$ and $d_k(v) = (p_1^{(k)}(v), p_2^{(k)}(v), \cdots, p_k^{(k)}(v))$ are fixed impartial decision rules which depend on v only through $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k$.

THEOREM 4. *Let $\mathfrak{D}$ be the class of all impartial decision rules $d(v)$. Then*

$$r(d_1 \mid \omega) = \sup_{d \varepsilon \mathfrak{D}} r(d \mid \omega), \qquad r(d_k \mid \omega) = \inf_{d \varepsilon \mathfrak{D}} r(d \mid \omega)$$

for all ω in Ω.

PROOF. Choose and fix an arbitrary impartial decision rule $d(v)$. Let $c > 0$ be any constant such that the subset $\Omega_c = \{\omega : \omega \,\varepsilon\, \Omega, \sigma = c\}$ is non-empty. Now, corresponding to each ω in Ω_c the probability density (with respect to n-dimensional Lebesgue measure) of the sample from π_i is of the form $h_c(u)g_i(\bar{x}:\omega)$, $i = 1, 2, \cdots, k$. Write $t(v) = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$. It follows from Theorem 2 that there exists an *impartial* decision rule based on $t(v)$ alone, say $d_c^*(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$, which is equivalent to $d(v)$ provided ω is restricted to Ω_c. From equivalence and (6), we have

$$r(d \mid \omega) = r(d_c^* \mid \omega) \tag{7}$$

for ω in Ω_c. Now, since for $i \neq j$ we have $\Pr(\bar{x}_1 = \bar{x}_j \mid \omega) = 0$ for all ω, it follows that (with probability equal to one for all ω) the representation of impartial decision rules based on $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k$ which is given by Theorem 3 coincides with the representation assumed in Theorem 1 of [1]. An application of this last theorem (cf. Example 1 in Section 6 of [1]) shows that

$$r(d_1 \mid \omega) \geqq r(d_c^* \mid \omega), \qquad r(d_k \mid \omega) \leqq r(d_c^* \mid \omega) \tag{8}$$

for all ω. It follows from (7) and (8) that $r(d_1 \mid \omega) \geqq r(d \mid \omega)$ and $r(d_k \mid \omega) \leqq r(d \mid \omega)$ for ω in Ω_c. Since both $d(v)$ and c are arbitrary, Theorem 4 is proved.

In conclusion, we describe two applications of Theorem 4. Suppose that v is the outcome of preliminary experiments on $\pi_1, \pi_2, \cdots, \pi_k$ and now it is desired to draw a total of N observations from the k populations in such a way that the mathematical expectation of the sum of the values obtained is as large as possible. Let $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ be a suitable decision rule and suppose that $Np_i(v)$ observations are drawn from π_i, $i = 1, 2, \cdots, k$. Then the mathematical expectation of the sum of the values obtained is $N \sum_1^k m_i E[p_i(v) \mid \omega]$. Since the maximum of this quantity is $N \max \{m_1, m_2, \cdots, m_k\}$, the expected loss in using $d(v)$ may be taken to be

$$(9) \qquad N\left[\max \{m_i\} - \sum_1^k m_i E[p_i(v) \mid \omega]\right].$$

The expected loss is of the form (6), with $g_i(\omega) = Nm_i$ for $i = 1, 2, \cdots, k$. It follows that in the class of all impartial decision rules the uniformly best rule is to drawn an equal number of observations from populations π_i such that $\bar{x}_i = \max \{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k\}$ and none from the others.

Suppose now that it is desired to select one of the populations π_i, the object being to select a population, π_r say, such that $m_r = \max \{m_1, m_2, \cdots, m_k\}$. As pointed out in the introductory section, the statistician may then employ a suitable decision rule, say $d(v) = (p_1(v), \cdots, p_k(v))$, in the following way: given v, he performs a random experiment whose outcome ρ takes on the values $1, 2, \cdots, k$ with $\Pr(\rho = i) = p_i(v)$ $(i = 1, 2, \cdots, k)$, and selects π_ρ. Write $m_{(k)} = \max \{m_1, m_2, \cdots, m_k\}$, and set

$$g_i(\omega) = \begin{cases} 1 & \text{if } m_i = m_{(k)}, \\ 0 & \text{otherwise} \end{cases} \qquad (i = 1, 2, \cdots, k).$$

Then it is readily seen from (5) and (6) that with the present convention for the manner in which a decision rule $d(v)$ is to be used, we have

$$(10) \qquad r(d \mid \omega) = \Pr(\text{incorrect selection using } d(v) \mid \omega).$$

It follows from Theorem 4 that in the class of all impartial decision rules, the rule $d_k(v)$ which is to assign equal probabilities of selection to populations π_i such that $\bar{x}_i = \max \{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k\}$ and zero probabilities to the rest, minimizes the probability of an incorrect selection uniformly for all ω in Ω.

The reader is referred to [5] for an investigation from a more general viewpoint of the problem of minimizing (9) or (10) in the case $k = 2$. The discussion in [5] does not presuppose samples of equal size, and the class of all decision rules is taken into consideration.

REFERENCES

[1] R. R. Bahadur, "On a problem in the theory of k populations," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 362–375.

[2] H. A. SIMON, "Symmetric tests of the hypothesis that the mean of one normal population exceeds that of another," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 149–154.

[3] P. R. HALMOS, *Measure Theory*, D. Van Nostrand Company, New York, 1950.

[4] J. R. HODGES AND E. L. LEHMANN, "Some problems in minimax point estimation," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 182–197.

[5] R. R. BAHADUR AND H. ROBBINS, "The problem of the greater mean," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 469–487.

[6] E. L. LEHMANN, "A general concept of unbiasedness," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 587–592.