THE PROBLEM OF THE GREATER MEAN

By Raghu Raj Bahadur and Herbert Robbins[1]

University of North Carolina

1. Introduction and summary. Let π_1, π_2 be normal populations with means m_1, m_2 respectively and a common variance σ^2, the parameter point $\omega = (m_1, m_2 : \sigma)$ which characterizes the two populations being unknown, and let Ω be an arbitrary given set of possible points ω. Random samples of fixed sizes n_1, n_2 are drawn from π_1, π_2 respectively, giving the combined sample point $v = (x_{11}, x_{12}, \cdots, x_{1n_1}; x_{21}, x_{22}, \cdots, x_{2n_2})$. For reasons which will be made clear later in connection with practical examples, any function $f(v)$ such that $0 \leq f(v) \leq 1$ is called a *decision function*, and for any such $f(v)$ the *risk function* is defined to be

$$r(f \mid \omega) = \max[m_1, m_2] - m_1 E[f \mid \omega] - m_2 E[1 - f \mid \omega] \geq 0, \tag{1}$$

where E denotes the expectation operator. A decision function $\bar{f}(v)$ is said to be (a) *uniformly better* than $f(v)$ if $r(\bar{f} \mid \omega) \leq r(f \mid \omega)$ for all ω in Ω, the strict inequality holding for at least one ω, (b) *admissible* if no decision function is uniformly better than $\bar{f}(v)$, and (c) *minimax* if

$$\sup_{\omega \epsilon \Omega} [r(\bar{f} \mid \omega)] = \inf_f \sup_{\omega \epsilon \Omega} [r(f \mid \omega)].$$

The "problem of the greater mean" is, for any given Ω, to determine the minimax decision functions, particularly those which are also admissible. Special interest attaches to the case in which there exists a *unique* minimax decision function $\bar{f}(v)$ (in the sense that if $f(v)$ is any minimax decision function then $f(v) = \bar{f}(v)$ for almost every v in the sample space); such an $\bar{f}(v)$ is automatically admissible.

The problem of the greater mean is, of course, a special problem in Wald's general theory of statistical decision functions [1]. Our results will, however, be derived by very simple direct methods which make no use of Wald's general theorems.

We cite without proofs a few examples in order to show how strongly the solution of the problem of the greater mean depends on the structure of Ω. In each case the minimax decision function is a function only of the two sample means $\bar{x}_1$, $\bar{x}_2$.

(i) Let Ω' consist of the two points $(a, b: \sigma)$ and $(b, a: \sigma)$, with $a < b$. Then

$$f^*(v) = \begin{cases} 1 \text{ if } n_1\bar{x}_1 - n_2\bar{x}_2 > (n_1 - n_2)(a + b)/2, \\ 0 \text{ otherwise,} \end{cases} \tag{2}$$

is the unique minimax decision function.

[1] This work was supported in part by the Office of Naval Research.

(ii) Let Ω'' consist of the two points $(c + h, c\colon \sigma)$ and $(c - h, c\colon \sigma)$, with $h > 0$. Then

$$f_c^0(v) = \begin{cases} 1 \text{ if } \bar{x}_1 > c, \\ 0 \text{ otherwise,} \end{cases} \tag{3}$$

is the unique minimax decision function.

(iii) Let Ω''' consist of the three points $(\frac{1}{2}, -\frac{1}{2}\colon 1)$, $(\frac{1}{2}, \frac{3}{2}\colon 1)$, $(-\frac{3}{2}, -\frac{1}{2}\colon 1)$, and let $n_1 = n_2 = n$. Then

$$f^{**}(v) = \begin{cases} 1 \text{ if } e^{-2n\bar{x}_1} + e^{2n\bar{x}_2} < \lambda, \\ 0 \text{ otherwise,} \end{cases} \tag{4}$$

where λ is a certain definite constant, is the unique minimax decision function.

The parameter spaces of two or three points specified in these examples are rather trivial, but in fact the corresponding decision functions (2), (3), (4) remain the unique minimax solutions of the decision problem with respect to much more general parameter spaces. Thus, for example, it is clear that $f^*(v)$ will remain the unique minimax decision function with respect to any Ω which contains Ω' and is such that

$$\sup_{\omega \in \Omega} [r(f^* \mid \omega)] = \sup_{\omega \in \Omega'} [r(f^* \mid \omega)].$$

Corresponding remarks apply to $f_c^0(v)$ and $f^{**}(v)$.

When $n_1 = n_2$, (2) reduces to

$$f^0(v) = \begin{cases} 1 \text{ if } \bar{x}_1 > \bar{x}_2, \\ 0 \text{ otherwise.} \end{cases} \tag{5}$$

This decision function is of particular interest when both the means m_1, m_2 are unknown. It will be shown that whether or not $n_1 = n_2$, $f^0(v)$ is the unique minimax decision function under certain conditions on Ω which are likely to hold in practice, at least when both n_1 and n_2 are sufficiently large (Theorem 3). Likewise, $f_c^0(v)$, which is the analogue of $f^0(v)$ when one of the means (m_2) is known exactly, is apt to be the unique minimax decision function in such cases, at least when n_1 is sufficiently large (Theorem 4). These results on $f^0(v)$ and $f_c^0(v)$ form the main results of the present paper.

So much by way of a general summary. We shall now give a practical illustration (another is given in Section 3) to show how the problem of the greater mean arises in applications.

Suppose that a consumer requires a certain number of manufactured articles which can be supplied at the same cost by each of two sources π_1 and π_2. The quality of an article is measured by a numerical characteristic x, and it is known that in the product of π_i, x is normally distributed with mean m_i and variance σ^2, but the values of these parameters are unknown. The consumer has obtained a random sample of n_1 and n_2 articles from π_1 and π_2 respectively, and has found the values of x to be $(x_{11}, x_{12}, \cdots, x_{1n_1}; x_{21}, x_{22}, \cdots, x_{2n_2}) = v$. What is the best way of ordering a total of N articles from the two sources?

The usual statistical theory, which confines itself to estimating the unknown parameters and to testing hypotheses of the form $H_0(m_1 = m_2)$, has at best an indirect bearing on the problem at hand. We therefore adopt Wald's point of view and investigate the consequences of any given course of action. If the consumer orders fN articles from π_1 and $(1 - f)N$ from π_2, where $0 \leq f \leq 1$, then the expectation of the sum of the x-values in the articles he obtains will be $N(m_1 f + m_2(1 - f))$. The maximum possible value of this quantity is $N \max [m_1, m_2]$, and the "loss" per article which he sustains may therefore be taken as

$$W(\omega, f) = \max [m_1, m_2] - m_1 f - m_2(1 - f) \geq 0,$$

where $\omega = (m_1, m_2 : \sigma)$ is the true parameter point.

The consumer wants to choose f so as to make W as small as possible. If he knew m_1 to be greater, or to be less, than m_2, then by choosing $f = 1$ or 0 respectively he could make $W = 0$. But since he does not know which m_i is the greater he will presumably choose f as some function of the sample point v. Suppose, therefore, that a "decision function" $f(v)$, such that $0 \leq f(v) \leq 1$ but not necessarily taking on only the values 0 and 1, is defined for all points v in the sample space and that the consumer sets $f = f(v)$.[2] In repeated applications of this procedure, the "risk" or expected loss (a double expectation is involved: the expected loss for a given f and the expected value of f in using the decision function $f(v)$) per article is given by (1), and the consumer will try to find an $f(v)$ which minimizes this risk. Since the value of the risk depends on ω it is necessary to specify which values of ω are to be regarded as possible in the given problem; let the set of all such ω be denoted by Ω. If the consumer agrees to adopt the "conservative" criterion of minimizing the maximum possible risk, then the statistician's problem is to find the minimax decision functions in the sense defined above. We have given the solutions of this problem for certain types of parameter spaces. The reader will observe that each of the minimax decision functions (2), (3), (4) was of the "all or nothing" type, with values 0 and 1 only. (Whether this remains true for every Ω we do not know.) By using one of these decision functions in a given instance one arrives at either the best possible decision or the worst. The attitudes of doubt sometimes associated with the non-rejection of the hypothesis $H_0(m_1 = m_2)$ are therefore

[2] One might say that the consumer should choose f in the light of what he can infer from v about the m_i. But this formulation as a problem in ordinary statistical inference (estimation and testing) is not relevant and may be misleading. For example, a plausible $f(v)$, based on the idea that the problem is one of testing hypotheses, is as follows: "Perform the two-tailed t test of $H_0(m_1 = m_2)$ at the five per cent level. If H_0 is rejected set $f = 0$ or 1 according as $\bar{x}_1$ is less than or greater than $\bar{x}_2$. If H_0 is not rejected set $f = \frac{1}{2}$." Another $f(v)$, based on the theory of estimation, according to which the $\bar{x}_i$ are the "best" estimates of the m_i, is as follows: "Set $f = 0$ or 1 according as $\bar{x}_1$ is less than or greater than $\bar{x}_2$." Actually, the latter procedure is, from the remarks above concerning (5), the "best" in a certain definite sense and under certain conditions, but this fact does not follow from the usual theory of estimation.

irrelevant to the problem of the greater mean in the examples cited. (Cf. footnote 2; also Example 1 in Section 3.)

The risk function (1) is but one of a general class R of risk functions, to be defined in Section 2, which are associated with the problem of the greater mean. The most important members of R are (1) and

$$\bar{r}(f \mid \omega) = P(\text{incorrect decision using } f(v) \mid \omega), \tag{6}$$

where "$m_1 \leq m_2$" and "$m_1 \geq m_2$" are the two possible decisions. The risk function (6) is relevant to applications of a purely "scientific" nature in which the statistician is asked merely to give his opinion as to which population has the greater mean. Although the problem of constructing a suitable decision function for (6) is akin in spirit to the problems considered in the now classical Neyman-Pearson theory of statistical tests, no satisfactory solutions seem to be available. It is easy to see, however, that (1) and (6) are quite similar. Of course, in the case of (1) a decision function $f(v)$ may take on any value between 0 and 1 inclusive, while for (6) we allow only functions which take on only the values 0 and 1, corresponding respectively to the decisions "$m_1 \leq m_2$" and "$m_1 \geq m_2$". We then have for any such $f(v)$,

$$\bar{r}(f \mid \omega) = \begin{cases} P(f(v) = 1 \mid \omega) = E[f \mid \omega] & \text{if } m_1 < m_2, \\ P(f(v) = 0 \mid \omega) = E[1 - f \mid \omega] & \text{if } m_1 > m_2, \\ 0 & \text{if } m_1 = m_2 \end{cases}, \tag{6'}$$

and by comparison with (1) we see that $r(f \mid \omega) = \mid m_1 - m_2 \mid \bar{r}(f \mid \omega)$ for all ω. Now, in the three examples (i), (ii), (iii) cited above the unique minimax decision functions happen to take on only the values 0 and 1, and $\mid m_1 - m_2 \mid$ is constant on each of the respective parameter sets. It follows that (2), (3), (4) are also the unique minimax decision functions relative to (6) and to Ω', Ω'', Ω''' respectively. The remarks above following Example (iii) also remain valid for the risk function (6).

We conclude this section with a remark on the methods of this paper. Any decision function relevant to (6) is equivalent to a test of the hypothesis $H_0(m_1 < m_2)$ against the alternative $H_1(m_1 > m_2)$, the region $\{v : f(v) = 1\}$ being the "critical region." Hence the Neyman-Pearson probability ratio method can be used to obtain the unique minimax decision function with respect to (6) and an Ω consisting of two (or more) points, and the result carries over to more general types of Ω in the manner already indicated. It turns out, however, that the dominant properties of the probability ratio tests are not confined to the class of tests alone, but extend to the class of all functions $f(v)$ such that $0 \leq f(v) \leq 1$. This result (Theorem 1) enables us to solve the problem of the greater mean for the risk function (1) as well as for (6). The reader who is interested in applications may turn to Section 3.

2. Theorems. We require the following slight generalization of a well-known result of Neyman and Pearson [2].

THEOREM 1. *Let* $\phi(v)$, $\phi_1(v)$, $\phi_2(v)$, $\cdots$, $\phi_r(v)$ *be summable functions defined on a measure space* E *with points* v *and measure* μ, $\mu(E) \leq \infty$, *let* $c_1, \cdots, c_r$ *be arbitrary constants, and let* $A \subseteq E$ *be such that*

$$\begin{cases} v \in A \text{ implies } \phi(v) \geq \sum_1^r c_i\phi_i(v), \\ v \in E - A \text{ implies } \phi(v) \leq \sum_1^r c_i\phi_i(v). \end{cases} \tag{7}$$

Set

$$\int_A \phi_i \, d\mu = a_i \qquad (i = 1, \cdots, r), \tag{8}$$

and let $f(v)$ *be any measurable function such that*

$$0 \leq f(v) \leq 1 \tag{9}$$

and such that

$$\int_E f\phi_i \, d\mu = a_i \qquad (i = 1, \cdots, r). \tag{10}$$

Then

$$\int_E f\phi \, d\mu \leq \int_A \phi \, d\mu. \tag{11}$$

PROOF.

$$\begin{aligned} \int_E f\phi \, d\mu &= \int_A f\phi \, d\mu + \int_{E-A} f\phi \, d\mu \\ &\leq \int_A f\phi \, d\mu + \int_{E-A} f\left(\sum_1^r c_i\phi_i\right) d\mu && \text{by (9), (7),} \\ &= \int_A f\phi \, d\mu + \sum_1^r c_i \int_{E-A} f\phi_i \, d\mu \\ &= \int_A f\phi \, d\mu + \sum_1^r c_i \left[\int_E f\phi_i \, d\mu - \int_A f\phi_i \, d\mu\right] \\ &= \int_A f\phi \, d\mu + \sum_1^r c_i \left[a_i - \int_A f\phi_i \, d\mu\right] && \text{by (10),} \\ &= \int_A f\phi \, d\mu + \sum_1^r c_i \left[\int_A (1 - f)\phi_i \, d\mu\right] && \text{by (8),} \\ &= \int_A \phi \, d\mu - \int_A (1 - f)\phi \, d\mu + \int_A (1 - f)\left(\sum_1^r c_i\phi_i\right) d\mu \\ &= \int_A \phi \, d\mu + \int_A (1 - f)\left(\sum_1^r c_i\phi_i - \phi\right) d\mu \\ &\leq \int_A \phi \, d\mu && \text{by (9), (7).} \end{aligned}$$

NOTE 1. *If the condition*

$$\mu\left\{v:\phi(v) = \sum_1^r c_i\phi_i(v)\right\} = 0 \tag{12}$$

holds, then in order that the equality hold in (11) *it is necessary and sufficient that*

$$f(v) = \chi_A(v) \qquad a.e.\ (\mu), \tag{13}$$

where $\chi_A(v)$ *is the characteristic function of the set* A,

$$\chi_A(v) = \begin{cases} 1 & \text{if } v \in A, \\ 0 & \text{if } v \in E - A. \end{cases}$$

PROOF. The sufficiency is obvious. To prove the necessity we observe from the proof of Theorem 1 that for equality to hold in (11) it is necessary that

$$f(v)\left(\phi\ (v) - \sum_1^r c_i\phi_i(v)\right) = 0 \qquad \text{a.e. } (\mu) \text{ in } E-A,$$

and that

$$(1 - f(v))\left(\phi(v) - \sum_1^r c_i\phi_i(v)\right) = 0 \qquad \text{a.e. } (\mu) \text{ in } A.$$

These relations and (12) imply (13).

NOTE 2. *If relations* (10) *are replaced by*

$$\int_E f\phi_i\, d\mu \leq a_i \qquad (i = 1, \cdots, r), \tag{10'}$$

and if each of the constants c_i *is non-negative, then Theorem* 1 *and Note* 1 *remain valid.*

Theorem 1 has applications to a number of decision problems of a certain type. In the present paper we consider only the "problem of the greater mean" for two normal populations with a common variance σ^2, where at least one of the means m_1, m_2 is unknown. The following assumptions and definitions will be valid henceforth.

(A) E_N is the $N = n_1 + n_2$ dimensional sample space of points $v = (x_{11}, x_{12}, \cdots, x_{1n_1}; x_{21}, x_{22}, \cdots, x_{2n_2})$. A measurable function $f(v)$ defined for all v in E_N is a *decision function* if $0 \leq f(v) \leq 1$. $f_1(v) \equiv f_2(v)$ means $f_1(v) = f_2(v)$ for almost every v in E_N.

(B) Ω is a given set of points $\omega = (m_1, m_2 : \sigma)$, $\sigma > 0$. Given ω in Ω, the probability measure in E_N is that generated by the distribution function

$$K(v \mid \omega) = \prod_{i=1}^{2}\prod_{j=1}^{n_i} G\left[(x_{ij} - m_i)/\sigma\right],$$

where

$$G(x) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{x} e^{-u^2/2}\,du.$$

Given any function $\phi = \phi(v)$ for which the integral exists we write

$$E[\phi \mid \omega] = \int_{E_N} \phi(v) \, dK(v \mid \omega).$$

(C) Let $\gamma(\omega) = (g_1, g_2)$ be a function defined for all ω in Ω, with values in E_2, and such that

$$m_i \leq m_j \text{ implies } g_i \leq g_j \qquad (i, j = 1, 2). \tag{14}$$

Given p, $0 \leq p \leq 1$, we define

$$W(\omega, p) = \max[g_1, g_2] - g_1 p - g_2(1 - p),$$

and given a decision function $f(v)$ we define the *risk function*

$$\begin{aligned} r(f \mid \omega) &= E[W(\omega, f) \mid \omega] = W(\omega, E[f \mid \omega]) \\ &= \max[g_1, g_2] - g_1 E[f \mid \omega] - g_2 E[1 - f \mid \omega]. \end{aligned} \tag{15}$$

The class of risk functions (15) corresponding to all functions $\gamma(\omega)$ which satisfy (14) is denoted by R. (The two most important members of R are (1), with

$$\gamma(\omega) = (m_1, m_2),$$

and (6), with

$$\gamma(\omega) = \begin{cases} (0, 1) & \text{if } m_1 < m_2, \\ (1, 0) & \text{if } m_1 > m_2, \\ (0, 0) & \text{if } m_1 = m_2. \end{cases}$$

The risk functions (1) and (6) appear in the examples in Section 3.) Throughout this section $r(f \mid \omega)$ will denote a fixed but arbitrary member of R. We shall use the notations

$$h(\omega) = | g_1 - g_2 |,$$

$$d(\omega) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} (m_1 - m_2)/\sigma,$$

$$\bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij} \qquad (i = 1, 2).$$

THEOREM 2. *Let* $\omega_1 = (m_1, m_2 : \sigma)$ *and* $\omega_2 = (\mu_1, \mu_2 : \sigma)$ *be two parameter points such that*

$$d(\omega_1) < 0, \qquad d(\omega_2) > 0, \qquad h(\omega_1)h(\omega_2) > 0.$$

For any λ, $-\infty \leq \lambda \leq \infty$, *let* $f_\lambda(v)$ *be the characteristic function of the set*

$$A_\lambda = \{v : n_1(\mu_1 - m_1)\bar{x}_1 + n_2(\mu_2 - m_2)\bar{x}_2 > \lambda\sigma\}. \tag{16}$$

Then

(i) *Corresponding to any decision function* $f(v)$, *there exists a* λ *such that*

$$r(f_\lambda \mid \omega_1) = r(f \mid \omega_1), \qquad r(f_\lambda \mid \omega_2) \leq r(f \mid \omega_2);$$

the inequality is strict unless $f(v) \equiv f_\lambda(v)$.

(ii) *Given any* λ, *if* $f(v)$ *is a decision function such that*

$$r(f \mid \omega_i) \leq r(f_\lambda \mid \omega_i) \qquad (i = 1, 2),$$

then

$$f(v) \equiv f_\lambda(v).$$

(iii) *There exists a unique* c *such that*

$$r(f_c \mid \omega_1) = r(f_c \mid \omega_2) = B \text{ say,} \tag{17}$$

and for any decision function $f(v)$ *we have*

$$B \leq \max\,[r(f \mid \omega_1), r(f \mid \omega_2)]; \tag{18}$$

the inequality is strict unless $f(v) \equiv f_c(v)$. *It follows that* $f_c(v)$ *is the unique minimax decision function corresponding to the two-point parameter space* $\Omega = (\omega_1, \omega_2)$.

PROOF.[3] (a) Let $\phi(v)$, $\phi_1(v)$ be the joint frequency functions of the sample point v corresponding to the parameter points ω_2, ω_1 respectively. It is readily seen that for any λ there exists a unique constant $c_1(\lambda)$, $0 \leq c_1(\lambda) \leq \infty$, such that

$$A_\lambda = \{v : \phi(v) > c_1\phi_1(v)\}$$

$(c_1(-\infty) = 0, c_1(\infty) = \infty)$. Moreover, since $\omega_1 \neq \omega_2$,

$$\mu\{v : \phi(v) = c_1\phi_1(v)\} = 0.$$

It follows from Theorem 1, Note 2, that if $f(v)$ is any decision function such that

$$E[f \mid \omega_1] \leq E[f_\lambda \mid \omega_1],$$

then

$$E[f \mid \omega_2] \leq E[f_\lambda \mid \omega_2],$$

and the strict inequality holds unless $f(v) \equiv f_\lambda(v)$.

(b) It is clear from the definition (16) that for any fixed parameter point ω the function

$$E[f_\lambda \mid \omega] = P(A_\lambda \mid \omega)$$

is continuous and strictly decreasing from 1 to 0 as λ varies from $-\infty$ to $+\infty$.

(c) For any decision function $f(v)$ and any parameter point ω we have by (C),

$$r(f \mid \omega) = \max\,[g_1, g_2] - g_1E[f \mid \omega] - g_2E[1 - f \mid \omega].$$

Hence

$$\begin{cases} r(f \mid \omega_1) = h(\omega_1)E[f \mid \omega_1], & h(\omega_1) > 0, \\ r(f \mid \omega_2) = h(\omega_2)E[1 - f \mid \omega_2], & h(\omega_2) > 0. \end{cases} \tag{19}$$

[3] Theorem 2 (as also Example (iii) of Section 1) could be derived from Wald's general results on the completeness of the class of Bayes solutions of statistical decision problems.

Since for any decision function $f(v)$, $0 \leq E[f \mid \omega_1] \leq 1$, we can by (b) choose λ so that

$$E[f_\lambda \mid \omega_1] = E[f \mid \omega_1], \tag{20}$$

and by (a) it follows that unless $f(v) \equiv f_\lambda(v)$,

$$E[f_\lambda \mid \omega_2] > E[f \mid \omega_2]. \tag{21}$$

(i). Follows from (19), (20) and (21).
(ii). Follows from (19) and (a).
(iii). (17) follows from (19) and (b). Then (18) follows from (17) and (ii).

Theorem 2 provides the solution of any problem of the greater mean when Ω consists of just two points ω_1, ω_2. For, the problem is trivial unless $d(\omega_1)\, d(\omega_2) < 0$ and $h(\omega_1)h(\omega_2) > 0$, and in the non-trivial case the unique minimax decision function is $f_c(v)$ defined by (17). Moreover, it follows at once from the definition that if $\bar{f}(v)$ is the unique minimax decision function with respect to some parameter set $\bar{\Omega}$, then it remains so with respect to any Ω such that $\Omega \supseteq \bar{\Omega}$ and

$$\sup_{\omega \in \Omega} [r(\bar{f} \mid \omega)] = \sup_{\omega \in \bar{\Omega}} [r(\bar{f} \mid \omega)].$$

By taking sets $\bar{\Omega}$ which consist of two points, Theorem 2 can therefore be used to obtain sufficient conditions for an $\bar{f}(v) = f_c(v)$ to be the unique minimax decision function with respect to a quite general Ω. (It is clear that results analogous to Theorem 2(iii) but pertaining to more than two parameter points can be derived from Theorem 1, and that these results can be exploited in a similar way. An instance of this procedure where $\bar{\Omega}$ consists of three points will be given at the end of this section.)

The theorems which follow exploit Theorem 2 in this way to obtain conditions on Ω under which the decision functions $f^0(v)$ and $f_c^0(v)$ defined by (5) and (3) are minimax. We consider $f^0(v)$ first. From (C) we have, after a simple computation,

$$r(f^0 \mid \omega) = h(\omega) \cdot G(-\mid d(\omega) \mid). \tag{22}$$

THEOREM 3. *Suppose that there exist sequences* $\{\omega_k\}$, $\{\omega_k'\}$ *of points* $\omega_k = (m_{1k}, m_{2k} : \sigma_k)$, $\omega_k' = (\mu_{1k}, \mu_{2k} : \sigma_k)$ *in* Ω *such that*

(i) $\lim_{k\to\infty} r(f^0 \mid \omega_k) = \sup_{\omega \in \Omega} [r(f^0 \mid \omega)] \qquad (\neq 0, \infty)$,

(ii) $d(\omega_k) = -d(\omega_k')$, $h(\omega_k) = h(\omega_k')$, *and* $n_1 m_{1k} + n_2 m_{2k} = n_1 \mu_{1k} + n_2 \mu_{2k}$ *for every* $k = 1, 2, \cdots$.

Then $f^0(v)$ *is an admissible minimax decision function. If there exist* $\omega_0 = (m_1, m_2 : \sigma)$, $\omega_0' = (\mu_1, \mu_2 : \sigma)$ *in* Ω *satisfying* (i) *and* (ii), *then* $f^0(v)$ *is the unique minimax decision function.*

PROOF. By (22) and (ii),

$$r(f^0 \mid \omega_k) = r(f^0 \mid \omega_k') \text{ for every } k. \tag{23}$$

Without loss of generality, we may assume the two sequences to be so chosen that $h(\omega_k) = h(\omega_k') > 0$ for every k. Then, by interchanging corresponding members if necessary, we may assume that

$$d(\omega_k) = -d(\omega_k') < 0 \text{ for every } k. \tag{24}$$

Consider the two points ω_k, ω_k' in Ω with arbitrary but fixed k. Writing ω_k, ω_k' for ω_1, ω_2 respectively, and using conditions (ii), a simple calculation shows that the set defined by (16) is

$$A_\lambda = \{v:\bar{x}_1 - \bar{x}_2 > L\}, \tag{25}$$

L being a strictly increasing function of λ.

Choose and fix an arbitrary decision function $f(v) \not\equiv f^0(v)$. Comparing (5) and (25), it follows from Theorem 2(iii) and (23) that

$$r(f^0 \mid \omega_k) = r(f^0 \mid \omega_k') < \max\,[r(f \mid \omega_k), r(f \mid \omega_k')]. \tag{26}$$

Clearly, $f(v)$ cannot be uniformly better than $f^0(v)$ in Ω. Again, from (26),

$$r(f^0 \mid \omega_k) < \sup_{\omega \in \Omega}\,[r(f \mid \omega)], \tag{27}$$

so that, since k is arbitrary,

$$\sup_{\omega \in \Omega}[r(f^0 \mid \omega)] = \lim_{k \to \infty} r(f^0 \mid \omega_k) \leq \sup_{\omega \in \Omega}\,[r(f \mid \omega)]. \tag{28}$$

Since $f(v) \not\equiv f^0(v)$ in the preceding argument is arbitrary, we have shown that (a) no $f(v)$ can be uniformly better than $f^0(v)$ and (b) $\sup_\omega\,[r(f^0 \mid \omega)] = \inf_f \sup_\omega\,[r(f \mid \omega)]$, i.e. that $f^0(v)$ is admissible and minimax. The last part of the theorem follows upon setting $\omega_k = \omega_0$ in (27). This completes the proof of Theorem 3.

The conditions on Ω for $f^0(v)$ to be the unique minimax decision function may be written as follows:

There exist $\omega_0 = (m_1, m_2 : \sigma)$, $\omega_0' = (\mu_1, \mu_2 : \sigma)$ *in* Ω *such that*

$$\text{(i) } r(f^0 \mid \omega_0)(= r(f^0 \mid \omega_0')) = \sup_{\omega \in \Omega}\,[r(f^0 \mid \omega)] \qquad (\neq 0, \infty),$$

$$\text{(ii) } \mu_1 = m_2 + \left(\frac{n_1 - n_2}{n_1 + n_2}\right)(m_1 - m_2), \qquad \mu_2 = m_1 + \left(\frac{n_1 - n_2}{n_1 + n_2}\right)(m_1 - m_2), \tag{29}$$

$$\text{(iii) } h(\omega_0) = h(\omega_0').$$

For the important risk functions (1) and (6), (29)(ii) implies (29)(iii) (i.e. $h(\omega)$ depends on $|m_1 - m_2|$ alone). Moreover, when $n_1 = n_2$, (29)(ii) becomes $\mu_1 = m_2$, $\mu_2 = m_1$. Thus for (1) and (6), when $n_1 = n_2$ the conditions (29) reduce simply to the condition that *at least two points in* Ω *at which the risk for* $f^0(v)$ *is maximum be image points of one another in the plane* $\{\omega: m_1 = m_2\}$. In particular, it follows that if $n_1 = n_2$ and if the given set Ω is "symmetric" in the sense that whenever $(m_1, m_2 : \sigma)$ is in Ω then $(m_2, m_1 : \sigma)$ is also in Ω, then $f^0(v)$ is the unique minimax decision function provided that it attains its maximum risk in Ω, the risk function in question beging (1) or (6). There are obvious modifications (involving two sequences of points in Ω) of these remarks which assert that $f^0(v)$ is at least an admissible minimax decision function in case $f^0(v)$ does not attain its maximum risk in Ω.

We shall now state the result analogous to Theorem 3 for the case when one of the means is known exactly, say $m_2 = c$. The decision function $f_c^0(v)$ is defined by (3).

THEOREM 4. *Suppose that there exist sequences* $\{\omega_k\}$, $\{\omega_k'\}$ *of points* $\omega_k = (c + a_k, c: \sigma_k)$, $\omega_k' = (c - a_k, c: \sigma_k)$ *in* Ω *such that*

(i) $\lim_{k\to\infty} r(f_c^0 \mid \omega_k) = \sup_{\omega\in\Omega} [r(f_c^0 \mid \omega)]. \qquad (\neq 0, \infty)$

(ii) $h(\omega_k) = h(\omega_k')$ *for every* $k = 1, 2, \cdots$.

Then $f_c^0(v)$ *is an admissible minimax decision function. If there exist* $\omega_0 = (c + a, c: \sigma)$, $\omega_0' = (c - a, c: \sigma)$ *in* Ω *satisfying* (i) *and* (ii), *then* $f_c^0(v)$ *is the unique minimax decision function.*

The proof (based on Theorem 2(iii)) is similar to that of Theorem 3 and will be omitted. Note that for the risk functions (1) and (6), condition (ii) is automatically satisfied.

The reader will have observed that results which may be obtained from Theorem 2(iii) in the manner of Theorems 3 and 4 will assert the optimal character of decision functions which are characteristic functions of sets of the type $\{v: a\bar{x}_1 + b\bar{x}_2 > c\}$. The following example, cited as Example (iii) of Section 1, shows that for arbitrary Ω the optimum decision function need not be of this type.

Suppose that $n_1 = n_2 = n$, that $\bar{\Omega}$ consists of the three points

$$\omega_0 = (\tfrac{1}{2}, -\tfrac{1}{2}: 1), \omega_1 = (\tfrac{1}{2}, \tfrac{3}{2}: 1), \omega_2 = (-\tfrac{3}{2}, -\tfrac{1}{2}: 1),$$

and that the risk function under consideration is given by (1) or (6). Then the unique minimax decision function is $f^{**}(v)$ given by (4), where $\lambda > 0$ is determined by

$$E[1 - f^{**} \mid \omega_0] = E[f^{**} \mid \omega_1]. \tag{30}$$

The proof follows. $f^{**}(v)$ is the characteristic function of the set $\{v: \phi(v) > c_1\phi_1(v) + c_2\phi_2(v)\}$, where ϕ, ϕ_1, ϕ_2 are the frequency functions of the probability distributions in E_{2n} corresponding to the parameter points ω_0, ω_1, ω_2 respectively, with $c_1 = c_2 = e^n/\lambda$. Since for all $\lambda > 0$,

$$E[f^{**} \mid \omega_1] = E[f^{**} \mid \omega_2],$$

and since a unique $\lambda > 0$ satisfying (30) certainly exists, it follows (cf. (19) and (C)) that

$$r(f^{**} \mid \omega_0) = r(f^{**} \mid \omega_1) = r(f^{**} \mid \omega_2) = B,$$

say. Let $f(v)$ be any decision function $\not\equiv f^{**}(v)$. We shall show that

$$B < \max\,[r(f \mid \omega_0), r(f \mid \omega_1), r(f \mid \omega_2)]. \tag{31}$$

Suppose not. Then

$$r(f \mid \omega_1) = E[f \mid \omega_1] \leq E[f^{**} \mid \omega_1] = r(f^{**} \mid \omega_1),$$

$$r(f \mid \omega_2) = E[f \mid \omega_2] \leq E[f^{**} \mid \omega_2] = r(f^{**} \mid \omega_2).$$

Then, by Theorem 1, Note 2, we must have $E[f \mid \omega_0] < E[f^{**} \mid \omega_0]$, so that

$$r(f \mid \omega_0) = 1 - E[f \mid \omega_0] > 1 - E[f^{**} \mid \omega_0] = r(f^{**} \mid \omega_0) = B,$$

contrary to hypothesis. Hence (31) holds, and since $f(v) \not\equiv f^{**}(v)$ is arbitrary our assertion is proved. (Note that

$$r(f^0 \mid \omega_0) = r(f^0 \mid \omega_1) = r(f^0 \mid \omega_2)$$

also, so that $f^{**}(v)$ is uniformly better than $f^0(v)$ in $\bar{\Omega}$.) We remind the reader that $f^{**}(v)$ remains the unique minimax decision function with respect to (1) or (6) and any Ω which contains ω_0, ω_1, ω_2, and is such that $\sup_{\omega \in \Omega} [r(f^{**} \mid \omega)] = B$. Whether a set Ω satisfies the last condition will in general depend on whether the risk function in question is (1) or (6).

3. Examples and discussion. In this section we shall discuss the relevance of Theorems 3 and 4 to two specific problems of the greater mean. The examples given are purely illustrative and the reader will readily construct others in which the statistician is faced with similar problems of decision.

EXAMPLE 1. A farmer F has tested two varieties π_1, π_2 of grain in a field experiment in which n_i plots were assigned to π_i, $i = 1, 2$, all plots being of equal area. The plot yields obtained were $y_{11}, y_{12}, \cdots, y_{1n_1}$ and $y_{21}, y_{22}, \cdots, y_{2n_2}$ bushels respectively. F gives this data to a statistician S for analysis. F is willing to assume that the yields per plot for each of the two varieties are normally distributed with unknown means μ_1, μ_2 and a common variance, also unknown. F says he is particularly interested in whether the two varieties are "significantly different."

S is well aware that F's interest in the varieties is not purely scientific—that is to say, F did not perform the field experiment for the sole purpose of estimating the unknown parameters or testing hypotheses concerning them. S also knows that it is very unlikely that μ_1 *is* equal to μ_2.

Suppose that in fact F wishes to decide which variety he should use next year on his land in order to make the maximum possible profit, and is afraid that if he were to act as if the observed mean yields $\bar{y}_1$, $\bar{y}_2$ were the true population mean yields, he might make a gross error. So F is willing to compromise between the two varieties (that is, he will assign some fraction f of his land to π_1 and the rest to π_2) in case S declares that there is no evidence of the two varieties being different.

If this is the case, S should ask F how much it costs him to use π_i and the price at which he expects to sell his grain. Supposing that these quantities are a_i dollars per acre and b dollars per bushel respectively, and that the area of each plot in the field experiment was c acres, S will set

$$m_i = \text{expected profit per acre in using variety } \pi_i$$

$$= (b/c)\mu_i - a_i \quad \text{dollars} \qquad (i = 1, 2),$$

$$\omega = (m_1, m_2 : \sigma), \ \sigma^2 \text{ being the variance of the profit per acre in using } \pi_i \qquad (i = 1, 2),$$

$$\gamma(\omega) = (m_1, m_2) \qquad \text{(see Section 2, (C))},$$

$$x_{ij} = (b/c)y_{ij} - a_i, \ \bar{x}_i = n_i^{-1} \cdot \sum_{j=1}^{n_i} x_{ij}, \qquad v = (x_{11}, \cdots, x_{1n_1}; x_{21}, \cdots, x_{2n_2}),$$

so that $r(f \mid \omega)$ is given by (1) and is equal to the expected loss (in terms or profit per acre) incurred by using the proportions $f(v)$, $1 - f(v)$ of the varieties π_1, π_2 as compared with using the variety with the greater mean for the whole of the land. Then if S is satisfied that the set Ω of possible points ω satisfies the conditions of Theorem 3 he should recommend that F use π_1 alone if $\bar{x}_1 > \bar{x}_2$, and π_2 alone if $\bar{x}_2 > \bar{x}_1$, this being the safest procedure in the sense that it is the minimax strategy (cf. Example 1 in [3]).

We shall illustrate by a simple example the obvious method of verifying whether $f^0(v)$ is the minimax decision function for a given Ω. We have by (22), using the risk function (1) obtained by setting $\gamma(\omega) = (m_1, m_2)$,

$$r(f^0 \mid \omega) = h(\omega)G(- \mid d(\omega) \mid) = \mid m_1 - m_2 \mid G\left(-\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \mid m_1 - m_2 \mid /\sigma\right). \tag{32}$$

Now suppose that

$$\Omega = \left\{\omega: a - \frac{l}{2} \leq m_1 \leq a + \frac{l}{2}, \ b - \frac{l}{2} \leq m_2 \leq b + \frac{l}{2} : \sigma_0 - \rho \leq \sigma \leq \sigma_0\right\}, \qquad l > \mid a - b \mid, \tag{33}$$

where a, b, l, σ_0, $\rho(\geq 0)$ are certain constants. By (32), the maximum risk occurs at some points in Ω for which $\sigma = \sigma_0$. We have

$$r(f^0 \mid \sigma = \sigma_0) = \sigma_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}} \cdot [xG(-x)], \tag{34}$$

where

$$x = x(\omega) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \mid m_1 - m_2 \mid /\sigma_0 .$$

If $a = b$ and $n_1 = n_2$ we see from the remark following (29) that $f^0(v)$ is the unique minimax decision function. Suppose therefore that $a \neq b$ or $n_1 \neq n_2$ or both. Now

$$\sup_x [xG(-x)] = x_0 G(-x_0) = .1700 \text{ (approx.)}, \tag{35}$$

where $x_0 = .7518$ (approx.). If m_1, m_2 were unrestricted, $r(f^0 \mid \sigma = \sigma_0)$ would be a maximum when $|m_1 - m_2| = \sigma_0 x_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}}$, by (34) and (35). Hence $f^0(v)$ will be the unique minimax decision function if these two lines intersect the square $\left\{a - \frac{l}{2} \leq m_1 \leq a + \frac{l}{2}, b - \frac{l}{2} \leq m_2 \leq b + \frac{l}{2}\right\}$ in such a way that at least two points lying on these lines and in the square satisfy (29)(ii). This will be the case if

$$l > \max\left[|a - b| + y_0, \max(|a - b|, y_0) + \left|\frac{n_1 - n_2}{n_1 + n_2}\right| y_0\right], \tag{36}$$

where

$$y_0 = x_0 \sigma_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}}.$$

We have assumed that $l > |a - b|$, for otherwise either $m_1 \leq m_2$ or $m_1 \geq m_2$ for all ω in Ω, and there is no problem. It is therefore clear that for n_1 and n_2 sufficiently large, $f^0(v)$ will be the unique minimax decision function. That (36) is not a very strong requirement may be seen by setting $a = b$, $n_1 = 2n_2$, in which case (36) reduces to

$$l > \sigma_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}} \qquad \text{(approx.)}.$$

We remark that $f^0(v)$ remains the unique minimax decision function for any n_1, n_2 "when $l = \infty$" so that Ω is given by

$$\Omega = \{\omega: -\infty < m_1 < \infty, -\infty < m_2 < \infty : \sigma_0 - \rho \leq \sigma \leq \sigma_0\}. \tag{33$'$}$$

It is of interest to consider the "one sample" case when one of the means is known, say $m_2 = c$. This will be the case (approximately) if π_2 is a standard variety which has been in use for some time and π_1 is a new variety. The analogue of the parameter space discussed above is then

$$\Omega = \left\{\omega: m_2 = c, a - \frac{l}{2} \leq m_1 \leq a + \frac{l}{2} : \sigma_0 - \rho \leq \sigma \leq \sigma_0\right\}, \quad \frac{l}{2} > |a - c|. \tag{37}$$

By using Theorem 4 it can be seen that $f_c^0(v)$ as defined by (3) is the unique minimax decision function if $c = a$ or if c is not necessarily equal to a, but

$$\frac{l}{2} - |a - c| > \sigma_0 x_0 \left(\frac{1}{n_1}\right)^{\frac{1}{2}}, \tag{38}$$

where x_0 is given by (35). Since the left-hand side of (38) is positive, it is clear that $f_c^0(v)$ will be the unique minimax decision function with respect to (37) if

n_1 is sufficiently large. Note that $f_c^0(v)$ is the unique minimax decision function for any n_1 when $l = \infty$ and Ω is given by

$$(37') \qquad \Omega = \{\omega : m_2 = c, -\infty < m_1 < \infty : \sigma_0 - \rho \le \sigma \le \sigma_0\}.$$

The reader may find it instructive to consider other plausible sets Ω which satisfy the conditions of Theorems 3 and 4 and also some which do not, assuming $\sigma = 1$ for simplicity. It should be observed that no matter what Ω may be, provided only that $\sigma \le \sigma_0$ for all ω in Ω, we shall have by (32) and (35)

$$\sup_{\omega \in \Omega} [r(f^0 \mid \omega)] \le .1700 \cdot \sigma_0 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}} \qquad \text{(approx.)}.$$

In a similar way it can be seen that for any Ω in which m_2 equals c and $\sigma \le \sigma_0$,

$$\sup_{\omega \in \Omega} [r(f_c^0 \mid \omega)] \le .1700 \cdot \sigma_0 \cdot \left(\frac{1}{n_1}\right)^{\frac{1}{2}} \qquad \text{(approx.)}.$$

Example 2. π_1 and π_2 are two soporific drugs, the random variables generated by them being the duration of sleep induced by a standard dose in an individual chosen at random. It is assumed that these two populations are normal with unknown means m_1, m_2 and a common variance σ^2, also unknown. In a series of independent trials in which n_1 individuals received the first drug and n_2 the second, the outcome was $v = (x_{11}, x_{12}, \cdots, x_{1n_1}; x_{21}, x_{22}, \cdots, x_{2n_2})$. The statistician S is required to say which is the more effective drug.

Here a reasonable risk function is (6), where $f(v)$ takes on only the values 0, 1, corresponding to the decisions "$m_1 \le m_2$" and "$m_1 \ge m_2$" respectively.[4] The problem of choosing $f(v)$ so as to minimize this risk was considered by Simon [4]. He showed that in case $n_1 = n_2$, $f^0(v)$ is the *uniformly best* decision function in the class of symmetric decision functions. (Given $n_1 = n_2 = n$, a decision function $f(v)$ is said to be symmetric if $f(x_{11}, x_{12}, \cdots, x_{1n}; x_{21}, x_{22}, \cdots, x_{2n}) \equiv 1 - f(x_{21}, x_{22}, \cdots, x_{2n}; x_{11}, x_{12}, \cdots, x_{1n})$. See also [3].) It is natural to confine oneself to the class of symmetric decision functions when the sample sizes are equal, but under the implicit assumption that if $\omega = (a, b: \sigma)$ is a possible parameter point, then $\omega' = (b, a: \sigma)$ is also (cf. the remarks following (29)). The illustrations in Section 1 show that if the sample sizes are unequal or if Ω is not symmetric in the sense just described, there may exist decision functions which are *uniformly better* than $f^0(v)$: in (i) we have a "symmetric" Ω but $n_1 \ne n_2$; in (iii), $n_1 = n_2$ but Ω is not "symmetric."

However, $f^0(v)$ is an admissible minimax decision function no matter what the sample sizes, provided only that Ω satisfies a certain not too restrictive condition. We have

$$(39) \qquad \bar{r}(f^0 \mid \omega) = \begin{cases} G(-\mid d(\omega) \mid) & \text{for } m_1 \ne m_2, \\ 0 & \text{for } m_1 = m_2. \end{cases}$$

[4] For some purposes it would be more appropriate to take (1) as the risk function for this problem, letting the decision functions $f(v)$ take on only the values 0 and 1. We have (essentially) discussed this case in the previous example.

It is clear that if $\{\omega_k\}$ is a sequence of points in Ω such that

$$\lim_{k\to\infty} d(\omega_k) = 0, \quad \text{then} \quad \lim_{k\to\infty} \bar{r}(f^0 \mid \omega_k) = \frac{1}{2} = \sup_{\omega\in\Omega} [\bar{r}(f^0 \mid \omega)].$$

Therefore, by Theorem 3, $f^0(v)$ is admissible and minimax if some point in the plane $\{\omega: m_1 = m_2\}$ is an interior point of the set Ω of possible parameter points (in fact it is sufficient if some plane $\sigma = \sigma_0(>0)$ intersects Ω in a set which has an interior point on the line $m_1 = m_2$). Hence if nothing much is known about the two drugs, S could regard the foregoing as a justification for asserting "$m_1 \geq m_2$" if $\bar{x}_1 > \bar{x}_2$ and "$m_1 \leq m_2$" otherwise.

We have given no criterion for the choice of a suitable decision function when two or more admissible minimax decision functions exist, and our diffidence in recommending the use of $f^0(v)$ in the present case is due to the fact that under the condition stated above there will exist decision functions other than $f^0(v)$ which are also admissible and minimax with respect to (6). Let us suppose that Ω is given by (33). Then $f^0(v)$ is admissible and minimax, by the preceding paragraph. However, it follows from Theorem 4 that each of

$$f^0_{c_1}(v) = \begin{cases} 1 \text{ if } \bar{x}_1 > c_1, \\ 0 \text{ otherwise,} \end{cases} \quad \text{and} \quad f^{c_2}_0(v) = \begin{cases} 0 \text{ if } \bar{x}_2 > c_2, \\ 1 \text{ otherwise,} \end{cases}$$

is also admissible and minimax, where c_1 and c_2 are arbitrary constants with $\max\,[a, b] - \frac{l}{2} \leq c_1, c_2 \leq \min\,[a, b] + \frac{l}{2}$.

There is, however, some reason for preferring $f^0(v)$ to other decision functions in the present case. S has been asked to give his opinion as to which is the better drug, and presumably no immediate consequences follow from the opinion which he might express. (This would not be the case if there were a sleepless individual on hand who had to be given a dose of one of the two drugs. Cf. footnote 4.) Although the problem *is* of a scientific nature, insistence upon literal exactitude in the interpretation of "incorrect decision" is meaningful only insofar as it is compatible with the physical situation. In view of the limited determinacy of unknown parameters in general, and of the limitations of experiments on soporific drugs in particular, it may be possible and even desirable to modify (6) in such a way that for any fixed σ the risk tends to zero with $|m_1 - m_2|$. Thus modified, the risk function would be essentially similar to (1). A rather drastic way of introducing this modification would be to agree that the assertion of equality of the two means does not constitute an error in case $|m_1 - m_2| < \epsilon$, where ϵ is some positive constant. S will then take

$$\bar{r}_\epsilon(f \mid \omega) = \begin{cases} \bar{r}(f \mid \omega) \text{ if } |m_1 - m_2| \geq \epsilon, \\ 0 \text{ otherwise,} \end{cases} \tag{40}$$

as the risk function. (Note that in using $\bar{r}_\epsilon(f \mid \omega)$ rather than $\bar{r}(f \mid \omega)$, S has in effect deleted the set $\{\omega: |m_1 - m_2| < \epsilon\}$ from the given set Ω by defining $\gamma(\omega) =$

(0, 0) there, instead of only when $m_1 = m_2$ as in the case of $\bar{r}(f \mid \omega)$. Cf. "zones of indifference," [5, pp. 27–30]). It follows from Theorem 3 that $f^0(v)$ is the unique minimax decision function with respect to (40) and (33) if $a = b$ and $n_1 = n_2$ and also if at least one of these conditions does not hold but

$$l > \max \left[\mid a - b \mid + \epsilon, \max (\mid a - b \mid, \epsilon) + \left| \frac{n_1 - n_2}{n_1 + n_2} \right| \epsilon \right].$$

Thus $f^0(v)$ will be the unique minimax decision function no matter what n_1, n_2, a, b or l may be, provided only that ϵ is sufficiently small. We shall leave other modifications of $\bar{r}(f \mid \omega)$ and discussion of $\bar{r}(f \mid \omega)$ with respect to other types of parameter spaces (e.g. (37)) to the reader.

We conclude this discussion with a remark on the proper choice of n_1 and n_2 in using $f^0(v)$ when the risk function belongs to the class R defined in Section 2, (C). (The risk functions (1) and (6) belong to R.) Suppose that before experimentation starts, it is agreed that one must have $n_1 + n_2 = 2k$, where k is a fixed integer. In that case, choosing $n_1 = n_2 = k$ will be the best choice of n_1, n_2 in the following sense. (a) For any fixed ω, $r(f^0 \mid \omega)$, which is the expected loss, then becomes a minimum. This follows immediately from (22), since

$$r(f^0 \mid \omega) = h(\omega)G(- \mid d(\omega) \mid), \qquad \mid d(\omega) \mid = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \mid m_1 - m_2 \mid / \sigma,$$

and $\mid d(\omega) \mid$ has its maximum when $n_1 = n_2 = k$. (b) For any fixed ω, the *variance* of the loss also becomes a minimum. In using $f^0(v)$, the loss takes the values 0 and $h(\omega)$ only, with $P(\text{loss} = h(\omega) \mid \omega) = G(- \mid d(\omega) \mid) = \alpha$ say. Therefore, the variance of the loss is $h^2\alpha(1 - \alpha)$. Since $\alpha \leq \frac{1}{2}$, this expression increases with increasing α, and so has its minimum when $n_1 = n_2 = k$. This remark is, of course, without prejudice to the question of whether $f^0(v)$ is admissible and minimax with respect to a given Ω for every n_1 and n_2 with $n_1 + n_2 = 2k$.

4. A remark on randomized decision functions. In the foregoing discussion we have confined attention to the class of non-randomized decision functions: the space of possible decisions being some subset of $0 \leq f \leq 1$, the statistician constructs (in advance) a suitable decision function $f(v)$, obtains a particular sample point v by sampling the two populations, and takes $f(v)$ as his decision. It is, however, of some theoretical interest to consider more general formulations in which the decision arrived at by the statistician may be a random function of the sample point v.

A randomized decision function can be defined in several ways. One definition is as follows. Let $\phi(z \mid v)$ be a function defined for all v in E_N and all real z such that for any fixed z it is a measurable function of v, and such that for any fixed v it is the distribution function of a random variable with values in $0 \leq z \leq 1$. We shall denote this random variable by $Z_\phi(v)$ and call it a (randomized) decision function. In using it, the statistician first obtains a particular point v by sampling the two populations, then performs a random experiment whose outcome Z

has the known distribution function $P(Z \leq z) = \phi(z \mid v)$, and takes Z as his decision. The class of all decision functions corresponding to all functions $\phi(z \mid v)$ will be denoted by $\{Z_\phi(v)\}$. It is clear that this class includes the class of non-randomized decision functions.

This definition of the structure of randomized decision functions follows the method described by Halmos and Savage in their interesting remarks ([6], pp. 239–241) on the value of sufficient statistics in statistical methodology. For any $Z_\phi(v)$, we have

$$P(Z_\phi(v) \leq z \mid \omega) = \int_{E_N} P(Z_\phi(v) \leq z \mid \omega, v)\, dK(v \mid \omega) = \int_{E_N} \phi(z \mid v)\, dK(v \mid \omega). \tag{41}$$

We shall now show that in all problems of the greater mean in which the methods of Section 2 can be applied to non-randomized decision functions, randomization cannot be recommended. More precisely, the following holds.

Theorem. *Let $\bar{f}(v)$ be a non-randomized decision function which takes on only the values 0 and 1 and which is the unique non-randomized decision function whose expected value $E[\bar{f} \mid \omega]$ satisfies a certain condition Q as a function of ω. Then $\bar{f}(v)$ is the unique decision function whose expected value satisfies the condition Q;* i.e. if *$Z_\phi(v)$ is a decision function such that $E[Z_\phi \mid \omega]$ satisfies Q, then*

$$P(\bar{f}(v) = Z_\phi(v) \mid \omega) = 1 \quad \textit{for all } \omega. \tag{42}$$

It follows in particular that Theorem 2 remains valid with the arbitrary non-randomized $f(v)$ replaced by an arbitrary $Z_\phi(v)$, and in consequence, Theorems 3 and 4 remain valid when the class of decision functions in question is $\{Z_\phi(v)\}$.

Proof. Let $Z_\phi(v)$ be a decision function whose expected value satisfies the condition Q. Now, by (41) and Theorem 5 of [7] we have

$$E[Z_\phi \mid \omega] = \int_{E_N} f^\phi(v)\, dK(v \mid \omega) = E[f^\phi \mid \omega], \tag{43}$$

where

$$f^\phi(v) = \int_0^1 z d_z\phi(z \mid v), \qquad 0 \leq f^\phi(v) \leq 1. \tag{44}$$

It is clear from (43) that $E[f^\phi \mid \omega]$ satisfies Q and so we must have

$$f^\phi(v) = \bar{f}(v) \text{ a.e.} \tag{45}$$

by hypothesis. Since $\bar{f}(v)$ takes on only the values 0 and 1, it follows from (44) and (45) that

$$\int_{\{z=\bar{f}(v)\}} d_z\,\phi(z \mid v) = 1 \text{ a.e.},$$

which implies (42). In order to verify the last part of the remark, consider any particular problem of the greater mean. The risk function of any decision function $Z_\phi(v)$ is, by (15),

$$r(Z_\phi \mid \omega) = W(\omega, E[Z_\phi \mid \omega]).$$

Hence a condition on the risk function of Z_ϕ is equivalent to a condition on $E[Z_\phi \mid \omega]$ as a function of ω, and the truth of the remark follows by appropriate definition of the condition Q in terms of the risk function.

REFERENCES

[1] A. WALD, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 165–205.

[2] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. I (1936), pp. 1–37.

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

[4] 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.

[5] A. WALD, *Sequential analysis*, John Wiley and Co., 1947.

[6] P. R. HALMOS AND L. J. SAVAGE, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 225–241.

[7] H. ROBBINS, "Mixture of distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 360–369.