ON MIXED SINGLE SAMPLE EXPERIMENTS[1]

By Leonard Cohen

College of the City of New York

1. Introduction and summary. William Kruskal [1], Howard Raiffa [2], J. L. Hodges, Jr. and E. L. Lehmann [4], have shown that in certain Neyman-Pearson type problems of testing a simple hypothesis against a simple alternative, determining the sample size by means of a chance device yields improvements over fixed sample size procedures. The purpose of this paper is not only to investigate the general problem of randomizing over fixed sample size tests of a simple hypothesis against a simple alternative, but also randomizing over other fixed sample size procedures in topics such as confidence interval estimation, the k-decision problem, etc.

In Section 2, a fixed sample size test of a simple hypothesis against a simple alternative is identified with an operating characteristic (α, β, n) where α denotes the probability of a type I error, β denotes the probability of a type II error, and n denotes the sample size. A mixed single sample test is defined as a sequence of quadruples.

$(\gamma_i, \alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^{\infty} \gamma_i = 1$, where $(\alpha_i, \beta_i, \eta_i)$ is a fixed sample size test and where γ_i is interpreted as the probability of using the fixed sample size test (α_i, β_i, n_i) for $i = 1, 2, \cdots$. A mixed single sample test is identified with an operating characteristic $(\alpha, \beta, n) = \sum_{i=1}^{\infty} \gamma_i(\alpha_i, \beta_i, n_i)$. For each non-negative integer n, the class A_n of admissible fixed sample size procedures of sample size n is defined in an obvious way. We define $A = \bigcup_{i=0}^{\infty} A_i$ and A^* as the convex hull of A. It is not necessarily true that A^* is closed. An example is given to show this. However, it is true that the lower boundary of A^* is a subset of A^* so that the lower boundary of A^* determines a minimally complete class, $\mathcal{A}$, of mixed single sample tests. The tests in $\mathcal{A}$ are characterized from a Bayes point of view and a technique for constructing the tests in $\mathcal{A}$ is given.

In Section 3, the technique is applied to tests on the mean of a normal distribution with known variance. It is shown that the tests in $\mathcal{A}$ are either

(a) fixed sample size tests, or

(b) mixtures of at most two fixed sample size tests.

It is shown that there exists a minimal subset $\mathcal{A}_0$ of A such that all improved randomized procedures are of the form $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ or $(\alpha, \beta, n) = \gamma(1, 0, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$, where $0 < \gamma < 1$ and where $(\alpha_0, \beta_0', n_0) \varepsilon \mathcal{A}_0$. It is then shown how to construct $\mathcal{A}_0$. The following problems (of the Neyman-Pearson type) are solved:

(a) Given α and β, how can we find the test in $\mathcal{A}$ with the given α and β?

Received January 2, 1958.

[1] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Faculty of Political Science, of Columbia University, 1957.

(b) Given α and n, how can we find the test in $\mathcal{A}$ with the given α and n? Numerical examples are worked out.

In Section 4, the technique is applied to tests on the mean of a binomial distribution. Although no general results were obtained, numerical examples of interest are given.

In Section 5, the technique is applied to tests on the range of a rectangular distribution (when one end point is known). It is shown that if $\alpha > 0$, $n > 0$, and $(\alpha, \beta, n) \varepsilon A_n$, then $(\alpha, \beta, n) \not\varepsilon \mathcal{A}$. The tests in $\mathcal{A}$ are characterized by a simple equation which makes it easy to

(a) determine whether a given point (α, β, n) belongs to $\mathcal{A}$, and

(b) construct any test in $\mathcal{A}$, given two of the three coordinates.

It is shown that if $(\alpha, \beta, n) \varepsilon A_n$, then there exists a test (α, β, n') in $\mathcal{A}$ such that $n' = (1 - \alpha)n$. Hence, the fractional saving in the expected sample size achieved by randomization is equal to α.

In Section 6, it is shown that in tests on the mean of a rectangular distribution (with known range), it never pays to randomize.

In Section 7, confidence intervals are evaluated in terms of confidence coefficient (α), expected length (L) and expected sample size (n). For the problem of obtaining a confidence interval for the mean of a normal distribution with known variance, "improved" randomized procedures exist and are of the form $(\alpha, L, n) = \gamma(0, 0, 0) + (1 - \gamma)(\alpha', L', n')$ where $0 < \gamma < 1$ and where (α', L', n') is a fixed sample size confidence interval procedure. Clearly, the randomized procedures obtained are of such a nature that the question of confidence intervals evaluated in terms of expected length and/or expected sample size is thrown open to discussion.

In Section 8, the k-decision problem is discussed. It is shown that improvements can be obtained by randomization.

In Section 9, the problem of applying mixed single sample tests of a composite hypothesis against a composite alternative is discussed.

In Section 10, mixed single sample procedures are compared to Wald's sequential probability ratio test in the problem of tests on the range of a rectangular distribution when one endpoint is known and are shown to be efficient in a certain sense.

In Section 11, the estimation problem is mentioned. It is shown that in most practical problems, fixed sample size procedures are optimal.

In Section 12, applications of mixed single sample tests are discussed.

2. Testing a simple hypothesis against a simple alternative. Let X denote a random variable with density function (or discrete probability function) $f(x, \theta)$. We wish to test the hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta = \theta_1$. In the sequel, we shall restrict ourselves exclusively to fixed sample size tests, both randomized and non-randomized, and mixtures of such tests. Any test of the preceding kinds will be identified with an operating characteristic (α, β, n), where α denotes the probability of a type I error, β denotes the probability of a

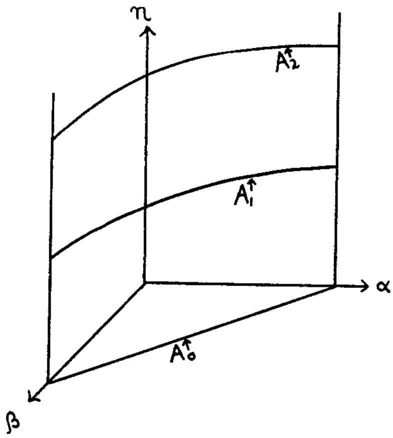

FIG. 1

type II error, and n denotes the expected number of observations. If two tests have the same operating characteristic, they will be considered equivalent.

Let $(x_1, x_2, \cdots, x_n)$ denote a sample of n independent observations on X. Let δ_n denote a real-valued, measurable function of n variables whose range is the closed interval (0, 1). The expression $\delta_n(x_1, x_2, \cdots, x_n)$ is interpreted as the probability of rejecting H_0 if $(x_1, x_2, \cdots, x_n)$ is observed. Let Δ_n denote the class of functions $\{\delta_n\}$ of the preceding type.

Definition 1. For any integer $n > 0$, let $S_n = \{(\alpha, \beta, n): \alpha = E(\delta_n \mid \theta_n), \beta = E(1 - \delta_n \mid \theta_1), \delta_n \varepsilon \Delta_n\}$. S_n is the class of tests of fixed sample size n. We define $S_0 = \{(\alpha, \beta, 0): 0 \leqq \alpha \leqq 1, \alpha + \beta = 1\}$.

Definition 2. For any integer $n \geqq 0$, let $A_n = \{(\alpha, \beta, n): a)(\alpha, \beta, n) \varepsilon S_n$, and $b)$ there exists no other test (α', β', n) belonging to S_n with the property that $\alpha' \leqq \alpha$, $\beta' \leqq \beta$, at least one of these inequalities being strict.$\}$

The set A_n is the class of admissible procedure based on samples of fixed size n, and is known to be complete. See Fig. 1.

Definition 3. Let $A = \bigcup_{i=0}^{\infty} A_i$.

Definition 4. Let $A^* = \{(\alpha, \beta, n): (\alpha, \beta, n) = \sum_{i=0}^{\infty} \gamma_i(\alpha_i, \beta_i, n_i)$ where $\gamma_i \geqq 0$, $\sum_{i=0}^{\infty} \gamma_i = 1$, and $(\alpha_i, \beta_i, n_i) \varepsilon A$ for $i = 0, 1, 2, \cdots \}$.

γ_i is interpreted as the probability of selecting the fixed sample size test (α_i, β_i, n_i) for $i = 0, 1, 2, \cdots$. A^* is the convex hull of A.

Definition 5. Let $\mathcal{A} = \{(\alpha, \beta, n): a)(\alpha, \beta, n) \varepsilon A^*$, and $b)$ there exists no other test (α', β', n') belonging to A^* with the property that $\alpha' \leqq \alpha$, $\beta' \leqq \beta$, $n' \leqq n$, at least one of these inequalities being strict.$\}$.

The set $\mathcal{A}$ is the class of admissible mixed single sample tests.

We next wish to show that $\mathcal{A}$ is complete, i.e., for any test (α', β', n') not in $\mathcal{A}$, there exists a test (α, β, n) in $\mathcal{A}$ such that $\alpha \leqq \alpha'$, $\beta \leqq \beta'$, $n \leqq n'$, at least one of these inequalities being strict. If, in general, A^* were closed, it would follow that $\mathcal{A}$ is complete. However, A^* is not necessarily closed, as the following example will illustrate.

Example. Let $f(x, \theta) = 1$ if $\theta \leqq X \leqq \theta + 1$

$= 0$ elsewhere.

We wish to test the hypothesis $H_0 : \theta = 0$ against the alternative $H_1 : \theta = \theta_1$, where $0 < \theta_1 < 1$. A simple calculation shows that

$$A_n = \{(\alpha, \beta, n) : 0 \leqq \alpha \leqq (1 - \theta_1)^n, \alpha + \beta = (1 - \theta_1)^n\}. \tag{1}$$

We define a sequence $\{(\alpha_k, \beta_k, n_k)\}$, where $(\alpha_k, \beta_k, n_k) = (1 - 1/k)(0, 1, 0) + (1/k)(0, (1 - \theta_1)^k, k)$. Clearly $\lim_{k\to\infty}(\alpha_k, \beta_k, n_k) = (0, 1, 1)$. However, $(0, 1, 1) \not\varepsilon A^*$. To prove this, assume $(0, 1, 1) \varepsilon A^*$. Then, since A^* is a three-dimensional convex set, $(0, 1, 1)$ can be expressed as a convex linear combination of at most four points in A, i.e., $(0, 1, 1) = \sum_{i=1}^4 \gamma_i(\alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^4 \gamma_i = 1$ and $(\alpha_i, \beta_i, n_i) \varepsilon A$ for $i = 1, 2, 3, 4$. Since $\sum_{i=1}^4 \gamma_i\beta_i = 1$, it follows that $\beta_i = 1$ if $\gamma_i > 0$. However, if $\beta_i = 1$, it follows from (1) that $n_i = 0$, contradicting the assumption $\sum_{i=1}^4 \gamma_i n_i = 1$. Q.E.D.

In order to show that $\mathfrak{A}$ is complete, we define $A_L^* = \{(\alpha, \beta, n)$: (a) (α, β, n) is a boundary point of A^*, and (b) there exists no test (α', β', n') belonging to A^* such that $\alpha' \leqq \alpha, \beta' \leqq \beta, n' \leqq n$, at least one of these inequalities being strict.$\}$.

The set A_L^* is the "lower" boundary of A^*. Clearly, $\mathfrak{A} \subset A_L^*$. We shall now prove an important theorem.

THEOREM 1. $A_L^* \subset A^*$.

PROOF. Suppose $(\alpha, \beta, n) \varepsilon A_L^*$. Then, since (α, β, n) is a boundary point of A^*, there exists a sequence of points $\{(\alpha_k, \beta_k, n_k)\}$ belonging to A^* such that

$$(\alpha, \beta, n) = \lim_{k\to\infty}(\alpha_k, \beta_k, n_k).$$

Since A^* is a three dimensional convex set, each point (α_k, β_k, n_k) of this sequence can be expressed as a convex linear combination of at most four points in A: i.e., for each k, there exist numbers $\gamma_{ik}, \alpha_{ik}, \beta_{ik}, n_{ik}$ such that $(\alpha_k, \beta_k, n_k) = \sum_{i=1}^4 \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik})$, where $\gamma_{ik} \geqq 0, \sum_{i=1}^4 \gamma_{ik} = 1$ and $(\alpha_{ik}, \beta_{ik}, n_{ik}) \varepsilon A$ for $i = 1, 2, 3, 4$. Without any loss of generality, we can assume that the sequences $\{\gamma_{ik}\}$, $\{\alpha_{ik}\}$ and $\{\beta_{ik}\}$ are convergent for $i = 1, 2, 3, 4$ as k tends to infinity. Let $\gamma_i = \lim_{k\to\infty} \gamma_{ik}$, $\alpha_i = \lim_{k\to\infty} \alpha_{ik}$, $\beta_i = \lim_{k\to\infty} \beta_{ik}$ for $i = 1, 2, 3, 4$. Clearly, $\gamma_i \geqq 0, \sum_{i=1}^4 \gamma_i = 1$ for $i = 1, 2, 3, 4$. Before proceeding with the proof of Theorem 1, we prove a useful lemma.

LEMMA 1. *If $\gamma_i > 0$, there exists a number N_i such that $n_{ik} \leqq N_i$ for all k.*

PROOF. Since $\lim_{k\to\infty} n_k = n$, there exists a positive integer K such that if $k > K$, $n_k < n + 1$. Furthermore, since $\lim_{k\to\infty} \gamma_{ik} = \gamma_i > 0$, there exists a number K_i such that if $k > K_i$, $\gamma_{ik} > \frac{1}{2}\gamma_i$. Let $M_i = \max(K, K_i)$. Then, if $k > M_i$, $\frac{1}{2}\gamma_i n_{ik} \leqq \sum_{i=1}^4 \gamma_{ik} n_{ik} = n_k < n + 1$. Thus, if $k > M_i$, $n_{ik} < 2(n + 1)/\gamma_i$. Then, $N_i = \max[n_{i1}, n_{i2}, \cdots n_{iM_i}, 2(n + 1)/\gamma_i]$ is the required number, proving the lemma.

We now proceed with the proof of Theorem 1. Consider four cases.

Case 1. $\gamma_i > 0, i = 1, 2, 3, 4$.

Let $N = \max(N_1, N_2, N_3, N_4)$ where N_i is defined in Lemma 1. Then, since

$0 \leqq n_{ik} \leqq N$, for $i = 1, 2, 3, 4$ and all k, the sequences $\{n_{ik}\}$ are bounded. Hence, for each i, there exists a convergent subsequence which we denote by $\{\bar{n}_{ik}\}$. Let $\{\bar{\alpha}_{ik}\}$, $\{\bar{\beta}_{ik}\}$ and $\{\bar{\gamma}_{ik}\}$ denote respectively the subsequences of $\{\alpha_{ik}\}$, $\{\beta_{ik}\}$ and $\{\gamma_{ik}\}$ corresponding to the convergent subsequence $\{\bar{n}_{ik}\}$ of $\{n_{ik}\}$. Let $\lim_{k\to\infty} \bar{n}_{ik} = n_i$. Clearly,

$$(\alpha, \beta, n) = \lim_{k\to\infty} (\alpha_k, \beta_k, n_k) = \lim_{k\to\infty} \sum_{i=1}^{4} \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik}) = \sum_{i=1}^{4} \lim_{k\to\infty} \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik})$$

$$= \sum_{i=1}^{4} \lim_{k\to\infty} \bar{\gamma}_{ik}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = \sum_{i=1}^{4} \gamma_i(\alpha_i, \beta_i, n_i).$$

Since $(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) \varepsilon A$ for all i and k, and since A is closed,

$$\lim_{k\to\infty}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = (\alpha_i, \beta_i, n_i) \in A \qquad \textit{for } i = 1, 2, 3, 4.$$

Furthermore, since $\gamma_i > 0$, $i = 1, 2, 3, 4$, and $\sum_{i=1}^{4} \gamma_i = 1$, $\sum_{i=1}^{4} \gamma_i(\alpha_i, \beta_i, n_i) \varepsilon A^*$. Hence $(\alpha, \beta, n) \varepsilon A^*$.

The more difficult case to prove is Case 2.

Case 2. Exactly one of the γ_i's is 0.

To fix ideas, suppose $\gamma_1 = 0$, $\gamma_2 > 0$, $\gamma_3 > 0$, $\gamma_4 > 0$. Let $N = \max(N_2, N_3, N_4)$. In a manner analogous to that used in Case 1, we define sequences $\{\bar{\alpha}_{ik}\}$, $\{\bar{\beta}_{ik}\}$, $\{\bar{n}_{ik}\}$ and $\{\bar{\gamma}_{ik}\}$ for $i = 2, 3, 4$. We define ne w sequences

$$\alpha_k' = \bar{\gamma}_{1k}(0) + \sum_{i=2}^{4} \bar{\gamma}_{ik}\, \bar{\alpha}_{ik},$$

$$\beta_k' = \bar{\gamma}_{1k}(1) + \sum_{i=2}^{4} \bar{\gamma}_{ik}\, \bar{\beta}_{ik},$$

$$n_k' = \bar{\gamma}_{1k}(0) + \sum_{i=2}^{4} \bar{\gamma}_{ik}\, \bar{n}_{ik},$$

where $\bar{\gamma}_{1k} = 1 - \sum_{i=2}^{4} \bar{\gamma}_{ik}$. It is easily seen that

$$\lim_{k\to\infty} \alpha_k' = \alpha,$$

$$\lim_{k\to\infty} \beta_k' = \beta,$$

$$\lim_{k\to\infty} n_k' \leqq n,$$

Since $(\alpha_k', \beta_k', n_k') \varepsilon A^*$ for each k, and since $(\alpha, \beta, n) \varepsilon A_L^*$, it follows that the inequality $\lim_{k\to\infty} n_k' < n$ cannot hold. Hence,

$$(\alpha, \beta, n) = \lim_{k\to\infty} (\alpha_k', \beta_k', n_k') = \lim_{k\to\infty} \sum_{i=2}^{4} \bar{\gamma}_{ik}, (\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik})$$

$$\sum_{i=2}^{4} \lim \bar{\gamma}_{ik}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = \sum_{i=2}^{4} \gamma_i(\alpha_i, \beta_i, n_i).$$

Using the argument in Case 1, we find that $(\alpha, \beta, n) \varepsilon A^*$.

Case 3. Two of the γ_i's are 0. The proof of Case 3 is analogous to the proof of Case 2.

Case 4. Three of the γ_i's are 0. The proof of Case 4 is analogous to the proof of Case 2.

COROLLARY 1. $\mathfrak{A} = A_L^*$.

COROLLARY 2. $\mathfrak{A}$ *is complete.*

PROOF. Let (α', β', n') be a test which does not belong to $\mathfrak{A}$.

Let

$$A_{(\alpha'\beta')} = \{(\alpha, \beta, n){:}a)\ \alpha = \alpha', \beta = \beta' \text{ and } (\alpha, \beta, n) \ \varepsilon\ A^*\}.$$

$A_{(\alpha',\beta')}$ is non-empty since $(\alpha', \beta', n') \ \varepsilon\ A_{(\alpha',\beta')}$. Let

$$N = N_{(\alpha',\beta')} = \inf_{\{n':(\alpha',\beta',n') \,\varepsilon\, A_{(\alpha',\beta')}\}} n'$$

Then $(\alpha', \beta', N) \ \varepsilon\ \mathfrak{A} = A_L^*$ where $N < n'$.

Note: It is possible to show that $\mathfrak{A}$ is complete using a different approach. If we define $S = \bigcup_{i=0}^* S_i$ and S^* as the convex hull of S, it can be shown that S^* is closed. This implies that $\mathfrak{A}$ is complete. However, to prove that S^* is closed requires a technique similar to that used in proving Theorem 1.

THEOREM 2. *If $f(x, \theta_0) = 0$ if and only if $f(x, \theta_1) = 0$, then a necessary and sufficient condition for (α, β, n) to belong to $\mathfrak{A}$ is that for some non-negative a and b and positive c, we have*

$$a\alpha + b\beta + cn = \min_{(\alpha',\beta',n')\varepsilon A^*} \{a\alpha' + b\beta' + cn'\}.$$

PROOF. To prove the sufficiency of the condition, we consider 4 cases.

Case 1. $a = 0, b = 0, c > 0$. Then $a\alpha' + b\beta' + cn' = cn'$ is minimized *only* by tests $(\alpha, \beta, 0)$ belonging to A_0. However, $A_0 \subset \mathfrak{A}$, proving the sufficiency of the condition if Case 1 holds.

Case 2. $a = 0, b > 0, c > 0$. Then, $a\alpha' + b\beta' + cn' = b\beta' + cn'$ is minimized only by the test $(1, 0, 0)$ which belongs to A_0.

Case 3. $a > 0, b = 0, c > 0$. (Similar to Case 2.)

Case 4. $a > 0, b > 0, c > 0$. Then, it is well known, and can be easily proved that any test (α, β, n) such that $a\alpha + b\beta + cn = \min_{(\alpha',\beta',n') \,\varepsilon\, A^*}(a\alpha' + b\beta' + cn')$ belongs to $\mathfrak{A}$.

To prove the necessity of the condition, we assume $(\alpha, \beta, n) \ \varepsilon\ \mathfrak{A}$.

(i) If $n = 0$, choose $a = 0, b = 0, c = 1$.

(ii) If $n > 0$, then it is well known in the theory of convex sets that there exist non-negative numbers a, b and c such that

$$a\alpha + b\beta + cn = \min_{(\alpha',\beta',n') \,\varepsilon\, A^*} (a\alpha' + b\beta' + cn').$$

It remains is to show that $c > 0$. Assume $c = 0$. Then

$$a\alpha + b\beta = \min_{(\alpha',\beta',n') \,\varepsilon\, A^*} (a\alpha' + b\beta') = 0.$$

Since $(\alpha, \beta, n) \varepsilon \mathcal{A}$, then there exist numbers $\gamma_i, \alpha_i, \beta_i, n_i$ such that $(\alpha, \beta, n) = \sum_{i=1}^4 \gamma_i(\alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^4 \gamma_i = 1$, $(\alpha_i, \beta_i, n_i) \varepsilon A$ for $i = 1, 2, 3, 4$.

Thus,

$$a\alpha + b\beta = a \sum_{i=1}^{4} \gamma_i \alpha_i + b \sum_{i=1}^{4} \gamma_i \beta_i = 0.$$

Since both a and b cannot equal 0, either $\alpha = 0$ or $\beta = 0$. Assume $\alpha = 0$. Then, if $\gamma_i > 0$, $\alpha_i = 0$. Using the fact that $f(x, \theta_0) = 0$ if and only if $f(x, \theta_1) = 0$, it follows that if $\alpha_i = 0$, $\beta_i = 1$. Hence, $(\alpha, \beta, n) = (0, 1, n)$. But, $(0, 1, n) \not\varepsilon \mathcal{A}$ since $(0, 1, 0)$ is preferred. Thus we are led to a contradiction of the fact that $(\alpha, \beta, n) \varepsilon \mathcal{A}$. If we assume $\beta = 0$, we are led to a similar contradiction. Therefore, the assumption $c = 0$ is false. Theorem 2 is thus proved.

Theorem 2 states, in effect, that the problem of generating $\mathcal{A}$ reduces to constructing tests (α, β, n) which minimize the expression $a\alpha + b\beta + cn$ for all choices of non-negative a and b and positive c. The cases where either a or b is 0 were discussed and disposed of in proving Theorem 2. The main problem, then, is to construct the tests (α, β, n) which minimize the expression $a\alpha + b\beta + cn$. We proceed as follows: without any loss of generality we may assume that $a + b = 1$ and write $a = \pi$ and $b = 1 - \pi$, where $0 < \pi < 1$. Then, we wish to find the tests (α, β, n) in $\mathcal{A}$ such that

$$\pi\alpha + (1 - \pi)\beta + cn = \min_{(\alpha', \beta', n') \varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn'].$$

Clearly,

$$\begin{aligned}
&\min_{(\alpha', \beta', n') \varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn'] \\
&\quad = \min_{\substack{[\gamma_i, \alpha_i, \beta_i, n_i : \gamma_i \geqq 0, \sum_{i=0}^{\infty} \gamma_i = 1, (\alpha_i, \beta_i, n_i) \varepsilon A] \\ i = 0, 1, 2, \cdots}} \cdot \left\{\pi \sum_{i=0}^{\infty} \gamma_i \alpha_i + (1 - \pi) \sum_{i=0}^{\infty} \gamma_i \beta_i + c \sum_{i=0}^{\infty} \gamma_i n_i\right\} \\
&\quad = \min_{\substack{[\gamma_i, \alpha_i, \beta_i, n_i : \gamma_i \geqq 0, \sum_{i=0}^{\infty} \gamma_i = 1, (\alpha_i, \beta_i, n_i) \varepsilon A] \\ i = 0, 1, 2, \cdots}} \cdot \left\{\sum_{i=0}^{\infty} \gamma_i[\pi\alpha_i + (1 - \pi)\beta_i] + c \sum_{i=0}^{\infty} \gamma_i n_i\right\} \\
&\quad = \min_{N \geqq 0} \left(cN + \left\{ + \min_{(\gamma_i, n_i : \sum_{i=0}^{\infty} \gamma_i n_i = N, \gamma_i \geqq 0, \Sigma \gamma_i = 1, i = 0, 1, 2, \cdots)} \cdot \sum_{i=0}^{\infty} \gamma_i \min_{(\alpha_i, \beta_i, n_i) \varepsilon A_{n_i}} [\pi\alpha_i + (1 - \pi)\beta_i]\right\}\right).
\end{aligned}$$

It should be noted that the operation "min"$_{N \geqq 0}$ is not restricted to integral values of N.

From the above, it is clear that the desired minimization can be accomplished in 3 steps, which we shall now describe in detail.

Step 1. We can, for each n_i, find the tests (α_i, β_i, n_i) belonging to A_{n_i} which minimize the expression $\pi\alpha_i + (1 - \pi)\beta_i$. For each n_i, let

$$R_\pi(n_i) = \min_{[(\alpha_i,\beta_i):(\alpha_i,\beta_i,n_i)\varepsilon A_{ni}]} \{\pi\alpha_i + (1 - \pi)\beta_i\}.$$

$R_\pi(n_i)$ may be interpreted as the Bayes risk for fixed s mple-size procedures of sample size n_i where π is the a priori probability that θ_0 is the true parameter and $1 - \pi$ is the a priori probability that θ_1 is the true parameter.

In particular,

$$R_\pi(0) = \min_{[\alpha,\beta:0 \leqq \alpha \leqq 1, \alpha+\beta=1]} \{\pi\alpha + (1 - \pi)\beta\} = \min(\pi, 1 - \pi).$$

If $0 < \pi < \frac{1}{2}$, $R_\pi(0) = \pi$. The only test $(\alpha, \beta, 0)$ belonging to A_0 satisfying the equation $\pi\alpha + (1 - \pi)\beta = \pi$ is the test $(1, 0, 0)$. Similarly, if $\frac{1}{2} < \pi < 1$, $R_\pi(0) = 1 - \pi$. The only test belonging to A_0 satisfying the equation $\pi\alpha + (1 - \pi)\beta = 1 - \pi$ is the test $(0, 1, 0)$. If $\pi = \frac{1}{2}$, $R_\pi(0) = \frac{1}{2}$. Then, any test belonging to A_0 satisfies the equation $\frac{1}{2}\alpha + \frac{1}{2}\beta = \frac{1}{2}$, since $\alpha + \beta = 1$.

We note that

$$\min_{(\alpha',\beta',n')\varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn']$$

$$= \min_{N \geqq 0} \left\{cN + \min_{[\gamma_i,n_i:\gamma_i \geqq 0, \sum_{i=0}^\infty \gamma_i = 1, \sum_{i=0}^\infty \gamma_i n_i = N]} \sum_{i=0}^\infty \gamma_i R_\pi(n_i)\right\}.$$

Step 2. Subject to the conditions

$$\gamma_i \geqq 0, i = 0, 1, 2, \cdots, \textstyle\sum_{i=0}^\infty \gamma_i = 1, \sum_{i=0}^\infty \gamma_i n_i = N,$$

we can, for each non-negative value of N choose the γ_i's so that $\sum_{i=0}^\infty \gamma_i R_\pi(n_i)$ is minimized. To this end, let

$$R_\pi = \bigcup_{k=0}^\infty (k, R_\pi(k)).$$

Let R_π^* denote the convex hull of R_π and let $\mathcal{R}_\pi$ denote the lower boundary of R_π^*, i.e., $\mathcal{R}_\pi = \{(k, r):$ (a) $(k, r) \varepsilon R_\pi^*$ and (b) there exists no point (k', r') belonging to R_π^* such that $k' \leqq k$, $r' \leqq r$, at least one of these inequalities being strict.$\}$.

Then, to accomplish Step 2 of the minimization, given $N \geqq 0$, we merely select the point (N, r) belonging to $\mathcal{R}_\pi$. Since (N, r) is a boundary point of a two dimensional convex set, (N, r) *can always be expressed as a convex linear combination of at most two points in* R. We define

$$r_\pi(N) = \min_{[\gamma_i,n_i:\gamma_i \geqq 0, \sum_{i=0}^\infty \gamma_i = 1, \sum_{i=0}^\infty \gamma_i n_i = N, i=0,1,2,\cdots]} \left\{\sum_{i=0}^\infty \gamma_i R_\pi(n_i)\right\}.$$

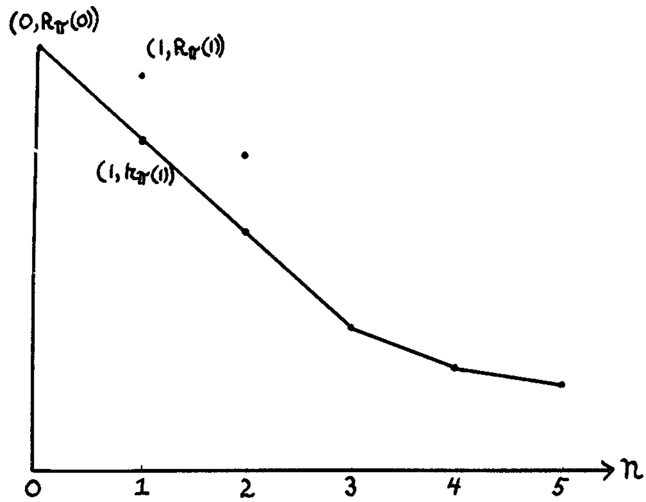

FIG. 2

See Fig. 2. We note that

$$\min_{(\alpha',\beta',n')\,\varepsilon\,A^*} [\pi\alpha' + (1 - \pi)\beta' + cn'] = \min_{N\geqq 0} [r_\pi(N) + cN].$$

Step 3. We now wish to choose $N \geqq 0$ to minimize the expression $r_\pi(N) + cN$. Since $r_\pi(N)$ is a strictly decreasing, convex and piecewise linear function of N, there exists at least one value of N and at most a finite interval of values of N which minimize $r_\pi(N) + cN$.

It should be noted that if we are given a specific value of N, then there exists a number $c > 0$ such that $r_\pi(N) + cN = \min_k [r_\pi(k) + ck]$. Therefore, for an arbitrary but fixed value of $N > 0$ any procedure obtained in Step 2 will be an admissible mixed single sample test so that Step 3 is inessential in constructing $\mathfrak{A}$.

We shall apply the technique in several problems in the following sections.

3. Testing the mean of a normal distribution when the variance is known. Let

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\},$$

where $\sigma > 0$ is known. We wish to test the hypothesis $H_0:\theta = \theta_0$ against the alternative $H_1:\theta = \theta_1$, $\theta_0 < \theta_1$. It can be shown that for any integer $n \geqq 0$,

$$A_n = \{(\alpha, \beta, n): \alpha = 1 - \Phi(t), \beta = \Phi(t - \sqrt{n}\delta) \text{ for } -\infty \leqq t \leqq \infty\},$$

where

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

and

$$\delta = \frac{\theta_1 - \theta_0}{\sigma}.$$

Step 1. We have already seen that $R_\pi(0) = \min(\pi, 1 - \pi)$. For any integer $n > 0$,

$$
\begin{aligned}
R_\pi(n) &= \min_{\{(\alpha,\beta):(\alpha,\beta,n)\,\varepsilon\, A_n\}} [\pi\alpha + (1-\pi)\beta] \\
&= \min_t \{\pi[1 - \Phi(t)] + (1-\pi)\Phi(t - \sqrt{n}\delta)\} \qquad (2)\\
&= \pi\left[1 - \Phi\left(\frac{\xi}{\sqrt{n}\delta} + \frac{\sqrt{n}\delta}{2}\right)\right] + (1-\pi)\,\Phi\left(\frac{\xi}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right)\Bigg],
\end{aligned}
$$

where $\xi = \log \pi/(1 - \pi)$. Furthermore, the test (α, β, n) such that

$$\alpha = 1 - \Phi\left(\frac{\xi}{\sqrt{n}\delta} + \frac{\sqrt{n}\delta}{2}\right) \quad \text{and} \quad \beta = \Phi\left(\frac{\xi}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right)$$

is unique. It should also be noted that for any π such that $0 < \pi < 1$, $R_\pi(n)$ is a strictly decreasing function of n. See Figure 3.

Step 2. To accomplish Step 2 of the minimization, we consider $R_\pi(n)$ formally as a function of a continuous variable n. We shall first show that there exists a number $n_i = n_i(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_i)$ and convex on the interval (n_i, ∞). To show the existence of n_i, we use the identities

(a) $\varphi(x - y) = e^{2xy}\varphi(x + y)$,

(b) $\varphi'(x) = -x\varphi(x)$, where $\varphi(x) = \Phi'(x)$.

A routine calculation shows that

(c) $R_\pi'(n) = \dfrac{-\pi}{2\sqrt{n}\delta}\,\varphi\left(\dfrac{\xi}{\sqrt{n}\delta} + \dfrac{\sqrt{n}\delta}{2}\right)$

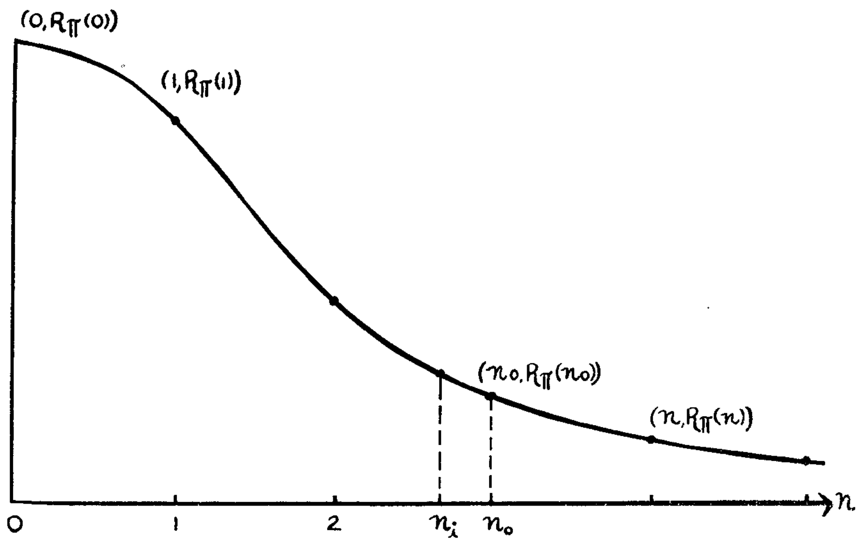

FIG. 3

and

$$\text{(d)}\quad R_\pi''(n) = (n^2\delta^4 + 4n\delta^2 - 4\xi^2)\,\frac{\pi\sqrt{n}\delta}{16n^3}\,\varphi\left(\frac{\xi}{\sqrt{n}\delta} + \frac{\sqrt{n}\delta}{2}\right).$$

Setting $R_\pi''(n)$ equal to 0, we find that

$$n_i = \frac{-2 + 2\sqrt{1+\xi^2}}{\delta^2}. \tag{3}$$

Therefore, (3) gives a unique inflection point of the function $R_\pi(n)$. See Fig. 3.

Since $R_\pi(n)$, defined in (2), is defined only for integral values of n, and since n_i in general is not an integer, we assert that there exists an integer $n_0 = n_0(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_0)$ and convex on the interval (n_0, ∞). See Fig. 3. It then follows that

$$r_\pi(N) = \begin{cases} \left(1 - \dfrac{N}{n_0}\right)R_\pi(0) + \dfrac{N}{n_0}R_\pi(n_0) & \text{if } N \leqq n_0 \\ ([N] + 1 - N)R_\pi([N]) + (N - [N])R_\pi([N] + 1) & \text{if } N > n_0 \end{cases}$$

Thus Step 2 of the minimization is achieved.

It now becomes clear that improved randomized procedures (α, β, n) exist and are of the form $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ or $(\alpha, \beta, n) = \gamma(1, 0, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ where $0 < \gamma < 1$ and where

$$n_0 = n_0(\pi), \qquad \alpha_0 = \alpha_0(\pi) = 1 - \Phi\left(\frac{\xi}{\sqrt{n_0}\delta} + \frac{\sqrt{n_0}\delta}{2}\right),$$

$$\beta_0 = \beta_0(\pi) = \Phi\left(\frac{\xi}{\sqrt{n_0}\delta} - \frac{\sqrt{n_0}\delta}{2}\right)$$

for some π such that $0 < \pi < 1$.

It also becomes clear that a test $(\alpha, \beta, n) \,\varepsilon\, A_n$ if and only if $n \geqq n_0(\pi)$, where π is defined by the equation

$$\beta = \Phi\left(\frac{\log\dfrac{\pi}{1-\pi}}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right).$$

This gives a complete answer to the general question of whether or not a fixed sample size procedure can be improved upon by means of randomization.

3.1. We now consider the following problem: Given α and β, how can we find the test in $\mathfrak{A}$ achieving the given α and β? To this end, consider two cases.

Case 1. $\alpha < \beta$. Let $\mathfrak{A}_0 = \{(\alpha_0, \beta_0, n_0) : n_0 = n_0(\pi), \beta_0 = \beta_0(\pi), \alpha_0 = \alpha_0(\pi) \text{ for } \tfrac{1}{2} < \pi < 1\}$.

Let (α, β, n) denote the test in $\mathfrak{A}$ with the given α and n.

From the discussion of Step 2, it is evident that (α, β, n) is an improved randomized procedure if and only if $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$, where $0 < \gamma < 1$ and where $(\alpha_0, \beta_0, n_0) \,\varepsilon\, \mathfrak{A}$. In this case, $\alpha = (1 - \gamma)\alpha_0$, $\beta =$

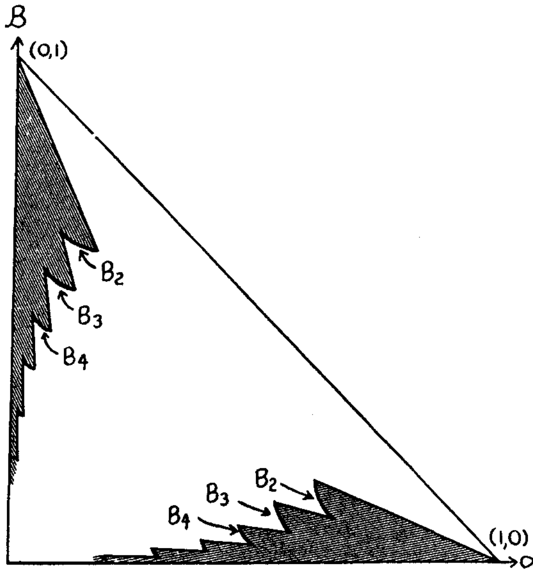

FIG. 4. Shaded region corresponds to the set of (α, β) for which the admissible test (α, β, n) is a randomized procedure; $B_i = \{(\alpha_0, \beta_0): n_0 = i\}$, $\bigcup_{i=2}^{\infty} B_i = P(\mathcal{A}_0 \mid \alpha, \beta)$.

$\gamma + (1 - \gamma)\beta_0$, $n = (1 - \gamma)n_0$. These equations imply that $\alpha/(1 - \beta) = \alpha_0/(1 - \beta_0)$ and $1 - \gamma = \alpha/\alpha_0$. The equation $\alpha/(1 - \beta) = \alpha_0/(1 - \beta_0)$ when interpreted geometrically means that the points $(0, 1)$, (α, β) and (α_0, β_0) are co-linear. The equation $1 - \gamma = \alpha/\alpha_0$ when interpreted geometrically means that (α, β) is between $(0, 1)$ and (α_0, β_0).

If (α, β, n) is not an improved randomized procedure, then

$$(\alpha, \beta, n) = \gamma(\alpha_1, \beta_1, [n]) + (1 - \gamma)(\alpha_2, \beta_2, [n] + 1)$$

where $0 \leqq \gamma \leqq 1$ and where $(\alpha_1, \beta_1, [n])$ and $(\alpha_2, \beta_2, [n] + 1) \varepsilon A$ and is of little interest.

We summarize the preceding as follows: Let $P(\mathcal{A}_0 \mid \alpha, \beta)$ denote the projection of $\mathcal{A}_0$ on the (α, β) plane. See Fig. 4. It was convenient to let $\delta = 1$. If (α, β) lies on a line segment joining $(0, 1)$ to one of the points (α_0, β_0) in $P(\mathcal{A}_0 \mid \alpha, \beta)$, then the test $(\alpha, \beta, n) = (1 - \alpha/\alpha_0)(0, 1, 0) + \alpha/\alpha_0(\alpha_0, \beta_0, n_0)$ is the test in $\mathcal{A}$ with the given α and β. Otherwise, (α, β, n) is achieved by randomizing over two fixed sample size procedures, one in $A_{[n]}$ and the other in $A_{[n]+1}$.

Case 2. $\alpha > \beta$. Similar to Case 1.

Table (1) shows the improvement in the expected sample size N which can be achieved for selected tests (α, β, n) belonging to $A_n - \mathcal{A}$. In this case, we let $\delta = .1$.

3.2. Consider next the following problem: Given α and n, how can we construct the test in $\mathcal{A}$ having the given α and n? We solve this problem geometri-

TABLE 1

α	β	Sample size, n, of admissible single sample tests achieving the given α and β	Expected sample size, N, of admissible mixed single sample test achieving the given α and β	Percent saving $\frac{n-N}{n} \times 100$
.005	.862	221	119	46
.005	.732	383	287	25
.01	.732	147	84	43
.01	.463	585	574	2
.05	.687	134	123	8
.05	.868	28	20	28

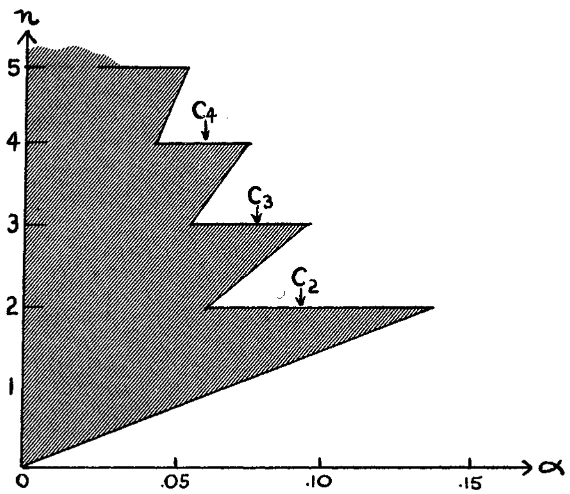

FIG. 5. Shaded region corresponds to the set of (α, n) for which the admissible test (α, β, n) is a randomized procedure; $C_i = \{(\alpha_0, n_0): n_0 = i\}$, $\bigcup_{i=2}^{\infty} C_i = P(\mathfrak{A}_0 \mid \alpha, n)$.

cally. Let $P(\mathfrak{A}_0 \mid \alpha, n)$ denote the projection of the set $\mathfrak{A}_0$ on the (α, n) plane. See Fig. 5. Then, draw a line of slope n/α through the origin. Determine the point of intersection (α_0, n_0) of this line and $P(\mathfrak{A}_0 \mid \alpha, n)$. Clearly, $n/\alpha = n_0/\alpha_0$. If $\alpha_0 > \alpha$, the test in $\mathfrak{A}$ having the given α and n is the mixture

$$\frac{\alpha}{\alpha_0}(\alpha_0, \beta_0, n_0) + \left(1 - \frac{\alpha}{\alpha_0}\right)(0, 1, 0).$$

If $\alpha_0 \leqq \alpha$, the test in $\mathfrak{A}$ having the given α and n is a mixture of two tests, one in $A_{[n]}$ and the other in $A_{[n]+1}$ and hence is of little interest.

4. Tests on the mean of a binomial distribution. Let

$$\begin{aligned} f(x, \theta) &= \theta^x(1-\theta)^{1-x} && \text{if } x = 0, 1, \\ &= 0 && \text{elsewhere, } 0 < \theta < 1. \end{aligned}$$

We wish to test the hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta = \theta_1$, $\theta_1 > \theta_0$. It is known that

$$A_n = \Bigg\{(\alpha, \beta, n) : \alpha = \sum_{i=0}^{n+1} \gamma_i \, \alpha_i, \beta = \sum_{i=0}^{n+1} \gamma_i \beta_i,$$

$$\gamma_0 = \gamma_1 = \gamma_{i-1} = \gamma_{i+2} = \cdots \gamma_{n+1} = 0, \gamma_i \geqq 0,$$

$$\gamma_{i+1} \geqq 0, \sum_{i=0}^{n+1} \gamma_i = 1,$$

$$\alpha_i = \sum_{i=1}^{n} \binom{n}{r} \theta_0^r (1 - \theta_0)^{n-r}, \qquad \beta_i = \sum_{r=0}^{i=1} \binom{n}{r} \theta_1^r (1 - \theta_1)^{n-r},$$

$$i = 0, 1, 2, \cdots n + 1 \Bigg\}.$$

Howard Raiffa [2] has pointed out that if we consider the projections of A_1 and A_2 on the (α, β) plane, there exists a test in A_2 whose operating characteristic is $(\theta_0, 1 - \theta_1, 2)$. However, there exists a test in A_1 whose operating characteristic is $(\theta_0, 1 - \theta_1, 1)$. Hence $(\theta_0, 1 - \theta_1, 2) \notin \mathcal{A}$. See Fig. 6. Furthermore, if π is such that

$$\frac{\pi}{1 - \pi} = \frac{\theta_0(1 - \theta_0)}{\theta_1(1 - \theta_1)},$$

then $R_\pi(1) = R_\pi(2)$.

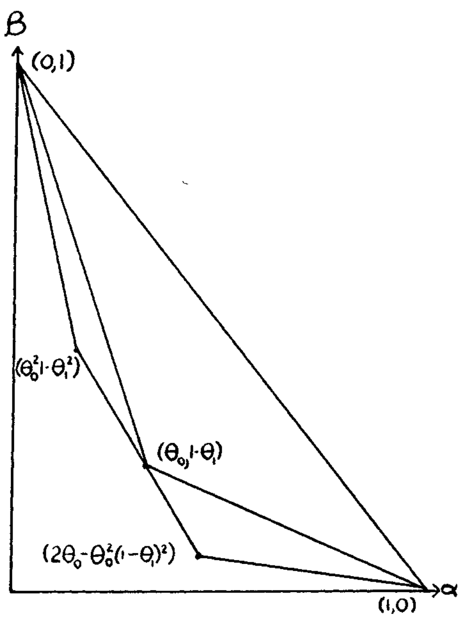

FIG. 6

TABLE 2

α	n	Probability of a type II error of admissible single sample test having the given α and n	Probability of a type II error of randomized test having the given α and n	Percent decrease
.512	30	.029	.018	38
.361	20	.112	.090	19
.098	20	.320	.286	11
.350	40	.030	.024	20

Unlike the normal distribution, there does not exist an integer $n_0(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_0(\pi))$ and convex on the interval $(n_0(\pi), \infty)$. Rather, it was found by numerical calculation that $R_\pi(n)$ has many inflection points. Thus, we do not generalize any further and present the following examples.

Example I. Let $\theta_0 = .04$, $\theta_1 = .15$. Table 2 shows the percent decrease in the probability of a type II error that randomization achieves over fixed sample size procedures for the given α and n. Since $R_\pi(n)$ was calculated for values of n where $n = 5k$ where k is a non-negative integer, it cannot be said with certainty that the improvements shown in Table 2 are optimal. However, the optimal improvements are at least as great as the ones recorded.

Example II. We again wish to test the hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$ where $\theta_0 < \theta_1$. Then, it is well known that any test $(\alpha, \beta, 1)$ such that

$$(\alpha, \beta, 1) = \gamma(0, 1, 1) + (1 - \gamma)(\theta_0, 1 - \theta_1, 1),$$

where $0 \leqq \gamma \leqq 1$ belongs to A_1. We shall now show that if we are given a test $(\alpha, \beta, 1)$ of the above type such that $\gamma \leqq 1 - \theta_1/2$ then there exists a mixed single sample test $(\alpha^*, \beta, 1)$ such that

$$\frac{\alpha - \alpha^*}{\alpha} = \frac{\gamma(1 - \alpha - \beta)}{(1 - \gamma)(1 + \beta - 2\gamma)}. \tag{4}$$

The expression $(\alpha - \alpha^*)/\alpha$ is interpreted as the fractional saving in α achieved by randomization.

To prove this, consider the test

$$(\alpha^*, \beta', n') = \frac{\gamma}{2 - \theta_1}(0, 1, 0) + \frac{\gamma}{2 - \theta_1}(\theta_0^2, 1 - \theta_1^2, 2) + \left(1 - \frac{2\gamma}{2 - \theta_1}\right)(\theta_0, 1 - \theta_1, 1).$$

Since $\gamma \leqq 1 - \theta_1/2$, the above test is a bonafide mixture. It is easily verified that $\beta' = \beta$, $n' = 1$ and that $(\alpha - \alpha^*)/\alpha$ has the value given in (4).

To illustrate the fractional saving in α which can be achieved, consider the test $H_0: \theta = .10$ against $H_1: \theta = .95$. Then, there exists a test $(\alpha, \beta, 1)$ in A_1 where $(\alpha, \beta, 1) = .5(0, 1, 1) + .5(.10, .05, 1) = (.05, .525, 1)$. Consider the test

$$(\alpha^*, \beta, 1) = \frac{.5}{2 - .95}(0, 1, 0) + \frac{.5}{2 - .95}(.01, .0975, 2) + \left(1 - \frac{1}{2 - .95}\right)(.10, .05, 1) = \left(\frac{1}{105}, .525, 1\right).$$

Then

$$\frac{\alpha - \alpha^*}{\alpha} = \frac{17}{21}.$$

5. Tests on the range of a rectangular distribution when one endpoint is known. Let

$$f(x, \theta) = \frac{1}{\theta} \quad \text{if } 0 \leqq x \leqq \theta,$$
$$= 0 \quad \text{elsewhere.}$$

We wish to test the hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta = \theta_1$, $\theta_1 > \theta_0$. It can be shown that

$$(5) \qquad \begin{aligned} A_n &= \left\{(\alpha, \beta, n): \alpha = \frac{\theta_0^n - t^n}{\theta_1^n}, \beta = \frac{t^n}{\theta_1^n}, 0 \leqq t \leqq \theta_0\right\} \\ &= \left\{(\alpha, \beta, n): 0 \leqq \alpha \leqq 1, \beta = \left(\frac{\theta_0}{\theta_1}\right)^n (1 - \alpha)\right\}. \end{aligned}$$

It should be noted that Theorem 2 does not hold since $f(x, \theta_0)$ and $f(x, \theta_1)$ do not vanish simultaneously for values of x such that $\theta_0 \leqq x \leqq \theta_1$. Hence, we shall alter our approach to generating $\mathfrak{A}$ by proving a theorem which will yield as a consequence a technique for constructing $\mathfrak{A}$.

THEOREM 3. *If $(\alpha, \beta, n) \,\varepsilon\, A_n$ where $\alpha > 0$ and $n > 0$, then $(\alpha, \beta, n) \not\varepsilon\, \mathfrak{A}$.*

PROOF. If $(\alpha, \beta, n) \,\varepsilon\, A_n$, then it follows from (5) that $\beta = (\theta_0/\theta_1)^n(1 - \alpha)$. Consider the test

$$\begin{aligned} (\alpha', \beta', n') &= \alpha(1, 0, 0) + (1 - \alpha)\left(0, \left(\frac{\theta_0}{\theta_1}\right)^n, n\right) \\ &= \left(\alpha, (1 - \alpha)\left(\frac{\theta_0}{\theta_1}\right)^n, (1 - \alpha)n\right) \\ &= (\alpha, \beta, (1 - \alpha)n). \end{aligned}$$

Clearly $(\alpha, \beta, (1 - \alpha)n)$ is preferred to (α, β, n). Theorem 3 states that all single sample tests (α, β, n) such that $0 < \alpha \leqq 1$ and $n > 0$ are inadmissible in the class of mixed single-sample tests. Consequently, the class $\mathfrak{T}$ can be generated by the test $(1, 0, 0)$ and the sequence of tests $\{(0, (\theta_0/\theta_1)^k, k)\}$, $k = 0, 1, 2, \cdots$. Since $(\theta_0/\theta_1)^n$ is a convex function of n, it can be shown that $(\alpha, \beta, n) \,\varepsilon\, \mathfrak{A}$ if and only if $(\alpha, \beta, n) = \gamma_1(1, 0, 0) + \gamma_2(0, \theta_0/\theta_1)^k, k) + \gamma_3(0, (\theta_0/\theta_1)^{k+1}, k + 1$ for some non-negative numbers $\gamma_1, \gamma_2, \gamma_3$ and some non-negative integer k where

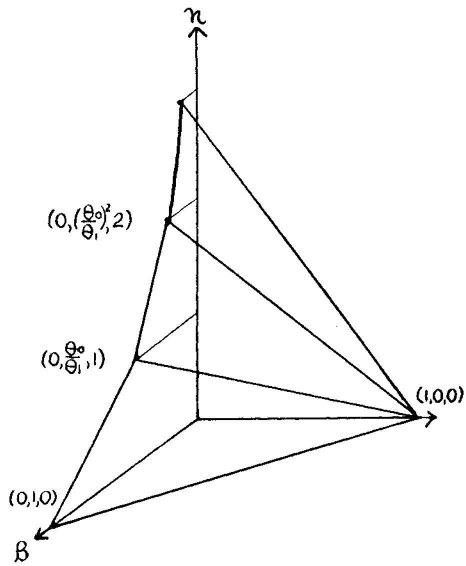

FIG. 7

$\sum_{i=1}^{3} \gamma_i = 1$. In fact it is easily verified that $k = [n/1 - \alpha]$, $\gamma_1 = \alpha$, $\gamma_2 = (1 - \alpha)([n/1 - \alpha] + 1) - n$ and $\gamma_3 = n - (1 - \alpha)\,[n/1 - \alpha]$. See Fig. 7.

COROLLARY 1. *If* $(\alpha, \beta, n) \,\varepsilon\, A_n$, *there exists a test* $(\alpha, \beta, n') \,\varepsilon\, \mathfrak{A}$ *where* $n' = (1 - \alpha)n$.

PROOF. From the preceding discussion, the test $(\alpha, \beta, n') = \alpha(1, 0, 0) + (1 - \alpha)(0, (\theta_0/\theta_1)^n, n) \,\varepsilon\, \mathfrak{A}$. Since $n' = (1 - \alpha)n$, the desired conclusion follows.

We note that the fractional saving in the expected number of observations obtained by randomization is equal to α, *i.e.*,

$$\frac{n - n'}{n} = \frac{n - (1 - \alpha)n}{n} = \alpha.$$

6. Tests on the mean of a rectangular distribution when the range is known. Let

$$\begin{aligned} f(x, \theta) &= 1 && \text{if } \theta < x < \theta + 1, \\ &= 0 && \text{elsewhere.} \end{aligned}$$

We wish to test the hypothesis $H_0\colon \theta = 0$ against the alternative $H_1\colon \theta = \theta_1$ where $0 < \theta_1 < 1$. A simple calculation shows that

$$\begin{aligned} A_n = \{(\alpha, \beta, n)\colon \alpha = (1 - t)^n, \quad \beta = (1 - \theta_1)^n - (1 - t)^n, \\ \theta_1 \leqq t \leqq 1\} = \{(\alpha, \beta, n)\colon 0 \leqq \alpha \leqq (1 - \theta_1)^n, \alpha + \beta = (1 - \theta_1)^n\}. \end{aligned}$$

See Fig. 8. Let $R_\pi(n) = \min_{(\alpha,b,n)\varepsilon A_n} [\pi\alpha + (1 - \pi)\beta] = \min\,[\pi(1 - \theta_1)^n, (1 - \pi)(1 - \theta_1)^n] = (1 - \theta_1)^n \min\,(\pi, 1 - \pi)$. Obviously $R_\pi(n)$ is a convex function of n. It follows that $A_n \subset \mathfrak{A}$. In other words, all fixed sample size tests are admissible in the class of mixed single sample tests.

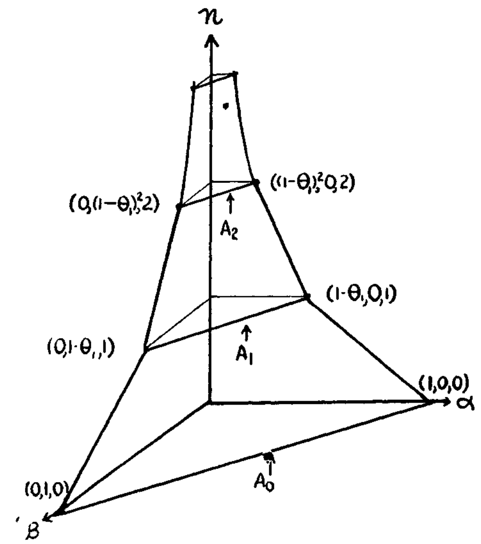

FIG. 8

7. Confidence interval estimation. We next wish to extend the notion of mixed single sample procedures to confidence interval estimation. Perhaps this purpose can best be served by an illustrative example.

Example. Let X denote a normally distributed random variable with unknown mean μ and known variance σ^2. (There is no loss in generality if we assume that $\sigma^2 = 1$, and we shall do so for the remainder of this section.) We wish to consider the problem of obtaining a confidence interval for μ. The standard procedure consists of

(a) choosing a number α between 0 and 1, called the confidence coefficient.
(b) calculating a number t using the equation $\alpha = 1 - 2\Phi(-t)$.
(c) drawing a sample of n independent observations on X and calculating $\bar{X}$, the sample mean.
(d) making the statement tha the interval $(\bar{X} - t/\sqrt{n}, \bar{X} + t/\sqrt{n})$ covers μ with confidence α.

A confidence interval procedure is evaluated in terms of a triple $(1 - \alpha, L, n)$ where $1 - \alpha$ denotes the probability that the confidence interval will not cover μ, L denotes the length of the confidence interval and n denotes the sample size.

We will now exploit the notion of randomizing over the sample size in confidence interval estimation using an approach similar to the one used in Section 2. For integral values of $n \geqq 1$, we let

$$A_n = \left\{(1 - \alpha, L, n) : \alpha = 1 - 2\Phi(t), L = \frac{2t}{\sqrt{n}}, 0 \leqq t < \infty\right\}$$

$$= \left\{(1 - \alpha, L, n) : \alpha = 1 - 2\Phi\left(-\frac{\sqrt{n}L}{2}\right)\right\}.$$

We define $A_0 = \{(1, 0, 0)\}$.

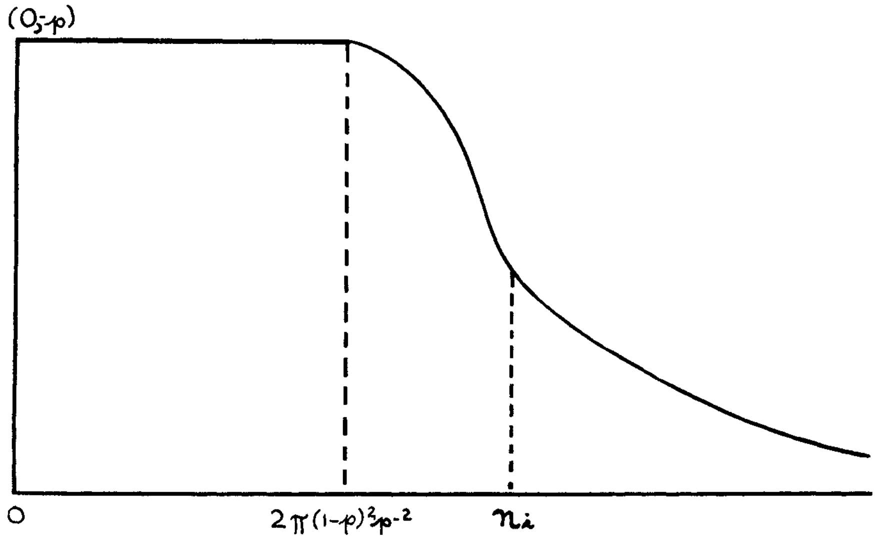

FIG. 9

As an analogue of the Bayes risk $R_\pi(n)$, we consider

$$R_p(n) = \min_{(1-\alpha, L, n) \in A_n} [p(1-\alpha) + (1-p)L].$$

A routine calculation shows that

$$R_p(n) = p \qquad \text{if } 0 \leqq n \leqq 2\pi(1-p)^2p^{-2},$$

$$= 2p\Phi\left(-\sqrt{\log\frac{np^2}{2\pi(1-p)^2}}\right) + \frac{2(1-p)}{n}\sqrt{\log\frac{np^2}{2\pi(1-p)^2}} \qquad \text{if } n > 2\pi(1-p)^2p^{-2},$$

See Fig. 9.

If we treat $R_p(n)$ as a function of a continuous variable n, we find that

$$R_p'(n) = 0 \qquad \text{if } n < \frac{1}{c},$$

$$R_p''(n) = -\frac{(1-p)}{2}\frac{1-3(\log cn)}{\sqrt{n^5}\sqrt{\log cn}} \qquad \text{if } n > \frac{1}{c},$$

where $c = p^2/[2\pi(1-p)^2]$. As in Section 3, there exists a non-negative number $n_i = n_i(p)$ such that $R_p(n)$ is concave on the interval $(0, n_i)$ and convex on the interval (n_i, ∞). In fact, $n_i = [2\pi(1-p)^2]/p^2e^{\frac{1}{3}}$. Using an argument similar to the one used in Section 3, it becomes clear that "improved" mixed confidence interval procedures exist and are of the form

$$(\alpha, L, n) = \gamma(0, 0, 0) + (1-\gamma)(\alpha', L', n'),$$

where $0 < \gamma$ 1 and (α', L', n') is a fixed sample size confidence interval procedure.

TABLE 3

Confidence coefficient α	Expected sample size n	Length of fixed sample size procedure having the given α and n	Expected length of randomized confidence interval procedure having the given α and n	Percent decrease in the expected length
.044	1	.110	.037	66
.392	9	.334	.329	1
.174	4	.220	.146	34

Table 3 gives some examples of admissible mixed single sample procedures and improvements which can be obtained in the expected length of a confidence interval if a mixing scheme is used.

Improved randomized confidence intervals are of such a nature that certain questions are brought to mind. First, how much "confidence" can we place in randomized confidence intervals? It is true that a confidence interval of the form $(\alpha, L, n) = \gamma(0, 0, 0) + (1 - \gamma)(\alpha', L', n')$ will cover μ 100 $\alpha\%$ of the time, will have average length L and will have expected sample size n. However, if we are given confidence interval (0, 0, 0), we no longer have confidence α that we are covering μ. On the other hand, if we are given the confidence interval $(\bar{X} - L'/2, \bar{X} + L'/2)$, we have confidence $\alpha' > \alpha$ that we are covering μ. Furthermore, if a statistician uses a mixed procedure and does not tell this to his customers, then his customers can have confidence α—unless, of course, they are given the procedure (0, 0, 0). (However, if we restrict ourselves to procedures where the sample size n is at least 1, then they could still have confidence α.) In other words, by withholding information from his customers, the statistician gives them confidence α. By giving them information, he either reduces their confidence to 0, or increases their confidence to α'.

This is not the only example of such a situation in statistical techniques. Take, for example, the Stein two sample procedure for finding a confidence interval (of fixed length 1 and confidence coefficient α) for the mean of a normal distribution with unknown variance. A sample of n_0 observations is taken and the sample variance S_0^2 is calculated. Then, an additional n_1 observations are taken where

$$n_1 = \max\left\{n_0, \left[\frac{S_0^2}{d}\right] + 1\right\} - n_0,$$

where d depends on α and 1. The two samples are then combined, the mean $\bar{X}$ of the combined samples is calculated and the confidence interval $\left(\bar{X} - \frac{l}{2}, \bar{X} + \frac{l}{2}\right)$ is given. Now, if it turns out that the variance S^2 of the combined samples is much larger than S_0^2, one is led to believe that the second sample size was not large enough. Thus, one's confidence of α might be reduced, given this information. However, if one did not have this information about S^2, then one's confidence would still be α. This situation is indeed similar to the preceding one.

Another peculiarity of mixed single sample confidence interval procedures is that we get short length only when we do not cover μ. This immediately brings to mind the question of average length as a criterion for a confidence interval procedure. It is clear that small length is desirable if μ is being covered. What one wants when μ is not covered is open to question. Clearly, we can agree that procedures which give small length when μ is not covered and large length when μ is covered are not desirable ones. Randomized procedures are of this nature.

8. The k decision problem. Let X denote a random variable with distribution function $F(x, \theta)$. Instead of considering only two possible values of θ, θ_0 and θ_1, as we did in the previous section, we now consider k possible values of θ. Let $\theta_1, \theta_2, \cdots, \theta_k$ denote the k possible values of θ. We assume that $\theta_1 < \theta_2 < \cdots < \theta_k$. For any fixed sample size decision rule δ_n, based on samples of size n, let $\alpha_i(\delta_n)$ denote the probability that θ_i will not be selected as the true value of θ when θ_i is the true value of θ if the decision rule δ_n is used. Every fixed sample size decision rule is then identified with an operating characteristic $(\alpha_1, \alpha_2, \cdots \alpha_k, n)$ where $\alpha_i = \alpha_i(\delta_n)$ for $i = 1, 2, \cdots, k$ and where n denotes the sample size. The classes S_n, A_n, A, A^* and $\mathfrak{A}$ are defined in an obvious way and the functions $R_\pi(n)$ and $r_\pi(n)$ are defined as in Section 2 where $\pi = (\pi_1, \pi_2, \cdots \pi_k)$, $\pi_i \geqq 0$ and $\sum_{i=1}^k \pi_i = 1$. We can then extend all the results obtained in Section 2 to the k decision problem.

In the particular case

$$F(x, \theta) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{t-\theta}{\sigma}\right)^2\right\} dt,$$

where $\sigma > 0$ is known, we shall show that it is possible to obtain improvements by randomization. For each positive integral value of n, an essentially complete class of decision rules, C_n, can be generated in the following way: Let $(x_1, x_2, \cdots x_n)$ denote a sample of n independent observations on X and let let $(t_0, t_1, \cdots, t_k)$ denote a partition of the real line such that $t_i \leqq t_{i+1}$, $i = 0, 1, \cdots k-1$. In particular, $t_0 = -\infty$ and $t_k = \infty$. Then any procedure which selects θ_i as the true value of θ whenever $t_{i-1} \leqq \bar{X} < t_i$ is called a monotone procedure. Let C_n denote the class of all monotone procedures. The class C_n is known to be essentially complete.

By definition,

$$\begin{aligned} R_\pi(n) &= \min_{(\alpha_1,\alpha_2,\cdots,\alpha_k)\varepsilon A_n} \sum_{i=1}^k \pi_i \alpha_i = \min_{(\alpha_1,\alpha_2,\cdots,\alpha_k)\varepsilon C_n} \sum_{i=1}^k \pi_i \alpha_i \\ &= \min_{(t_1,t_2,\cdots,t_{k-1})} \sum_{i=1}^k \pi_i \left[1 - \Phi\left(\sqrt{n}\,\frac{(t_i - \theta_i)}{\sigma}\right) + \Phi\left(\sqrt{n}\,\frac{(t_{i-1} - \theta_i)}{\sigma}\right)\right] \\ &= \sum_{i=1}^k \pi_i \left[1 - \Phi\left(\frac{\xi_i}{\sqrt{n}\delta_i} - \frac{\sqrt{n}\delta_i}{2}\right) + \Phi\left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right)\right], \end{aligned}$$

where

$$\xi_i = \log \frac{\pi_{i+1}}{\pi_i}, \qquad \delta_i = \frac{\theta_i - \theta_{i+1}}{\sigma}, \qquad i = 1, 2, \cdots, k-1,$$

$$\frac{\xi_0}{\sqrt{n}\delta_0} + \frac{\sqrt{n}\delta_0}{2} = -\infty \quad \text{and} \quad \frac{\xi_k}{\sqrt{n}\delta_k} - \frac{\sqrt{n}\delta_k}{2} = \infty.$$

Considering $R_\pi(n)$ as a function of a continuous variable n, we find

$$\text{(a)} \quad R'_\pi(n) = \sum_{i=2}^{k} \pi_i \, \rho\left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right) \cdot \frac{\delta_{i-1}}{\sqrt{n}}.$$

$$\text{(b)} \quad R''_\pi(n) = \sum_{i=2}^{k} \frac{-1}{\delta_n^{5/2}\,\delta_{i-1}} \left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right)(\delta_{i-1}^4 n^2 + 4\delta_{i-1}^2 n - 4\xi_{i-1}^2).$$

For each value of $i = 2, 3, \cdots k$, the function $f_i(n) = \delta_{i-1}^4 n^2 + 4\delta_{i-1}^2 n - 4\xi_{i-1}^2$ is a quadratic function of n. Since the only non-negative root of the equation $f_i(n) = 0$ is

$$n_i = \frac{-2 + 2\sqrt{1 + \xi_{i-1}^2}}{\delta_{i-1}^2},$$

it follows that

$$f_i(n) \leqq 0 \qquad \text{if } 0 \leqq n \leqq n_i,$$

$$f_i(n) > 0 \qquad \text{if } n > n_i.$$

Then, since $\delta_{i-1} < 0$ for $i = 2, 3, \cdots, k$, it follows that

$$R''_\pi(n) < 0 \qquad \text{if } n < \min_i (n_i)$$

and

$$R''_\pi(n) > 0 \qquad \text{if } n > \max_i (n_i)$$

Hence, if we let $a = \min_i (n_i)$ and $b = \max_i (n_i)$, it follows that $R_\pi(n)$ is concave on the interval $(0, a)$ and convex on the interval (b, ∞). Clearly, $a \leqq b$.

Thus, for certain values of $\pi_1, \pi_2, \cdots \pi_k$ and $\theta_1, \theta_2, \cdots, \theta_k$, it is possible to achieve improvements by randomization.

9. Testing a composite hypothesis against a composite alternative. We next wish to extend the notion of mixed single sample tests to the problem of testing a composite hypothesis againxt a composite alternative. To fix ideas, let

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\,\delta} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\},$$

where $\sigma > 0$ is known. We wish to test the hypothesis $H_0: \theta \leqq \theta_0$ against the alternative $H_1: \theta > \theta_1, \theta_1 > \theta_0$. If we are given α and n, the "best" fixed sample size test of level α and size n is obtained by using the best fixed sample size test of $H'_0: \theta = \theta_0$ against $H'_1: \theta = \theta_1$ corresponding to the given α and n. The resultant fixed sample size test has the desirable property that its power function $P(\theta \mid \alpha, n)$ tends to 1 as θ tends to infinity.

Can we construct, for given α and n, a "good" mixed single sample test of level α and *expected* sample size n in an analogous way? Clearly, if the best mixed single sample test of H'_0 against H'_1 is a bona fide mixture, it is not even true that its power function, $P(\theta)$, approaches 1 as θ approaches infinity. For, in this case, the fixed sample size test (0, 1, 0) will be chosen with probability λ, say, where $0 < \lambda < 1$, so that $P(\theta) \leqq 1 - \lambda$ for all θ.

However, it should be noted that the fact that $P(\theta)$ does not tend to 1 as θ tends to infinity is not always undesirable for we know, in certain cases, that the set of possible values of θ is bounded, e.g., in testing the mean height θ of American soldiers, we know that $\theta \leqq 6$ feet 2 inches. Consequently, a test procedure which does not have high power at $\theta = 7$ feet is not necessarily undesirable.

Finally, we note that if we restrict ourselves to randomizing over fixed sample size tests of sample size $n > 1$, then $P(\theta) \to 1$ as $\theta \to \infty$.

10. Comparison with the Wald Sequential Probability Ratio Test. In general, it is difficult to compare the improvements attainable by using the Wald Sequential Probability Ratio Test with improvements attainable by randomizing over fixed sample size procedures. For, every test will now be identified with a quadruple $(\alpha, \beta, E_{\theta_0}(n), E_{\theta_1}(n))$. $E_{\theta_0}(n)$ and $E_{\theta_1}(n)$ are usually difficult to calculate. However, in the case of mixed single sample tests, $E_{\theta_0}(n) = E_{\theta_1}(n)$ and do not depend on the unknown value of θ. In some special cases it is easy to make a comparison and this we shall do.

Example.

$$\begin{aligned} f(x, \theta) &= \frac{1}{\theta} \qquad && \text{if } \theta \leqq x < \theta, \\ &= 0 && \text{elsewhere.} \end{aligned}$$

It can be shown that if we use Wald's test, only two types of tests are attainable. They are the test (1, 0, 0, 0) or tests of the form

$$\left(0, \left(\frac{\theta_0}{\theta_1}\right)^k, k, \frac{1 - \left(\frac{\theta_0}{\theta_1}\right)^k}{1 - \left(\frac{\theta_0}{\theta_1}\right)}\right),$$

where k is a non-negative integer. However, using mixed single sample tests, we can attain the test (1, 0, 0, 0) and tests of the form $(0, (\theta_0/\theta_1)^k, k, k)$ where k is a non-negative integer, and mixtures of such tests. Since

$$\lim_{\frac{\theta_0}{\theta_1} \to 1} \frac{k}{\left[\dfrac{1 - \left(\frac{\theta_0}{\theta_1}\right)^k}{1 - \left(\frac{\theta_0}{\theta_1}\right)}\right]} = 1,$$

it is clear that if θ_0/θ_1 is close to 1, then mixed single sample procedures are almost as good as Wald procedures.

11. Estimation. Can mixing fixed sample estimation procedures yield improvements in estimation techniques? If we evaluate a fixed sample size estimator t_n in terms of a pair of numbers $\{E[L(t_n, \theta)], n\}$, where $E[L(t_n, \theta)]$ denotes the expected loss if the estimator t_n is used when θ is the true parameter and where n denotes the sample size, then mixing over fixed sample size procedures will not yield improvements since in all problems of practical interest $E[L(t_n, \theta)]$ is a convex function of n. For example, if we wish to estimate the mean θ of a distribution with finite variance σ^2, then, if $t_n = \bar{X}$ and if $L(t_n, \theta) = k(\bar{X} - \theta)^2$, we find that $E[L(t_n, \theta)] = k\,\sigma^2/n$. Thus, it will not pay to randomize.

12. Conclusion. In what situations is a mixed single sample procedure justifiable? In order to answer this question, we must first realize that throughout this paper, we have been judging a test δ by its operating characteristic (α, β, n). If this triple is our only means of evaluating a test procedure, then it is true that single sample procedures would not be justifiable since a sequential probability ratio test achieving the given α and β would be better. However, practical considerations might limit one to a single stage of sampling, e.g., in agricultural experiments, one might not wish to use more than one stage of sampling; or, if one is testing electric light bulbs, one might not wish to test the bulbs sequentially. Other examples could be given.

One could reasonably ask why fixed sample size procedures should not always be used in these situations. Presumably, if the experiment were a so called "one shot affair", i.e., if the experiment were never to be repeated, then one might reasonably insist on a non-randomized fixed sample size procedure (although, of course, this position is not universally held). However, if one repeats the experiment often, it would be reasonable to use a mixed sample size procedure. To illustrate this point, consider Example II in Section 4. In this example, suppose θ_0 represents the probability that a person who has been contaminated with a certain disease will respond positively to a certain test and θ_1 represents the probability that a person who has not been contaminated will respond positively to this same test. Then, if several thousand people are to be classified as either contaminated or non-contaminated according to this test, then the mixed test (1/101, .525, 1) would be preferred to the test (.05, .525, 1) since the mixed test will falsely classify less than 1 percent of the contaminated people whereas the fixed sample size procedure will misclassify 5 percent of the contaminated people. On the other hand, both tests will misclassify the same percentage of non-contaminated people, and both procedures will use on the average of one test per person.

At this point, one could raise strenuous objections to mixed single sample tests on grounds similar to those raised in Section 7, i.e., if one is told which single sample test is actually used, the conditional probabilities of misclassification are no longer α and β. For example, consider a mixed test of the form

$$(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha', \beta', n').$$

Now, suppose that a person is told that he has been classified according to the test (0, 1, 0). Such a person would of course be most unhappy. On the other hand, if he is not told which of the tests was used, he would maintain his confidence in the procedure used. In other words, *by withholding information, one can influence a person's willingness to accept a result.* Some feel that axiomatically this is an untenable policy.

13. Acknowledgements. The author is indebted to Allan Birnbaum for suggesting the Bayes approach exploited in this paper. The author is also indebted to Howard Raiffa for his many helpful suggestions and constructive criticisms.

REFERENCES

[1] W. Kruskal, Private communication.
[2] H. Raiffa, "Statistical decision theory approach to item selection for dichotomous test criterion variables," School of Aviation Medicine Series in Statistics, Report No. 56–139, Randolph Field, 1957.
[3] A. Birnbaum, "Applications of vector-valued risk functionals," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 544–545 (abstract).
[4] J. L. Hodges, Jr., and E. L. Lehmann, "The efficiency of some non-parametric competitors of the t-test", *Ann. Math. Stat.*, Vol. 27, No. 2 (1956) pp. 324–335.