SOME PROPERTIES OF A CLASS OF BAYES TWO-STAGE TESTS

By Morris Skibinsky

Purdue University[1]

Summary. Let X be a normal random variable with variance one and mean either $\pm a/2$, where a is a given positive constant. Let $C_m(m \geqq 0)$ denote the class of all two-stage rules with first sample size of m for deciding, after two successive samples of independent observations on X, which of the two mean values is correct. This paper investigates the class of Bayes rules in C_m, parametrized by *a priori* probabilities on the hypotheses, and simple wrong decision losses. The cost per observation is taken throughout to be unity.

Section 1 gives some general properties of Bayes rules in C_m for decisions between any two continuous densities for X; Sections 2, 3, and 4 concern the densities specified above. Section 2 consists of a detailed development of Bayes second sample size properties in terms of the Bayes parameters and first sample outcomes. For example, Theorem 2.1 gives non-trivial lower and upper bounds for positive values of the Bayes second sample size corresponding to any fixed value for the minimum wrong decision loss. In Section 3, sufficient conditions are given under which the losses may be chosen so as to obtain Bayes rules with preassigned invariant error probabilities. (Invariance is taken with respect to changes in the prior probabilities.) It is shown how this result leads to rules which minimize the maximum expected sample size among rules in C_m with error probabilities less than or equal to specified values. An illustrative example is considered for the case when these specified bounds are equal. The selection of an optimum first sample size for this example is treated in Section 4. The resulting rule has the above described good property among all two-stage rules (of any first sample size) subject to these bounds. Tables are included giving optimum first and second sample sizes and the values of auxiliary functions when this common bound on the error probabilities is .05 and .01.

1. Introduction. Let f_0 and f_1 be two given probability densities, and suppose that X is a random variable which has a probability density known, *a priori*, to be either f_0 or f_1. By the decision i we shall mean the decision to accept f_i as the true density of X. Unless otherwise explicitly noted, the index i will always take on the values 0, 1. Our problem is parametrized by four positive numbers, g_i, W_i, with

$$g_0 + g_1 = 1. \tag{1.1}$$

It will be helpful to regard g_i as the "*a priori* probability" of f_i, and W_i, as the loss incurred when f_i is the density of X, and the decision $1 - i$ is made, although this interpretation is not essential.

Received, October 10, 1958; revised August 3, 1959.

[1] Work on Section 4 was done at the Brookhaven National Laboratory.

The *a priori* probabilities, in view of (1.1), are uniquely determined by their ratio, $g = g_0/g_1$. The losses are uniquely determined by their ratio and minimum, $W = W_0/W_1$, $M = \min(W_0, W_1)$. These three numbers in turn uniquely determine the *a priori* probabilities and losses. i.e.

$$g_0 = g/(1 + g), g_1 = 1/(1 + g), \quad \text{and} \quad W_0 = M\psi(W), W_1 = M\psi(1/W),$$

where

$$\psi(\zeta) = \begin{cases} 1, & \zeta \leqq 1 \\ \zeta, & \zeta \geqq 1 \end{cases}$$

We shall refer as convenience requires, sometimes to one, (g_0, g_1, W_0, W_1), sometimes to the other, (g, W, M), of these two equivalent sets of parameters.

Let $X_1, X_2, \cdots$ denote a sequence of independent random variables identically distributed with X. We shall refer to the members of this sequence as observations. Denote by $\mathbf{X}_j$ the vector of the first j observations. Let $E_i\cdot$ denote the expectation operator under f_i; $E_i(\cdot \mid \mathbf{X}_k)$, the conditional expectation, under f_i, given the first k observations; and $E(\cdot)$, the operator, $g_0E_0(\cdot) + g_1E_1(\cdot)$.

We shall consider the class, C_m, of two-stage rules S for deciding between f_0 and f_1 which depend on a non-negative first sample size m; a non-negative second sample size, $\nu_m(\mathbf{X}_m)$, dependent on the outcome of the first sample; and a randomized terminal decision probability, $D_n(\mathbf{X}_n)$, $n = m + \nu_m(\mathbf{X}_m)$, dependent on the outcome of both samples, where we must accept f_1 with this probability or f_0 with one minus this probability. We suppose, for $m > 0$, that the second sample size and terminal decision functions are always measurable with respect to m and n dimensional Borel sets, respectively, and that the expectations, defined below, exist. It can be shown that no loss in generality derives, in the present case, from failure to consider procedures which randomize first and second sample sizes.

The expected overall sample size, under f_i, associated with a rule S in C_m, is $\mathcal{E}_i(S) = m + E_i\nu_m(\mathbf{X}_m)$. The probability, under f_i, that it will lead to decision $1 - i$, may be written

$$Q_i(S) = E_iV_{im}(\mathbf{X}_m, \nu_m(\mathbf{X}_m)),$$

where

$$V_{im}(\mathbf{X}_m, \nu) = i + (1 - 2i)E_i[D_{m+\nu}(\mathbf{X}_{m+\nu}) \mid \mathbf{X}_m]$$

is the conditional probability, under f_i, given the first sample observations (and second sample size equal to ν), that the rule will lead to decision, $1 - i$.

The average risk associated with a rule, S, in C_m is defined to be

$$R(g, W, M \mid S) = \sum_{i=0}^{1} g_i[\mathcal{E}_i(S) + W_iQ_i(S)]. \tag{1.2}$$

We define a Bayes two-stage rule (with respect to g, W, M), for deciding between f_0 and f_1 to be a rule S in C_m for which ν_m and D_n minimize the average risk.

It is easy to show that a terminal decision function which minimizes the average risk with respect to any given triplet of parameters, g, W, M, for any overall sample size n and any overall sample $\mathbf{X}_n$ is given by

$$D_n^*(\mathbf{X}_n, gW) = (1, \tfrac{1}{2}, 0) \tag{1.3}$$

according as $\prod_{j=1}^n f_1(X_j)/f_0(X_j)(>, =, <)gW$, where the choice of $\frac{1}{2}$ in the case of equality has been made simply for convenience in later definitions, and the product is defined to be one, when $n = 0$.

We may express the average risk (1.2) in the form

$$R(g, W, M \mid S) = m + EL_m(\mathbf{X}_m, \nu_m(\mathbf{X}_m), g, W, M),$$

where

$$L_m(\mathbf{X}_m, \nu, g, W, M) = \nu + \sum_{i=0}^{1} W_i g_{im}(\mathbf{X}_m) V_{im}(\mathbf{X}_m, \nu) \tag{1.4}$$

is the conditional average expected loss associated with S, given the first sample observations, $\mathbf{X}_m$, and second sample size equal to ν, and

$$g_{im}(\mathbf{X}_m) = g_i \prod_{j=1}^{m} f_i(X_j) \Big/ \sum_{k=0}^{1} g_k \prod_{j=1}^{m} f_k(X_j)$$

is the "*a posteriori* probability," given the first sample observations, that f_i is the density of X.

Since the values of L_m are always bounded below by ν, there must exist a second sample size which minimizes the average risk with respect to any given triplet of parameters, g, W, M, and first sample outcome $\mathbf{X}_m$. When the terminal decision function is given by (1.3), we shall denote the values of such a function by $\nu_m^*(\mathbf{X}_m, g, W, M)$. The form of the terminal decision function (1.3) then implies that $0 \leqq \nu_m^*(\mathbf{X}_m, g, W, M) \leqq \min [W_0 g_{0m}(\mathbf{X}_m), W_1 g_{1m}(\mathbf{X}_m)] \leqq M$. We shall refer to ν_m^* as a Bayes second sample size function. In an analogous way it can be shown that a Bayes first sample size exists and must always lie in the interval $[0, 2M]$.

Let $S^*(g, W, M)$ denote the rule in C_m with second sample size function, ν_m^*, and terminal decision function (1.3). Clearly, $S^*(g, W, M)$ is a Bayes rule. We state the following immediate consequence of this fact for later reference.

LEMMA 1.1. *If S is any rule in C_m such that $Q_i(S) \leqq Q_i(S^*(g, W, M))$, then* $\sum_{i=0}^1 g_i \mathcal{E}_i(S^*(g, W, M)) \leqq \sum_{i=0}^1 g_i \mathcal{E}_i(S)$.

2. The Bayes second sample size function. In that which follows, we shall be concerned with the class of two-stage rules $S^*(g, W, M)$, defined above, for an arbitrary value of g, as W and M are allowed to vary. The problem will be to distinguish specifically between the two densities

$$f_i(x) = (2\pi)^{-1} \exp \{-\tfrac{1}{2}[x + (\tfrac{1}{2} - i)a]^2\} \tag{2.1}$$

(where a is some given positive constant). We shall throughout use the notation

$$\phi(\lambda) = (\sqrt{2\pi})^{-1} \int_\lambda^\infty \exp [-\tfrac{1}{2}z^2]\, dz$$

The present section contains an outline of basic functional properties associated with the Bayes second sample size and related functions, arranged as a sequence of lemmas and theorems in the order of their dependence. Proofs are mostly omitted due to space requirements. Some of greater interest or importance are sketched.

In view of the specified densities f_i, we may write our terminal decision function (1.3) as $D_n^*(\mathbf{X}_n, gW) = (1, \frac{1}{2}, 0)$ according as $an\bar{X}_n(>, =, <) \ln gW$, where $\bar{X}_n = \sum_{j=1}^n X_j/n$, $n > 0$, $\bar{X}_0 = 0$.

The Bayes second sample size $\nu_m^*(\mathbf{X}_m, g, W, M)$ associated with the rule $S^*(g, W, M)$ is, for fixed values of its arguments, a value of $\nu \geqq 0$, with respect to which (1.4) is minimum. For convenience in treatment, we may (again in view of the specified densities) express the function (1.4) in the following form.

$$L_m(\mathbf{X}_m, \nu, g, W, M) \equiv a^{-2}\mathcal{L}(a^2\nu, am\bar{X}_m - \ln gW, W, a^2M),$$

where

$$(2.2) \qquad \mathcal{L}(y, t, \zeta, \mu) = y + \mu\psi(\zeta)[\alpha(y, t) + e^t\alpha(y, -t)]/(1 + \zeta e^t),$$

for $0 \leqq y, \zeta, \mu < \infty$, $-\infty < t < \infty$, and

$$(2.3) \qquad \alpha(y, t) = \phi(.5\sqrt{y} - t/\sqrt{y}), \; y > 0,$$

$$\alpha(0, t) = \lim_{y \to +0} \alpha(y, t) = (0, \tfrac{1}{2}, 1) \text{ as } t\ (<, =, >)0.$$

Observe that

$$(2.4) \qquad \mathcal{L}(y, -t, 1/\zeta, \mu) \equiv \mathcal{L}(y, t, \zeta, \mu).$$

LEMMA 2.1. *For fixed, positive* ζ, μ, *the equation* $\partial\mathcal{L}/\partial y = 0$,

(a) *has* 2 *positive roots in* y *whenever* $-\hat{t}(1/\zeta, \mu) < t < \hat{t}(\zeta, \mu)$, $t \neq 0$,

(b) 1 $\quad t = 0$,

(c) 1 $\quad t = -\hat{t}(1/\zeta, \mu)$ *or* $t = \hat{t}(\zeta, \mu)$,

(d) 0 $\quad$ *otherwise*,

where $\hat{t}(\zeta, \mu)$ *is the unique positive root in* t *of the equation*

$$(\partial\mathcal{L}/\partial y)_{y=G(t)} = 0, \qquad G(t) = 2(\sqrt{t^2 + 1} - 1).$$

With respect to its argument y the function $\mathcal{L}$ has, in case (a), a unique relative maximum at the smaller, and a unique relative minimum at the larger of the two roots; in case (b), it has a unique absolute minimum at the single root; in case (c), $\mathcal{L}$ is strictly increasing in y, except at the single root; in case (d), it is strictly increasing in y.

We note that the function $G(t)$, which is defined in the above lemma, is for all ζ, $\mu > 0$, and all t, the unique root in y of $\partial^2\mathcal{L}/\partial y^2 = 0$.

LEMMA 2.2. $\hat{t}(\zeta, \mu)$ *is*

(1) *for fixed* $\mu > 0$, *a positive, bounded function of* $\zeta > 0$, *strictly monotonic to either side of a minimum at* $\zeta = 1$.

(2) *for fixed* $\zeta > 0$, *a strictly increasing function of* μ *which tends to* 0 *as* $\mu \to 0$, *and to* ∞, *as* $\mu \to \infty$ (*uniformly for all* $\zeta > 0$).

(3) *a continuous function of non-negative* ζ *and positive* μ.

LEMMA 2.3.

(1) *For any fixed positive* ζ, *bounded away from zero*, $\lim_{\mu\to\infty} [\hat{t}(\zeta, \mu)/\ln \mu] = 1$.

(2) $\lim_{\mu\to\infty} [\lim_{\zeta\to 0} \hat{t}(\zeta, \mu)/\mu^2] = 1/16\pi$.

Let

$$
\begin{aligned}
A &= \{(t, \zeta, \mu): -\infty < t < \infty, \zeta, \mu > 0\},\\
A_0 &= \{(t, \zeta, \mu): t = 0, \zeta, \mu > 0\},\\
\hat{A} &= \{(t, \zeta, \mu): -\hat{t}(1/\zeta, \mu) < t < \hat{t}(\zeta, \mu), \zeta, \mu > 0\},\\
\hat{A}_0 &= \{(t, \zeta, \mu): t = -\hat{t}(1/\zeta, \mu) \text{ or } t = \hat{t}(\zeta, \mu), \zeta, \mu > 0\},\\
\hat{A}_1 &= \hat{A} \cup \hat{A}_0 .
\end{aligned}
$$

For all $(t, \zeta, \mu) \varepsilon \hat{A}_1$, we define $\hat{y}(t, \zeta, \mu)$ to be either the larger or the unique root in y of the equation: $\partial\mathcal{L}/\partial y = 0$. By Lemma 2.1, $\mathcal{L}$ has a unique relative minimum with respect to y at $y = \hat{y}$, for all $(t, \zeta, \mu) \varepsilon \hat{A}$. On the other hand, $\mathcal{L}$ is absolutely minimum at $y = 0$, for all $(t, \zeta, \mu) \varepsilon A - \hat{A}$.

Now for all $(t, \zeta, \mu) \varepsilon \hat{A}_1$, let

$$(2.5) \qquad U(t, \zeta, \mu) = \mathcal{L}(\hat{y}(t, \zeta, \mu), t, \zeta, \mu) - \mathcal{L}(0, t, \zeta, \mu).$$

and define

$$(2.6) \qquad y^*(t, \zeta, \mu) = \begin{cases} \hat{y}(t, \zeta, \mu), & (t, \zeta, \mu) \varepsilon \hat{A}_1 \text{ and } U(t, \zeta, \mu) < 0,\\ 0, & \text{elsewhere in } A. \end{cases}$$

Observe that by (2.4),

$$(2.7) \qquad \hat{y}(t, \zeta, \mu) \equiv \hat{y}(-t, 1/\zeta, \mu),$$

and that this symmetry holds also for U and y^*.

LEMMA 2.4.

(1) $\mathcal{L}$ *has a unique absolute minimum with respect to* y *at* $y = y^*$, *for all* $(t, \zeta, \mu) \varepsilon A$ *with the exception of those points for which* $U(t, \zeta, \mu) = 0$. *When* $U(t, \zeta, \mu) = 0$, $\mathcal{L}$ *is absolutely minimum with respect to* y *at both* $y = y^* = 0$ *and* $y = \hat{y} > 0$.

(2) $y^*(t, \zeta, \mu)$ *is bounded above by* $\mathcal{L}(0, t, \zeta, \mu) = \mu\psi(\zeta)e^{.5(t-|t|)}/(1 + \zeta e^t) < \mu$.

(3) $U(t, \zeta, \mu) \leqq 0$ *implies that* $(t, \zeta, \mu) \varepsilon \hat{A}$.

(4) $U(0, \zeta, \mu) < 0$.

It will be convenient for our purpose from this point on to regard any second sample size ν as ranging over the non-negative real numbers rather than restricting it to 0 and the positive integers. As a consequence of this relaxation, it follows from (2.1) and part (1) of the above lemma, that except when

$$U(am\bar{X}_m - \ln gW, W, a^2M) = 0, \tag{2.8}$$

the Bayes second sample size function is unique and indeed

$$\nu_m^*(\mathbf{X}_m, g, W, M) = a^{-2}y^*(am\bar{X}_m - \ln gW, W, a^2M). \tag{2.9}$$

Thus, in effect, the Bayes second sample size is reduced to dependence on only three arguments. In the exceptional case, which will be shown always to have probability zero, we may, by part (1) of the above lemma, choose ν_m^* to be either 0 or $a^{-2}\hat{y}$.

Let $v(t, \zeta) = 2t[2(1 - \zeta e^t)^{-1} - 1]$. By straightforward calculation, we find that for ζ positive but not equal to one, and every positive y, μ, $\partial\mathcal{L}/\partial y$ is monotonic in t to either side of a minimum at the unique value of t for which $v(t, \zeta) = y$. When $\zeta = 1$, and y and μ are positive, $\partial\mathcal{L}/\partial y$ is monotonic in t to either side of a minimum at $t = 0$.

We are now in a position to describe the behavior of the value of t for which $\hat{y}$ is maximum. Thus, we introduce, below, the function T. (See statement one of Lemma 2.6).

LEMMA 2.5. *For all* $\zeta, \mu > 0, \zeta \neq 1$, *the equation,*

$$\partial\mathcal{L}/\partial y \,|_{y=v(t,\zeta)} = 0,$$

has a unique root in t. *This root,* $t = T(\zeta, \mu)$, *say, has the following properties.*

(1) (a) $T(\zeta, \mu) \equiv -T(1/\zeta, \mu)$.
 (b) $0 < T(\zeta, \mu) < \min[-\ln\zeta, \hat{t}(\zeta, \mu)]$, *whenever* $0 < \zeta < 1$.
 (c) $\lim_{\zeta\to 0} T(\zeta, \mu) = \mu^2/16\pi$.
 (d) $\lim_{\mu\to 0} T(\zeta, \mu) = 0$, $\lim_{\mu\to\infty} T(\zeta, \mu) = -\ln\zeta$.

(2) *Define* $T(1, \mu) = 0$, *then* $T(\zeta, \mu)$ *is*
 (a) *for fixed* $\mu > 0$, *a strictly decreasing function of* $\zeta > 0$.
 (b) *for fixed* ζ, $0 < \zeta < 1$, *a strictly increasing function of* $\mu > 0$.
 (c) *continuous in* $\zeta, \mu > 0$.

LEMMA 2.6. *On the set,* $\hat{A}_1$, *over which it is defined,* $\hat{y}(t, \zeta, \mu)$ *is positive and*

(1) *for fixed* ζ, μ, *strictly monotonic in* t *to either side of a maximum at* $t = T(\zeta, \mu)$.
(2) *for fixed* t, μ, *bounded in* ζ *and strictly decreasing for* $\zeta < 1$.
(3) *for fixed* t, ζ, *unbounded and strictly increasing in* μ.
(4) *continuous in* t, ζ, μ.

By part 4 of the above lemma, $U(t, \zeta, \mu)$, defined in (2.5), is continuous in its arguments over its domain of definition, $\hat{A}_1$. Differentiating U with respect to t, we find that

$$\partial U/\partial t \gtrless 0, \qquad t \gtrless 0. \tag{2.10}$$

By part 3 of Lemma 2.4, $U(-\hat{t}(1/\zeta, \mu), \zeta, \mu) > 0$, $U(\hat{t}(\zeta, \mu), \zeta, \mu) > 0$. Hence, by part 4 of that lemma and (2.10), the equation, $U(t, \zeta, \mu) = 0$, has, for each positive ζ and μ, a unique negative and a unique positive root in t. If we denote the positive root by $t = t^*(\zeta, \mu)$, then by the symmetry (2.4), (2.7), the corresponding negative root must be $t = -t^*(1/\zeta, \mu)$. Hence, by (2.6), for each positive ζ and μ, we may write

$$(2.11) \qquad y^*(t, \zeta, \mu) = \begin{cases} \hat{y}(t, \zeta, \mu), & -t^*(1/\zeta, \mu) < t < t^*(\zeta, \mu), \\ 0, & \text{otherwise.} \end{cases}$$

It is apparent that

$$(2.12) \qquad 0 < t^*(\zeta, \mu) < \hat{t}(\zeta, \mu), \qquad \zeta, \mu > 0.$$

Also, we may now write the event (2.8) in the form

$$am\bar{X}_m = [\ln gW - t^*(1/W, a^2M)] \quad \text{or} \quad [\ln gW + t^*(\mathrm{W}, a^2M)],$$

for which it is clear that the probability of this event is always zero. We recall that the Bayes second sample size, (2.9), is unique with the exception of the above event. Hence, the Bayes rules, $S^*(g, W, M)$ in C_m are unique up to sets of probability zero.

LEMMA 2.7. *$G(\hat{t}(1, \mu)) \leqq \hat{y}(t, \zeta, \mu) < \mu^2/8\pi$, for all $(t, \zeta, \mu) \varepsilon \hat{A}_1$. ($G$ is defined in Lemma 2.1.)*

PROOF. The lower bound follows from the first parts of Lemmas 2.6 and 2.2, the fact that G is symmetric about $t = 0$ and strictly increasing in $|t|$, and the fact that by the definition of $\hat{t}$ in Lemma 2.1, $\hat{y}(t, \zeta, \mu) \equiv G(t)$ for all $(t, \zeta, \mu) \varepsilon \hat{A}_0$. The upper bound follows from Lemma 2.5 and part one of Lemma 2.6, together with the fact that $\hat{y}(T(1/\zeta, \mu), 1/\zeta, \mu) \equiv \hat{y}(T(\zeta, \mu), \zeta, \mu)$, is for fixed $\mu > 0$, strictly monotonic to either side of a minimum at $\zeta = 1$, and

$$\lim_{\zeta\to 0} \hat{y}(T(\zeta, \mu), \zeta, \mu) = \lim_{\zeta\to 0} v(T(\zeta, \mu), \zeta) = \mu^2/8\pi$$

We note that for any $\mu > 0$, the lower bound of the above Lemma is attained when $\zeta = 1$ and $t = \pm\hat{t}(1, \mu)$, the upper bound, for $t = T(\zeta, \mu)$, in the limit as $\zeta \to 0$ or ∞.

Combining the above lemma with part 2 of Lemma 2.4, and (2.9), we have, immediately the following.

THEOREM 2.1. *$a^{-2}G(\hat{t}(1, a^2M)) < \nu_m^*(\mathbf{X}_m, g, W, M) < \min[M, a^2M^2/8\pi]$, whenever*

$$[\ln gW - t^*(1/W, a^2M)] < am\bar{X}_m < [\ln gW + t^*(W, a^2M)],$$

i.e. whenever the Bayes second sample size is positive.

LEMMA 2.8.

(1) *For all finite t and for all $\zeta > 0$, $\lim_{\mu\to\infty} [\hat{y}(t, \zeta, \mu)/\ln \mu] = 8$.*

(2) *Let δ be an arbitrary, fixed, positive number less than 1, and define $t_\delta(\mu) = (1 - \delta) \ln \mu$. Then for all positive ζ bounded away from zero,*

$$\lim_{\mu\to\infty} [\hat{y}(t_\delta(\mu), \zeta, \mu)/ \ln \mu] = 2(1 + \sqrt{\delta})^2.$$

PROOF. The identity,

$$\partial\mathcal{L}/\partial y \,|_{y=\,(t,\zeta,\mu)} \equiv 0, \qquad (t, \zeta, \mu) \; \varepsilon \; \hat{A}_1,$$

may be written in the form

$$\hat{y}(t, \zeta, \mu)[1 - \rho(t, \zeta, \mu)] - 8 \ln \mu \equiv 0,$$

where

$$\rho(t, \zeta, \mu) = 4\{ \ln \hat{y}(t, \zeta, \mu) + 2 \ln [2\sqrt{2\pi}(1 + \zeta e^t)/\psi(\zeta)] - t + t^2/\hat{y}(t, \zeta, \mu)\}/\hat{y}(t, \zeta, \mu).$$

By part 3 of Lemma 2.6, for all bounded t and all $\zeta > 0$, $\rho(t, \zeta, \mu)$ tends to zero as $\mu \to \infty$, which proves part (1). When $t = t_\delta(\mu)$, the identity may be put in the form

$$[1 + 4\rho_{1\delta}(\zeta, \mu)]r_\delta^2(\zeta, \mu) - 4(1 + \delta)[1 - \rho_{2\delta}(\zeta, \mu)]r_\delta(\zeta, \mu) + 4(1 - \delta)^2 \equiv 0,$$

where

$$r_\delta(\zeta, \mu) = \hat{y}(t_\delta(\mu), \zeta, \mu)/ \ln \mu,$$

$$\rho_{1\delta}(\zeta, \mu) = \ln \hat{y}(t_\delta(\mu), \zeta, \mu)/\hat{y}(t_\delta(\mu), \zeta, \mu),$$

$$\rho_{2\delta}(\zeta, \mu) = 2 \ln [2\sqrt{2\pi}(\zeta + \mu^{\delta-1})/\psi(\zeta)]/(1 + \delta) \ln \mu.$$

Now let ζ be positive and bounded away from zero. By Lemma 2.3, when μ is sufficiently large, $t_\delta(\mu) < \hat{t}(\zeta, \mu)$. Hence $(t_\delta(\mu), \zeta, \mu) \; \varepsilon \; \hat{A}_1$ for μ sufficiently large. We may now use the lower bound of Lemma 2.7 (which tends to ∞ with μ) to show that $\rho_{1\delta}(\zeta, \mu)$ tends to zero as $\mu \to \infty$. $\rho_{2\delta}(\zeta)$ obviously tends to zero as $\mu \to \infty$. Thus, for μ sufficiently large, the quadratic equation in r,

$$[1 + 4\rho_{1\delta}(\zeta, \mu)]r^2 - 4(1 + \delta)[1 - \rho_{2\delta}(\zeta, \mu)]r + 4(1 - \delta)^2 = 0,$$

has, two real roots in r, one of which, by the identity, must be equal to $r_\delta(\zeta, \mu)$. Hence, in the limit, as $\mu \to \infty$, $r_\delta(\zeta, \mu)$ must be equal to one or the other of $2(1 \pm \sqrt{\delta})^2$. By Lemmas 2.7 and 2.3,

$$\lim_{\mu\to\infty} r_\delta(\zeta, \mu) \geqq \lim_{\mu\to\infty} G(\hat{t}(1, \mu))/ \ln \mu = 2.$$

Hence, the conclusion of part 2 follows.

LEMMA 2.9. $t^*(\zeta, \mu)$ is

(1) *for fixed $\mu > 0$, a positive, bounded function of $\zeta > 0$, strictly monotone to either side of a minimum at $\zeta = 1$.*

(2) *for fixed $\zeta > 0$, a strictly increasing function of μ which tends to 0 as $\mu \to 0$ and to ∞ as $\mu \to \infty$ (uniformly, for all $\zeta > 0$).*

(3) *a continuous function of positive ζ and μ.*

PROOF. That t^* is, for fixed $\mu > 0$, a positive, bounded function of $\zeta > 0$, and tends to zero with μ, uniformly for all $\zeta > 0$, follows from (2.12) and Lemma 2.2.

By (2.5) and (2.2), $U(t, \zeta, \mu) = \hat{y}(t, \zeta, \mu) + \mu\psi(\zeta)H(t, \zeta, \mu)/(1 + \zeta e^t)$, where for $t \geqq 0$, $H(t, \zeta, \mu) = e^t\alpha(\hat{y}(t, \zeta, \mu), -t) + \alpha(\hat{y}(t, \zeta, \mu), t) - 1$. H depends on ζ and μ only through $\hat{y}$. By Lemma 2.6, $\hat{y}$, and hence U, is continuous over $\hat{A}_1$. Thus, using (2.10) and the lower bound of Lemma 2.7, it may be shown that for every positive ζ and μ, there exists a positive number k, such that $0 \leqq t \leqq t^*(\zeta, \mu) + k \Rightarrow \partial U/\partial\zeta \gtrless 0$, $\zeta \lessgtr 1$. It then follows that for any fixed $\mu > 0$, t^* is strictly monotone to either side of a minimum at $\zeta = 1$. For suppose not, then there exist ζ', ζ'' such that either $0 < \zeta' < \zeta'' \leqq 1$ or $1 \leqq \zeta'' < \zeta' < \infty$ such that $t^*(\zeta'', \mu) \geqq t^*(\zeta', \mu)$, and hence we would have that

$$0 = U(t^*(\zeta', \mu), \zeta', \mu) < U(t^*(\zeta', \mu), \zeta'', \mu) \leqq U(t^*(\zeta'', \mu), \zeta'', \mu) = 0,$$

a contradiction. In a similar way, it may be shown that for every fixed $\zeta > 0$, $t^*(\zeta, \mu)$ is strictly increasing with μ. To complete the proof of part 2, it is sufficient to prove that for any fixed positive $\delta < 1$,

$$t^*(1, \mu) > (1 - \delta) \ln \mu = t_\delta(\mu), \qquad \mu \text{ sufficiently large.} \tag{2.13}$$

To show this, we need only show that

$$U(t_\delta(\mu), 1, \mu) < 0, \qquad \mu \text{ sufficiently large.}$$

Using Lemma 2.8 and a well known inequality on Mill's ratio [1], we find that $\lim_{\mu\to\infty} H(t_\delta(\mu), 1, \mu) = -1$. Hence, applying Lemma 2.8 once more, we find that

$$\lim_{\mu\to\infty} [U(t_\delta(\mu), 1, \mu)/\ln \mu] = 2(1 + \sqrt{\delta}) - \lim_{\mu\to\infty} [\mu/(1 + \mu^{1-\delta})] = -\infty.$$

This completes the proof of part 2. Continuity may be proved using a device employed by Wald and Wolfowitz in [6]. Let ζ, μ be arbitrary, fixed, positive. Let K_1, K_2 be any two numbers such that

$$0 < K_1 < t^*(\zeta, \mu) < K_2, \qquad K_2 - K_1 < \Delta,$$

where Δ is an arbitrarily small positive number. By (2.10),

$$U(K_1, \zeta, \mu) < 0 < U(K_2, \zeta, \mu).$$

Let $\Delta\zeta$, $\Delta\mu$ be non zero increments in ζ and μ, respectively, which tend to zero with Δ, and such that $\zeta + \Delta\zeta > 0$, $\mu + \Delta\mu > 0$, then, since U is a continuous function of its arguments, we have, for Δ sufficiently small, that

$$U(K_1, \zeta + \Delta\zeta, \mu + \Delta\mu) < 0 < U(K_2, \zeta + \Delta\zeta, \mu + \Delta\mu).$$

Hence, for Δ sufficiently small, $K_1 < t^*(\zeta + \Delta\zeta, \mu + \Delta\mu) < K_2$. This completes the proof.

3. Bayes rules with preassigned invariant error probabilities. Below, we consider the error probabilities associated with the Bayes rules, $S^*(g, W, M)$, in C_m in terms of their dependence on the Bayes parameters and the first sample size m. (a, the distance between the means of f_0 and f_1 is always regarded as arbitrary, fixed, positive.) The properties developed lead to sufficient conditions under which the parameters, W and M, may be chosen, for arbitrary, positive m and g, so that the error probabilities take on preassigned, fixed values. This leads to a class of rules, parametrized by m and g and the preassigned error probabilities, each member of which minimizes the average (g) expected number of observations among all rules in C_m with error probabilities less than or equal to those preassigned. It is pointed out how rules, within the same subclass, which minimize the maximum expected sample size, may be obtained by proper selection of g.

Let $p_i(t \mid z, \lambda) = (\sqrt{2\pi z})^{-1} \exp\{-[t + (\frac{1}{2} - i)z + \ln \lambda]^2/2z\}$. By the results obtained in the preceding section, and in particular, by (2.9), we may write the expected overall sample size, under f_i, required by the Bayes rule, $S^*(g, W, M)$, in the form

$$\mathcal{E}_i(S^*(g, W, M)) = a^{-2}\mathcal{E}_i^*(a^2m, gW, W, a^2M), \tag{3.1}$$

where

$$\mathcal{E}_i^*(z, \lambda, \zeta, \mu) = z + \int_{-\infty}^{\infty} y^*(t, \zeta, \mu)p_i(t \mid z, \lambda)\, dt, \qquad z, \lambda, \zeta, \mu > 0,$$

$$\mathcal{E}_i^*(0, \lambda, \zeta, \mu) = y^*(-\ln \lambda, \zeta, \mu), \qquad \lambda, \zeta, \mu > 0.$$

Similarly, the probability, under f_i, that $S^*(g, W, M)$ will lead to decision $1 - i$, may be written in the form

$$Q_i(S^*(g, W, M)) = Q_i^*(a^2m, gW, W, a^2M), \tag{3.2}$$

where

$$Q_i^*(z, \lambda, \zeta, \mu) \doteq \int_{-\infty}^{\infty} \alpha(y^*(t, \zeta, \mu,), (1 - 2i)t)p_i(t \mid z, \lambda)\, dt, \quad z, \lambda, \zeta, \mu > 0,$$

$$Q_i^*(0, \lambda, \zeta, \mu) = \alpha(y^*(-\ln \lambda, \zeta, \mu), (2i - 1) \ln \lambda), \quad \lambda, \zeta, \mu > 0. \tag{3.3}$$

Note that the symmetry, (2.7), implies the corresponding symmetries

$$\mathcal{E}_1^*(z, \lambda, \zeta, \mu) \equiv \mathcal{E}_0^*(z, 1/\lambda, 1/\zeta, \mu), \qquad Q_1^*(z, \lambda, \zeta, \mu) \equiv Q_0^*(z, 1/\lambda, 1/\zeta, \mu) \tag{3.4}$$

These imply, as a special case, that the rules, $S^*(1, 1, M)$ are minimax in C_m with respect to wrong decision losses that are both equal to M. This fact was noted by Wald ([7], pp. 151–156) for the case, $M = 1$.

LEMMA 3.1. *Let δ be an arbitrary, fixed, positive number less than one, then* $\lim_{\mu \to \infty} Q_0^*(z, \lambda, \zeta, \mu) = 0$, uniformly for all z, λ, ζ such that $z \geqq 0, \zeta > 0, \lambda \geqq \delta$.

PROOF. By (3.3), (2.3), (2.11), whenever $z, \lambda, \zeta, \mu > 0$,

$$Q_0^*(z, \lambda, \zeta, \mu) = \int_{t^*(\zeta,\mu)}^{\infty} p_0(t \mid z, \lambda)\, dt + \int_{-t^*(1/\zeta,\mu)}^{t^*(\zeta,\mu)} \phi(h(t, \zeta, \mu))p_0(t \mid z,\lambda)\, dt, \tag{3.5}$$

where $h(t, \zeta, \mu) = .5(\hat{y}(t, \zeta, \mu))^{\frac{1}{2}} - t/(\hat{y}(t, \zeta, \mu))^{\frac{1}{2}}$. By (2.13), for μ sufficiently large (uniformly for all z, λ, $\zeta > 0$), the right hand side of (3.5) is bounded above by

$$\int_{(1-\delta)\ln\mu}^{\infty} p_0(t \mid z, \lambda)\, dt + \int_{-t^*(1/\zeta,\mu)}^{(1-\delta)\ln\mu} \phi(h(t, \zeta, \mu))p_0(t \mid z, \lambda)\, dt.$$

For μ sufficiently large (uniformly for all z, $\zeta > 0$, $\lambda \geqq \delta$, the first of these terms is bounded above by $\phi((2[(1 - \delta) \ln \mu + \ln \delta])^{\frac{1}{2}})$ which tends to zero as $\mu \to \infty$. On the other hand, by Lemma 2.7, the second term is, for μ sufficiently large (uniformly for all z, λ, $\zeta > 0$), bounded above by

$$\phi(.5\sqrt{G(\hat{t}(1, \mu))} - (1 - \delta) \ln \mu/\sqrt{G(\hat{t}(1, \mu))}). \tag{3.6}$$

By Lemma 2.3, $\lim_{\mu\to\infty} [G(\hat{t}(1, \mu))/ \ln \mu] = 2$. Thus, the argument of ϕ in (3.6) is, for large μ, asymptotically equivalent to $\delta(\frac{1}{2} \ln \mu)^{\frac{1}{2}}$. Hence (3.6) tends to zero as $\mu \to \infty$.

Finally, for μ sufficiently large (uniformly for all $\zeta > 0, \lambda \geqq \delta$), $Q_0^*(0, \lambda, \zeta, \mu)$ is also bounded above by (3.6). This completes the proof of the lemma.

Define $Q_i^*(z, \lambda, \zeta, 0) = \lim_{\mu\to 0} Q_i^*(z, \lambda, \zeta, \mu)$, $z \geqq 0$, $\lambda, \zeta > 0$. Note that the symmetry of (3.4) continues to hold in the limit as $\mu \to 0$. By (3.5) and Lemma 2.9,

$$Q_0^*(z, \lambda, \zeta, 0) = \phi(.5\sqrt{z} + \ln \lambda/\sqrt{z}), \qquad z, \lambda, \zeta > 0. \tag{3.7}$$

Observe that the above expression is independent of $\zeta > 0$ and is a continuous function of positive z and λ. For $z = 0$, we have by (3.3) and (2.3), that $Q_0^*(0, \lambda,\zeta,0) = 1$, .5, or 0, according as $\lambda <, =,$ or >1.

LEMMA 3.2 *Let K be an arbitrary, fixed, positive number, then*

$$\lim_{\lambda\to 0} Q_0^*(z, \lambda, \zeta, \mu) = 1, \qquad \lim_{\lambda\to\infty} Q_0^*(z, \lambda, \zeta, \mu) = 0,$$

uniformly for all z, ζ, μ such that $0 \leqq z, \mu \leqq K, \zeta > 0$.

PROOF. When $\mu = 0$, it is obvious from the definition of $Q_0^*(z, \lambda, \zeta, 0)$ that the limits hold uniformly for $0 \leqq z \leqq K$, $\zeta > 0$. When $z = 0$, it follows from (3.3), (2.3), (2.11), that the limits hold uniformly for $\zeta > 0, 0 < \mu \leqq K$. Finally, for any z, λ, ζ, μ such that $0 < z, \mu \leqq K; \lambda, \zeta > 0$, we have by (3.5), that

$$\int_{t^*(\zeta,\mu)}^{\infty} p_0(t \mid z, \lambda)\, dt < Q_0^*(z, \lambda, \zeta, \mu) < \int_{-t^*(1/\zeta,\mu)}^{\infty} p_0(t \mid z, \lambda)\, dt.$$

For sufficiently small $\lambda > 0$, the left hand side of the above inequality is bounded below by $\phi(.5\sqrt{K} + [\ln \lambda + t^*(\zeta, K)]/\sqrt{K})$. Since, by Lemma 2.9, $t^*(\zeta, K)$ is bounded, this lower bound tends to one, as $\lambda \to 0$, uniformly for all $\zeta > 0$. On the other hand, for sufficiently large λ, the right hand side of the inequality is bounded above by $\phi([\ln \lambda - t^*(1/\zeta,K)]/\sqrt{K})$, which tends to zero, as $\lambda \to \infty$, uniformly for all $\zeta > 0$. This completes the proof.

For any positive r and z, the equation,

$$Q_0^*(z, \lambda, \zeta, 0) = rQ_1^*(z, \lambda, \zeta, 0), \tag{3.8}$$

has, by (3.7), (3.4), a unique positive root in λ. We shall denote, by $\xi(r, z)$, the value common to both sides of (3.8), when λ is equal to this root. Recall that both sides of the above equation are independent of ζ. In the following lemma, we state, without proof, several properties of the function, ξ.

LEMMA 3.3. *$\xi(r, z)$ is*

(1) *for fixed $z > 0$, a strictly increasing function of $r > 0$.*
(2) *for fixed $r > 0$, a strictly decreasing function of $z > 0$.*
(3) *a continuous function of positive r and z.*
(4) $\lim_{r\to 0} \xi(r, z) = 0$, $\xi(1, z) = \phi(.5\sqrt{z})$, $\lim_{r\to\infty} \xi(r, z) = 1$.
(5) $\lim_{z\to 0} \xi(r, z) = r/(1 + r)$, $\lim_{z\to\infty} \xi(r, z) = 0$.

We shall require, finally, for the proof of our theorem, the use of a lemma which is proved by T. Rado and P. V. Reichelderfer in [5], Lemma 16, p. 390. The lemma is paraphrased, below,

LEMMA 3.4. *Given*

(a) *A bounded, simply connected Jordan region, J, in the complex plane.*
(b) *The arbitrarily oriented boundary curve, c, of J.*
(c) *A continuous, real or complex valued function, $s(\omega)$ in J which is different from zero in J.*

Then V_c Argument $s(\omega) = 0$, i.e. the variation in the argument of $s(\omega)$ on c is zero.

THEOREM 3.1. *Let δ be an arbitrary positive number less than* 1, *and define the set*

$$\Lambda_\delta = \{(\beta, r, z, \gamma):\ 0 < \beta \leqq \xi(r, z), r > 0, z > 0, \delta < \gamma < 1/\delta\},$$

Then for each point, $\theta = (\beta, r, z, \gamma)$ in Λ_δ, there exist numbers $\zeta^*(\theta) > 0$ and $\mu^*(\theta) \geqq 0$ such that

$$Q_0^*(z, \gamma\zeta^*(\theta), \zeta^*(\theta), \mu^*(\theta)) = rQ_1^*(z, \gamma\zeta^*(\theta), \zeta^*(\theta), \mu^*(\theta)) = \beta.$$

PROOF. For any point in Λ_δ such that $\beta = \xi(r, z)$, the conclusion is obvious.

Let (β, r, z, γ) be an arbitrary, but fixed point in Λ_δ such that $\beta < \xi(r, z)$. For this proof only, the letter i will be used to denote the imaginary unit. We define the complex variable, $\omega = \zeta + i\mu$, and the complex function of this complex variable, $s(\omega) = s_0(\omega) + is_1(\omega)$, where

$$s_0(\omega) = Q_0^*(z, \gamma\zeta, \zeta, \mu) - rQ_1^*(z, \gamma\zeta, \zeta, \mu),$$

and $s_1(\omega) = Q_0^*(z, \gamma\zeta, \zeta, \mu) - \beta$. Since Lemma 3.1 is obviously true with its δ replaced by the square of the δ of the present theorem, we have by that lemma and by the symmetry (3.4), that there exists a positive value of μ, μ_δ, say, such that

$$\mu \geqq \mu_\delta \Rightarrow \begin{cases} s_1(\omega) < 0, & \delta \leqq \zeta < \infty, \\ s_1(\omega) < s_0(\omega), & 0 < \zeta \leqq 1/\delta. \end{cases} \tag{3.9}$$

By Lemma 3.2, taking $K = \mu_\delta$, and by the symmetry (3.4) there exists a positive value of ζ, ζ_δ, say, which is less than one, such that

$$\begin{aligned} &0 < \zeta \leqq \zeta_\delta \Rightarrow s_0(\omega) > 0, \\ &1/\zeta_\delta \leqq \zeta < \infty \Rightarrow s_0(\omega) < 0, \end{aligned} \qquad \text{for all } \mu,\ 0 \leqq \mu \leqq \mu_\delta . \tag{3.10}$$

Let c_j, $j = 1, 2, 3, 4$, denote the line segments with complex endpoints, as follows.

$$c_1 : \zeta_\delta + i\mu_\delta,\ \zeta_\delta \qquad c_2 : \zeta_\delta,\ 1/\zeta_\delta$$

$$c_3 : 1/\zeta_\delta,\ 1/\zeta_\delta + i\mu_\delta \qquad c_4 : 1/\zeta_\delta + i\mu_\delta,\ \zeta_\delta + i\mu_\delta .$$

The rectangle c composed of these four line segments is the boundary of a simply connected Jordan region in the complex ω plane. By (3.10), $s_0(\omega)$, which is the real part of $s(\omega)$, is positive everywhere on c_1, including the endpoints of this line segment. Thus, the image of c_1 under s lies entirely to the right of the imaginary axis. Similarly, by (3.10), the image of c_3 under s lies entirely to the left of the imaginary axis. As ω moves on c_2, from ζ_δ to $1/\zeta_\delta$, by (3.7) and the symmetry (3.4), $s_0(\omega)$ decreases monotonically from its positive value at $\omega = \zeta_\delta$ to its negative value at $\omega = 1/\zeta_\delta$, taking on the value zero precisely once at the value of ζ for which $\gamma\zeta$ is equal to the unique root in λ of the equation (3.8). Since $s_1(\omega)$, which is the imaginary part of $s(\omega)$ is equal to $\xi(r, z) - \beta$, at this point, and since $\beta < \xi(r, z)$, the image of c_2 crosses the imaginary axis precisely once, at a point *above* the real axis. On the other hand, by (3.5), the symmetry (3.4), and the lemmas of section 2, $s_0(\omega)$ is a continuous function of ζ everywhere on c_4. Hence, it follows that as ω moves from $1/\zeta_\delta + i\mu_\delta$ to $\zeta_\delta + i\mu_\delta$, on c_4, $s_0(\omega)$, starting negative and ending positive, must take on the value zero at least once. By (3.9), however, each time that it does, $s_1(\omega)$ must be negative. Thus, the image of c_4 under s, must cross the imaginary axis an odd number of times and each time it does so, the crossing must be made *below* the real axis. It is evident, that as ω describes a path about c in either direction and returns to the initial point, that the argument of $s(\omega)$ must increase or decrease by 2π. But this contradicts the conclusion of Lemma 3.4. Hence $s(\omega)$ must have at least one zero inside of c. This proves the theorem.

We remark that by (3.3), (2.3), (2.11),

$$Q_i^*(0, 1, 1, \mu) = \phi(\tfrac{1}{2}\sqrt{\bar{g}(0, 1, \mu)}). \tag{3.11}$$

Hence, by Lemmas 2.6, 2.7, the conclusion of the above theorem holds also for the points $\theta = (\beta, 1, 0, 1)$, with $0 < \beta \leqq \frac{1}{2}$.

COROLLARY 3.1. *Let $\theta^* = (\beta_0, \beta_0/\beta_1, a^2m, g)$, where a and g are arbitrary, but fixed, positive numbers, m is a positive first sample size, and β_0 and β_1 are any two preassigned numbers such that*

$$0 < \beta_0 \leqq \xi(\beta_0/\beta_1, a^2m), \qquad 0 < \beta_1 < 1.$$

Let $W^* = \zeta^*(\theta^*)$, $M^* = a^{-2}\mu^*(\theta^*)$, *then*

$$Q_i(S^*(g, W^*, M^*)) = \beta_i, \qquad i = 0, 1, \tag{3.12}$$

and if S *is any rule in* C_m *such that*

$$Q_i(S) \leqq \beta_i, \qquad i = 0, 1, \tag{3.13}$$

then

$$\sum_{i=0}^{1} g_i \mathcal{E}_i(S^*(g, W^*, M^*)) \leqq \sum_{i=0}^{1} g_i \mathcal{E}_i(S).$$

Proof. The first conclusion follows immediately from the theorem and (3.2). The second conclusion follows from Lemma 1.1.

By the remark concerning (3.11), the corollary may be extended to a zero first sample size when $\beta_0/\beta_1 = g = 1$.

If we can now *choose* g so that $\mathcal{E}_0(S^*(g, W^*, M^*)) = \mathcal{E}_1(S^*(g, W^*, M^*))$, the resulting rule will minimize the maximum expected overall sample size among all rules in C_m with error probabilities less than or equal to the ones which it possesses.

For example, if we take $\beta_0 = \beta_1 = \beta$, say, where

$$0 < \beta \leqq \xi(1, a^2 m) = \phi(\tfrac{1}{2}a\sqrt{m}) \leqq \tfrac{1}{2}, \tag{3.14}$$

and we choose $g = 1$, then by (3.4), we may take $\zeta^*(\beta, 1, a^2m, 1) = 1$, and Lemma 3.1, (3.7), and (3.11) will ensure the existence of the corresponding value $\mu^*(\beta, 1, a^2m, 1)$. Thus, if we let

$$\hat{S}(a, \beta, m) = S^*(1, 1, a^{-2}\mu^*(\beta, 1, a^2m, 1)), \tag{3.15}$$

we have, subject to (3.14), that $Q_0(\hat{S}(a, \beta, m)) \equiv Q_1(\hat{S}(a, \beta, m)) \equiv \beta$. But in addition, by (3.1), (3.4),

$$\mathcal{E}_0(\hat{S}(a, \beta, m)) = \mathcal{E}_1(\hat{S}(a, \beta, m)) = a^{-2}\hat{\mathcal{E}}(\beta, a^2m), \text{ say.} \tag{3.16}$$

Hence, by the corollary and the remark which follows Theorem 3.1, $\hat{S}(a, \beta, m)$ has, for any non-negative first sample size m, the property that

$$\max_{i=0,1} \mathcal{E}_i(\hat{S}(a, \beta, m)) \leqq \max_{i=0,1} \mathcal{E}_i(S), \tag{3.17}$$

for all S in C_m such that

$$Q_i(S) \leqq \beta, \quad i = 0, 1. \tag{3.18}$$

The selection of an optimum first sample size for this example is considered in Section 4.

If it were possible to choose W^* and M^* so as to satisfy (3.12) and in addition also to satisfy the requirement that $\mathcal{E}_i(S^*(g, W^*, M^*))$, be independent of g, the resulting rules would minimize the expected overall sample size simultaneously under both densities (2.1), among all rules in C_m which satisfy (3.13). It is conjectured that in the present case, this requirement is impossible to fulfill. As shown by Wald and Wolfowitz in [6], the analogous result in the sequential case is actually achieved by the sequential probability ratio test.

In conclusion, we mark that the above results do not overlap with those of [2]. In [2], a Bayes rule in C_m is found for testing the composite hypothesis that the mean of a normal distribution with known variance is positive against the composite hypothesis that it is less than or equal to zero. The prior distribution of the mean is taken to be its fiducial distribution based on the outcome of the first sample. The loss is taken negatively proportional to the mean when a correct choice is made, and zero otherwise. Cost is proportional to the number of observations. The solution is shown to be admissible with respect to the loss function chosen.

4. Two-stage rules which minimize total expected sample size. Continuing our example of the preceding section, we have by (3.16), (3.1) that

$$\hat{\mathcal{E}}(\beta, z) = \mathcal{E}_i^*(z, 1, 1, \mu^*(\beta, 1, z, 1)). \tag{4.1}$$

Using the expressions which follow (3.1), we find that

$$\mathcal{E}_i^*(z, 1, 1, \mu) = z + \sqrt{2\pi/z} \int_0^{t^*(1,\mu)} I(t, \zeta, \mu)\, dt, \qquad z, \mu > 0, \tag{4.2}$$

where $I(t, z, \mu) = \hat{y}(t, 1, \mu) \cosh(t/2) \exp[-(z/8) - (t^2/2z)]$;

$$\mathcal{E}_i^*(0, 1, 1, \mu) = \hat{y}(0, 1, \mu), \qquad \mu > 0; \tag{4.3}$$

$$\mathcal{E}_i^*(z, 1, 1, 0) = \lim_{\mu \to 0} \mathcal{E}_i^*(z, 1, 1, \mu) = z, \qquad z > 0. \tag{4.4}$$

In a similar way we may find simplified expressions for $Q_i^*(z, 1, 1, \mu)$.

To motivate the lemma which follows and with reference to (3.14) we note that by part 4 of Lemma 3.3, the inequalities $0 < \beta \leqq \xi(1, z) \leqq \frac{1}{2}$ are equivalent to the inequalities

$$0 \leqq z \leqq 4\lambda_\beta^2, \qquad 0 < \beta \leqq \tfrac{1}{2}, \tag{4.5}$$

where λ_β is the unique root of the equation $\phi(\lambda) = \beta$.

LEMMA 4.1. $\hat{\mathcal{E}}(\beta, 0) = \hat{\mathcal{E}}(\beta, 4\lambda_\beta^2) = 4\lambda_\beta^2$.

PROOF. By (3.7)

$$Q_0^*(z, 1, 1, 0) = \phi(\tfrac{1}{2}\sqrt{z}).$$

Hence we may take $\mu^*(\beta, 1, 4\lambda_\beta^2, 1) = 0$. Thus, by (4.1), (4.4),

$$\hat{\mathcal{E}}(\beta, 4\lambda_\beta^2) = 4\lambda_\beta^2.$$

On the other hand, by (3.3), (2.3),

$$Q_i^*(0, 1, 1, \mu) = \phi(\tfrac{1}{2}\sqrt{\hat{y}(0, 1, \mu)}).$$

Hence, by (4.1), (4.3),

$$\hat{\mathcal{E}}(\beta, 0) = \hat{y}(0, 1, \mu^*(\beta, 1, 0, 1)) = 4\lambda_\beta^2.$$

THEOREM 4.1.

1. *For each $a > 0$ and each β, $0 < \beta \leqq \frac{1}{2}$, there exists a zero or positive integral value of $m < 4a^{-2}\lambda_\beta^2$, call it $\hat{m}(a, \beta)$, such that the rule*

$$\hat{S}(a, \beta, \hat{m}(a, \beta)) = S^*(1, 1, a^{-2}\mu^*(\beta, 1, a^2\hat{m}(a, \beta), 1)) \tag{4.6}$$

minimizes the maximum expected total sample size among all two-stage rules S, with integral first sample size, which satisfy (3.18).

2. *Whenever $a \geqq 2\lambda_\beta$, we may take $\hat{m}(a, \beta) = 0$.*

PROOF. Clearly, if $a^2m \geqq 4\lambda_\beta^2$, then any two-stage rule in C_m will have, under either hypothesis, an expected total sample size $\geqq 4a^{-2}\lambda_\beta^2$. By (3.16), (3.17), and Lemma 4.1, we may thus restrict our consideration to the rules $\hat{S}(a, \beta, m)$, with $a^2m < 4\lambda_\beta^2$. Since only finitely many multiples of a^2 are bounded above by $4\lambda_\beta^2$, part 1 of the theorem follows. Part 2 is immediate.

It is of interest to note that by (2.9), (3.1) and Lemma 4.1, the second sample size specified by (4.6) when $\hat{m}(a, \beta) = 0$ is $4a^{-2}\lambda_\beta^2$, which (rounded to the following integer) is the sample size of the corresponding optimum one-stage procedure. Thus, the optimum one and two-stage rules are identical whenever $a \geqq 2\lambda_\beta$, $0 < \beta \leqq \frac{1}{2}$.

For fixed β, $\hat{\mathcal{E}}(\beta, z)$ appears to be continuous in z over the interval $[0, 4\lambda_\beta^2]$ and monotonic to either side of a unique minimizing $z(= \hat{z}(\beta)$, say. See Tables I and II.

TABLE I

$\beta = .05$

	z	μ^*	$\hat{\mathcal{E}}(0.5, z)$	$t^*(1, \mu^*)$	$\hat{t}(1, \mu^*)$
	0.	127.588	10.8222		
	0.5	128.067	10.7742	2.2106	2.4618
	1.	130.937	10.5598	2.2285	2.4809
	1.5	132.888	10.1762	2.2405	2.4936
	2.	132.086	9.7148	2.2356	2.4884
	3.	122.987	8.8159	2.1778	2.4271
	4.	107.516	8.1567	2.0694	2.3120
	5.	89.428	7.8222	1.9221	2.1552
	5.4854	80.494	7.7774	1.8386	2.0662
	5.5293	79.691	7.7770	1.8307	2.0577
$\hat{z}(.05) =$	5.5393	79.510	7.7769	1.8289	2.0558
	5.5504	79.307	7.7770	1.8269	2.0537
	6.	71.216	7.8081	1.7423	1.9632
	7.	54.297	8.0742	1.5322	1.7377
	8.	39.298	8.5698	1.2902	1.4757
	9.	26.252	9.2458	1.0070	1.1649
	10.	14.504	10.0617	0.6497	0.7632
	10.5	8.205	10.5138	0.3962	0.4701
$4\lambda_{.05}^2 =$	10.8222	0.	10.8222	0.	0.

TABLE II

$\beta = .01$

	z	μ^*	$\hat{\hat{\mathcal{E}}}(.01, z)$	$t^*(1, \mu^*)$	$\hat{t}(1, \mu^*)$
	0.	698.285	21.6476		
	6.	657.699	16.2500	3.5699	3.8941
	8.	526.323	14.5568	3.3815	3.6963
	10.	385.531	13.8947	3.1198	3.4214
$\hat{z}(.01) =$	10.3781	360.422	13.8787	3.0634	3.3621
	10.4479	355.874	13.8793	3.0528	3.3509
	12.	262.659	14.1471	2.7999	3.0846
	14.	167.700	15.0849	2.4302	2.6943
$4\lambda^2_{.01} =$	21.6476	0.	21.6476	0.	0.

Let us assume the existence of this minimum and the monotonicity and continuity of $\hat{\mathcal{E}}$ to either side of it. It then follows that $\hat{m}(a, \beta)$ must be the integer either immediately preceding or immediately following $a^{-2}\hat{z}(\beta)$, when this latter is non-integral; otherwise, $\hat{m}(a, \beta) = a^{-2}\hat{z}(\beta)$. Also it follows that

$$\lim_{a \to 0} a^2\hat{m}(a, \beta)/\hat{z}(\beta) = 1. \tag{4.7}$$

The computations seem also to indicate that $\mu^*(\beta, 1, z, 1)$ is continuous in z over the above interval and assuming this is true it follows that

$$\lim_{a \to 0} \mu^*(\beta, 1, a^2\hat{m}(a, \beta), 1) = \mu^*(\beta, 1, \hat{z}(\beta), 1).$$

It should be noted that the existence of an optimum integral first sample size $\hat{m}(a, \beta)$ rests entirely upon Theorem 4.1 and does *not* depend upon the above assumptions which are of computational origin. These assumptions, by virtue of the implications which proceed from them, allow us as indicated below to approximate the rule (4.6) by the rule (4.8) when a is small, and otherwise enable us to locate $\hat{m}(a, \beta)$ with reference to $a^{-2}\hat{z}(\beta)$ by only two computations of $\hat{\mathcal{E}}$. Without these assumptions, a finite number of computations of $\hat{\mathcal{E}}$ would be required to make certain that the desired minimum had been attained.

In Tables I and II, for $\beta = .05$ and .01, respectively, we have given values of the functions μ^*, $\hat{\mathcal{E}}$, t^*, and $\hat{t}$ associated with the rules $S^*(1, 1, a^{-2}\mu^*(\beta, 1, z, 1))$ for selected z in the interval $[0, 4\lambda_\beta^2]$. All entries are rounded in the last place given. More extensive calculations not tabulated here ensure that despite the relative insensitivity of $\hat{\mathcal{E}}$ in a neighborhood of the minimizing argument $\hat{z}(\beta)$, the values of $\hat{z}(.05)$ and $\hat{z}(.01)$ are accurate to the first three decimal places with perhaps a slight error in the fourth. The abbreviation $\mu^* = \mu^*(\beta, 1, z, 1)$ is employed in the table headings.

In Table III, we have given values of $\hat{m}(a, \beta)$ for some values of $a \geq \frac{1}{2}$ and for $\beta = .05$ and .01.

TABLE III

a	$\hat{m}(a, .05)$	a	$\hat{m}(a, .01)$
		$\geqq 4.6527$	0
$\geqq 3.2897$	0	4.	1
3.	1	3.	1
2.	1	2.	3
1.	6	1.	10
0.5	22	0.5	41

If we admit any non-negative real number as a first sample size (We have already done this for second sample sizes), then clearly

$$S^*(1, 1, a^{-2}\mu^*(\beta, 1, \hat{z}(\beta), 1)) \tag{4.8}$$

with first sample size $a^{-2}\hat{z}(\beta)$ possesses the property attributed to (4.6) in the wider class of rules where non-integral first sample sizes are allowed. The ratio of the expected total sample size of (4.8) to the sample size of the corresponding optimum one-stage procedure is $\hat{\mathcal{E}}(\beta, \hat{z}(\beta))/4\lambda_\beta^2 = .7186, .6411$, for $\beta = .05$ and .01 respectively. The first of these figures is a clear improvement upon .7569, the corresponding ratio for an intuitive two-stage rule proposed by Owen [4].

TABLE IV

$\beta = .05$

t	$\hat{y}(t, 1, \mu^{**})$	t	$\hat{y}(t, 1, \mu^{**})$		t	$\hat{y}(t, 1, \mu^{**})$
0.	8.1655	0.6573	7.7345		1.3145	6.4300
0.0286	8.1647	0.6858	7.6963		1.3431	6.3516
0.0572	8.1622	0.7144	7.6564		1.3717	6.2710
0.0857	8.1581	0.7430	7.6149		1.4003	6.1883
0.1143	8.1524	0.7716	7.5718		1.4288	6.1034
0.1429	8.1451	0.8002	7.5270		1.4574	6.0161
0.1715	8.1361	0.8287	7.4806		1.4860	5.9264
0.2000	8.1255	0.8573	7.4326		1.5146	5.8342
0.2286	8.1132	0.8859	7.3829		1.5431	5.7393
0.2572	8.0994	0.9145	7.3315		1.5717	5.6416
0.2858	8.0839	0.9430	7.2785		1.6003	5.5409
0.3143	8.0667	0.9716	7.2238		1.6289	5.4370
0.3429	8.0480	1.0002	7.1674		1.6575	5.3297
0.3715	8.0276	1.0288	7.1093		1.6860	5.2187
0.4001	8.0056	1.0573	7.0495		1.7146	5.1036
0.4287	7.9820	1.0859	6.9879		1.7432	4.9840
0.4572	7.9567	1.1145	6.9246		1.7718	4.8594
0.4858	7.9298	1.1431	6.8595		1.8003	4.7291
0.5144	7.9013	1.1716	6.7927	$t^*(1,\mu^{**}) =$	1.8289	4.5922
0.5430	7.8712	1.2002	6.7239			
0.5715	7.8395	1.2288	6.6533			
0.6001	7.8061	1.2574	6.5809			
0.6287	7.7711	1.2860	6.5064	$\hat{t}(1,\mu^{**}) =$	2.0558	2.5722

TABLE V

$\beta = .01$

t	$\hat{y}(t, 1, \mu^{**})$	t	$\hat{y}(t, 1, \mu^{**})$		t	$\hat{y}(t, 1, \mu^{**})$
0.	17.2623	1.1009	16.0875		2.2018	12.8353
0.0479	17.2600	1.1488	15.9863		2.2497	12.6498
0.0957	17.2531	1.1967	15.8812		2.2976	12.4603
0.1436	17.2417	1.2445	15.7724		2.3454	12.2665
0.1915	17.2256	1.2924	15.6598		2.3933	12.0683
0.2393	17.2051	1.3403	15.5435		2.4412	11.8655
0.2872	17.1800	1.3881	15.4235		2.4890	11.6579
0.3351	17.1503	1.4360	15.2999		2.5369	11.4452
0.3829	17.1162	1.4839	15.1726		2.5848	11.2271
0.4308	17.0776	1.5317	15.0418		2.6326	11.0033
0.4787	17.0346	1.5796	14.9074		2.6805	10.7732
0.5265	16.9872	1.6275	14.7694		2.7284	10.5364
0.5744	16.9353	1.6753	14.6279		2.7762	10.2922
0.6223	16.8792	1.7232	14.4829		2.8241	10.0399
0.6701	16.8187	1.7711	14.3343		2.8720	9.7785
0.7180	16.7539	1.8189	14.1822		2.9198	9.5067
0.7659	16.6850	1.8668	14.0266		2.9677	9.2230
0.8137	16.6118	1.9147	13.8674		3.0156	8.9254
0.8616	16.5345	1.9625	13.7046	$t^*(1,\mu^{**}) =$	3.0634	8.6109
0.9095	16.4531	2.0104	13.5382			
0.9573	16.3677	2.0583	13.3681			
1.0052	16.2782	2.1061	13.1943			
1.0531	16.1848	2.1540	13.0167	$\hat{t}(1,\mu^{**}) =$	3.3621	5.0153

For small values of a (say $\leqq \frac{1}{2}$), we may with an error at most unity in the first sample size and relatively small error in the second sample size function employ (4.8) as a substitute for (4.6), taking first and second sample sizes to the nearest integer. Due to lack of space, the only second sample size functions which are tabulated here are those which correspond to the rules (4.8) for $\beta = .05$ and .01 (Tables IV and V, respectively). For convenience, we use the abbreviation $\mu^{**} = \mu^*(\beta, 1, \hat{z}(\beta), 1)$. Recall that by (2.7), $\hat{y}(t, 1, \mu)$ is symmetric about $t = 0$. By (2.9), (2.11) the second sample size specified by (4.8) corresponding to a first sample with mean $\bar{X}$ is

$$a^{-2}\hat{y}(a^{-1}\hat{z}(\beta)\bar{X}, 1, \mu^{**}), \qquad \text{if} \qquad |\bar{X}| < at^*(1, \mu^{**})/\hat{z}(\beta),$$

and zero, otherwise.

A program for use on the Datatron prepared by the author with the aid of the Purdue Compiler [3] is available (routine library of computing laboratory at Purdue) for calculation of $\hat{c}$ for any z and β which satisfy (4.5) as well as for any of the auxiliary functions tabulated here. The individual computations of which the program is composed may be determined directly from the definitions of the functions tabulated. Once $\hat{m}(a, \beta)$ has been determined or approximated by $a^{-2}\hat{z}(\beta)$, computations of corresponding optimum second sample size values to the degree of accuracy attained in Tables IV and V require about fifteen minutes of datatron time. Location of $\hat{z}(\beta)$ may take several hours or more depending on the accuracy desired. If only $\hat{m}(a, \beta)$ is wanted, for particular arguments, the time may be considerably reduced. For example, $\hat{m}(2, .05)$ would require only two non-trivial calculations of $\hat{\mathcal{E}}$.

Let us now consider a specific example with $a = .1$, $\beta = .05$. Using the rule (4.8), we take a first sample of $100 \times 5.5393 \cong 554$ observations. If $|\bar{X}_{554}| \geqq .1 \cdot 1.8289/5.5393 = .0330$, we take no additional observations and choose f_0 or f_1 according as $\bar{X}_{554}$ is $<$ or $>$ 0. If $|\bar{X}_{554}| < .0330$, we take $100\hat{y}(10 \cdot 5.5393\bar{X}_{554}, 1, \mu^{**})$ additional observations. For example, if $\bar{X}_{554}$ came out equal to .0191, we would (using Table IV) take 705 additional observations. We would then choose f_0 or f_1 according as the overall mean of both samples is negative or positive. We toss a coin to decide, if the overall mean equals zero.

The optimum rule outlined above has an expected total sample size of 777.69 (a possible error in the decimal may exist due to the use of integral sample sizes). The rule requires a minimum of 554 observations (when no second sample is required), and can call for at most 1082 observations in the second sample. Under either hypothesis, Prob $\{|\bar{X}_{554}| < .0330\} = .3192$, i.e. there is less than one chance in three that the rule will require a second sample. 1082 observations would be required by the corresponding one-stage rule.

REFERENCES

[1] R. D. GORDON, "Value of Mills ratio of area to bounding ordinate of the normal probability integral for large values of the argument," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 364–6.

[2] P. M. GRUNDY, M. J. R. HEALY, AND D. H. REES, "Economic choice of the amount of experimentation," *J. Roy. Stat. Soc., Series B*, Vol. 18 (1956), pp. 32–55.

[3] SYLVIA ORGEL, *Purdue Compiler, General Description*, Purdue Research Foundation, July, 1958.

[4] DONALD B. OWEN, "A double sample test procedure," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 449–457.

[5] T. RADO, AND P. V. REICHELDERFER, *Continuous Transformations in Analysis*, Springer-Verlag, Berlin, 1955.

[6] A. WALD AND J. WOLFOWITZ, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–39.

[7] ABRAHAM WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.