BOUNDS ON THE MAXIMUM SAMPLE SIZE OF A BAYES SEQUENTIAL PROCEDURE[1]

By S. N. Ray

University of North Carolina and Stanford University

1. Introduction and summary. Sufficient conditions for a Bayes sequential procedure to be truncated contained in Wald [12] and in Blackwell and Girshick [3] have been used by Sobel [11] and Mikhalevich [7] to show that under certain conditions a Bayes sequential test is truncated when the cost of observation does not depend on the unknown parameter. When this latter condition is not satisfied as illustrated by some sequential testing problems considered by Kiefer and Weiss [5] and Weiss [14], and by Anscombe [2] (in connection with ethical cost of medical trials) the sufficient condition referred to (see Corollary 3.1) is not always applicable. When it is applicable it generally provides an upper bound on the maximum sample size that a Bayes sequential procedure can take. For the purpose of computation of the Bayes procedure by backward induction it is desirable to know the exact stage of truncation as close as possible.

Section 3 contains two theorems (Theorem 3.2 and 3.3) which provide better sufficient conditions for a Bayes sequential procedure to be truncated and simultaneously provide better bounds on the exact stage of truncation. Sufficient conditions are also provided for the upper bounds involved to be exact. Section 4 contains some results with the help of which the theorems of Section 3 may be applied to some special sequential testing problems in Section 5. Finally Section 6 contains proofs of the results of Sections 3 and 4.

2. Statement of the problem. Let the pair $(\Omega, \mathcal{A})$ be an abstract space and a σ-field of subsets. Let X_n: $n = 1, 2, \cdots$ be a sequence of random variables on $(\Omega, \mathcal{A})$. Let $\mathfrak{J}$ be the space of unknown states of nature such that for each $\theta \varepsilon \mathfrak{J}$ there corresponds a probability P_θ on Ω so that $\{X_n\}$ are independent and identically distributed with respect to P_θ. Let Ξ be the set of all probability measures on a fixed Borel field $\mathcal{B}$ of subsets of $\mathfrak{J}$. Let $F_\theta(x)$ be the distribution function of X_1 with respect to P_θ and $f_\theta(x)$ be the corresponding density with respect to some σ-finite measure μ. We assume $F_\theta(x)$ and $f_\theta(x)$ are jointly measurable in (θ, x). For any integrable function $g(\theta, x)$ and for any $\xi \varepsilon \Xi$, $E_\xi g(\Theta, X)$ denotes the expectation of $g(\Theta, X)$ with respect to the joint probability distribution of X and Θ where the conditional distribution of X given $\Theta = \theta$ is given by $F_\theta(x)$ and the marginal distribution of Θ is given by ξ. If g involves only Θ then the expectation is over the marginal distribution ξ. For any $\omega \varepsilon \Omega$, let $Q_\xi(\omega) = E_\xi P_\theta(\omega)$ denote the integrated probability measure on $(\Omega, \mathcal{A})$. The terms "almost surely" (abbreviated as a.s.) or its equivalent "with probability one" (abbreviated as

Received 25 June 1964; revised 28 November 1964.

[1] This work was initiated with the support of NIH Grant No. 325-NIH-187 and completed with the support of NSF Grant No. GP-2220.

w.p.1) and "with positive probability" that are used in this paper, are with respect to this integrated probability measure.

Let A be the action space and $L(\theta, a)$, $\theta \varepsilon \mathfrak{J}$, $a \varepsilon A$, a non-negative real valued function on $\mathfrak{J} \times A$ be the loss incurred due to taking action a when θ is the true state of nature. Let $c(\theta)$ denote the cost of a single observation, which is non-negative and is allowed to depend on θ. Let Δ be the class of all measurable randomized sequential procedures. Let $r(\theta, \delta)$, $\theta \varepsilon \mathfrak{J}$, $\delta \varepsilon \Delta$ be the risk (expected loss plus expected sampling cost) when procedure δ is used and the true state of nature is θ. Let $r(\xi, \delta) = E_\xi r(\Theta, \delta)$. Let $\rho(\xi)$ denote the infimum of the average risk $r(\xi, \delta)$ over the class Δ. If there exists a procedure δ_ξ such that $r(\xi, \delta_\xi) = \rho(\xi)$ then δ_ξ is called a *Bayes procedure* with respect to ξ and $\rho(\xi)$ the *Bayes* (average) *risk* in the class Δ.

In order not to distract attention from the main object we assume whatever conditions are required in order that the following procedure exist and be Bayes. This procedure is to stop whenever the stopping risk is equal to the optimum *a posteriori* risk. When we stop we take that action which minimizes the *a posteriori* stopping risk. Among the risk equivalent Bayes procedures relative to ξ, this is the one which leads to termination at the earliest time.

DEFINITION. A sequential procedure δ is said to be truncated if there exists a non-negative integer n such that $N \leqq n$ with probability one where N denotes the (random) sample size associated with δ.

The smallest of such integers, say, n^0, is called the *exact stage of truncation* of the procedure δ.

The purpose of this paper is:

(i) to study conditions for which the Bayes procedure δ_ξ in the class Δ is truncated; and

(ii) if the Bayes procedure is truncated, to determine upper bounds on its exact stage of truncation $n^0(\xi)$ [henceforth denoted simply by n^0].

3. Statements of main theorems. For any positive integer n and for any vector $\mathbf{x}_n = (x_1, \cdots, x_n)$ and for any *a priori* $\xi \varepsilon \Xi$, let $\tau_{\mathbf{x}_n}\xi$ denote the *a posteriori* distribution after observing $X_i = x_i$, $i = 1, 2, \cdots, n$. For consistency we set $\tau_{\mathbf{x}_0}\xi \equiv \xi$. We note that in our case of independent identically distributed variables τ_x is a commutative and associative operator from Ξ to Ξ. In particular,

$$\tau_{\mathbf{x}_{n+1}}\xi = \tau_{x_1}(\tau_{x_2}(\cdots(\tau_{x_{n+1}}\xi)\cdots)) = \tau_{x_{n+1}}(\tau_{\mathbf{x}_n}\xi).$$

For any integrable function $g(\theta)$ and for any $\xi \varepsilon \Xi$, we denote

$$g(\xi) = E_\xi g(\Theta).$$

Let

$$\rho_0(\xi) = \min_{a \varepsilon A} L(\xi, a).$$

THEOREM 3.1. *A sufficient condition for a Bayes sequential procedure with respect to ξ to be truncated is that there exists a non-negative integer n' such that*

$$\rho_0(\tau_{x_n}\xi) \leqq c(\tau_{x_n}\xi) \quad \text{a.s.}$$

for every $n \geqq n'$.

If $\dot{n}_0$ *denotes the smallest such integer* n' *then* $\dot{n}_0$ *is an upper bound to* n^0.

In other words, if the stopping risk must be less than or equal to the expected cost of one additional observation for every observation after the n'th, irrespective of the data, then the Bayes procedure is truncated at n'.

We omit the proof of this theorem which is a trivial generalization of the following special case which is essentially contained in Wald [12] and in Blackwell and Girshick [3].

COROLLARY 3.1. *If the cost of an observation* $c(\theta)$ *is independent of* θ, *then a sufficient condition for a Bayes procedure to be truncated is*

$$\lim_{n\to\infty} \text{essup}_{x_n} \rho_0(\tau_{x_n} \xi) = 0. \tag{3.1}$$

Corollary 3.1 has been the main tool so far in the literature for determining whether or not a Bayes sequential procedure is truncated. When the cost depends on θ there are situations (see Example 5.3) where the sufficient condition of Theorem 3.1 is not satisfied although the Bayes procedure in question is truncated. We are thus led to seek better sufficient conditions than those given in Theorem 3.1, which lead to smaller upper bounds on n^0. The simplest such condition is given by the following theorem which also states sufficient conditions for the upper bound to be exact. Before stating the theorem we need the following notations.

For any non-negative integer n, and for any $\xi \varepsilon \Xi$, let

$$\rho_n(\xi) = \min_{\delta\varepsilon\Delta_n} r(\xi, \delta) \tag{3.2}$$

denote the Bayes risk in the subclass Δ_n of Δ for which δ takes at most n observations. Note that this notation is consistent with $\rho_0(\xi)$ defined earlier. Also, for any $\xi \varepsilon \Xi$, let

$$\lambda(\xi) = \rho_0(\xi) - E_\xi\rho_0(\tau_X\xi) \tag{3.3}$$

denote the advantage of stopping after one free observation rather than stopping immediately.

THEOREM 3.2. (i) *With the assumption that*

$$\lim_{n\to\infty} \rho_n(\xi) = \rho(\xi) \tag{3.4}$$

there is a Bayes sequential procedure with respect to ξ *that is truncated if there exists an integer* n' *such that*

$$\lambda(\tau_{x_n}\xi) \leqq c(\tau_{x_n}\xi) \quad \text{a.s.} \tag{3.5}$$

for every $n \geqq n'$.

If $\bar{n}_0$ *denote the smallest such integer* n' *then* $n^0 \leqq \bar{n}_0 \leqq \dot{n}_0$.

(ii) *A sufficient condition for* $\bar{n}_0 = n^0$ *is that there exists a set of* $\{x_1, x_2, \cdots\}$

with positive probability such that

$$\lambda(\tau_{x_i}\xi) > c(\tau_{x_i}\xi) \tag{3.6}$$

for each $i = 0, 1, 2, \cdots, \bar{n}_0 - 1$.

The proof of this theorem involves the following property of a Bayes sequential procedure whose risk can be approximated arbitrarily closely by that of a Bayes truncated sequential procedure. If, at a certain stage in the history of the sampling, the expected reduction in the stopping risk derivable from taking one more observation cannot, at any future time, be compensated for by the expected cost of that additional observation, then the Bayes procedure tells one to stop. If such a situation must arise by stage n' irrespective of data, then the Bayes procedure must be truncated at n'.

Although the theorem above has not been stated explicitly as such in the literature, its verbal rendition given above is intuitively so obvious that many authors have used this sort of result without verifying some such condition as (3.4) (see for instance, Wald ([12], p. 166) and Wetherill [13]).

The second part of the theorem says that if there exists a set of sample paths such that if at any stage of following such a path the expected reduction in the stopping risk derivable from taking one more observation exceeds the expected cost of that additional observation then the Bayes procedure tells one to continue. This is a special case (viz, looking ahead one step) of the modified Bayes rule of Amster [1] which is known to stop earlier than the Bayes procedure. We shall, however, give a proof of this part too in Section 6, partly for the sake of completeness and partly for the sake of a stepping stone to the proof of the less obvious Theorem 3.3(ii).

For typical problems with continuous random variable X it is possible to find a set of $\{x_1, x_2, \cdots,\}$ with positive probability satisfying (3.6), as illustrated in Example 5.1, so that n^0 is given exactly by $\bar{n}_0$. With a discrete random variable X, however, this may not be the case. An illustration is Example 5.2 where $n^0 < \bar{n}_0$. In such cases it is possible to obtain a better bound to n^0 by using a more elaborate sufficient condition than that provided by Theorem 3.2(i). In fact Lemma 6.3, used in proving Theorem 3.2(i) and Theorem 3.3(i) gives a characterization of a non-increasing sequence $\{\bar{n}_k\}$ of upper bounds to n^0, of which only $\bar{n}_0$ and $\bar{n}_1$ will later be used in the illustrative examples. The second part of the next theorem states sufficient conditions for $\bar{n}_1$ to coincide n^0. These conditions will be shown to be satisfied in some symmetrical binomial problem considered in Example 5.3 thus enabling us to obtain n^0 explicitly. Before we state the next theorem we need the following notation.

For any $\xi \in \Xi$, let

$$\beta(\xi) = [(\lambda - c)(\xi) + E_\xi\{(\lambda - c)_+(\tau_X\xi)\}]_+ - (\lambda - c)_+(\xi) \tag{3.7}$$

where $_+$ denotes the positive part.

THEOREM 3.3. (i) *With the assumption that*

$$\lim_{n\to\infty} \rho_n(\xi) = \rho(\xi) \tag{3.8}$$

there is a Bayes sequential procedure with respect to ξ *that is truncated if there exists an integer* n' *such that*

$$\beta(\tau_{\mathbf{x}_n}\xi) = 0 \quad \text{a.s.} \tag{3.9}$$

for every $n \geqq n'$.

If $\bar{n}_1$ *denotes the smallest such integer then* $n^0 \leqq \bar{n}_1 \leqq \bar{n}_0$.

(ii) *A sufficient condition for* $\bar{n}_1$ *to coincide with* n^0 *is that there exists a set of* $\{x_1, x_2, \cdots\}$ *with positive probability such that*

$$\lambda(\tau_{\mathbf{x}_i}\xi) > c(\tau_{\mathbf{x}_i}\xi) \tag{3.10}$$

for every $i = 0, 2, \cdots, [\frac{1}{2}(\bar{n}_1 - 1)] - 1$; *and*

$$\beta(\tau_{\mathbf{x}_i}\xi) > 0 \tag{3.11}$$

for every $i = 1, 3, \cdots, [\frac{1}{2}(\bar{n}_1 - 1)]$.

A verbal paraphrase of this theorem is more involved than that of Theorem 3.2 and will not be attempted. For the same reason this theorem is less intuitively obvious than Theorem 3.2. However, the motivation behind this theorem will be well understood if the reader first goes through the Example 5.3.

Sufficient conditions for (3.4) or (3.8) to hold is given by Hoeffding [4] which we quote below for the sake of completeness.

Theorem 3.4. [Hoeffding] *In order that*

$$\lim_{n\to\infty} \rho_n(\xi) = \rho(\xi) \tag{3.12}$$

it is sufficient that either

$$\lim_{n\to\infty} E_\xi \rho_0(\tau_{\mathbf{x}_n}\xi) = 0 \tag{3.13}$$

or, in case $L(\theta, a)$ *is bounded,*

$$\xi\{c(\theta) > 0\} = 1. \tag{3.14}$$

We note that (3.1) is sufficient for (3.13) and hence for (3.12).

4. Specialization to two-action (hypothesis testing) problem. Let us define

$$L(\theta) = L(\theta, a_1) - L(\theta, a_2) \tag{4.1}$$

and call it, following Mikhalevich [7], the *loss characteristic*. If, for every θ, there is an action for which the loss is zero, then the optimum stopping risk depends on the losses through their difference $L(\theta)$ only as seen in the following lemma.

Lemma 4.1. *If for each* $\theta \in \mathfrak{I}$ *there is an* a_θ *such that* $L(\theta, a_\theta) = 0$ *then for any* ξ

$$2\rho_0(\xi) = E_\xi |L(\Theta)| - |E_\xi L(\Theta)| \leqq \{\text{Var}_\xi L(\Theta)\}^{\frac{1}{2}}. \tag{4.2}$$

We note that the hypothesis of Lemma 4.1 is redundant if we want to minimize the expected *regret* instead of the expected loss where regret is defined as

$$R(\theta, a) = L(\theta, a) - \min_a L(\theta, a).$$

Lemma 4.1 may be used as a tool for the application of Theorem 3.1, Corollary 3.1 and Theorem 3.4 to two-action problems. For example, when Lemma 4.1 applies

$$\text{essup}_{x_n} \text{Var}_{\tau_{x_n}\xi} \{L(\Theta)\} \to 0 \quad \text{as} \quad n \to \infty \tag{4.3}$$

can be used to replace (3.1) in Corollary 3.1, and also to replace (3.13) in Theorem 3.4.

We present Corollary 4.1 to treat the important special case of *linear loss* which is defined for a one-dimensional real parameter θ as follows:

Definition. The loss function L is said to be a linear loss function if

$$\begin{aligned} L(\theta, a_1) &= 0 && \text{if} \quad \theta \leqq \theta_0, \\ &= k_1(\theta - \theta_0) && \text{if} \quad \theta > \theta_0; \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} L(\theta, a_2) &= k_2(\theta_0 - \theta) && \text{if} \quad \theta \leqq \theta_0, \\ &= 0 && \text{if} \quad \theta > \theta_0; \end{aligned} \tag{4.4$'$}$$

where k_1 and k_2 are certain positive constants. θ_0 is called the *break-even* point. We shall refer to the case where $k_1 = k_2 = k$ as the case of *symmetric linear loss*.

Corollary 4.1. *When the cost per observation is a constant independent of θ, a Bayes sequential procedure for testing $\theta \leqq \theta_0$ against $\theta > \theta_0$ with linear loss, is truncated if*

$$\lim_{n\to\infty} \text{essup}_{x_n} \text{Var}_{\tau_{x_n}\xi}\, \Theta = 0. \tag{4.5}$$

It may easily be verified that (4.5) is satisfied in the following special cases involving common statistical distributions for X and their *natural conjugate priors* for ξ. By a natural conjugate prior we mean an *a priori* distribution such that all possible *a posteriori* distributions belong to the same family as the *a priori*. For a fuller definition we refer to Raiffa and Schlaifer [9]. Sometimes such a family of *a priori* distributions has been called *closed under sampling* (see Wetherill [13]).

Case (i) $X \sim N(\theta, 1), \quad -\infty < \theta < \infty;$
$\xi \sim N(\mu, \sigma^2)$

(ii) $X \sim \text{Poisson}(\theta), \quad \theta > 0$
$d\xi(\theta) \propto \theta^{\lambda-1}e^{-\alpha\theta}\, d\theta, \quad \alpha > 0, \lambda \geqq 1.$

(iii) $X \sim \text{Binomial}(\theta), \quad 0 < \theta < 1;$
$d\xi(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}, \quad a > 0, b > 0.$

(iv) $X \sim \text{Trinomial}(\pi_1, \pi_2, 1 - \pi_1 - \pi_2), \quad 0 < \pi_i < \pi_1 + \pi_2 \leqq 1, \; i = 1, 2,$
$d\xi(\pi_1, \pi_2) \propto \pi_1^{a-1}\pi_2^{b-1}(1 - \pi_1 - \pi_2)^{c-1}, \quad a > 0, b > 0, c > 0;$
$\theta \equiv \pi_1 - \pi_2.$

Just as Lemma 4.1 is useful in the application of Theorem 3.1 so is the following lemma useful in the application of Theorem 3.2 and Theorem 3.3 to two-action problems.

LEMMA 4.2. *For any* ξ,

$$2\lambda(\xi) = E_\xi |L(\tau_X\xi)| - |L(\xi)|. \tag{4.6}$$

In the rest of this section we consider only densities $f_\theta(x)$ belonging to the *one-dimensional exponential family.* Let

$$f_\theta(x) = \psi(\theta)e^{\theta x} \tag{4.7}$$

where

$$\psi(\theta) = [\int_{-\infty}^{\infty} e^{\theta x}\, d\mu(x)]^{-1} > 0 \quad \text{for all} \quad \theta \in \mathfrak{I}$$

and $\mathfrak{I}$ is an interval $[\underline{\theta}, \bar{\theta}]$ which may be finite or infinite, open, half-open or closed.

For any positive integer n, the *a posteriori* distribution $\tau_{\mathbf{x}_n}\xi$ depends on $\mathbf{x}_n$ through the sufficient statistic $(n, v \equiv \sum_{i=1}^n x_i)$, so that we may write it as $\tau_{n,v}\xi$. The set of all *a posteriori* distributions arising from a fixed *a priori* ξ may thus be identified with a certain subset of the (n, v)-plane. Fixing our attention to a particular ξ, we may, for simplicity of notation, replace $\tau_{n,v}\xi$ by the pair (n, v) or even simply by v_n if no confusion arises. If, however, we are interested in a class of Bayes procedures corresponding to a class of natural conjugate priors then we may denote ξ by (n', v') and $\tau_{n,v}\xi$ by $(n' + n, v' + v)$.

We assume that $L(\theta)$ is non-decreasing in θ and takes on both positive and negative values with positive ξ-probability. Since the family $\{\tau_{n,v}\xi\}$ of distributions, written explicitly as

$$d\tau_{n,v}\xi(\theta) = [\psi(\theta)]^n e^{v\theta}\, d\xi(\theta)/[\int_{\mathfrak{I}} [\psi(\theta)]^n e^{v\theta}\, d\xi(\theta)] \tag{4.8}$$

forms a monotone likelihood ratio family it follows from Lemma 6.6 that $L(n, v) = E_{\tau_{n,v}\xi}L(\theta)$ is non-decreasing in v and takes both positive and negative values for every positive integer n. We now assume that $\mu(x)$ is absolutely continuous with respect to the Lebesgue measure so that $L(n, v)$ is continuous in v for each n. The monotonicity and continuity of $L(n, v)$, which has its range on both sides of zero, ensure the existence of a unique root, v_n^0, say, of $L(n, v) = 0$ for each n. Following Wetherill [13], we shall call the set of points $\{n, v_n^0\}$ the *neutral boundary,* since for such points the stopping risk is the same for either action.

The Condition (3.5) of Theorem 3.2 may be written in this special situation as follows.

$$\operatorname{essup}_{v_n} [\lambda(v_n)/c(v_n)] \leqq 1. \tag{4.9}$$

For the case of constant (independent of θ) cost of observation, we are thus led to investigate when and where $\lambda(v_n)$ attains its essential supremum. We now state a theorem which states that $\lambda(v_n)$ attains its supremum on the neutral boundary. Theorem 3.2, properly interpreted in this special case implies that sampling can continue longest on the neutral boundary.

THEOREM 4.1. *If the density $f_\theta(x)$ belongs to the one-dimensional exponential family given by* (4.9) where μ *is absolutely continuous with respect to the Lebesgue measure and if the loss characteristic $L(\theta)$ is non-decreasing in θ then $\lambda(v_n)$ attains* *its supremum on the neutral boundary* (n, v_n^0). *Moreover this supremum is given by*

(4.10) $\quad \operatorname{essup}_{v_n} \lambda(v_n) = \lambda(v_n^0) = |\int_{\mathfrak{Z}} L(\theta) F_\theta(v_{n+1}^0 - v_n^0)\, d\tau_{n,v_n^0}\xi(\theta)|$.

Using this theorem we may define $\bar{n}_0$ as the smallest positive integer for which $\lambda(v_n^0) \leqq c$ for every $n \geqq \bar{n}_0$. If the cost depends on θ, but $c(v_n)$ attains its infimum at v_n^0 then the left hand side of (4.9) is still given by

$$\chi_n \equiv \lambda(v_n^0)/c(v_n^0).$$

Hence $\bar{n}_0$ is given in such cases by the smallest positive integer for which $\chi_n \leqq 1$ for every $n \geqq \bar{n}_0$.

5. Examples.

Example 5.1 (Normal). Suppose X has a normal density with mean θ and variance 1. Suppose the loss $L(\theta, a_1)$ and $L(\theta, a_2)$ are so defined that

(5.1) $$L(\theta) \equiv L(\theta, a_1) - L(\theta, a_2) = k(\theta - \theta_0)$$

where k is a positive constant. Equation (5.1) holds for *symmetric linear loss* (see (4.4) and (4.4′) where $k_1 = k_2 = k$). We take the *a priori* distribution ξ to be normal with mean μ_0 and variance σ_0^2. Then it may be verified that the *a posteriori* distribution $\tau_{n,v_n}\xi$ after n observations is also normal with mean μ_n and variance σ_n^2 where

(5.2) $$\mu_n = (v_n + \mu_0\sigma_0^{-2})/(n + \sigma_0^{-2}), \qquad \sigma_n^2 = 1/(n + \sigma_0^{-2}).$$

The neutral boundary is given by (n, v_n^0) where

(5.3) $$v_n^0 = [(\theta_0 - \mu_0)/\sigma_0^2] + n\theta_0 .$$

Hence $\tau_{n,v_n^0}\xi$ is normal with mean θ_0 and variance σ_n^2. It follows from Theorem 4.1 that

(5.4) $$\operatorname{essup}_{v_n} \lambda(v_n) = \lambda(v_n^0) = |\int_{-\infty}^{\infty} k(\theta - \theta_0)\Phi(\theta_0 - \theta)\, d\Phi[(\theta - \theta_0)/\sigma_n]|$$

where Φ is the standard normal probability integral. Evaluating this integral, we derive

(5.5) $$\lambda(v_n^0) = [k/(2\pi)^{\frac{1}{2}}][\sigma_n^2/(1 + \sigma_n^2)^{\frac{1}{2}}] = [k/(2\pi)^{\frac{1}{2}}][(n + \sigma_0^{-2})^2 + (n + \sigma_0^{-2})]^{-\frac{1}{2}},$$

a decreasing function of n.

We shall consider two types of cost, viz. $c(\theta) \equiv 1$ and $c(\theta) = |\theta - \theta_0|$ called the *absolute deviation cost.* The later type of costs arises in connection with ethical cost in medical trials (see Anscombe [2]).

(i) *Constant cost,* $c = 1$. Applying Theorem 3.2, all the conditions of which may easily be verified, the exact stage of truncation n^0 for this Bayes test is given by the smallest non-negative integer for which $\lambda(v_n^0) \leqq 1$ i.e., using (5.5), the smallest non-negative integer greater than or equal to $(k^2/2\pi + \frac{1}{4})^{\frac{1}{2}} - \frac{1}{2} - 1/\sigma_0^2$.

For convenience of notation let us denote

$$
\begin{aligned}
[x]^* &= 0 && \text{if } x \leqq 0\\
&= [x] && \text{if } x > 0 \text{ and integral}\\
&= [x] + 1 && \text{otherwise}
\end{aligned}
$$

where $[x]$ denotes the largest integral part in x.

Using the above notation we display n^0 as follows:

$$n^0 = [((k^2/2\pi) + \tfrac{1}{4})^{\frac{1}{2}} - \tfrac{1}{2} - (1/\sigma_0^2)]^*. \tag{5.6}$$

This result is also given by Wetherill [13].

(ii) *Absolute deviation cost*, $c = |\theta - \theta_0|$. In this case the expected cost $c(v_n)$ of one observation at the nth stage is given by the expected value of $|\theta - \theta_0|$ where the expectation is with respect to the normal distribution with mean μ_n and variance σ_n^2. It follows from (5.2) and (5.3) that $\mu_n = \theta_0$ if and only if $v_n = v_n^0$. Now since the mean deviation is minimum from the median (mean in this case) we readily see that $c(v_n)$ attains its minimum on the neutral boundary, i.e.,

$$\operatorname{essup}_{v_n} \lambda(v_n)/c(v_n) = \lambda(v_n^0)/c(v_n^0)$$

$$c(v_n^0) = \int_{-\infty}^{\infty} |\theta - \theta_0| \, d\Phi[(\theta - \theta_0)/\sigma_n] = (2/\pi)^{\frac{1}{2}}\sigma_n .$$

$$\lambda(v_n^0)/c(v_n^0) = \tfrac{1}{2}k[\sigma_n/(1 + \sigma_n^2)^{\frac{1}{2}}] \tag{5.7}$$

a decreasing function of n. Applying Theorem 3.2, the exact stage of truncation is given in this case by the smallest positive integer for which (5.7) $\leqq 1$. Thus

$$n^0 = [(k^2/4) - 1 - (1/\sigma_0^2)]^*. \tag{5.8}$$

Comparing (5.8) with (5.6) we note that with absolute deviation cost, the exact stage of truncation is approximately quadratic in k while it is approximately linear in k for constant cost.

Example 5.2. (Binomial). Consider the problem of the Bayes sequential test for $p \leqq p_0$ versus $p > p_0$ where p is a binomial parameter and $0 < p_0 < 1$. Suppose the losses $L(p, a_1)$ and $L(p, a_2)$ are so defined that

$$L(p) \equiv L(p, a_1) - L(p, a_2) = k(p - p_0) \tag{5.9}$$

where k is a positive constant. Equation (5.9) holds for *symmetric linear loss* as defined in (4.4) and (4.4′) if $k_1 = k_2 = k$.

We consider an *a priori* distribution ξ which is beta with parameters (a_0, b_0). Specifically,

$$d\xi(p) = p^{a_0-1}(1 - p)^{b_0-1}\, dp/B(a_0, b_0), \qquad 0 \leqq p \leqq 1, \quad a_0 > 0, \quad b_0 > 0.$$

It is well known that the *a posteriori* distribution $\tau_{n,v_n}\xi$ after n observations is also beta with parameters.

$$
\begin{aligned}
a &= a_0 + v_n\\
b &= b_0 + n - v_n .
\end{aligned} \tag{5.10}
$$

Thus without any ambiguity we shall sometimes write $\xi_{a,b}$ for $\tau_{n,v_n}\xi$. As in the previous example we shall consider two types of cost, viz. (i) $c(p) \equiv 1$ and, (ii) the absolute deviation cost $c(p) = |p - p_0|$ per observation.

We now use Lemma 4.2 to compute λ. For this example we have

$$(5.11) \quad 2\lambda(\xi_{a,b}) = |L(\xi_{a+1,b})| \, E_{\xi_{a,b}}p + |L(\xi_{a,b+1})| \, E_{\xi_{a,b}}(1 - p) - |L(\xi_{a,b})|$$

where

$$L(\xi_{a,b}) = E_{\xi_{a,b}}[k(p - p_0)] = k[(a/(a + b)) - p_0].$$

For every (a, b) such that $L(\xi_{a+1,b})$, $L(\xi_{a,b+1})$ and $L(\xi_{a,b})$ have the same sign, $\lambda(\xi_{a,b}) = 0$ as may be verified using (6.26). Considering furthermore, all possible situations where these three quantities are not of the same sign we finally arrive at the following:

$$(5.12) \quad \begin{aligned} \lambda(\xi_{a,b}) &= k[ab/((a + b)^2(a + b + 1))], && \text{for } (a, b) \in \Gamma_0, \\ &= k[b/(a + b)] \\ &\quad \cdot[((a + 1)/(a + b + 1)) - p_0], && \text{for } (a, b) \in \Gamma_1, \\ &= k[b/(a + b)][p_0 - (a/(a + b + 1))], && \text{for } (a, b) \in \Gamma_2, \\ &= 0, && \text{otherwise,} \end{aligned}$$

where

$$\begin{aligned} \Gamma_0 &= \{(a, b) | [a/(a + b)] = p_0\}, \\ \Gamma_1 &= \{(a, b) | [a/(a + b)] < p_0 < (a + 1)/(a + b + 1)\}, \\ \Gamma_2 &= \{(a, b) | [a/(a + b + 1)] < p_0 < [a/(a + b)]\}. \end{aligned}$$

Γ_0 is the neutral boundary, i.e., the set of points for which either action is equally preferable when we stop. We may describe the set Γ_j, $j = 1, 2$ as set of points which, though not on the neutral boundary, are so near to it that the next observation has the potential to change the current optimal action a_j to the other one. Thus we may call the set

$$\Gamma_0 = \Gamma_0 \cup \Gamma_1 \cup \Gamma_2$$

as the *extended neutral boundary.*

The neutral boundary has previously been defined as the set of points (n, v_n^0) in the sample space of the sufficient statistic (n, v_n). Using (5.10) and writing $n_0 = a_0 + b_0$, $q_0 = 1 - p_0$, we may write (5.12) alternatively as follows.

$$(5.13) \quad \begin{aligned} \lambda(v_n) &= [k/(n_0 + n + 1)]p_0q_0, && \text{for } v_n = v_n^0, \\ &= [k/(n_0 + n + 1)][p_0 - ((v_n^0 - v_n)/(n_0 + n))] \\ &\quad \cdot[q_0 - (v_n^0 - v_n)], && \text{for } v_n^0 - q_0 < v_n < v_n^0, \\ &= [k/(n_0 + n + 1)][p_0 - (v_n - v_n^0)] \\ &\quad \cdot[q_0 - ((v_n - v_n^0)/(n_0 + n))], && \text{for } v_n^0 < v_n < v_n^0 + p_0, \\ &= 0, && \text{otherwise.} \end{aligned}$$

It follows from (5.13) that

$$\lambda(v_n) \leqq (kp_0q_0)/(n_0 + n + 1) \tag{5.14}$$

with equality if and only if $v_n = v_n^0$. We note at this point that a result similar to (5.13) holds for any arbitrary *a priori* distribution ξ. In fact it is possible to show (see [10]) in this binomial problem that for any $\xi \varepsilon \Xi$, and any $L(\theta)$,

$$\lambda(\xi) \leqq \operatorname{Var}_\xi [L(\Theta)] \tag{5.15}$$

with equality if and only if ξ is such that $L(\xi) = 0$. Thus if v_n^0 is an integer then we may write in this case

$$\sup_{v_n} \lambda(v_n) = \lambda(v_n^0) = \operatorname{Var}_{\tau_{n,v_n^0}\xi}[L(\Theta)],$$

corresponding to (5.4) for the normal case.

Applying Theorem 3.2(i) in this example, we find that for the case of constant cost, $\bar{n}_0$ is given by the smallest positive integer such that $\lambda(v_n^0) = kp_0q_0/(n_0 + n + 1) \leqq 1$ for every $n \geqq \bar{n}_0$ i.e.,

$$\bar{n}_0 = [kp_0q_0 - n_0 - 1]^*. \tag{5.16}$$

For the case of *absolute deviation cost*, we have

$$\begin{aligned} c(\xi_{a,b}) &= \int_0^1 |p - p_0|\, d\xi_{a,b}(p) \\ &= 2[p_0 I_{p_0}(a, b) - [a/(a + b)]I_{p_0}(a + 1, b)] + [a/(a + b)] - p_0 \end{aligned} \tag{5.17}$$

where $I_x(a, b)$ denotes the incomplete beta integral. Because of (5.12) we need evaluate this integral only for (a, b) on the extended neutral boundary. Using (5.12) and (5.17) it is feasible to compute $\bar{n}_0$ by calculating the ratio $\lambda(a, b)/c(a, b)$ along the extended neutral boundary for decreasing $a + b$ starting from a sufficiently large $a + b$. A good approximation for $\bar{n}_0$ is obtained by using the following asymptotic (large $a + b$) relation

$$c(\xi_{a,b}) \sim [(2p_0q_0/\pi)(1/(a + b))]^{\frac{1}{2}} \qquad \text{for} \quad (a, b) \,\varepsilon\, \Gamma. \tag{5.18}$$

Writing (5.18) as

$$c(v_n) \sim [(2p_0q_0/\pi)(1/(n_0 + n))]^{\frac{1}{2}} \qquad \text{for } v_n^0 - q_0 < v_n < v_n^0 + p_0$$

we see that

$$\lambda(v_n)/c(v_n) \lesssim \tfrac{1}{2}k(2\pi p_0q_0)^{\frac{1}{2}}[n_0 + n]^{-\frac{1}{2}}.$$

Thus $\bar{n}_0$ is given approximately as

$$\bar{n}_0 \approx \tfrac{1}{2}k^2\pi p_0q_0 - n_0 \tag{5.19}$$

which compares with (5.16) as (5.8) does with (5.6) in the normal case. We note that because of the discrete nature of the observations it is impossible (for both cost functions), to find a sequence of observations satisfying the Condition (3.6) of Part (ii) of Theorem 3.2. Thus $\bar{n}_0 > n^0$ where n^0 is the exact stage of truncation. It is possible, however, to apply Theorem 3.3 to obtain n^0 exactly, in the following special symmetric case of the problem.

Example 5.3. (Symmetric Binomial). $p_0 = \frac{1}{2}$, $a_0 = b_0$, $L(p) = k(p - \frac{1}{2})$. We assume for simplicity that a_0 is an integer. In this case it may be verified that Γ_1 and Γ_2 are empty and the neutral boundary Γ_0 consists of points on the diagonal $a = b$ in the (a, b)-plane. Thus

$$(5.20)\qquad \lambda(\xi_{a,b}) = k/4(2a + 1) \quad \text{for } a = b$$
$$= 0 \quad \text{for } a \neq b.$$

In order to determine $\bar{n}_0$ we need evaluate $c(a, b)$ only for $a = b$. Now

$$c(\xi_{a,a}) = 1 \quad \text{for constant cost}$$
$$= \int_0^1 |p - \tfrac{1}{2}|\, d\xi_{a,a}(p) \quad \text{for absolute deviation cost.}$$

We may easily evaluate

$$(5.21)\qquad \int_0^1 |p - \tfrac{1}{2}|\, d\xi_{a,a}(p) = \binom{2a}{a} 2^{-2a-1}.$$

We now apply Theorem 3.2 (i) in this example to obtain the following:

In the case of *constant cost*, $c = 1$,

$$(5.22)\qquad \bar{n}_0 = 2(a^0 - a_0) - 1$$

where

$$(5.23)\qquad a^0 = [k/8 - \tfrac{1}{2}]^* \approx k/8.$$

This result is also given by Moriguti and Robbins [8].

In the case of *absolute deviation cost*, $c = |p - \frac{1}{2}|$,

$$\bar{n}_0 = 2(a^0 - a_0) - 1$$

where a^0 is the smallest non-negative integer such that

$$(5.24)\qquad (2a + 1)\binom{2a}{a} 2^{-2a+1} \geqq k, \quad \text{for every} \quad a \geqq a^0.$$

Standard binomial tables may be used to obtain a^0 from (5.24). For large k, we have

$$a^0 \approx \pi k^2/16$$

which corresponds, using (5.22), to (5.19) with $p_0 = \frac{1}{2}$.

We are now prepared to apply Theorem 3.3. We notice from (5.20) that $(\lambda - c)_+(\xi_{a,b}) = 0$ for $|a - b| > 0$. Thus

$$E_{\xi_{a,b}}\{(\lambda - c)_+(\tau_X \xi_{a,b})\}$$
$$= (\lambda - c)_+(\xi_{a+1,b}) E_{\xi_{a,b}} p + (\lambda - c)_+(\xi_{a,b+1}) E_{\xi_{a,b+1}}(1 - p)$$
$$= 0 \quad \text{for } |a - b| > 1.$$

Hence $\beta(\xi_{a,b}) = 0$ for $|a - b| > 1$ and for $a = b$. Because of symmetry, $\beta(\xi_{a,a-1}) = \beta(\xi_{a-1,a})$. Now

$$\beta(\xi_{a,a-1}) = \max\{0, -c(\xi_{a,a-1}) + [(a - 1)/(2a - 1)][\lambda(\xi_{a,a}) - c(\xi_{a,a})]_+\}.$$

From the definition of a^0, $\lambda(\xi_{a,a}) \leqq c(\xi_{a,a})$ for every $a \geqq a^0$. Thus $\beta(\xi_{a,a-1}) = 0$ for $a \geqq a^0$. Also from symmetry one easily verifies that $c(\xi_{a,a-1}) = c(\xi_{a-1,a-1})$. We may thus apply Theorem 3.3 (i) to obtain $\bar{n}_1 = 2(a^1 - a_0) - 1$ where a^1 is the smallest positive integer such that

$$-c(\xi_{a-1,a-1}) + [\lambda(\xi_{a,a}) - c(\xi_{a,a})][(a-1)/(2a-1)] \leqq 0 \quad \text{for every} \quad a \geqq a^1.$$

It follows from the above inequality that $a^1 < a^0$ and hence $\bar{n}_1 < \bar{n}_0$. It follows from (5.20) that in the case of *constant cost* $c = 1$, a^1 is given by the smallest positive integer such that

$$(5.25) \quad 4(2a+1)[(3a-2)/(a-1)] \geqq k \quad \text{for every} \quad a \geqq a^1.$$

Using (5.21), in addition, it follows that in the case of *absolute deviation cost* $c = [p - \frac{1}{2}|$, a^1 is given by the smallest positive integer such that

$$(5.26) \quad (2a+1)\binom{2a}{a}2^{-2a+1}(3a-1)/(a-1) \geqq k \quad \text{for every} \quad a \geqq a^1.$$

Comparing (5.25) with (5.23) and (5.26) with (5.24) we note that in either case

$$(5.27) \qquad a^1 \sim a^0/9.$$

For either cost function c, it may be verified that a sequence of observations that alternates with 0 and 1 satisfies the Condition (3.10) and (3.11) of Theorem 3.3 (ii) so that $\bar{n}_1$ coincides with n^0, the exact stage of truncation. It follows from Lemma 6.3 that every $\bar{n}_k$, $k \geqq 1$, must coincide with n^0. We note that for obvious reasons $(a^1 - 1, a^1 - 1)$ has sometimes been called the *last reachable continuation point* while $(a^0 - 1, a^0 - 1)$ has been called the *last continuation point*. For computation of the Bayes boundary by backward induction it is sufficient to know the last reachable point only, beyond which the Bayes continuation set consists of a long neck of points on the diagonal $a = b$ up to $(a^0 - 1, a^0 - 1)$. It is seen from (5.27) that considerable saving may be effected if a^1 is known precisely. Since a^0 is approximately proportional to k for the case of constant cost and to k^2 for the case of absolute deviation cost, the saving effected is more pronounced for the latter case than with the former.

6. Proofs.

LEMMA 6.1. *For any positive integer n and for any $\xi \in \Xi$,*

$$(6.1) \qquad 0 \leqq (\rho_n - \rho_{n+1})(\xi) \leqq E_\xi\{(\rho_{n-1} - \rho_n)(\tau_X\xi)\}.$$

PROOF. By definition $\rho_m(\xi)$ satisfies the following well known recursion relation:

$$(6.2) \qquad \rho_m(\xi) = \min[\rho_0(\xi), c(\xi) + E_\xi\rho_{m-1}(\tau_X\xi)], \qquad m \geqq 1.$$

Substituting $m = n$ and $m = n + 1$ in (6.2), subtracting, and using the trivial fact that $\rho_m(\xi)$ is non-increasing in m three cases occur:

(i) For ξ such that $\rho_0(\xi) \leqq c(\xi) + E_\xi\rho_n(\tau_X\xi)$,

$$(\rho_n - \rho_{n+1})(\xi) = 0.$$

(ii) For ξ such that $c(\xi) + E_\xi \rho_{n-1}(\tau_X \xi) \leqq \rho_0(\xi)$,

$$(\rho_n - \rho_{n+1})(\xi) = E_\xi\{(\rho_{n-1} - \rho_n)(\tau_X \xi)\}. \tag{6.3}$$

(iii) For ξ such that

$$c(\xi) + E_\xi \rho_n(\tau_X \xi) \leqq \rho_0(\xi) \leqq c(\xi) + E_\xi \rho_{n-1}(\tau_X \xi), \tag{6.4}$$

$$(\rho_n - \rho_{n+1})(\xi) = \rho_0(\xi) - c(\xi) - E_\xi \rho_n(\tau_X \xi),$$

$$(\rho_n - \rho_{n+1})(\xi) \leqq E_\xi\{(\rho_{n-1} - \rho_n)(\tau_X \xi)\} \tag{6.5}$$

using the second inequality in (6.4).

Combining all three cases the lemma follows.

Let

$$\mu_n^{(k)}(\xi) \equiv E_\xi\{(\rho_k - \rho_{k+1})(\tau_{\mathbf{X}_{n-k}} \xi)\}, \qquad k = 0, 1, 2, \cdots, n. \tag{6.6}$$

Lemma 6.2. $(\rho_n - \rho_{n+1})(\xi) \equiv \mu_n^{(n)}(\xi) \leqq \mu_n^{(n-1)}(\xi) \leqq \cdots \leqq \mu_n^{(1)}(\xi) \leqq \mu_n^{(0)}(\xi)$.

Proof. In (6.1) substitute k for n, $\tau_{\mathbf{x}_{n-k}} \xi \equiv \xi_{n-k}$ for ξ, and X_{n-k+1} for X. Then we have

$$(\rho_k - \rho_{k+1})(\xi_{n-k}) \leqq E_{\xi_{n-k}}\{(\rho_{k-1} - \rho_k)(\tau_{X_{n-k+1}} \xi_{n-k})\}, \quad 1 \leqq k \leqq n. \tag{6.7}$$

Since (6.7) is true for every $\xi_{n-k} \varepsilon \Xi$, and hence for $\tau_{\mathbf{x}_{n-k}} \xi$ for every $\mathbf{x}_{n-k}$, taking expectations over $\mathbf{X}_{n-k}$, we have,

$$\mu_n^{(k)}(\xi) \leqq \mu_n^{(k-1)}(\xi), \tag{6.8}$$

completing the proof of the lemma.

Let $\bar{n}_k$ denote the smallest non-negative integer such that $\mu_n^{(k)}(\xi) = 0$ for every $n \geqq \bar{n}_k$.

Lemma 6.3. *If*

$$\lim_{n \to \infty} \rho_n(\xi) = \rho(\xi), \tag{6.9}$$

then

$$n^0 \leqq \cdots \leqq \bar{n}_k \leqq \cdots \leqq \bar{n}_1 \leqq \bar{n}_0 .$$

Proof. Since $\mu_n^{(k)}(\xi)$ is an upper bound on $\rho_n(\xi) - \rho_{n+1}(\xi)$, it follows from the definition of $\bar{n}_k$ that

$$\rho_n(\xi) = \rho_{n+1}(\xi) \qquad \text{for every } n \geqq \bar{n}_k.$$

This along with (6.9) implies that

$$\rho_{\bar{n}_k}(\xi) = \rho(\xi).$$

Thus there is a Bayes procedure which is truncated at $\bar{n}_k$. The monotonicity of $\bar{n}_k$ follows from that of $\mu_n^{(k)}(\xi)$, completing the proof of the lemma.

We now investigate conditions under which the bounds are exact. Now by definition of n^0, it is the largest positive integer for which $\rho_{n^0-1}(\xi) > \rho_{n^0}(\xi)$.

Thus the upper bound $\bar{n}_k$ coincides with n^0 if and only if $\rho_{\bar{n}_k-1}(\xi) > \rho_{\bar{n}_k}(\xi)$. We are thus led to investigate conditions under which

$$\rho_n(\xi) > \rho_{n+1}(\xi)$$

for any non-negative integer n. A simple sufficient condition is given by the following lemma.

LEMMA 6.4. *For any non-negative integer n, a sufficient condition for*

$$\rho_n(\xi) > \rho_{n+1}(\xi) \tag{6.10}$$

is that there exists a set of $\mathbf{x}_n$ with positive probability such that

$$\rho_0(\tau_{\mathbf{x}_i}\xi) > \rho_1(\tau_{\mathbf{x}_i}\xi) \tag{6.11}$$

for each $i = 0, 1, \cdots, n$.

PROOF. The inequality (6.11) and the monotonicity of ρ_m imply that

$$\rho_0(\tau_{\mathbf{x}_i}\xi) > \rho_{n-i}(\tau_{\mathbf{x}_i}\xi)$$

and

$$\rho_0(\tau_{\mathbf{x}_i}\xi) > \rho_{n+1-i}(\tau_{\mathbf{x}_i}\xi)$$

where $i \leqq n - 1$. It follows that (6.3) applies and hence

$$(\rho_{n-i} - \rho_{n+1-i})(\tau_{\mathbf{x}_i}\xi) = E_{\tau_{\mathbf{x}_i}\xi}\{(\rho_{n-1-i} - \rho_{n-i})(\tau_X(\tau_{\mathbf{x}_i}\xi))\}, \tag{6.12}$$

for $i = 0, 1, \cdots, n - 1$. Combining the equalities in (6.12), starting from $i = 0$ up to $i = n - 1$ successively, we obtain

$$(\rho_n - \rho_{n+1})(\xi) = E_\xi\{(\rho_0 - \rho_1)(\tau_{\mathbf{X}_n}\xi)\}. \tag{6.13}$$

Now the quantity within the braces of the right hand side of (6.13) is non-negative by the monotonicity of ρ_m and is strictly positive, by (6.11) with $i = n$, for a set of $\mathbf{x}_n$ with positive probability. Hence the left hand side of (6.13) is strictly positive, completing the proof of the lemma.

LEMMA 6.5. *For any non-negative integer m, a sufficient condition for*

$$\rho_{2m}(\xi) > \rho_{2m+1}(\xi)$$

is that there exists a set of $\mathbf{x}_{2m}$ with positive probability such that

$$\rho_0(\tau_{\mathbf{x}_{2j}}\xi) > \rho_1(\tau_{\mathbf{x}_{2j}}\xi) \tag{6.14}$$

for each $j = 0, 1, \cdots, m - 1$, and

$$\rho_1(\tau_{\mathbf{x}_{2j+1}}\xi) > \rho_2(\tau_{\mathbf{x}_{2j+1}}\xi) \tag{6.15}$$

for each $j = 0, 1, \cdots, m - 1$.

PROOF. It follows just as in Lemma 6.4 that (6.14) implies

$$(\rho_{2m-2j} - \rho_{2m+1-2j})(\tau_{\mathbf{x}_{2j}}\xi) = E_{\tau_{\mathbf{x}_{2j}}\xi}\{(\rho_{2m-1-2j} - \rho_{2m-2j})(\tau_X(\tau_{\mathbf{x}_{2j}}\xi))\}, \tag{6.16}$$

for $j = 0, 1, \cdots, m - 1$. Again, the Inequality (6.15) implies by the monotonicity of ρ_m that

$$(\rho_0(\tau_{\mathbf{x}_{2j+1}}\xi) > \rho_{2m-1-2j}(\tau_{\mathbf{x}_{2j+1}}\xi)$$

and

$$\rho_0(\tau_{\mathbf{x}_{2j+1}}\xi) > \rho_{2m-2j}(\tau_{\mathbf{x}_{2j+1}}\xi)$$

where $j \leqq m - 2$. It follows just as in (6.12) that

$$(6.17) \quad (\rho_{2m-2j-1} - \rho_{2m-2j})(\tau_{\mathbf{x}_{2j+1}}\xi) = E_{\tau_{\mathbf{x}_{2j+1}}\xi}\{(\rho_{2m-2(j+1)} - \rho_{2m+1-2(j+1)})(\tau_X(\tau_{\mathbf{x}_{2j+1}}\xi))\}, \quad \text{for } j = 0, 1, \cdots, m-2.$$

Combining (6.16) and (6.17) alternately and successively we have

$$(6.18) \qquad (\rho_{2m} - \rho_{2m+1})(\xi) = E_\xi\{(\rho_1 - \rho_2)(\tau_{\mathbf{x}_{2m-1}}\xi)\}.$$

Now the quantity within the braces of the right hand side of (6.18) is non-negative by the monotonicity of ρ_m and is strictly positive, by (6.15) with $j = m - 1$, for a set of $\mathbf{x}_{2m}$ with positive probability. Hence the left hand side of (6.18) is strictly positive completing the proof of the lemma.

Proof of Theorem 3.2. Applying (3.3) and (6.2) we have

$$(6.19) \qquad \rho_0(\xi) - \rho_1(\xi) = \max\{0, \lambda(\xi) - c(\xi)\} \equiv (\lambda - c)_+(\xi).$$

Using (6.19), Part (i) follows from Lemma 6.3 with $k = 0$ while Part (ii) follows from Lemma 6.4.

Proof of Theorem 3.3. Since

$$\rho_2(\xi) = \min[\rho_0(\xi), c(\xi) + E_\xi\rho_1(\tau_X\xi)],$$

we have

$$\rho_0(\xi) - \rho_2(\xi) = \max[0, \rho_0(\xi) - E_\xi\rho_1(\tau_X\xi) - c(\xi)].$$

Adding and subtracting $E_\xi\rho_0(\tau_X\xi)$ in the second part of the right hand side and using (3.3) and (6.19) we have

$$\rho_0(\xi) - \rho_2(\xi) = [(\lambda - c)(\xi) + E_\xi\{(\lambda - c)_+(\tau_X\xi)\}]_+ .$$

Thus

$$(6.20) \qquad \rho_1(\xi) - \rho_2(\xi) = \beta(\xi).$$

Using (6.20), Part (i) follows from Lemma 6.3 with $k = 1$ while Part (ii) follows from Lemma 6.5.

Proof of Lemma 4.1. By definition of $\rho_0(\xi)$,

$$\rho_0(\xi) = \min[L(\xi, a_1), L(\xi, a_2)]$$

$$(6.21) \qquad = \tfrac{1}{2}[L(\xi, a_1) + L(\xi, a_2)] - \tfrac{1}{2}|L(\xi, a_1) - L(\xi, a_2)|$$

$$= \tfrac{1}{2}E_\xi\{\min[L(\Theta, a_1), L(\Theta, a_2)] + |L(\Theta)|\} - \tfrac{1}{2}|E_\xi L(\Theta)|. \tag{6.22}$$

By the hypothesis of the lemma, $\min[L(\theta, a_1), L(\theta, a_2)] = 0$ for every $\theta \varepsilon \mathfrak{J}$. Thus

$$\rho_0(\xi) = \tfrac{1}{2}E_\xi|L(\Theta)| - \tfrac{1}{2}|E_\xi L(\Theta)|. \tag{6.23}$$

Now by the Cauchy-Schwarz inequality

$$\begin{aligned} E_\xi|L(\Theta)| &\leqq [E_\xi L^2(\Theta)]^{\frac{1}{2}} \\ &\leqq [\operatorname{Var}_\xi L(\Theta)]^{\frac{1}{2}} + |E_\xi L(\Theta)|. \end{aligned} \tag{6.24}$$

Equation (4.2) follows immediately from (6.23) and (6.24).

Corollary 4.1 follows from Corollary 3.1 and Lemma 4.1 noting that the hypothesis of Lemma 4.1 is satisfied by linear loss.

PROOF OF LEMMA 4.2. From the Definition (3.3) of λ, we have using (6.22),

$$2\lambda(\xi) = [E_\xi|g(\Theta)| - E_\xi E_{\tau_X\xi}|g(\Theta)|] + E_\xi|E_{\tau_X\xi}L(\Theta)| - |E_\xi L(\Theta)| \tag{6.25}$$

where $g(\theta) = \min[L(\theta, a_1), L(\theta, a_2)] + |L(\theta)|$. Using the fact that the average of the random measure, $\tau_X\xi$, on $\mathfrak{J}$, averaged over the marginal distribution of X, is ξ itself, it follows that for any integrable function $g(\theta)$,

$$E_\xi E_{\tau_X\xi}g(\Theta) = E_\xi g(\Theta). \tag{6.26}$$

Hence, the first part in the right hand side of (6.25) vanishes, completing the proof of the lemma.

LEMMA 6.6. *If $f_\theta(x)$ is a family of densities on the real line $\mathfrak{R}$ with monotone likelihood ratio in x, and if g is a non-decreasing function of x, then $E_\theta g(X)$ is a non-decreasing function of θ.*

We refer to Lehmann ([6], p. 74) for a proof of this lemma.

NOTE. It may easily be verified that the conclusion of the lemma holds if g is a non-decreasing function of both x and θ.

LEMMA 6.7. *For a one-dimensional exponential family $\{f_\theta(x), \theta \varepsilon \mathfrak{J}\}$ of densities given by (4.9) and for each non-negative integer n, the family of densities*

$$f_{n,v}(x) = E_{\tau_{n,v}\xi}\{f_\Theta(x)\} \tag{6.27}$$

has a monotone likelihood ratio in x, where $\tau_{n,v}\xi$ is given by (4.8).

PROOF. For $v_1 < v_2$,

$$\begin{aligned} f_{n,v_2}(x)/f_{n,v_1}(x) = {} & [\textstyle\int_{\mathfrak{J}} [\psi(\theta)]^{n+1} \exp[(v_2 + x)\theta]\, d\xi(\theta)/ \\ & \textstyle\int_{\mathfrak{J}} [\psi(\theta)]^{n+1} \exp[(v_1 + x)\theta]\, d\xi(\theta)] \cdot k_n(v_1)/k_n(v_2) \end{aligned} \tag{6.28}$$

where $k_n(v_i) = \int_{\mathfrak{J}} [\psi(\theta)]^n \exp[v_i\theta]\, d\xi(\theta)$. Let

$$d\eta_x(\theta) = \{[\psi(\theta)]^{n+1} \exp[(v_1 + x)\theta]\, d\xi(\theta)\}/\{\textstyle\int_{\mathfrak{J}} [\psi(\theta)]^{n+1} \exp[(v_1 + x)\theta]\, d\xi(\theta)\}$$

and

$$g(\theta) = \exp[(v_2 - v_1)\theta].$$

Then the first factor on the right hand side of (6.28) may be written as

$$E_x g(\Theta) \equiv \int_{\mathfrak{I}} g(\theta)\, d\eta_x(\theta). \tag{6.29}$$

Now for $x_1 < x_2$,

$$d\eta_{x_2}(\theta)/d\eta_{x_1}(\theta) = \exp\,[(x_2 - x_1)\theta]\cdot[k_{n+1}(v_1 + x_1)/k_{n+1}(v_1 + x_2))$$

is a non-decreasing function of θ. Thus $\{d\eta_x(\theta)\}$ has a monotone likelihood ratio in θ. Since $g(\theta)$ is non-decreasing in θ it follows from Lemma 6.6 that $E_x g(\Theta)$ is non-decreasing in x. Hence $\{f_{n,v}(x)\}$ has a monotone likelihood ratio in x for every n, completing the proof of the lemma.

Proof of Theorem 4.1. It follows from Lemma 4.2, that

$$2\lambda(v_n) = E_{\xi_n}|L(n + 1, v_n + X)| - |L(n, v_n)| \tag{6.30}$$

where (n, v_n) in the argument of L represents $\xi_n \equiv \tau_{n,v_n}\xi$. Let

$$U_a(y) = 0, \qquad y - a \leqq 0$$
$$= 1, \qquad y - a > 0.$$

From the definition of v_{n+1}^0 and the monotonicity of $L(n + 1, v)$ we have

$$|L(n + 1, v)| = L(n + 1, v)U_{v_{n+1}^0}(v) + [-L(n + 1, v)][1 - U_{v_{n+1}^0}(v)]. \tag{6.31}$$

Replacing v by $v_n + X$ and taking expectation with respect to X, we have $E_{\xi_n}|L(n + 1, v_n + X)| = h_1(v_n) + h_2(v_n)$ where

$$h_1(v_n) = E_{\xi_n}\{L(n + 1, v_n + X)U_{v_{n+1}^0}(v_n + X)\} \tag{6.32}$$

and

$$h_2(v_n) = E_{\xi_n}\{[-L(n + 1, v_n + X)][1 - U_{v_{n+1}^0}(v_n + X)]\}. \tag{6.33}$$

Now

$$h_1(v_n) - h_2(v_n) = E_{\xi_n}\{L(n + 1, v_n + X)\} = E_{\xi_n}\{E_{\tau_X\xi_n}L(\Theta)\}$$

$$h_1(v_n) - h_2(v_n) = E_{\xi_n}\{L(\Theta)\} \equiv L(n, v_n) \tag{6.34}$$

by (6.26).

It follows from the above equations and the monotonicity of $L(n, v)$ in v that

$$\lambda(v_n) = h_1(v_n) \qquad \text{for } v_n \leqq v_n^0$$
$$= h_2(v_n) \qquad \text{for } v_n \geqq v_n^0.$$

Hence to show $\lambda(v_n)$ attains its supremum at v_n^0, it is sufficient to show that $h_1(v_n)$ is non-decreasing and $h_2(v_n)$ is non-increasing in v_n. Now we may write

$$h_j(v_n) = \int_{\mathfrak{R}} g_j(v_n, x) f_{n,v_n}(x)\, d\mu(x)$$

where $f_{n,v_n}(x)$ is given by (6.27) and

$$g_1(v_n, x) = L(n + 1, v_n + x)U_{v_{n+1}^0}(v_n + x)$$

and

$$g_2(v_n, x) = -L(n + 1, v_n + x)[1 - U_{v_{n+1}^0}(v_n + x)]$$

are easily seen to be non-decreasing and non-increasing in $v_n + x$ respectively. Also by Lemma 6.7 the family $\{f_{n,v_n}(x)\}$ has a monotone likelihood ratio in x (where v_n is identified with the parameter). It follows from the note after Lemma 6.6, that h_1 is non-decreasing and h_2 is non-increasing completing the proof of the fact that

$$\text{essup}_{v_n} \lambda(v_n) = \lambda(v_n^0).$$

We now derive the other part of Equation (4.10),

$$\begin{aligned}\lambda(v_n^0) = |h_2(v_n^0)| &= |E_{\xi_n^0}[L(n + 1, v_n^0 + X) \mid X \leqq v_{n+1}^0 - v_n^0]| \\ &= |\textstyle\int_{-\infty}^{v_{n+1}^0 - v_n^0} L(n + 1, v_n^0 + x) f_{n,v_n^0}(x)\, d\mu(x)\, dx|.\end{aligned}$$

Now

$$\begin{aligned}L(n + 1, v_n^0 + x) &= E_{\tau_x \xi_n^0} L(\Theta) \\ &= \textstyle\int_{\mathfrak{Z}} L(\theta) f_\theta(x)\, d\xi_n^0(\theta) / \int_{\mathfrak{Z}} f_\theta(x)\, d\xi_n^0(\theta) \\ &= \textstyle\int_{\mathfrak{Z}} L(\theta) f_\theta(x)\, d\xi_n^0(\theta) / f_{n,v_n^0}(x).\end{aligned}$$

Thus

$$\begin{aligned}\lambda(v_n^0) &= |\textstyle\int_{\mathfrak{Z}} L(\theta)[\int_{-\infty}^{v_{n+1}^0 - v_n^0} f_\theta(x)\, d\mu(x)\, dx]\, d\xi_n^0(\theta)| \\ &= |\textstyle\int_{\mathfrak{Z}} L(\theta) F_\theta(v_{n+1}^0 - v_n^0)\, d\xi_n^0(\theta)| \\ &= E_{\tau_{n,v_n^0}\xi}\{L(\Theta) F_\Theta(v_{n+1}^0 - v_n^0)\}.\end{aligned}$$

7. Acknowledgments. The author is much indebted to Professor W. J. Hall of the University of North Carolina, who initiated the author to these problems. The author also wishes to express his deep gratitude to Professor C. Stein and to Professor H. Chernoff for many stimulating discussions that led to substantial improvement on the original version of the paper. The author is thankful to the referee for pointing out an error in the original manuscript.

REFERENCES

[1] Amster, S. J. (1963). A modified Bayes stopping rule. *Ann. Math. Statist.* **34** 1404–1413.
[2] Anscombe, F. J. (1963). Sequential medical trials. *J. Amer. Statist. Assoc.* **58** 365–383.
[3] Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.
[4] Hoeffding, W. (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.* **31** 352–368.
[5] Kiefer, J. and Weiss, L. (1957). Some properties of generalized sequential probability ratio tests. *Ann. Math. Statist.* **28** 57–74.
[6] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.
[7] Mikhalevich, V. S. (1956). Sequential Bayes solutions and optimal methods of statistical acceptance control. *Theor. Probability Appl.* **1** 395–421.

[8] MORIGUTI, S. and ROBBINS, H. (1962). A Bayes test of '$p \leqq \frac{1}{2}$' versus '$p > \frac{1}{2}$'. *Rep. Statist. Appl. Res. Un. Jap. Sci. Engrs.* **9** 1–22.

[9] RAIFFA, H. AND SCHLAIFER, R. (1961). *Applied Statistical Decision Theory.* Graduate School of Business Administration, Harvard Univ.

[10] RAY, S. N. (1963). Some sequential Bayes procedures for comparing two binomial parameters when observations are taken in pairs. *Institute of Statistics Mimeograph Series No. 318*, Chapel Hill.

[11] SOBEL, M. (1953). An essential complete class of decision functions for certain standard sequential procedures. *Ann. Math. Statist.* **24** 319–337.

[12] WALD, A. (1950). *Statistical Decision Functions.* Wiley, New York.

[13] WETHERILL, G. B. (1961). Bayesian sequential analysis. *Biometrika* **48** 281–292.

[14] WEISS, L. (1962). On sequential tests which minimize the maximum expected sample size. *J. Amer. Statist. Assoc.* **57** 551–566.