The Annals of Statistics
1974, Vol. 2, No. 1, 170–176

NOTE ON ANDERSON'S SEQUENTIAL PROCEDURES WITH TRIANGULAR BOUNDARY[1]

By Václav Fabian

Michigan State University

Simple expressions are given for the probability of a correct decision for the Anderson sequential procedure with a triangular boundary, for some values of parameters. This can be used to improve bounds of the probability of correct decision for the Paulson sequential ranking procedures and change slightly the comparison of these with the Bechhofer-Kiefer-Sobel procedures.

1. Introduction. The triangular stopping regions were, among some others, considered by Armitage (1957), Donelly (1957) and Anderson (1960) for testing a mean of a normal population. Later they were used to construct a ranking procedure by Paulson (1964).

Suppose the triangular region is given by the lower and upper boundaries

$$L_l(n) = -a + \lambda_1 n \,, \qquad L_u(n) = a - \lambda_2 n$$

with $a > 0$, $\lambda_1 \geqq 0$, $\lambda_2 \geqq 0$; $\{S_n\}$ is a sequence of partial sums of independent normal (μ, σ^2) random variables; and N is the first index n for which S_n is not in the continuation interval $I_n = (L_l(n), L_u(n))$. A decision d_2 is taken if $S_N \leqq L_l(N)$ and I_N is non-empty, or if $S_N \leqq \frac{1}{2}(L_l(N) + L_u(N))$ and I_N is empty. Otherwise another decision d_1 is selected. Note that if I_N is empty then N lies in the interval $[n_0, n_0 + 1)$ with $n_0 = 2a(\lambda_1 + \lambda_2)^{-1}$. It is then of interest to know the probabilities of selecting d_1 and d_2 respectively; let us denote by $p(\mu, \lambda_1, \lambda_2, a, \sigma^2)$ the probability of selecting d_2.

It is easy to verify that $p(\mu, \lambda_1, \lambda_2, a, \sigma^2) = p(\mu/\sigma, \lambda_1/\sigma, \lambda_2/\sigma, a/\sigma, 1)$, $p(\mu, \lambda_1, \lambda_2, a, 1) = p(0, \lambda_1 - \mu, \lambda_2 + \mu, a, 1) = p(\mu_0, \lambda, \lambda, a, 1)$ if $\lambda = (\lambda_1 + \lambda_2)/2$, $\mu_0 = \mu + \lambda - \lambda_1$. Thus we may restrict our attention to cases where $\lambda_1 = \lambda_2$, $\sigma^2 = 1$ and denote $p(\mu, \lambda, \lambda, a, 1)$ simply by $p(\mu, \lambda, a)$.

It will help to think about $p(\mu, \lambda, a)$, for $\mu \geqq 0$, as a probability of an incorrect decision.

Analogously let $q(\mu, \lambda, a)$ denote the probability that a continuous Brownian motion X on $[0, +\infty)$, with $EX_t = \mu t$, Var $X_t = t$ exits $I_t = (-a + \lambda t, a - \lambda t)$ through the lower boundary L_l. (The event that $X_t \in I_t$ for all t such that $I_t \neq \emptyset$ is null.)

Received May 1972; revised March 1973.

[1] This research was supported in part by NFS research grant number GP31123X.

AMS 1970 *subject classifications*. Primary 62L10; Secondary 62F07.

Key words and phrases. Triangular stopping regions, Paulson's sequential ranking procedure, Bechhofer-Kiefer-Sobel procedure, probability of correct decision, independent normal random variables, Brownian motion.

The number $q(\mu, a, \lambda)$ is an approximation to $p(\mu, a, \lambda)$; Anderson (1960, page 170) remarks that it is actually an upper bound for $p(\mu, a, \lambda)$ if $\mu \geqq 0$.

Donelly (1957) and Anderson (1960) have obtained expressions for $q(\mu, \lambda, a)$ which, unfortunately, are not easy to use for computation. Paulson (1964), when using the sequential test in his sequential ranking procedure, does not use Anderson's expressions but derives an upper bound $e^{-2(\mu-\lambda)a}$ for $p(\mu, \lambda, a)$ with $\mu \geqq \lambda$. The proof is essentially based on the Wald fundamental identity.

The purpose of this note is to show that the standard argument[2] of interchanging a density by another, on a set where the ratio of the two densities is bounded by a constant, gives an improvement to the bound used by Paulson, namely

$$p(\mu, \lambda, a) \leqq e^{-2(\mu-\lambda)a}(1 - p(\mu - 2\lambda, \lambda, a)) \qquad \text{if} \quad \mu \geqq \lambda\,. \tag{1}$$

For example, if $\lambda = \frac{1}{2}\mu$, this gives

$$p(\mu, \tfrac{1}{2}\mu, a) \leqq \tfrac{1}{2}e^{-2(\mu-\lambda)a} = \tfrac{1}{2}e^{-\mu a}\,. \tag{2}$$

The same technique, applied to $q(\mu, a, \lambda)$, yields an equality in (1) instead of the inequality, and makes possible an evaluation of $q(\mu, \lambda, a)$ for all $\lambda = \mu/(2j)$, $j = 1, 2, \cdots$.

I do not know whether the results concerning $q(\mu, \lambda, a)$ (Theorem 2.3) can be deduced easily from Anderson's (1960) results, but even if they can, the present derivation is much simpler. It is possible, however, that suitable numerical methods may make Anderson's results (Corollary 4.4, in particular) not so difficult to use.

Our formulas make it possible to compute $q(\mu, \lambda, a)$ only for μ and λ such that μ/λ is an even integer. Interpolation might complete the results for other values of μ/λ in some practical considerations. There is also some evidence how λ should be chosen. In two examples considered by Anderson (1960, Section 3), of testing $\mu = -1$ against $\mu = 1$ at the level .05 and .01, the optimal λ (mimimizing EN for $\mu = 0$) was found to be approximately .33 and .41. In general, when testing that the normal mean θ is μ against $\theta = -\mu$ at a prescribed level, asymptotically optimal value of λ (making $E_0 N$ minimal) is $\lambda = \mu/2$ (this is easy to see by using the strong law of large numbers, or using, e.g., Perng's (1969) Theorem 2.2). The same results hold when the procedure is used as a component in the Paulson ranking procedure (1966); here there are Monte Carlo studies by Ramberg (1966).

It is known that the triangular stopping region is not asymptotically optimal, and asymptotically optimal stopping regions, e.g., that of a truncated Wald sequential probability ratio test, are known. However, the probability of error for the latter are difficult to estimate. Also, based on Hoeffding's (1960) results,

[2] For a recent use of this argument leading to a considerable simplification of previous proofs, see Robbins (1970).

it is possible to see that the procedures from the Anderson example are nearly optimal (cf. Anderson (1960), page 173).

Ramberg (1966) shows that the Paulson procedure, for some values of parameters, is better than the Bechhofer–Kiefer–Sobel procedures; an improved bound for the probability of correct decision would increase the set of such parameter values (cf. [11]).

2. The results. The results are formulated in Theorems 2.2 and 2.3 using the notations $p(\mu, \lambda, a)$ and $q(\mu, \lambda, a)$ from Section 1. Through most of the discussion a and λ are fixed ($a > 0$, $\lambda \geqq 0$) and then we abbreviate $p(\mu, \lambda, a)$ as p_μ and $q(\mu, \lambda, a)$ as q_μ.

In the lemmas and proofs we shall write $P_\mu([\xi_1, \xi_2, \cdots] \in A)$ for the probability of the indicated event when $\xi_1, \xi_2, \cdots$ is a sequence of independent normal $(\mu, 1)$ random variables.

2.1. LEMMA. *Let N be a stopping time determined by a sequence $\{M_n\}$ of Borel subsets of R^n such that*

$$\{N = n\} = \{[\xi_1, \cdots, \xi_n] \in M_n\}\,.$$

Then, for any $\Delta \in R$, any closed subinterval I of R and any subintervals B_n of I,

$$P_\Delta(\textstyle\sum_{i=1}^N \xi_i \in B_N) = \kappa P_{-\Delta}(\sum_{i=1}^N \xi_i \in B_N)$$

with a $\kappa \in \{e^{2\Delta y};\, y \in I\}$.

PROOF. Denote by f_Δ the normal $(\Delta, 1)$ density so that $(f_\Delta/f_{-\Delta})(x) = e^{2\Delta x}$ and set

$$H_n = \{[x_1, \cdots, x_n];\ \textstyle\sum_{i=1}^n x_i \in B_n\}\,.$$

Then

$$\begin{aligned} P_\Delta(\textstyle\sum_{i=1}^N \xi_i \in B_N) &= \textstyle\sum_{n=1}^\infty P_\Delta(\sum_{i=1}^n \xi_i \in B_n,\ N = n) \\ &= \textstyle\sum_{n=1}^\infty \int_{H_n \cap M_n} \prod_{i=1}^n f_\Delta(x_i)\, dx_1 \cdots dx_n \\ &= \textstyle\sum_{n=1}^\infty \int_{H_n \cap M_n} \exp[2\Delta \sum_{j=1}^n x_j] \prod_{i=1}^n f_{-\Delta}(x_i)\, dx_1 \cdots dx_n \end{aligned}$$

and the assertion follows.

2.2. THEOREM. *If $\mu \geqq \lambda$ then*

$$p_\mu \leqq e^{-2a(\mu-\lambda)} p_{2\lambda-\mu} = e^{-2a(\mu-\lambda)}(1 - p_{\mu-2\lambda})\,; \tag{1}$$

in particular

$$p(\mu, \tfrac{1}{2}\mu, a) \leqq \tfrac{1}{2} e^{-a\mu}\,, \tag{2}$$

$$p(\mu, 0, a) \leqq e^{-2a\mu}/(1 + e^{-2a\mu})\,. \tag{3}$$

PROOF. Write $p_\Delta = P\{\sum_{i=1}^N (Z_i - \lambda) \leqq -a_N\}$ with $Z_1, Z_2, \cdots$ independent normal $(\Delta, 1)$, $a_n = \max\{a, \lambda n\}$ and with N the first time $-a < \sum_{i=1}^n (Z_i - \lambda) < a - 2\lambda n$ is violated. Then

$$p_\Delta = P_{\Delta-\lambda}(\textstyle\sum_{i=1}^N \xi_i \leqq -a_N)\,,$$

and applying this with $\Delta = \mu$ first, with $\Delta = 2\lambda - \mu$ later, and using Lemma 2.1 in between, we obtain

$$p_\mu \leqq \exp[-2a(\mu - \lambda)]P_{\lambda-\mu}(\sum_{i=1}^N \xi_i \leqq -a_N)$$
$$= \exp[-2a(\mu - \lambda)]p_{2\lambda-\mu} .$$

Since $p_\Delta = 1 - p_{-\Delta}$, this establishes (1). Inequality (2) obtains from (1) upon observing that $p_0 = \frac{1}{2}$. For $\lambda = 0$, (1) implies $p_\mu \leqq e^{-2a\mu}(1 - p_\mu)$ and this is equivalent to (3) (which, of course, concerns the Wald sequential probability ratio test and is well known).

The inequality (1) in the above theorem is due to the excess over the boundary. In the case of Brownian motion there is no excess over the boundary. This roughly corresponds to the possibility of replacing $I = (-\infty, -a]$ in the preceding proof by $[-a, -a]$. Lemma 2.1 then yields an equality instead of an inequality and we obtain the result of the next theorem. In fact, Lemma 2.1 does not apply directly to the continuous case; since the main idea is obvious the proof itself is postponed to Section 3.

We shall use the Kronecker symbol $\delta_{ij} = \chi_{\{i\}}(j)$.

2.3. THEOREM. *If* $\mu \geqq \lambda$ *then*

(1)
$$q_\mu = e^{-2a(\mu-\lambda)}q_{2\lambda-\mu} = e^{-2a(\mu-\lambda)}(1 - q_{\mu-2\lambda}) ;$$

in particular, for $\mu = 2j\lambda$

(2)
$$q(\mu, \lambda, a) = \sum_{s=1}^{j} (-1)^{s+1}\beta^{(2j-s)s/(2j-1)}[1 - \delta_{sj}\tfrac{1}{2}]$$

where

(3)
$$\beta = e^{-2a(\mu-\lambda)} .$$

2.4. REMARK. For μ an even multiple of λ, relation (2.3.2) gives a direct formula for the evaluation of the error probability $q(\mu, \lambda, a)$. Often, with a preassigned α and given μ we want to determine λ and a so that

(1)
$$q(\mu, \lambda, a) = \alpha .$$

If we also choose λ then the value of a satisfying (1) is uniquely determined and can be obtained by setting

(2)
$$a = \frac{1}{2}\frac{\log \beta^{-1}}{\mu - \lambda}$$

where β is the solution of $f(\beta) = \alpha$ with $f(\beta)$ the right-hand side of (2.3.2).

Some of the values of β are given in Table 1; a solution to $f(\beta) = \alpha$ can be easily obtained by starting with $\beta_0 = \alpha$ and using an iterative procedure to improve β_0. Notice that the Paulson's bound $q_\mu \leqq \exp[-2a(\mu - \lambda)]$ yields a given by (2) with β replaced by α.

For μ not equal to an even multiple of λ, formula (2) with β obtained by interpolation could be used.

3. Proof of Theorem 2.3. First we strengthen Theorem 2.2 in the following way.

3.1. LEMMA. *If* $\mu \geqq \lambda > 0$ *and* $c > \mu - \lambda$ *then*

$$e^{-2(a+\lambda+c)(\mu-\lambda)}(p_{2\lambda-\mu} - r) \leqq p_\mu \leqq e^{-2a(\mu-\lambda)}p_{2\lambda-\mu}$$

with

$$r = (a\lambda^{-1} + 1)(\lambda - \mu + c)^{-1}\exp[-\tfrac{1}{2}(\lambda - \mu + c)^2] .$$

PROOF. Only the left-hand inequality has to be proved as the rest is contained in Theorem 2.2. As in the proof there

$$p_\mu = P_{\mu-\lambda}(\textstyle\sum_{i=1}^N \xi_i \leqq -a_N) \geqq P_{\mu-\lambda}(\sum_{i-1}^N \xi_i \in [-a_N - c, -a_N]) .$$

An application of Lemma 2.1 to the last term gives, since $a_N < a + \lambda$,

$$\begin{aligned} p_\mu &\geqq \exp[-2(a + \lambda + c)(\mu - \lambda)]P_{\lambda-\mu}(\textstyle\sum_{i=1}^N \xi_i \in [-a_N - c, -a_N]) \\ &\geqq \exp[-2(a + \lambda + c)(\mu - \lambda)][p_{2\lambda-\mu} - P_{\lambda-\mu}(\textstyle\sum_{i=1}^N \xi_i < -a - c)] \end{aligned}$$

and it remains to prove that the second term in the brackets is less or equal to r. Since $\sum_{i=1}^{N-1} \xi_i > -a$, and $N < a\lambda^{-1} + 1$, this probability is less or equal to

$$P_{\lambda-\mu}\left\{\xi_i < -c \text{ for some } 1 \leqq i \leqq \frac{a}{\lambda} + 1\right\} \leqq \left(\frac{a}{\lambda} + 1\right) \int_{-\infty}^{-c-\lambda+\mu} e^{-t^2/2}\, dt$$

and this is less or equal to r since $\int_{-\infty}^{-z} e^{-t^2/2}\, dt \leqq z^{-1}e^{-z^2/2}$ for $z > 0$.

3.2. PROOF OF THEOREM 2.3. Let X_t be a continuous Brownian motion, $EX_t = \Delta t$, $\text{Var } X_t = t$, let τ and τ_m be defined as the times of exit of

$$X_t \qquad \text{from} \quad (-a + \lambda t, a + \lambda t)$$

with t unrestricted for τ, and t restricted to the set $\{1/m, 2/m, 3/m, \cdots\}$ for τ_m. Set

$$A = \{X_\tau \leqq -a + \lambda\tau\} , \qquad A_m = \{X_{\tau_m} \leqq -a + \lambda\tau_m\} .$$

We want to show that

$$P(A_m) \to P(A) . \tag{1}$$

According to the law of the iterated logarithm at 0 (cf., e.g., Breiman, Section 12.9), for every number c, with probability 1, every $(0, \varepsilon)$ with $\varepsilon > 0$ contains a t_1 and t_2 such that $X(t_1) < -ct_1$, $X(t_2) > ct_2$. Applying this to $X_1(t) = X(t + \tau) - X(\tau)$ we obtain that, with probability one, $(0, \tau + \varepsilon)$ contains points at which X is outside the continuation interval, for every $\varepsilon > 0$.

Also, if $\lambda > 0$, $P\{X_{a/\lambda} \neq 0\} = 1$. Let Ω_1 be the intersection of the two events of probability 1.

Thus if $\omega \in \Omega_1 \cap A$ and if N is a neighborhood of $\tau(\omega)$, then by continuity of $X(\omega)$ there is an open interval $(u, v) \subset N$ on which $X(\omega)$ is below the lower boundary and such that on $[0, v]$, $X(\omega)$ does not touch the upper boundary. But then $\omega \in \Omega_1 \cap A_m$ for all $m > (v - u)^{-1}$.

Similarly, if $\omega \in \Omega_1 \cap (\Omega - A)$ then $\omega \in \Omega_1 \cap (\Omega - A_m)$ for m sufficiently large. Thus $\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} (\Omega_1 \cap A_m) \subset \Omega_1 \cap A \subset \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} (\Omega_1 \cap A_m)$ and, since $P(\Omega_1) = 1$, (1) holds. Of course $P(A) = q(\mu, \lambda, a)$ and $P(A_m) = p(\Delta/m, \lambda/m, \lambda/m, a, 1/m) = P(\Delta m^{-\frac{1}{2}}, \lambda m^{-\frac{1}{2}}, am^{\frac{1}{2}})$ and (1) means, that for every Δ, $\lambda \geqq 0$, $a > 0$

$$P\left(\frac{\Delta}{m^{\frac{1}{2}}}, \frac{\lambda}{m^{\frac{1}{2}}}, am^{\frac{1}{2}}\right) \to q(\Delta, \lambda, a) . \tag{2}$$

Now suppose $\lambda > 0$. Lemma 3.1 states that

$$\begin{aligned} \exp[-2(a + \lambda m^{-1} + c_m m^{-\frac{1}{2}})(\mu - \lambda)] \left[p\left(\frac{2\lambda - \mu}{m^{\frac{1}{2}}}, \frac{\lambda}{m^{\frac{1}{2}}}, am^{\frac{1}{2}}\right) - r_m \right] \\ \leqq p\left(\frac{\mu}{m^{\frac{1}{2}}}, \frac{\lambda}{m^{\frac{1}{2}}}, am^{\frac{1}{2}}\right) \\ \leqq \exp[-2a(\mu - \lambda)] p\left(\frac{2\lambda - \mu}{m^{\frac{1}{2}}}, \frac{\lambda}{m^{\frac{1}{2}}}, am^{\frac{1}{2}}\right) \end{aligned}$$

if $c_m > (\mu - \lambda)/m^{\frac{1}{2}}$, with

$$r_m = [a\lambda^{-1}m + 1]\left(\frac{\lambda - \mu}{m^{\frac{1}{2}}} + c_m\right)^{-1} \exp\left[-\frac{1}{2}\left(\frac{\lambda - \mu}{m^{\frac{1}{2}}} + c_m\right)^2\right].$$

Taking $c_m = m^{\frac{1}{4}}$ we obtain $r_m \to 0$ and using (2) and taking limits in the inequalities above we obtain the assertion (2.3.1) for $\lambda > 0$.

For $\lambda = 0$ one obtains the result by showing that $q(\mu, \lambda, a) \to q(\mu, 0, a)$ as $\lambda \downarrow 0$, using similar considerations as in proving (1). Relation (2.3.2) follows from (2.3.1) by induction and using $q_0 = \frac{1}{2}$.

Acknowledgment. I am thankful to Mr. W. Allard who prepared a program and supervised the computation of the table on a computer.

TABLE 1
Values of β for which $q(\mu, \lambda, a) = \alpha$

μ/λ \ α	.1	.05	.01	.005	.001	.001
2	.2	.1	.02	.010	.002	.0002
4	.13443	.06237	.011262	.005483	.0010536	.00010240
6	.12367	.05742	.010549	.005182	.0010142	.00010037
8	.11957	.05571	.010354	.005108	.0010070	.00010014
10	.11745	.05487	.010272	.005080	.0010046	.00010008
12	.11617	.05439	.010229	.005065	.0010035	.00010005
$+\infty$	.11111	.05263	.010101	.005025	.0010010	.00010001

REFERENCES

[1] ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

[2] ARMITAGE, P. (1957). Restricted sequential procedures. *Biometrika* **44** 9–26.

[3] BECHHOFER, E. ROBERT, KIEFER, JACK and SOBEL, MILTON (1968). *Sequential Identification and Ranking Procedures.* Univ. of Chicago Press.

[4] Breiman, Leo (1968). ***Probability***. Addison-Wesley, Reading, Mass.

[5] Donelly, T. G. (1957). A family of truncated sequential tests. Ph. D. dissertation, Univ. of North Carolina.

[6] Hoeffding, Wassily (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.* **31** 352–368

[7] Paulson, Edward (1964). A sequential procedure for selecting the population with the largest mean from k normal populations. *Ann. Math. Statist.* **35** 174–180.

[8] Perng, S. K. (1969). A comparison of the asymptotic expected sample sizes of two sequential procedures for ranking problem. *Ann. Math. Statist.* **40** 2198–2202.

[9] Ramberg, John S. (1966). A comparison of the performance characteristics of two sequential procedures for ranking means of normal populations. Technical Report No. 4, Department of Industrial Engineering, Operations Research, Cornell Univ.

[10] Robbins, Herbert (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41** 1397–1409.

Added in proof.

[11] Ramberg, John S. and Bechhofer, Robert E. (1973). A comparison of two multiple decision procedures for ranking means of normal populations with common known variance (abstract 139-20). *IMS Bull.* **2** 108.

Department of Statistics and Probability
Michigan State University
East Lansing, Michigan 48823