EFFICIENT RECURSIONS FOR TRUNCATION OF THE SPRT

S. M. Pollock
Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109

D. Golhar
Management Department
Western Michigan University
Kalamazoo, Michigan 49008

EFFICIENT RECURSIONS FOR TRUNCATION OF THE SPRT

Stephen M. Pollock

Department of Industrial
and Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109

Damodar Y. Golhar

Management Department
Western Michigan University
Kalamazoo, Michigan 49008

ABSTRACT

Existing methods for characterizing truncated sequential probability ratio tests are either a) conservative, in that they may produce a truncation point m that is larger than that needed to guarantee given α and β values; b) inefficient in that the computational effort needed to find m is proportional to m^3; or c) subject to considerable round-off errors as m increases. We present an efficient recursive method for computing m that is linear in m, does not accumulate round-off errors, and produces the smallest conservative value of m. An example is given for i.i.d. normal observations.

INTRODUCTION

Suppose we wish to test H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$ by sequentially observing independent random variables X_1, X_2,...., X_n

with density $f_i(x_i|\theta_j)$, $j = 0, 1$, but with a limit on the maximum number of samples.

Wald (8) originally proposed the SPRT, with <u>truncation point</u> m, as follows: First define the single sample log-likelihood ratio

$$z_i = \ln \left\{ \frac{f_i(x_i|\theta_1)}{f_i(x_i|\theta_0)} \right\} , \quad i = 1, 2, \ldots n$$

and the total log-likelihood ratio up to and including the nth observation

$$\underline{z}_n = \sum_{i=1}^{n} z_i$$

Then, the truncated SPRT rule is:

reject H_0 if $\underline{z}_n \geq b$ for $n = 1, 2, \ldots, m$

accept H_0 if $\underline{z}_n \leq a$ for $n = 1, 2, \ldots, m$

and take one more observation if

$a < \underline{z}_n < b$ for $n = 1, 2, \ldots, m-1$.

If the experiment has not stopped at or before m observations, then

reject H_0 if $b > \underline{z}_m \geq 0$

and accept H_0 if $a < \underline{z}_m < 0$. \hfill (1)

The stopping bounds a and b are given by the approximation

$$a \doteq \ln \left\{ \frac{\beta}{1-\alpha} \right\} \quad \text{and} \quad b \doteq \ln \left\{ \frac{1-\beta}{\alpha} \right\} \hfill (2)$$

α and β in relation (2) represent the <u>desired</u> error probabilities of type I and type II respectively. For any fixed value of m, this procedure will produce actual errors $\tilde{\alpha}$ and $\tilde{\beta}$ that will, in general, differ from α and β.

In this paper we find the smallest truncation point m^* such that the actual error probabilities $\tilde{\alpha}$ and $\tilde{\beta}$ are less than or equal to the desired error probabilities α and β. To do so, we explore the relationship between m^* and the parameters of the density functions $f_i(x_i|\theta_j)$, in order to help an experimenter choose (in advance) a truncation point guaranteeing any desired error probabilities. We strictly restrict ourselves to using

Wald's stopping bounds (relation (2)), since they are easily computable and well understood. We thus do not, for example, consider the more efficient but less transparent non-constant bound techniques of Anderson (1960) and Armitage (1957).

Wald (1947) suggested setting the truncation point m "large enough" such that the effect of truncation on the actual error probabilities would be minimal. In this spirit, Johnson (1961) gives a working rule for choosing a non-integer truncation point m^* when the observations are independent and identically distributed normal random variables. This Johnson rule sets m^* = 3 $\sup_j E(N|\theta_j)$, where $E(N|\theta_j)$ is the expected number of observations for the <u>untruncated</u> SPRT when θ_j is the true state of nature, for j = 0, 1. An integer point m^{**} is then obtained by rounding m^* up to the next higher integer. It has been shown by Golhar and Pollock (1983) that the resulting integer truncation point m^{**} tends to be conservative, in that smaller truncation points can still guarantee errors no worse than α and β.

Aroian and Robison (1969) proposed obtaining a truncation point by first finding the operating characteristic function

$$L_j = L(\theta_j) = \text{prob.}\{\text{accept } H_0|\theta = \theta_j\}.$$

This is gotten from the p.d.f. of the random variable $\underline{Z}_n$ given $\bigcap_{k=1}^{n-1} \{a < \underline{z}_k < b\}$.

To obtain this p.d.f., let

$$P_j(z,n) = \text{prob.}\{(\underline{z}_n \leq z) \bigcap_{k=1}^{n-1} (a < \underline{z}_k < b)|\theta = \theta_j\}$$

be the c.d.f. for the random variable Z_n after having observed n samples and let $p_j(z,n)$ be the p.d.f. of Z, i.e., the derivative of $P_j(z,n)$ with respect to z.

Successive convolutions can then be used to calculate $p_j(z,n)$:

$$p_j(z, 1) = f_1(z|\theta_j) \quad \text{(by definition)} \tag{3}$$

$$p_j(z,n) = \int_a^b p_j(u, n-1) f(z-u|\theta_j) \, du \quad \text{for } n > 1. \tag{4}$$

These relationships in turn produce the operating characteristic function

$$L_j(m) = \sum_{n=1}^{m-1} [\int_{-\infty}^{a} p_j(z,n)\, dz] + \int_{-\infty}^{o} p_j(z,m)\, dz, \qquad j = 0, 1 \quad (5)$$

We can now recursively use equations (3-5) for $m = 1, 2, \ldots,$ and find the minimum value of m such that both $L_1(m) \leq \beta$ and $L_0(m) \geq 1-\alpha$ hold. Thus, the actual error probabilities do not exceed the desired error probabilities.

The major difficulty with this approach is that the numerical methods needed to evaluate relations (3)-(5) are tractable only if m is small. However, for large m, as shown by Golhar (1983), the number of grid points needed to carry out the required numerical integration increases on the order of m^3. Hence both computing time and memory increase on the order of km^3, where the constant k depends on the degree of accuracy desired. Computing time may thus become prohibitive for large m, particularly if one is interested in microcomputer implementation. In addition, the accumulation of round-off errors is exacerbated by the recursive nature of equation (4).

We present here a more efficient method using a different recursive formulation, one for which computing time and memory only are on the order of km. We also, by example, develop a simple relationship between m^* and the distribution parameters for IID normal random variables, when $\alpha = \beta$.

THE RECURSION METHOD

Let m be the maximum number of observations to be taken. Let $\phi_k(t)$ be the type I error probability with at most k more observations, when the present log-likelihood ratio (i.e., up to and including the $(m-k)_{th}$ observation) is t. If $a < t < b$ and $k > 0$, then one more observation is taken.

If z is the log-likelihood ratio computed after this next observation then,

$$\phi_k(t) = \begin{cases} 1 & \text{if } t + z \geq b \\ \phi_{k-1}(t+z) & \text{if } a < t + z < b \\ 0 & \text{if } t + z \leq a \end{cases}$$

with initial conditions:

$$\phi_0(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{if } t < 0. \end{cases}$$

Hence, one can recursively compute ϕ_k:

$$\phi_k(t) = \int_{b-t}^{\infty} f_i(z|\theta_0)\, dz + \int_{a-t}^{b-t} \phi_{k-1}(t+z)\, f_i(z|\theta_0)\, dz \qquad (6)$$

for $1 < k \leq m$,

where $f_i(z|\theta_0)$ is the density function for the ith observation when θ_0 is the true state of nature, where $i = m-k$.

Equation (6) holds since a type I error occurs either if the next observation causes the likelihood ratio to go above b (term 1), or the process continues but eventually produces a type I error (term 2).

A similar expression holds for $\psi_k(t)$, the type II error probability with at most k more observations, when the present log-likelihood ratio is t:

$$\psi_k(t) = \int_{-\infty}^{a-t} f_i(z|\theta_1)\, dz + \int_{a-t}^{b-t} \psi_{k-1}(t+z)\, f_i(z|\theta_1)\, dz$$

The truncation point m^* is then obtained by finding the smallest n $(=m^*)$ such that both $\phi_n(0) \leq \alpha$ and $\psi_n(0) \leq \beta$.

Example: Normal IID Observations

Let X_i be IID Normal random variables with mean θ_j and variance σ^2 for $j = 0$, 1 and $i \geq 1$. The log-likelihood ratio that corresponds to the ith observation $X_i = x$ is given by

$$z = \ln \frac{f(x|\theta_1,\, \sigma^2)}{f(x|\theta_0,\, \sigma^2)} = \frac{\theta_1-\theta_0}{\sigma^2}\, x - \frac{\theta_1^2-\theta_0^2}{2\sigma^2} \qquad (7)$$

Since Z is linear in X, it is easy to show that, when $\theta = \theta_0$,

$$Z \sim N\left(-\frac{1}{2}d^2,\ d^2\right) \text{ where } d = (\theta_0 - \theta_1)/\sigma.$$

For a given value of α $(= \beta)$ and d, equation (6) can be solved iteratively for n = 1, 2, 3,..... We then find (by interpolation if necessary) that particular smallest value of n $(= m^*)$ for which $\phi^n(0) \leq \alpha$. Table 1 shows the results for values

of $\alpha = \beta$ between .01 and .15 and values of d between .20 and 1.5. The values of m^* are very close to those obtained by Aroian and Robison (1969) and Golhar (1983). Finally, we note that an integer truncation point m^{**} can be obtained by rounding up the solution to (9), which of course guarantees satisfying the error probabilities desired since $m^{**} > m^*$.

Figure 1 is a plot of $\ln(m^*)$ vs. $\ln(d)$ for different values of $\alpha = \beta$: an essentially linear relationship. Table 2 shows the slopes of these lines (and standard error) as determined by a standard least-squares fit. Note that the slope is fairly constant for different α, and roughly equal to -2.16. This suggests that for IID Normal variates m^* and d have the approximate relationship:

$$m^* = k(\alpha) \; d^{-2.16} \tag{8}$$

where $\ln k(\alpha)$ is the intercept of the curves in Figure 1 at d = 1.

To obtain $k(\alpha)$, m^* was plotted against α for d = 1 as shown as Figure 2. This curve is well represented by the equation: $k(\alpha) = a + b(\alpha)^C$. In order to find constants, a, b and c the following procedure was adopted:

1.   For d = 1, using the three data points (α, m^*) of (.01, 24.3); (.075, 9.1); and (.15, 4.6) the value of c = -.1096 was obtained.

2.   Letting $y = (\alpha)^{-.1096}$ the general equation is transformed into the linear relationship: $k(\alpha) = a + by$. A least squares criterion using all points gives values of a = -53.17 and b = 46.9.

Thus, equation (8) can be written:

$$\ln(m^*) = \ln\{-53.17 + 46.9 \; \alpha^{-.1096}\} -2.16 *\ln(d) \tag{9}$$

The simple relationship (9) gives the m^* values shown in Table 3, which are very close to the (actual) m^* values of Table 1.

CONCLUSION

The recursive method presented here is very efficient compared to the methods proposed elsewhere in the literature.

Computing time and memory space requirements are of the order of
m. We have also established a simple relationship between a
useful truncation point m^*, the desired error probability a and
the discrimination factor d for testing Normal IID variables
using Wald's bounds. It has been shown by Golhar and Pollock
(1983) that the m^* thus obtained gives a smaller expected number
of observations than the truncation point from Johnson's working
rule, yet still gives actual error probabilities within the
desired values.

TABLE I
Computed (actual) values of a^* for given d and a

| d\a	0.01	0.02	0.025	0.03	0.04	0.05	0.06	0.07	0.075	0.08	0.09	0.1	0.125	0.15
0.20	775.1	615.0	562.8	520.0	452.2	399.8	357.1	321.3	305.3	290.5	263.8	240.1	191.3	153.2
0.25	479.0	378.9	346.4	319.7	277.6	245.1	218.7	196.6	186.8	177.7	161.2	146.7	116.7	93.4
0.30	323.3	255.1	233.0	214.9	186.4	164.4	146.6	131.7	125.1	118.9	107.9	98.1	78.0	62.4
0.40	173.9	136.7	124.6	114.8	99.4	87.6	78.0	70.0	66.5	63.2	57.3	52.1	41.4	33.1
0.50	107.6	84.3	76.8	70.7	61.1	53.8	47.9	43.0	40.8	38.8	35.1	31.9	25.4	20.3
0.75	45.0	35.1	32.0	29.4	25.4	22.3	19.9	17.8	16.9	16.1	14.6	13.2	10.6	8.5
1.00	24.3	19.0	17.3	15.9	13.7	12.0	10.7	9.6	9.1	8.7	7.9	7.2	5.7	4.6
1.25	15.2	11.8	10.8	9.9	8.6	7.5	6.7	6.0	5.7	5.5	4.9	4.5	3.6	2.9
1.50	10.4	8.0	7.3	6.8	5.8	5.1	4.6	4.1	3.9	3.7	3.4	3.1	2.5	2.0
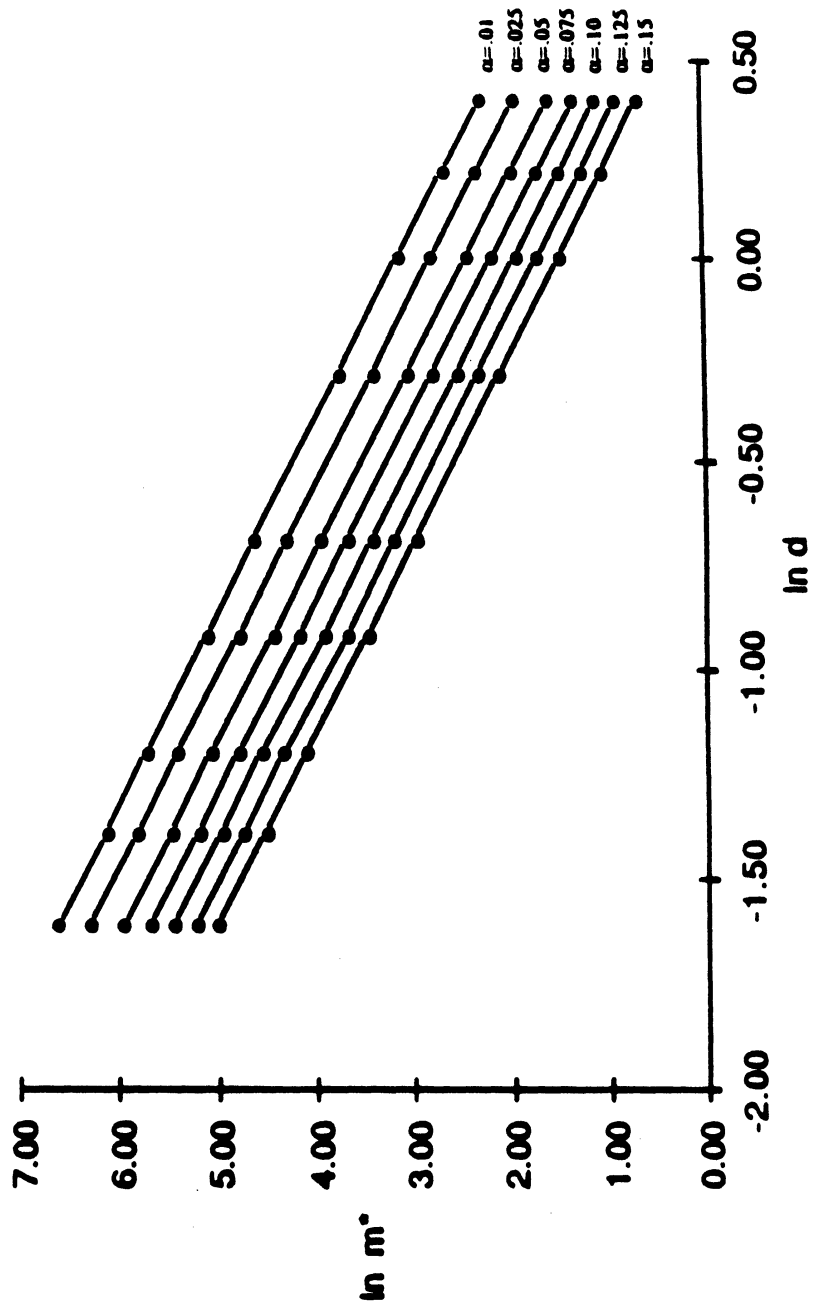
Figure 1: Relation between m* and d for independent Normal hypotheses

TABLE II

Slope of $\ln (m^{*})$ as a function of $\ln(d)$ for different values of α. The average slope for the values shown is -2.16. S_b = standard error.

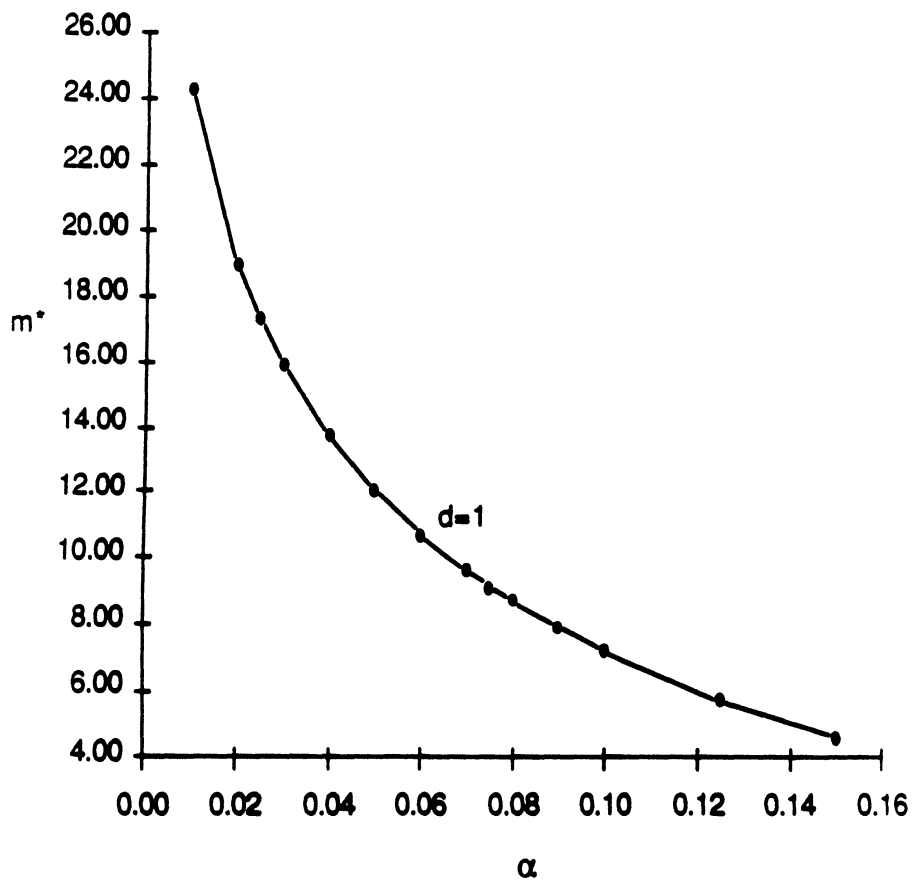| α	Slope	S_b
0.01	-2.143	0.0035
0.02	-2.154	0.0033
0.025	-2.157	0.0035
0.03	-2.169	0.0043
0.04	-2.162	0.0049
0.05	-2.177	0.0049
0.06	-2.164	0.0065
0.07	-2.168	0.0059
0.075	-2.168	0.0065
0.08	-2.164	0.0076
0.09	-2.164	0.0083
0.10	-2.163	0.0086
0.125	-2.159	0.0098
0.15	-2.157	0.0096

Figure 2: m* for d=1 as a function of α

TABLE III

Values of m^{*} from equation (9)

d\α	0.01	0.02	0.025	0.03	0.04	0.05	0.06	0.07	0.075	0.08	0.09	0.1	0.125	0.15
0.20	793.4	609.5	553.2	508.2	439.1	386.9	345.2	310.6	295.3	281.1	225.4	232.7	185.5	147.8
0.25	489.9	376.3	341.6	313.8	271.1	238.9	213.1	191.7	182.3	173.5	157.7	143.7	114.5	91.2
0.30	330.4	253.8	230.3	211.6	182.8	161.1	143.7	129.3	122.9	117.0	106.3	96.9	77.2	61.5
0.40	177.4	136.3	123.7	113.6	98.2	86.5	77.2	69.4	66.0	62.8	57.1	52.0	41.4	33.0
0.50	109.5	84.1	76.4	70.1	60.6	53.4	47.6	42.8	40.7	38.8	35.2	32.1	25.6	20.4
0.75	45.6	35.0	31.8	29.2	25.2	22.2	19.8	17.8	16.9	16.1	14.6	13.3	10.6	8.5
1.00	24.5	18.8	17.0	15.6	13.5	11.9	10.6	9.5	9.1	8.6	7.8	7.1	5.7	4.5
1.25	15.1	11.6	10.5	9.6	8.3	7.3	6.5	5.9	5.6	5.3	4.8	4.4	3.5	2.8
1.50	10.2	7.8	7.1	6.5	5.6	4.9	4.4	3.9	3.7	3.6	3.2	2.9	2.3	1.9

ACKNOWLEDGEMENT

BIBLIOGRAPHY

1. Anderson, T. W., "A Modification of the Sequential
    Probability Ratio   Test to Reduce the Sample Size",
    Ann. Math. Stat., 1960, Vol. 31, pp. 165-196.

2. Armitage, P., "Restricted Sequential Procedures",
    Biometrika, 1957, Vol. 44, pp. 9-26.

3. Aroian, L. A. and Robison, D. E., "Direct Methods for Exact
    Truncated Sequential Tests of the Mean of a Normal
    Distribution", Technometrics, 1969, Vol. 11, No. 4, pp.
    661-675.

4. Golhar, D. Y., "Sequential Analysis: Non-stationary
    Processes and Truncation", a Ph.D. dissertation
    submitted to and accepted by Industrial and Operations
    Engineering Department at The University of Michigan,
    Ann Arbor, 1983.

5. Golhar, D. Y. and Pollock, S. M., "Truncated SPRT for Non-
    stationary Processes: Sensitivity of Assumptions",
    Sequential Analysis, Vol. 3., No. 3-4, pp. 251-266,
    1984.

6. Golhar, D. Y., and Pollock, S. M., "Improved Truncation for
    SPRT with Normal Observations", Technical Report 83-20,
    Industrial and Operations Engineering Department, The
    University of Michigan, Ann Arbor, 1983.

7. Johnson, N. L., "Sequential Analysis: A Survey", Journal of
    Royal Statistical Society, Series A, 1961, Vol. 124,
    pp. 372-411.

8. Wald, A., Sequential Analysis, Dover Publications, Inc., New
    York, 1947.