Initial Sample Size Problem in the Sequential Test for the Mean of a Normal Distribution

S. C. Choi*

1. Introduction

Consider a sequential decision procedure for testing the following composite hypotheses about the mean of a normal distribution with unknown variance σ^2 with preassigned error probabilities of α_0 and α_1:

$$H_0: \mu=\mu_0, \ \sigma>0 \text{ vs. } H_1: \mu=\mu_1, \ \sigma>0.$$

The sequential t-test proposed by Wald [14] and, in a modified version by others, for example Rushton [13], would not be appropriate if certain absolute differences in the mean are of interest as it often happens in applications, irrespective of σ.

One approach to the problem is due to Baker [2]. The procedure consists of taking a preliminary sample of fixed size m to estimate σ and to choose the boundaries for the sequential probability ratio test (SPRT) accordingly, and then sample one at a time until a terminal decision is reached. In order to make the procedure efficient in applications it is necessary to have some idea as to the first-stage sample size. In brief, Baker ignored the information available in the first-stage sample about the population mean, but solely for estimating the variance. This information was incorporated in [9]. In this paper we shall be concerned with the optimal sample size problem based on the test denoted by T_0 which consists of "resampling" the first-stage sample.

Let $X_1, X_2, \cdots$ be a random sample from $N(\mu, \sigma^2)$ and without the loss of generality assume $\mu_1>\mu_0$. The first stage of the test T_0 consists of taking a preliminary sample of size m, $m>1$, to compute the usual unbiased sample variance S_m^2. As in [2] let A_m and B_m denote the SPRT boundaries for the log-likelihood ratio given

---

* Associate Professor of Biostatistics, School of Medicine and Professor of Statistics, School of Engineering, Washington University, St. Louis, Mo., U.S.A. The computer time was made available through NIH Grant RR 05389.

m. These boundaries can be obtained from the implicit equations involving A_m, B_m, α_0, α_1 and the density of S_m^2, and are extracted in part in Table 1. The alternative boundaries used in [9] are too conservative especially when α_0, α_1 and m are small. After S_m^2 is obtained the test T_0 is carried out as follows: sample starting from the first-stage sample on hand if

$$S_m^2 B_m/(\mu_1-\mu_0)+k(\mu_0+\mu_1)/2<\sum_{i=1}^{k} X_i<S_m^2 A_m/(\mu_1-\mu_0)+k(\mu_0+\mu_1)/2 \qquad (1)$$

and accept H_0 or H_1 according to whether the left-hand or the right-hand inequality is the first not satisfied. Note that if $k\leq m$ at the time of a terminal decision, then no additional sample is needed beyond the first-stage sample while if the decision is not reached by $k\leq m$, then additional observations are required. Theoretically, the test based on (1) may be subjected to question since $\sum_{i=1}^{k} X_i$ is not independent of S_m^2 if $k<m$. The possible effect, however, would be negligible since m is going to be small relative to the average sample number (ASN). Nevertheless, in order to circumvent the theoretical difficulty, we shall assume that $k\geq m$ although it may happen that $k<m$ in practice. A simulation study is employed to ascertain the negligible effect of possible inaccuracy in Section 4.

2. Conditional ASN and Its Upper Bound

Let

$$\gamma=(\mu_1-\mu_0)/\sigma,\ \ t=S_m^2/\sigma^2,$$

$$C_m=(e^{A_m t}-1)/(e^{A_m t}-e^{-A_m t}),$$

and $E(N|S_m^2)$ denote the conditional expectation of the sample size when a given S_m^2 is used in the SPRT. Under H_0 or H_1, and for $\alpha_0=\alpha_1$; and γ small, we have from [14] that

$$E(N|S_m^2)\approx 2tA_m(2C_m-1)/\gamma^2.$$

The conditional ASN, $E(N|m)$, when the preliminary sample of size m is used to compute S_m^2 is then

$$E(N|m)=\int_0^\infty E(N|S_m^2)p(S_m^2)dS_m^2, \qquad (2)$$

where $p(S_m^2)$ is the density function of S_m^2. The approximation of the ASN of the test T_0 by $E(N|m)$, of course, is valid only if the test is not likely to terminate with the first stage. Using the values of the A_m obtained in [2], the estimates $g(m)$ of $\gamma^2E(N|m)$ are computed from (2) by the numerical integration for $m=3(2)21$ and $m=21(10)41$ when $\alpha_0=\alpha_1=0.01$ and when $\alpha_0=\alpha_1=0.05$. These appear in Table 1. It can be seen that the value of $g(m)$ approaches the corresponding value of the SPRT (with σ known) *i.e.*, 5.6 and 9.1 as m increases. Hall [9] also computed $g(m)$ for $m=16$ and $m=31$ but the values differ slightly from Table 1 resulting from the slightly more conservative boundaries in the development.

Table 1. Termination Boundaries and the Estimate $g(m)$ of $\gamma^2E(N|m)$ when $\alpha_0=\alpha_1=\alpha=0.05$ and $\alpha=0.01$

	$A_m(=-B_m)$ α		$g(m)$ α	
m	0.05	0.01	0.05	0.01
3	13.60	67.00	27.1	134.4
5	5.84	15.93	11.3	31.8
7	4.65	10.24	8.9	20.4
9	4.12	8.31	7.8	16.5
11	3.85	7.35	7.3	14.6
13	3.68	6.78	6.9	13.4
15	3.56	6.40	6.7	12.7
17	3.48	6.13	6.5	12.1
19	3.41	5.94	6.3	11.7
21	3.36	5.79	6.2	11.4
31	3.22	5.35	5.9	10.5
41	3.15	5.15	5.8	10.1

The work of Baker [2], Bhate [3], Kemp [11], and Page [12] regarding the ASN all indicated that the Wald's approximation can underestimate the true ASN. In particular, Baker's experimental study showed that a better approximation is given by the upper bound of the ASN or by the mean of the ASN and the upper bound. Hence, it will be useful to obtain the analogous upper bound for $\gamma^2E(N|m)$. Let

$$f(\gamma)=\gamma[\gamma/2+y(\gamma/2)/\Phi(-\gamma/2)],$$

where $$y(x)=(2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$$

and $\Phi(x)$ denotes the standard normal distribution function. Then for both H_0 and H_1, it follows from [14, p. 170] that

$$E(N|S_m^2)\leq 2tA_m(2C_m-1)/\gamma^2+2f(\gamma)C_m/\gamma^2. \tag{3}$$

Before we proceed to obtain the upper bound for $\gamma^2E(N|m)$, consider the function $H(m,\mu)$ defined by

$$H(m,\mu)=\int_0^\infty\{[k(m,\mu)-1]/[k(m,\mu)-k^{-1}(m,\mu)]\}p(S_m^2)dS_m^2, \tag{4}$$

where $$k(m,\mu)=\exp[h(\mu)A_mt]$$

with $$h(\mu)=(\mu_1+\mu_0-2\mu)/(\mu_1-\mu_0).$$

The function $H(m,\mu)$ gives the operating characteristic function. The function is also given by [2] but with a slight error. (Also note the error in (34) of the same paper.) The expansion for (4) is given by

$$H(m,\mu)=(v/2)^{\frac{v}{2}}\sum_{j=0}^{\infty}[(2jh(\mu)A_m+v/2)^{-\frac{v}{2}}-2jh(\mu)A_m+v/2+h(\mu)A_m)^{-\frac{v}{2}}] \tag{5}$$

where $v=m-1$. The function $H(m,\mu_0)$ can be used to assess the adequacy of the stopping boundaries, A_m and B_m. Using (5) the values of $H(m,\mu_0)$ for $m=3(2)21$ were computed. The range obtained was 0.9508 to 0.9520 for $\alpha_0=\alpha_1=0.05$ and 0.9901 to 0.9904 when $\alpha_0=\alpha_1=0.01$, so the boundaries given in Table 1 appear to be remarkably good.

In order to obtain the upper bound for $\gamma^2E(N|m)$ we substitute (3) into (2) to obtain

$$\gamma^2E(N|m)\leq g(m)+2H(m,\mu_0)f(\gamma),$$

and hence the upper bound for $\gamma^2E(N|m)$ denoted by $g_u(m)$ is approximated by

$$g_u(m)\approx g(m)+2(1-\alpha_0)f(\gamma). \tag{6}$$

3. The First-Stage Sample Problem

As has been remarked the approximation of the ASN by $g(m)/\gamma^2$ or $g_u(m)|\gamma^2$ is valid only if the test is not likely to terminate with the first stage. Let $F(N/m)$ denote the ASN of the test T_0 irrespective of whether the test terminates with the first stage or not. If $P_m(N=j)$ denotes the probability that a decision is reached at the jth stage and not before when (A_m, B_m) are used as the stopping limits, then

$$F(N|m)=mP_m(N\leq m)+E(N|m)-\sum_{j=1}^{m} jP_m(N=j)$$

$$=E(N|m)+\sum_{j=1}^{m-1}(m-j)\ P_m(N=j). \tag{7}$$

Note that if m is very large then $F(N|m)=m$ is as it should be. The optimal sample size m_0 of the first-stage of T_0 can be defined as the value of m that minimizes $F(N|m)$.

The methods of determining the distribution of the decisive sample number (DSN) for the SPRT when σ is known have been studied by Kac [10], Bhate [3], Ghosh [8] and Chanda [4]. These methods are too complicated to be facilitated in practice. For the normal distribution Cox and Roseberry [6] showed that the variance of the DSN under H_0 or H_1 is approximately proportional to the square of the ASN, in agreement with the sample experiment they performed earlier [5]. It would, therefore, seem that $P_m(N=j)$ can be reasonably approximated a lognormal distribution. Using the standard correction term for continuity the desired approximation is

$$P_u(N=j)\approx\Phi[\log(j+0.5)-\mu_i/\sigma_i]-\Phi[\log(j-0.5)-\mu_i)/\sigma_i],\ j=1,2,\ldots, \tag{8}$$

where μ_i and σ_i^2 are the mean and the variance of $\log N$. In order to estimate μ_i and σ_i^2, let $V_i(N|m)$ denote the conditional variance of the DSN under H_i, $i=0,1$, when m is the first-stage sample size. If $\alpha_0=\alpha_1$, it follows from [13] that

$$V_i(N|m)\approx 4E(N|m)/\gamma^2-4\alpha_i(1-\alpha_i)(A_m-B_m)^2/\gamma^4. \tag{9}$$

Using the property of the lognormal distribution [1] we obtain

$$\sigma_i^2=\log[1+V_i(N|m)/E^2(N|m)], \tag{10}$$

and $$\mu_i=\log[E(N|m)]-\sigma_i^2/2. \tag{11}$$
Therefore, only the estimate of $E(N|m)$ is needed to approximate the distribution of the DSN. Note that m_0 would depend only on γ aside from α_0 and α_1. The calculations on Baker's experimental data indicated that (8) gives a reasonable approximation of the distribution if we use the mean of $g(m)$ and $g_u(m)$ as the estimate of $\gamma^2E(N|m)$ at least when $\alpha_0=\alpha_1=\alpha=0.05$ and when $\alpha=0.01$. See the Appendix for an example of goodness-of-fit for the distribution by the lognormal approximation.

4. Numerical Results

By substituting (8) into (7) we can determine numerically the optimal first-stage sample size m_0. Table 2 gives m_0 and the corresponding $F(N|m_0)$ for $\gamma=0.4(0.1)1.0$ and $\gamma=1.2$ when $\alpha=0.01$ and when $\alpha=0.05$. For the purpose of rough comparisons, the sample size required by the Student t-test and the approximate ASN of the sequential t-test are also given in Table 2. The ASN of the sequential t-test is computed as $(1+0.5\gamma^2)$ times the ASN of the σ known SPRT, the asymptotic result due to Cox [7].

It was noted that both the optimal first-stage sample size m_0 and the ASN are decreasing on γ although they become fairly stable as γ increases. Comparisons of the two-stage sequential test with the t-test and with the sequential t-test would be unfair to the first since the latter two require that the alternative hypothesis

Table 2. The Optimal First-Stage Sample Size (in Parentheses), the Corresponding ASN and Sample Size Required For t-Test and ASN of Sequential t-Test

γ	SPRT α 0.05	SPRT α 0.01	t-Test α 0.05	t-Test α 0.01	Sequential t-Test α 0.05	Sequential t-Test α 0.01
0.4	(23) 42.5	(35) 68.3	70	136	35.7	60.8
0.5	(18) 29.3	(27) 46.8	45	90	23.9	40.5
0.6	(13) 22.0	(21) 34.9	32	63	17.3	29.5
0.7	(12) 17.4	(18) 27.4	24	47	14.0	23.8
0.8	(10) 14.3	(16) 22.5	19	37	11.0	18.6
0.9	(9) 12.1	(14) 19.0	15	29	9.1	15.6
1.0	(8) 10.7	(13) 16.5	13	25	8.0	13.5
1.2	(7) 8.4	(11) 13.0	10	18	6.4	10.8

be specified in σ-units. Even so, the relative efficiency of the two-stage test to the t-test is appreciable especially when γ is small. It would appear that the two-stage test requires roughly 15 to 30 percent more observations on the average than the sequential t-test does.

It was encouraging to observe that the ASN function is fairly flat in the neighborhood of m_0 and that, even if the initial sample size does differ from m_0 by, say 4, the resulting loss in the ASN would be relatively small as it can be noted in part from Table 3.

The argument for the optimal first-stage sample size involves a number of approximations and assumptions. However, it is unlikely that the result will produce any serious error in the most practical application mainly because the ASN function $F(N|m)$ appears to be fairly constant in the neighborhood of m_0. In order to provide further assurance as to the efficacy of the results, simulation study was performed. For this purpose we considered the test of hypothesis $H_0: \mu=0$ against alternative $H_1: \mu=1$ when $\alpha_0=\alpha_1=0.05$ and when $\alpha_0=\alpha_1=0.01$ with $\gamma=0.4(0.2)1.0$ under H_0.

For each case m random samples from the corresponding normal distribution were generated to compute S_m^2. Then we proceeded according to the stopping rule described in Section 1. We performed 400 independent repetitions of such experiment for vairous values of m, and the ASN was obtained as the average of DSNs. The results for $\gamma=0.6$ and $\gamma=1.0$ are summarized in Table 3 together with the corresponding ASN obtained from the approximate method. From the simulation experiment performed, first, it would appear that the ASN is slightly underestimated by the approximate method as is the case with Wald's approximation of the ASN for the SPRT: even so the optimal initial sample size computed from the method seems to be quite satisfactory. Secondly, it was also reassuring to observe that the empirically estimated actual probability of Type I error ranges from 0.02 to 0.04 when $\alpha_0=0.05$ and 0.005 to 0.012 when $\alpha_0=0.01$. Note that because of the symmetry there is no need for performing the simulation when H_1 is true.

The results, of course, apply strictly only to the case where H_0 or H_1 is true and when $\alpha_0=\alpha_1$. Since m_0 is monotone decreasing on γ, it is clear that the optimum should be greater than m_0 if $\mu\in(\mu_0, \mu_1)$ and smaller than m_0 if $\mu\notin(\mu_0, \mu_1)$. Unfortunately, but as might be expected, m_0 depends on γ so that some idea about σ is required to choose m_0. Thus, further work is warranted to extend the applicability

Table 3. ASN Obtained by Approximation and by Simulation

γ	$\alpha_0(=\alpha_1)$	First-stage Sample Size	Approximation	Simulation
0.6	0.05	9	23.8	24.5
		11	22.7	23.7
		13	22.0	23.5
		15	22.3	22.6
		17	22.8	24.3
	0.01	17	36.2	36.5
		19	35.3	36.0
		21	34.9	35.1
		23	35.2	34.9
		25	35.4	36.5
1.0	0.05	3	28.3	28.4
		5	12.9	13.3
		7	10.8	12.1
		9	10.7	11.7
		11	11.9	12.8
	0.01	9	18.2	19.2
		11	16.7	18.1
		13	16.5	17.1
		15	17.1	17.9
		17	18.2	18.6

and to mitigate the difficulty. The following procedure is suggested purely on intuitive considerations although it may be difficult to justify on theortical grounds: draw a pilot sample of small size, say about $m_1=6$, to estimate σ. Then determine m_0 and take m_0-m_1 more observations as the first-stage sample. This paper at least provides some insight into the problem of the first-stage sample size.

APPENDIX

Table 4 presents an example comparing the fitted distribution based on a lognormal assumption and the DSN frequencies observed in [2] in the SPRT of H_0: $\mu=0$ versus H_1 : $\mu=1$ with $\sigma=1$ and $\alpha_0=\alpha_1=0.01$.

Table 4. Observed and Fitted Distributions of DSN

Sample Size	Observed	Fitted	Sample Size	Observed	Fitted
1	0	0.8	13	83	91.5
2	7	16.0	14	78	77.2
3	76	62.5	15	58	64.8
4	100	115.8	16	65	54.3
5	157	157.8	17	45	45.3
6	187	180.0	18	38	37.8
7	183	184.5	19	31	31.6
8	177	176.8	20	29	26.4
9	157	161.9	21-25	92	79.5
10	139	144.1	26-30	39	33.4
11	119	125.4	≥ 31	36	27.9
12	107	107.7			

From Table 4 we compute that $\chi^2=20.83$ with 21 degrees of freedom. (The first and the second frequencies were grouped.) Thus we find that there is agreement between the fitted and observed frequencies.

REFERENCES

[1] Aitchison, J. and Brown, J.A., *The Lognormal Distribution*, Cambridge: Cambridge Univ. Press, 1966.

[2] Baker, A. G., "Properties of Some Tests in Sequential Analysis," *Biometrika*, 37 (March 1950), 334—46.

[3] Bhate, D. H., "Approximation to the Distribution of Sample Size for Sequential Tests," *Biometrika*, 46 (June 1959), 130—8.

[4] Chanda, K. C., "Asymptotic Distribution of the Sample Size for a Sequential Probability Ratio Test," *Journal of the American Statistical Association*, 66 (March 1971), 178—83.

〔5〕 Cox, C. P. and Roseberry, T. D., "A Note on the Variance of the Distribution of Sample Number in Sequential Probability Ratio Tests," *Technometrics*, 8 (November 1966), 700—4.

〔6〕 __________ "A letter to the editor," *Technometrics*, 9 (May 1967), 349—50.

〔7〕 Cox, D. R., "Large Sample Sequential Tests for Composite Hypotheses," *Sankhya*, 25 (March 1963), 5—12.

〔8〕 Ghosh, B. K., "Moments of the Distribution of Sample Size in at SPRT," *Journal of the American Statistical Association*, 64 (December 1969), 1560—75.

〔9〕 Hall, W. J., "Some Sequential Analogs of Stein's Two-Stage Test," *Biometrika*, 49 (March 1962), 367—78.

〔10〕 Kac, M., "Random Walk in the Process of Absorbing Barriers," *The Annals of Mathematical Statistics*, 14 (March 1945), 62—7.

〔11〕 Kemp, K. W., "Formulae for Calculating the Operating Characteristic and the Average Sample Number of Some Sequential Tests," *Journal of the Royal Statistical Society*, B, 20 (1958), 379—86.

〔12〕 Page, E. S., "An Improvement to Wald's Approximation for Some Properties of Sequential Tests," *Journal of the Royal Statistical Society*, B, 16(1954), 136—9.

〔13〕 Rushton, S., "On a Sequential *t*-Test," *Biometrika*, 37 (March 1950), 326—33.

〔14〕 Wald, A., *Sequential Analysis*, New York: John Wiley and Sons, Inc., 1947.

SUMMARY

The two-stage sequential test, suggested by Baker 〔2〕 for testing hypotheses H_0: $\mu=\mu_0$ and H_1 : $\mu=\mu_1$ of $N(\mu,\sigma^2)$ with the unknown σ^2 would not be amenable for applications unless some clues on the choice of the first-stage sample size are available. The study in this paper is intended to shed some light on the size of the first-stage sample. An approximate method is used to estimate an optimal initial sample size that minimizes the average sample number. In brief, the optimal size is a strictly monotone decreasing function of the quantity $(\mu_1-\mu_0)/\sigma$. Empirical and simulation results are used to ascertain the negligible effect of possible errors due to approximations and assumptions used.