December, 1965

REMARKS ON SEQUENTIAL HYPOTHESIS TESTING

by H. Robbins and N. Starr*
Columbia University and The University of Minnesota

Technical Report No. 68

University of Minnesota
Minneapolis, Minnesota

*Research supported in part by National Science Foundation Grant NSF-GP3813 at The University of Minnesota.

1. Introduction and summary. Let $x_1, x_2, \ldots$ be independent $N(\mu,\sigma^2)$ with μ and σ^2 unknown, and let α, β, δ be given positive constants. We wish to test the hypothesis $H_0 : \mu = \mu_0$, where μ_0 is a given constant, in such a way that

(1) when $\mu = \mu_0$, $P(\text{Reject } H_0) \leq \alpha$ for all $0 < \sigma^2 < \infty$,

(2) when $|\mu-\mu_0| \geq \delta$, $P(\text{Accept } H_0) \leq \beta$ for all $0 < \sigma^2 < \infty$.

C. Stein [6] has given a two-stage procedure for accomplishing this; it involves choosing an initial sample size n. If n is chosen poorly in relation to the unknown σ^2, the expected sample size of Stein's procedure will be large in comparison to the sample size which could be used if σ^2 were known; moreover, two stage sampling is asymptotically inefficient as $\sigma \to \infty$, irrespective of the choice of n. We give a sequential procedure which, while satisfying (1) and (2) only approximately, is asymptotically efficient and seems to be reasonably efficient (with expected sample size about the same as in Stein's procedure when the optimal value of n is used) for all finite σ^2.

2. The case σ^2 known. For purposes of comparison we sketch the usual procedure to accomplish (1) and (2) when σ^2 is known. Let $f(x)$ denote the normal (0,1) p.d.f. and define a, b, N_0 by

$$(3) \qquad \int_{-a}^{a} f(x)dx = 1-\alpha, \qquad \int_{-a+b}^{a+b} f(x)dx = \beta, \qquad N_0 = \frac{b^2\sigma^2}{\delta^2} .$$

Reject H_0 iff $|\bar{x}_{N_0} - \mu_0| > \frac{a\delta}{b}$ (for simplicity we ignore the fact that N_0 need not be an integer), where by definition

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (n \geq 1).$$

It is easily seen that (1) and (2) hold. Thus as a possible measure of efficiency for any test which accomplishes (1) and (2) when σ^2 is unknown, we may use the ratio of N_0 to the expected sample size of the test in question.

3. Stein's procedure when σ^2 is unknown. Let $f_n(x)$ denote the p.d.f. of Student's $\underset{\sim}{t}$ with n-1 d.f. and define a_n, b_n by

$$(4) \qquad \int_{-a_n}^{a_n} f_n(x)dx = 1-\alpha, \quad \int_{-a_n+b_n}^{a_n+b_n} f_n(x)dx = \beta; \qquad \text{then } a_n \downarrow a,\ b_n \downarrow b.$$

Let n be any fixed integer $\geqq 2$ and set

$$(5) \qquad N = \text{least integer} \geqq \max\left(n, \frac{b_n^2 s_n^2}{\delta^2}\right),$$

where

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \qquad (n \geqq 2).$$

Reject H_0 iff $|\bar{x}_N - \mu_0| > \frac{a_n s_n}{\sqrt{N}}$.

It is well-known that (1) and (2) hold. However, from (4) and (5)

$$(6) \qquad EN \geqq E \max\left(n, \frac{b_n^2 s_n^2}{\delta^2}\right) \geqq \frac{b_n^2 \sigma^2}{\delta^2} > N_0 ;$$

hence $\frac{N_0}{EN} < 1$ for all σ, and in particular

$$(7) \qquad \lim_{\sigma \to \infty} \frac{N_0}{EN} = \frac{b^2}{b_n^2} < 1.$$

The expected sample size depends in a computable way on n and σ/δ; we shall later give a table showing this for $\alpha = .05$, $\beta = .025$. It will be seen that if n is poorly chosen, the ratio N_0/EN may be quite small.

4. A sequential procedure. Let

$$(8) \qquad N = \text{first odd integer} \geqq n_0 \text{ such that } s_n^2 \leqq \frac{d^2 n}{a_n^2},$$

with $n_0 \geqq 3$ a fixed odd integer, and where we have set

$$(9) \qquad d = \frac{a\delta}{b},$$

and let

(10) $$N^* = N + m,$$

where m is any fixed integer ≥ 0. Reject H_0 iff $|\bar{x}_{N^*} - \mu_0| > d$. It follows easily from [1] and [2] that

(11) $$\lim_{\sigma\to\infty} \frac{N_0}{EN^*} = 1,$$

(12) $$\lim_{\sigma\to\infty} P(\text{Reject } H_0) = \alpha \qquad \text{if } \mu = \mu_0$$

(13) $$\lim_{\sigma\to\infty} P(\text{Accept } H_0) \leq \beta \qquad \text{if } |\mu-\mu_0| \geq \delta$$

(in fact, these asymptotic properties hold even if the common distribution of the x_i is not normal), so that this procedure is asymptotically satisfactory as $\sigma \to \infty$. It remains to examine its performance for finite values of σ and to compare it with the procedure of Section 3.

We have by [3] and [5],

(14) $$P(\text{Accept } H_0) = P(|\bar{x}_{N^*} - \mu_0| \leq d) = \sum_n P(N = n) \cdot \int_{\frac{\sqrt{n+m}}{\sigma}(-d+\mu_0-\mu)}^{\frac{\sqrt{n+m}}{\sigma}(d+\mu_0-\mu)} f(x)dx ,$$

where the summation is over all odd values of $n \geq n_0$. The probability distribution of N depends on the parameter $\lambda = \sigma/d$ and has been computed in [5] using a method given in [3]. Thus

(15) $$\text{when } \mu = \mu_0, \quad P(\text{Reject } H_0) = 2\sum_n P(N = n) \cdot \int_{\frac{\sqrt{n+m}}{\lambda}}^{\infty} f(x)dx = \varphi(\lambda), \text{ say.}$$

From computations in [5] for the case $\alpha = .05$, $\beta = .025$, $a = 1.96$, $b = 2a = 3.92$, $d = \delta/2$, $n_0 = 3$, $m = 4$ it seems that

(16) $$\sup_{0<\lambda<\infty} \varphi(\lambda) = .05127.$$

(Actually, $\varphi(\lambda) = .05127$ for $\lambda = 3.5$, and from that value up to $\lambda = 6.75$ $\varphi(\lambda)$ is increasing; its limit as $\lambda \to \infty$ is from (12) equal to .05. A rigorous proof of (16) is lacking.) Thus presumably (1) holds almost exactly.

Concerning (2), we have from (14)

$$\text{(17)} \qquad \text{when } |\mu - \mu_0| \geqq \delta, \quad P(\text{Accept } H_0) \leqq \sum_n P(N = n) \cdot \int_{\frac{\sqrt{n+m}}{\sigma}(-d+\delta)}^{\frac{\sqrt{n+m}}{\sigma}(d+\delta)} f(x)dx$$

$$\leqq \sum_n P(N = n) \cdot \int_{\frac{\sqrt{n+m}}{\sigma}(-d+\delta)}^{\infty} f(x)dx \quad .$$

For the case under consideration, in which $\delta = 2d$, we therefore have by (16),

$$\text{(18)} \qquad \text{when } |\mu - \mu_0| \geqq \delta, \quad P(\text{Accept } H_0) \leqq \sum_n P(N = n) \int_{\frac{\sqrt{n+m}}{\lambda}}^{\infty} f(x)dx = \varphi(\lambda)/2$$

$$\leqq \tfrac{1}{2} \sup_{0<\lambda<\infty} \varphi(\lambda) = .02564.$$

(A more exact computation, using the first sum in (15) gives .02545 as the upper bound.) Thus presumably (2) holds almost exactly.

It remains to consider the expected sample size EN^* of this procedure, which depends on the parameter $\lambda = \sigma/d$. This is given in the accompanying table, reproduced from [5], for values of λ up to 5.0. A more exhaustive table of EN^* and $\left(\frac{N_0}{EN^*}\right)^{-1}$ may be found in [5].

Expected sample size using Stein's two-stage procedure, condensed and revised from [4]. EN^* and N_0 are included for comparison with the sequential method and the fixed sample size case, respectively.

$\alpha = .05,\ \beta = .025,\ d = \delta/2,\ n_0 = 3,\ m = 4$

$\lambda = \sigma/d$		1.0	1.25	1.667	2.0	2.5	3.333	5.0	∞
n	241	241	241	241	241	241	241	241	
	121	121	121	121	121	121	121	121	
	81	81	81	81	81	81	81	101	
	61	61	61	61	61	61	61.1	100	
	51	51	51	51	51	51	52.6	101	
	41	41	41	41	41	41.1	47.9	102	$\sim \frac{b_n^2\sigma^2}{\delta^2}$
	31	31	31	31	31	32	46.6	104	
	21	21	21	21	21	28.2	48.4	109	
	11	11	11	15.1	20.2	31.1	55.1	124	
	6	7.9	10.9	18.5	46.4	41.3	73.4	165	
EN^*		10.6	12.4	16.5	20.7	29.0	47.5	101.1	$\sim \frac{b^2\sigma^2}{\delta^2}$
N_0		3.8	6.0	10.7	15.4	24.0	42.7	96.0	$= \frac{b^2\sigma^2}{\delta^2}$

5. <u>An alternative sequential procedure</u>. Let

$$N = \text{first odd integer } n \geq n_0 \text{ such that } s_n^2 \leq \frac{\delta^2 n}{b_n^2}, \tag{19}$$

and reject H_0 iff $|\bar{x}_{N^*} - \mu_0| \geq \frac{a_{N^*}\, s_{N^*}}{\sqrt{N^*}}$ where again $N^* = N + m$, m a fixed integer ≥ 0. (This is a natural modification of the test of Section 3, just as the test of Section 4 was a natural modification of the test of Section 2.) We remark that since

$$s_n^2 = \frac{n-2}{n-1} \cdot \frac{1}{n-2} \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2 + \frac{(x_n - \bar{x}_n)^2}{n-1} \geq \frac{n-2}{n-1} s_{n-1}^2 \qquad (n \geq 3),$$

we have

$$\frac{\delta^2(N^*-2)}{b_{N^*-1}^2} < s_{N^*}^2 \leq \frac{\delta^2 N^*}{b_{N^*}^2}, \tag{20}$$

where the left-hand inequality holds only for $N > n_0$.

Since $\frac{\sqrt{N^*}(\bar{x}_{N^*}-\mu)}{\sigma}$ is N(0,1) given N = n, when H_0 is true we have

$$\begin{aligned}
P(\text{Reject } H_0) &= \sum_n P(N = n) \cdot P\left(\frac{\sqrt{N^*}|\bar{x}_{N^*}-\mu_0|}{\sigma} \geq \frac{a_{N^*}\, s_{N^*}}{\sigma} \middle| N = n\right) \\
&\leq P(N=n_0) + \sum_{n>n_0} P(N = n) \cdot P\left(\frac{\sqrt{N^*}|\bar{x}_{N^*}-\mu_0|}{\sigma} \geq \frac{a_{N^*}\sqrt{N^*-2}\,\delta}{\sigma b_{N^*-1}} \middle| N = n\right) \\
&= P(N=n_0) + 2 \sum_{n>n_0} P(N = n) \int_{\frac{a_{n+m}\,\delta\sqrt{n+m-2}}{\sigma b_{n+m-1}}}^{\infty} f(x)dx ,
\end{aligned} \tag{21}$$

and when $|\mu - \mu_0| \geq \delta$ we have

$$\begin{aligned}
P(\text{Accept } H_0) &= \sum_n P(N = n) \cdot P\left(\frac{-a_{N^*}\, s_{N^*}}{\sigma} + (\mu_0-\mu)\frac{\sqrt{N^*}}{\sigma} \leq \frac{\sqrt{N^*}}{\sigma}(\bar{x}_{N^*}-\mu) \leq \frac{a_{N^*}\, s_{N^*}}{\sigma} + (\mu_0-\mu)\frac{\sqrt{N^*}}{\sigma} \middle| N = n\right) \\
&\leq \sum_n P(N = n) \cdot \int_{\frac{\delta\sqrt{n+m}}{\sigma}\left(\frac{-a_{n+m}}{b_{n+m}}+1\right)}^{\frac{\delta\sqrt{n+m}}{\sigma}\left(\frac{a_{n+m}}{b_{n+m}}+1\right)} f(x)dx .
\end{aligned} \tag{22}$$

The asymptotic relations (11) - (13) hold for this test also. The performance for finite values of σ^2 is being computed, and results will be available shortly.

It should be observed that the tests of this and the preceding section are sequential only in that they merely attempt to estimate sequentially the nuisance parameter σ^2 of the test of Section 2, and not in the sense of Wald's sequential probability ratio test, which has an entirely different motivation. The ultimate sequential test that will satisfy (1) and (2) while reducing the expected sample size as much as possible remains to be devised.

Bibliography

[1] Chow, Y. S. and Robbins, H. On the asymptotic theory of fixed-width sequential confidence intervals for the mean. Ann. Math. Stat., 36 (1965), pp. 457-462.

[2] Gleser, L. J., Robbins, H. and Starr, N. Some asymptotic properties of fixed-width sequential confidence intervals, Columbia University Report, April 24, 1964.

[3] Robbins, H. Sequential estimation of the mean of a normal population. Probability and Statistics (Harald Cramér Volume), Almqvist and Wiksell, Uppsala, Sweden, pp. 235-245.

[4] Seelbinder, B. M. On Stein's two-stage sampling scheme. Ann. Math. Stat., 24 (1953), pp. 640-649.

[5] Starr, N. The performance of a sequential procedure for the fixed-width interval estimation of the mean. To appear: Ann. Math. Stat., February 1966.

[6] Stein, C. A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Stat., 16 (1945) pp. 243-258.