A SEQUENTIAL ANALOGUE OF THE BEHRENS-FISHER PROBLEM

By H. Robbins, G. Simons, and N. Starr

Purdue University, Stanford University, and Carnegie Institute of Technology

1. Introduction. In [1] and [4] a sequential procedure for the fixed-width interval estimation of the mean of a single population was investigated. We consider here an analogous procedure for estimating the difference of the means of two populations.

Let $x_1, x_2, \cdots$ and $y_1, y_2, \cdots$ be two independent sequences of rv's, the x's iid $N(\mu_1, \sigma_1^2)$ and the y's iid $N(\mu_2, \sigma_2^2)$. The four parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ are assumed unknown. We want to find a confidence interval I of width $2d$ and with coverage probability $\geqq \alpha$ for the parameter $\Delta = \mu_1 - \mu_2$, where $0 < d < \infty$ and $0 < \alpha < 1$ are preassigned constants.

If σ_1, σ_2 were known we could proceed as follows. Take r observations on x and s observations on y, and let

$$\bar{x}_r = 1/r \textstyle\sum_1^r x_i, \qquad \bar{y}_s = 1/s \sum_1^s y_j$$

be the respective sample means. If

$$I = [\bar{x}_r - \bar{y}_s - d, \bar{x}_r - \bar{y}_s + d] \tag{1}$$

is the interval of width $2d$ centered at $\bar{x}_r - \bar{y}_s$, then

$$P(\Delta \, \varepsilon \, I) = 2\Phi(d/(\sigma_1^2/r + \sigma_2^2/s)^{\frac{1}{2}}) - 1,$$

where Φ denotes the normal $(0, 1)$ df. Hence, defining the constants a and b by

$$2\Phi(a) - 1 = \alpha, \qquad b = (a/d)^2,$$

we have $P(\Delta \, \varepsilon \, I) \geqq \alpha$ providing that r, s satisfy the inequality

$$\sigma_1^2/r + \sigma_2^2/s \leqq 1/b. \tag{2}$$

Regarding r, s as continuous variables, the pair (r^*, s^*) which satisfies (2) and for which $n = r + s$ is a minimum is given by

$$r^* = b\sigma_1(\sigma_1 + \sigma_2), \qquad s^* = b\sigma_2(\sigma_1 + \sigma_2). \tag{3}$$

For this pair

$$r^*/s^* = \sigma_1/\sigma_2, \tag{4}$$

and the total sample size is

$$n^* = r^* + s^* = b(\sigma_1 + \sigma_2)^2. \tag{5}$$

We shall now give a sequential procedure when σ_1, σ_2 are unknown for determining r, s as random variables in such a way that (3) will hold approximately with

Received 18 November 1966; revised 2 March 1967.

high probability. The procedure consists of (a) a *sampling scheme* which tells us at each stage whether to take the next observation on x or y, and (b) a *stopping rule* which determines r and s and therefore I by (1).

2. The sequential procedure:

(a) Let

$$u_i^2 = (i-1)^{-1} \sum_{k=1}^{i} (x_k - \bar{x}_i)^2, \qquad v_j^2 = (j-1)^{-1} \sum_{k=1}^{j} (y_k - \bar{y}_j)^2$$

be the usual estimates of σ_1^2 and σ_2^2, for which $u_i \to \sigma_1$, $v_j \to \sigma_2$ a.s. as $i, j \to \infty$. We take $n_0 \geqq 2$ observations on x and on y to begin with. Then if at any stage we have taken i observations on x and j on y, with $n = i + j \geqq 2n_0$, we take the next observation on x or on y according as

$$i/j \leqq u_i/v_j \quad \text{or} \quad i/j > u_i/v_j .$$

This procedure generates an infinite sequence of observations and does not depend on the value of α or d. We shall show in the next section that $i/j \to \sigma_1/\sigma_2$ a.s. as $n \to \infty$.

(b) We now give three more or less equivalent stopping rules. Let $\{a_n\}$ be a given sequence of positive constants such that $a_n \to a$ as $n \to \infty$, and put $b_n = (a_n/d)^2$.

R_1 : Stop with the first $n \geqq 2n_0$ such that, if r observations on x and s observations on y have been taken, with $r + s = n$,

$$n \geqq b_n(u_r + v_s)^2 \qquad \text{(cf. (5))}. \tag{6}$$

R_2 : The same, with (6) replaced by

$$u_r^2/r + v_s^2/s \leqq 1/b_n \qquad \text{(cf. (2))}. \tag{7}$$

R_3 : The same, with (6) replaced by

$$r \geqq b_n u_r(u_r + v_s) \quad \textit{and} \quad s \geqq b_n v_s(u_r + v_s) \qquad \text{(cf. (3))}. \tag{8}$$

It is easy to check that the sample sizes n_k determined by these three rules R_k are such that $n_1 \leqq n_2 \leqq n_3$.

3. Asymptotic optimality of the sequential procedure. Let μ_1, μ_2, σ_1, σ_2, α, n_0, $\{a_n\}$ be fixed and let $d \to 0$, so that $n^* \to \infty$ where n^* is the optimal fixed sample size defined by (5). Denote by $n = r + s$ the sample size determined by any one of the stopping rules R_k, $k = 1, 2, 3$, and let I be the interval (1) for r, s.

THEOREM. *As* $d \to 0$

$$n/b(\sigma_1 + \sigma_2)^2 \to 1 \quad \text{a.s.,} \quad En/b(\sigma_1 + \sigma_2)^2 \to 1, \tag{9}$$

and

$$P(\Delta \,\varepsilon\, I) \to \alpha. \tag{10}$$

PROOF. We begin with a non-stochastic

LEMMA. *Given constants* $c_{i,j}(i, j = 1, 2, \cdots)$ *such that* $0 < c_{i,j} \to c > 0$ *as*

$i, j \to \infty$ and any integer $n_0 \geqq 1$, define $i(2n_0) = j(2n_0) = n_0$ and for $n \geqq 2n_0$ let

(I) $i(n+1) = i(n) + 1, \qquad j(n+1) = j(n) \quad \textit{if} \quad i(n)/j(n) \leqq c_{i(n),j(n)}$,

(II) $i(n+1) = i(n), \qquad j(n+1) = j(n) + 1 \quad \textit{if} \quad i(n)/j(n) > c_{i(n),j(n)}$.

Then $i(n)/j(n) \to c$ as $n \to \infty$.

Proof of the Lemma. Clearly $i(n) \to \infty, j(n) \to \infty$ as $n \to \infty$. Call an integer $n \geqq 2n_0$ of type I or II according as the right hand side of (I) or (II) holds for n. Then for all sufficiently large n there exist a largest integer n', $2n_0 \leqq n' < n$, of type I, and a largest integer n'', $2n_0 \leqq n'' < n$, of type II, and $n', n'' \to \infty$ as $n \to \infty$. Then

$$i(n)/j(n) \leqq (i(n') + 1)/j(n') \leqq c_{i(n'),j(n')} + 1/j(n') \to c,$$
$$i(n)/j(n) \geqq i(n'')/(j(n'') + 1) = i(n'')/j(n'')(1 - 1/(j(n'') + 1))$$
$$\geqq c_{i(n''),j(n'')}(1 - 1/j(n'') + 1) \to c,$$

and the lemma follows.

In the present case we put $c_{i,j} = u_i/v_j$ and obtain the result

$$i(n)/j(n) \to \sigma_1/\sigma_2 \quad \text{a.s.} \quad \text{as} \quad n \to \infty.$$

For any of our stopping rules, since $n \to \infty$ a.s. as $d \to 0$, it follows that as $d \to 0$,

$$r/s \to \sigma_1/\sigma_2 \quad \text{a.s.} \tag{11}$$

We shall now prove (9) for $R = R_3$, where we stop sampling with the first $n = r + s \geqq 2n_0$ such that

$$r \geqq b_n u_r(u_r + v_s) \quad \text{and} \quad s \geqq b_n v_s(u_r + v_s), \tag{12}$$

where

$$b_n = (a_n/d)^2, \qquad b = (a/d)^2.$$

Suppose that $r > n_0$ and that just before the rth observation on x there were ($r - 1$ observations on x and) j observations on y. Then by the sampling rule (a) of Section 2,

$$(r-1)/j \leqq u_{r-1}/v_j.$$

This implies that $r - 1 \leqq b_{r-1+j}u_{r-1}(u_{r-1} + v_j)$, for otherwise we would have

$$j \geqq (r-1)v_j/u_{r-1} > b_{r-1+j}v_j(u_{r-1} + v_j),$$

and sampling would have stopped at the $(r-1, j)$ stage. Hence even if $r = n_0$,

$$r \leqq b_{r-1+j}u_{r-1}(u_{r-1} + v_j) + n_0. \tag{13}$$

As $d \to 0$, $r \to \infty$, $r/j \to \sigma_1/\sigma_2$, so $j \to \infty$ a.s., and hence

$$\limsup_{d \to 0} r/b \leqq \sigma_1(\sigma_1 + \sigma_2) \quad \text{a.s.}$$

The reverse inequality for the lim inf is obvious from (12), so

$$(14) \qquad \lim_{d\to 0} r/b = \sigma_1(\sigma_1 + \sigma_2), \qquad \lim_{d\to 0} s/b = \sigma_2(\sigma_1 + \sigma_2) \quad \text{a.s.},$$

which implies the first part of (9).

The second part of (9) will follow from the dominated convergence theorem if we can bound the rv r/b by an integrable rv which is independent of d. Let

$$u = \sup_{i\geqq 2} u_i, \qquad v = \sup_{j\geqq 2} v_j ;$$

then

$$u^2 \leqq 2 \sup_{i\geqq 2} \{1/i \textstyle\sum_{k=1}^{i} (x_k - \mu_1)^2\},$$

and the right hand side is integrable (by Wiener's dominated ergodic theorem [5]) since the fourth moment of x is finite. Thus $Eu^2 < \infty$, and similarly $Ev^2 < \infty$. But from (13), for $d \leqq a$,

$$r/b \leqq C\cdot u(u + v) + n_0 ,$$

so we can apply the dominated convergence theorem to conclude from (14) that $Er < \infty$, $Es < \infty$ and

$$(15) \qquad \lim_{d\to 0} Er/b = \sigma_1(\sigma_1 + \sigma_2), \qquad \lim_{d\to 0} Es/b = \sigma_2(\sigma_1 + \sigma_2),$$

whence the second part of (9) follows.

(10) follows from (14) by a simple extension of the proof of Rényi's theorem to the case of two populations; cf. [2]. This completes the proof of the theorem for $R = R_3$. Because $n_1 \leqq n_2 \leqq n_3$, the other cases follow too.

We remark that the theorem just proved remains valid for non-normal populations; even the requirement that the population fourth moments be finite can be relaxed (see Section 5).

4. Small-sample behavior. It seems hopeless to try to find exact values of $P(\Delta \varepsilon I)$ and En for finite values of n^* by analytic methods. Instead, we present the results of an experiment using pseudo-random normal deviates for the values

$$\alpha = .95, \qquad a = 1.96, \qquad n_0 = 5, \qquad a_n^2 = (n + 4)/(n - 4)(1.96)^2,$$

for which (5) becomes

$$n^* = \{1.96\lambda(\Lambda + 1)\}^2$$

where

$$\Lambda = \sigma_1/\sigma_2 , \qquad \lambda = \sigma_2/d.$$

Values $\Lambda = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$ and $n^* = 10, 20, \cdots, 200$ were used, and 2,000 sequences of x and y were generated for each entry (except that for $n^* = 175$ and 200 only 1,000 sequences were generated). We denote by $\eta^{(k)}$ the average value of n and by $\nu^{(k)}$ the coverage frequency of I using the stopping rule R_k ($k = 1, 2, 3$).

$\Lambda = 1$

λ	n^*	$\eta^{(1)}$	$\eta^{(2)}$	$\eta^{(3)}$	$\nu^{(1)}$	$\nu^{(2)}$	$\nu^{(3)}$
.80670	10	15.0	15.1	15.5	.977	.977	.979
1.14085	20	24.1	24.1	24.7	.953	.954	.956
1.39725	30	34.2	34.2	34.8	.949	.949	.952
1.61341	40	44.0	44.0	44.6	.946	.947	.949
1.80384	50	53.7	53.7	54.4	.945	.945	.947
1.97601	60	64.1	64.3	64.9	.949	.949	.950
2.13434	70	74.8	74.8	75.5	.950	.950	.951
2.28170	80	85.0	85.1	85.6	.951	.951	.951
2.42011	90	94.9	94.9	95.5	.951	.951	.952
2.55102	100	105.0	105.0	105.6	.951	.951	.952
2.85213	125	130.7	130.7	131.3	.952	.952	.952
3.12435	150	155.8	155.8	156.4	.952	.952	.952
3.37468	175	179.7	179.7	180.4	.951	.951	.952
3.60769	200	206.6	206.6	207.2	.952	.952	.952

$\Lambda = \frac{1}{2}$

λ	n^*	$\eta^{(1)}$	$\eta^{(2)}$	$\eta^{(3)}$	$\nu^{(1)}$	$\nu^{(2)}$	$\nu^{(3)}$
1.07560	10	15.4	15.6	16.1	.976	.978	.980
1.31481	20	24.4	24.4	24.9	.955	.956	.959
1.86300	30	34.1	34.1	34.7	.950	.950	.952
2.15121	40	44.1	44.1	44.7	.948	.948	.950
2.40513	50	54.2	54.2	54.8	.948	.948	.949
2.63468	60	63.9	63.9	64.7	.947	.947	.949
2.84578	70	74.2	74.2	74.9	.949	.949	.950
3.04227	80	84.9	84.9	85.6	.951	.951	.952
3.22681	90	94.8	94.9	95.5	.950	.950	.951
3.40136	100	104.9	104.9	105.5	.951	.951	.952
3.80284	125	130.5	130.5	131.2	.952	.952	.953
4.16580	150	155.8	155.8	156.5	.952	.952	.952
4.49958	175	180.3	180.3	181.0	.952	.952	.952
4.81025	200	206.2	206.2	206.9	.952	.952	.952

$\Lambda = \frac{1}{4}$

λ	n^*	$\eta^{(1)}$	$\eta^{(2)}$	$\eta^{(2)}$	$\nu^{(1)}$	$\nu^{(2)}$	$\nu^{(3)}$
1.29073	10	15.4	16.1	16.9	.967	.973	.978
1.82536	20	25.0	25.5	26.1	.956	.960	.964
2.23560	30	34.7	31.9	35.6	.952	.953	.957
2.58115	40	44.5	44.6	45.2	.952	.952	.954
2.88165	50	53.9	54.0	54.7	.949	.949	.951
3.16162	60	64.1	64.2	64.8	.949	.950	.951
3.41494	70	74.4	74.5	75.1	.950	.950	.951
3.65072	80	84.8	84.8	85.8	.951	.951	.952
3.87218	90	94.7	94.7	95.3	.950	.950	.951
4.08163	100	104.9	105.0	105.6	.951	.951	.951
4.56340	125	130.5	130.5	131.2	.951	.951	.952
4.99896	150	155.1	155.1	155.7	.951	.951	.951
5.39949	175	180.0	180.0	180.6	.950	.950	.951
5.77230	200	204.5	204.5	205.2	.951	.951	.951

$\Lambda = \frac{1}{8}$

λ	n^*	$\eta^{(1)}$	$\eta^{(2)}$	$\eta^{(3)}$	$\nu^{(1)}$	$\nu^{(2)}$	$\nu^{(3)}$
1.43414	10	15.3	17.0	17.7	.953	.968	.973
2.02828	20	25.2	26.5	27.4	.947	.958	.967
2.48400	30	35.4	36.3	37.2	.953	.959	.962
2.86828	40	45.3	45.7	46.5	.953	.955	.958
3.20683	50	54.7	55.1	55.8	.951	.953	.955
3.51291	60	64.6	64.9	65.5	.951	.953	,954
3.79438	70	75.2	75.3	76.0	.951	.951	.953
4.05636	80	84.7	84.8	85.4	.952	.952	.953
4.30242	90	94.5	94.6	95.3	.952	.952	.952
4.53515	100	104.7	104.8	105.3	.950	.951	.951
5.07045	125	129.7	129.8	130.4	.950	.951	.951
5.55440	150	154.6	154.7	155.3	.951	.951	.951
5.99944	175	180.3	180.5	181.1	.950	.951	.951
6.41367	200	205.5	205.6	206.2	.951	.951	.951

$\Lambda = \frac{1}{16}$

λ	n^*	$\eta^{(1)}$	$\eta^{(2)}$	$\eta^{(3)}$	$\nu^{(1)}$	$\nu^{(2)}$	$\nu^{(3)}$
1.51850	10	15.1	17.6	18.1	.940	.963	.967
2.14748	20	24.7	27.4	28.3	.933	.955	.959
2.63012	30	34.7	37.1	38.2	.939	.954	.958
3.03700	40	45.0	46.8	47.8	.944	.953	.956
3.39547	50	55.9	57.0	58.0	.951	.956	.959
3.71955	60	65.4	66.3	67.2	.951	.954	.956
4.01758	70	75.3	76.1	76.9	.951	.954	.955
4.29497	80	86.1	86.7	87.5	.952	.955	.956
4.55550	90	95.4	95.8	96.6	.952	.953	.954
4.80192	100	105.4	105.7	106.4	.953	.953	.954
5.36871	125	129.8	130.0	130.6	.951	.951	.952
5.88113	150	154.9	155.0	155.6	.951	.951	.952
6.35234	175	179.7	179.7	180.3	.950	.950	.951
6.79097	200	204.0	204.4	205.0	.950	.950	.951

REMARKS. 1. Although (9) merely asserts that $En = n^* + O(n)$, the tables suggest that the difference $En - n^*$ may be bounded for all $d > 0$. That this is so will be shown in the next section.

2. As is to be expected, for values of $\Lambda < \frac{1}{8}$ the rule R_3 is somewhat more successful in keeping the coverage frequency $\geqq .95$ for small values of n^*.

3. We have no proof that the minimum coverage probability for all n^* is in fact attained in the computed range $n^* = 10, \cdots, 200$, although we hope so.

4. To test $H_0: \mu_1 = \mu_2$ so as to guarantee a Type I error $\leqq 1 - \alpha$ and a Type II error $\leqq (1 - \alpha)/2$ for alternatives such that $|\mu_1 - \mu_2| \geqq 2d$, irrespective of σ_1 and σ_2, we may reject H_0 iff $|\bar{x}_r - \bar{y}_s| > d$ (cf. [3] for the corresponding one population test of $H_0: \mu = 0$).

5. The cost of ignorance. Assume that, as in Section 4, $a_n = a + O(n^{-1})$

as $n \to \infty$. We shall show that for R_3 (and a fortiori for R_1 and R_2), $En - n^*$ is less than some finite constant for all $d > 0$.

PROOF. We assemble some facts that will be used in what follows. For some $0 \leqq M < \infty$,

$$b_n = (a_n/d)^2 \leqq b(1 + Mn^{-1}), \quad \text{where} \quad b = (a/d)^2.$$

Defining

$$U_i^2 = (i-1)u_i^2, \qquad V_j^2 = (j-1)v_j^2,$$

we have

$$U_{i+1}^2 \geqq U_i^2, \qquad V_{j+1}^2 \geqq V_j^2,$$

and

$$EU_r^2 \leqq E\{\textstyle\sum_{k=1}^r (x_k - \mu_1)^2\} = \sigma_1^2 Er, \qquad EV_r^2 \leqq \sigma_2^2 Es.$$

Finally, the function

$$g(r, n) = (r - n_0)^2 r^2 (r + M)^{-1} n^{-1}$$

is convex in r and n for $r \geqq n_0 > 0, n > 0, M \geqq 0$.

As in the proof of (9), suppose that $r > n_0$ and that just before the rth observation on x there were j observations on y. Then (as before)

$$(r-1)/j \leqq u_{r-1}/v_j, \qquad r - 1 \leqq b_{r-1+j} u_{r-1}(u_{r-1} + v_j),$$

and hence

$$r - 1 \leqq b(1 + Mn^{-1})u_{r-1}(u_{r-1} + ju_{r-1}/(r-1))$$
$$\leqq b(1 + Mr^{-1})U_r^2(n-1)(r-1)^{-1}(r-2)^{-1}.$$

Since $n_0 \geqq 2$ this implies that

$$(r - n_0)^2 r^2/(r + M)n \leqq bU_r^2,$$

and this holds even if $r = n_0$. By Jensen's inequality,

$$E^2(r - n_0)E^2 r/(Er + M)En \leqq bEU_r^2 \leqq b\sigma_1^2 Er.$$

Hence by (9) and (15),

$$(16) \qquad E^2(r - n_0) \leqq b\sigma_1^2\{En + MEnE^{-1}r\} \leqq b\sigma_1^2\{En + 0(1)\}.$$

Similarly,

$$(17) \qquad E^2(s - n_0) \leqq b\sigma_2^2\{E_n + 0(1)\}.$$

Since $n - 2n_0 = (r - n_0) + (s - n_0)$, it follows that

$$E^2(n - 2n_0) \leqq b(\sigma_1 + \sigma_2)^2\{En + 0(1)\},$$

and hence

$$En \leqq b(\sigma_1 + \sigma_2)^2 + O(1) = n^* + O(1),$$

which was to be proved. From this, (16), and (17) we obtain also that for all $d > 0$

$$Er \leqq r^* + 0(1), \qquad Es \leqq s^* + 0(1). \tag{18}$$

These results hold whenever the distributions of x and y are such that $\sigma_1^2 < \infty$, $\sigma_2^2 < \infty$ and $En < \infty$ for all $d > 0$; as we have noted, finite fourth moments for x and y will guarantee this. In fact, by truncating n and using a modification of the preceding argument we can prove that the second part of (9) holds, and hence that $En < \infty$, under the sole condition that $\sigma_1, \sigma_2 < \infty$. Consequently, all the results of this paper hold when $\sigma_1, \sigma_2 < \infty$.

Quite small upper bounds for the constants in (18) can be obtained in the case $a_n = a$; they depend only on n_0. In the case $a_n = a + 0(n^{-1})$, useful estimates of these constants depending on n_0, M, σ_1, σ_2 could perhaps be found by improving the above argument.

Acknowledgment. We wish to express our gratitude to Mrs. Elaine Frankowski of the Department of Statistics, University of Minnesota, for her assistance with the programming, and to the staff of the Numerical Analysis Center for making available machine time.

REFERENCES

[1] Chow, Y. S., and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.
[2] Rényi, A. (1957). On the asymptotic distribution of the sum of a random number of random variables. *Acta Math. Acad. Sci. Hung.* **8** 193–199.
[3] Robbins, H., and Starr, N. (1965). Remarks on sequential hypothesis testing. Univ. of Minnesota Technical Report No. 68.
[4] Starr, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* **37** 36–50.
[5] Wiener, N. (1939). The ergodic theorem. *Duke Math. J.* **5** 1-18.