PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 2, Fasc. 2 (1982), p. 149-156

A SEQUENTIAL CONFIDENCE INTERVAL FOR THE ODDS RATIO

BY

D. SIEGMUND (STANFORD, CALIFORNIA)

Abstract. In this paper*, a sequential fixed width confidence interval is proposed for the log odds ratio of a (2×2)-table. It is shown that the proposed interval has asymptotically the correct coverage probability and is asymptotically efficient uniformly in the unknown parameters.

1. Introduction. For $i = 1, 2$ let s_{in_i} and $f_{in_i} = n_i - s_{in_i}$ be the numbers of successes and failures, respectively, in n_i independent Bernoulli trials with constant success probability p_i on each trial. A simple large sample approximate confidence interval for the log odds ratio, $\log(p_1 q_2/p_2 q_1)$, is

$$\log(s_{1n_1} f_{2n_2}/s_{2n_2} f_{1n_1}) \pm z_\alpha [n_1/s_{1n_1} f_{1n_1} + n_2/s_{2n_2} f_{2n_2}]^{1/2}, \tag{1}$$

where

$$\int_{z_\alpha}^{\infty} (2\pi)^{-1/2} \exp(-x^2/2)\, dx = \alpha/2$$

(see [2], p. 35). The confidence coefficient $1-\alpha$ is asymptotically correct for fixed p_1 and p_2 as $\min(n_1, n_2) \to \infty$.

These intervals have two defects when p_1 and p_2 are near 0 or 1. On the one hand, the rate of approach to normality can be very slow, so that the use of asymptotic theory is questionable. More importantly, however, even with exact calculations, no fixed sample size design will permit one to estimate the log odds ratio by an interval of preassigned width in these boundary cases.

* Prepared with partial support of ONR Contract N00014-77-C-0306, NSF Grant MCS 77-16974, and the Deutsche Forschungsgemeinschaft.

For one binomial population with success probability p, Robbins and Siegmund [7] proposed a sequential scheme for obtaining approximately a confidence interval of preassigned width for $\log(p/q)$. However, they do not consider the question of the uniformity of their procedure for p near 0 or 1, when a sequential procedure would presumably be of the greatest value.

The purpose of this paper is to consider the two-population analogue of the procedure of Robbins and Siegmund. The procedure will be seen to attain asymptotically the required coverage probability and to be asymptotically efficient uniformly in $0 < p_1, p_2 < 1$.

In Section 2 the one-population case is reviewed, and the results of Robbins and Siegmund are appropriately strengthened to provide the tools for the two-population problem. It is also shown that the modification of the empirical odds ratio, suggested by Haldane [3] as a bias reducing device, is inappropriate in the sequential case.

Section 3 is concerned with the case of two populations. Remarks about further extensions are collected in Section 4.

2. One population. Let $x_1, x_2, \ldots$ be independent with $P\{x_j = 1\} = p$, $P\{x_j = 0\} = q = 1-p$ $(j = 1, 2, \ldots)$. Let $s_n = x_1 + \ldots + x_n$ and $f_n = n - s_n$. For large n, $\log(s_n/f_n)$ is approximately normally distributed with mean $\log(p/q)$ and variance $1/(npq)$. Hence to find a confidence interval for $\log(p/q)$ of preassigned width or, equivalently, in large samples to estimate $\log(p/q)$ by an estimator with preassigned variance $1/c$, Robbins and Siegmund [7] define

$$T = \inf\{n: s_n f_n > nc\}. \tag{2}$$

They propose estimating $\log(p/q)$ by

$$\log[(s_T + \tfrac{1}{2})/(f_T + \tfrac{1}{2})], \tag{3}$$

which they show is asymptotically normally distributed with mean $\log(p/q)$ and variance $1/c$ as $c \to \infty$. The modification of the empirical log odds by adding $\frac{1}{2}$ to numerator and denominator was originally suggested by Haldane [3] as a bias reducing device in the fixed sample case. Robbins and Siegmund also show that $ET \sim c/pq$ as $c \to \infty$. This may be interpreted as showing that their procedure is asymptotically efficient in the sense of requiring asymptotically about the same number of observations as a fixed sample procedure chosen to be appropriate for a value p_0 which happens to be the actual value of p.

In this section it is shown that the asymptotic normality of (3) holds uniformly over $0 < p < 1$. This is in marked contrast with the fixed sample case, as was noted in the Introduction. It will also be shown that the analogue of Haldane's bias reducing device in this sequential context is to *subtract* $\frac{1}{2}$ from numerator and denominator of the empirical odds ratio.

However, since the effect of this bias reduction on the asymptotic distribution is unknown, and because the appropriate modification for the two-sample case is likewise unknown, in most of what follows only the unmodified empirical odds ratio is considered.

The main result of this section is Theorem 1. Lemma 1, which was obtained by Robbins and Siegmund [7], is of interest in its own right. It asserts that the asymptotic efficiency of (2) is uniform in $0 < p < 1$. The repeated use will be made of the algebraic identity

$$s_n f_n/n = (q-p)(s_n-np)+npq-(s_n-np)^2/n. \tag{4}$$

THEOREM 1. *For the stopping rule T defined in* (2), *uniformly in $0 < p < 1$*

$$\lim_{c\to\infty} P\{c^{1/2}[\log(s_T/f_T)-\log(p/q)] \leqslant x\} = \Phi(x),$$

where

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-u^2/2)\,du.$$

The proof utilizes the following lemmas. For the simple proof of Lemma 1 based on (4), see [7].

LEMMA 1. $c < pq\,\mathrm{E}T < (c+1)/[1-(4c)^{-1}]$.

LEMMA 2. *There exists a c_0 such that for all $c \geqslant c_0$ and all p $(0 < p < 1)$*

$$(pq)^2\,\mathrm{E}(T-c/pq)^2 \leqslant 7c.$$

LEMMA 3. *For each ε $(0 < \varepsilon < 1)$ and $c \geqslant c_0$, where c_0 is defined as in Lemma 2,*

$$P\{|s_T-pT| \geqslant \varepsilon pqT\} < \varkappa/\varepsilon^2 c,$$

where $\varkappa$ does not depend on ε or c.

Proof of Lemma 2. Squaring (4) gives

$$(s_n f_n/n-c)^2 = (q-p)^2(s_n-np)^2+(pq)^2(n-c/pq)^2+(s_n-np)^4/n^2+$$
$$+2\{(q-p)(s_n-np)(pqn-c)-(q-p)(s_n-np)^3/n-(npq-c)(s_n-np)^2/n\}.$$

By the Schwarz inequality and Wald's second moment identity,

$$|\mathrm{E}\{(s_T-pT)(T-c/pq)\}| \leqslant \{pq\,\mathrm{E}(T)\,\mathrm{E}(T-c/pq)^2\}^{1/2}.$$

Hence, since $(s_T f_T/T-c)^2 < 1$, Wald's second moment identity yields

$$1 > (q-p)^2 pq\,\mathrm{E}T+(pq)^2\,\mathrm{E}(T-c/pq)^2-2pq|q-p|\{pq\,\mathrm{E}(T)\,\mathrm{E}(T-c/pq)^2\}^{1/2}-$$
$$-2|q-p|pq\,\mathrm{E}T-2(pq)^2\,\mathrm{E}T$$

or

$$(pq)^2\,\mathrm{E}(T-c/pq)^2-2pq|q-p|\{pq\,\mathrm{E}(T)\,\mathrm{E}(T-c/pq)^2\}^{1/2}+pq(q-p)^2\,\mathrm{E}T$$
$$< 1+2pq\,\mathrm{E}T.$$

Taking square roots in this expression, then rearranging terms and squaring yields

$$(pq)^2 \mathrm{E}(T-c/pq)^2 \leqslant \{(pq\,\mathrm{E}T)^{1/2}+(1+2pq\,\mathrm{E}T)^{1/2}\}^2$$
$$\leqslant 2(1+3pq\mathrm{E}T) \leqslant 2+6(c+1)/(1-1/4c),$$

where the last inequality follows from Lemma 1. This completes the proof.

Proof of Lemma 3. Let $0 < \delta < 1$ and $n_0 = c/pq$. By Lemma 2

$$P\{|T-n_0| > \delta c/pq\} \leqslant (\delta c)^{-2}(pq)^2 \mathrm{E}(T-n_0)^2 \leqslant 7/\delta^2 c.$$

Hence, by Wald's lemma for the second moment and Lemma 1,

$$P\{|s_T-pT| > \varepsilon pqT\} \leqslant 7/\delta^2 c + P\{|s_T-pT| > \varepsilon pqT, |T-n_0| \leqslant \delta c/pq\}$$
$$\leqslant 7/\delta^2 c + P\{|s_T-pT| > \varepsilon(1-\delta)c\}$$
$$\leqslant 7/\delta^2 c + \mathrm{E}(s_T-pT)^2/\varepsilon^2(1-\delta)^2 c^2$$
$$\leqslant 7/\delta^2 c + 2/[\varepsilon^2(1-\delta)^2 c].$$

Proof of Theorem 1. From the mean value theorem one obtains

$$c^{1/2}[\log(s_T/f_T)-\log(p/q)]$$
$$= c^{1/2}(s_T-pT)/pqT + c^{1/2}[(s_T-pT)/pqT]\left(\frac{pq}{\eta_T(1-\eta_T)}-1\right)$$
$$= n_0^{1/2}(s_T-pT)/(pq)^{1/2}\,T + n_0^{1/2}[(s_T-pT)/(pq)^{1/2}\,T]\left(\frac{pq}{\eta_T(1-\eta_T)}-1\right),$$

where $|\eta_T-p| \leqslant |T^{-1}s_T-p|$, and as before $n_0 = c/pq$. Hence it suffices to show that uniformly in $0 < p < 1$

$$\lim_{c\to\infty} P\{n_0^{1/2}(s_T-pT)/(pq)^{1/2}\,T \leqslant x\} = \Phi(x) \quad \text{and} \quad pq/\eta_T(1-\eta_T) \xrightarrow{P} 1.$$

The second statement follows easily from Lemma 3, and the first may be obtained by minor modifications in the standard proof of Anscombe's theorem (e.g., [6], p. 390).

An asymptotically more precise approximation to $\mathrm{E}T$ than that provided by Lemma 1, although one which is decidedly not uniform in p, is

$$pq\,\mathrm{E}T = c+\tfrac{1}{2}(p-q)^2+\tfrac{3}{2}pq+o(1) \quad (c\to\infty), \tag{5}$$

which is valid for all p for which $(p/q)^2$ is irrational. This result follows easily from (4) and Theorem 2 of Lai and Siegmund [4].

As an estimator of $\log(p/q)$, Haldane [3] considered $\log\{(s_n+a)/(f_n+a)\}$ and showed by a Taylor series expansion that the choice of a minimizing the asymptotic bias of this estimator is $a = \frac{1}{2}$. The following heuristic calculation shows that $a = -\frac{1}{2}$ is appropriate in the present context. The machinery for justifying this calculation may be found in [5]. It should

be noted that this result is appropriate for the stopping rule T defined by (2). It does not carry over to the two-population case discussed in Section 3.

A two-term Taylor series expansion gives

$$\log\{(s_T+a)/(f_T+a)\}-\log(p/q) = (s_T-pT+a)/pT-(f_t-qT+a)/qT- \\ -(s_T-pT)^2/2(pT)^2+(f_T-qT)^2/2(qT)^2+o_P(c^{-1}).$$

Since $T \overset{P}{\sim} c/pq$, and hence

$$\mathrm{E}\{(s_T-pT)^2/T^2\} \sim (pqc^{-1})^2\,\mathrm{E}(s_T-pT)^2 = (pqc^{-1})^2\,pq\,\mathrm{E}T \sim (pq)^2/c,$$

one obtains

(6)

$$\mathrm{E}[\log\{(s_T+a)/(f_T+a)\}]-\log(p/q) \sim \mathrm{E}\{(s_T-pT)/pqT\}+c^{-1}(q-p)(a-\tfrac{1}{2}).$$

It is proved in the sequel that

$$\mathrm{E}\{(s_T-pT)/T\} \sim c^{-1}pq(q-p) \quad (c\to\infty),$$

which shows that the right-hand side of (6) is $\sim c^{-1}(q-p)(a+\frac{1}{2})$, leading to the optimal choice $a=-\frac{1}{2}$.

Let $\xi_T = s_T f_T/T-c$. By (4) and Taylor expansions, one obtains

$$(s_T-pT)/T = (q-p)^{-1}(c+\xi_T)\{pqc^{-1}-(pqc^{-1})^2(T-c/pq)+ \\ +(pqc^{-1})^3(T-c/pq)^2+\ldots\}-(q-p)^{-1}pq+(q-p)^{-1}(s_T-pT)^2/T^2.$$

It is easy to see from (4) that $c+\mathrm{E}\xi_T = pq\,\mathrm{E}T-pq+o(1)$; and Robbins and Siegmund [7] have obtained $\mathrm{E}(T-c/pq)^2 = (q-p)^2c/(pq)^2+O(1)$. Hence, by the asymptotic independence of ξ_T and $c^{-1/2}(T-c/pq)$ (see [4]),

$$\begin{aligned}\mathrm{E}\{(s_T-pT)/T\} &= (q-p)^{-1}(c+\mathrm{E}\,\xi_T)\{pqc^{-1}-(pqc^{-1})^2(\mathrm{E}\,\xi_T/pq+1)+ \\ &\quad +(pqc^{-1})^3(q-p)^2c/(pq)^2\}- \\ &\quad -(q-p)^{-1}pq+(pq)^2/(q-p)c+o(c^{-1}) \\ &\sim c^{-1}pq(q-p),\end{aligned}$$

as claimed.

3. Two populations. Consider again the two-population case described in the Introduction and suppose that observations are taken in pairs, one from each population, so $n_1 = n_2 = n$, say. This restriction is stronger than necessary, but it simplifies the subsequent analysis. It is easy to modify the results to accommodate the case in which observations are taken from the two populations in an arbitrary fixed ratio. It seems possible to achieve a slight reduction in the total expected sample size by choosing the sampling rates adaptively, but the fairly small improvement seems not to be worth the considerable complication in analysis.

The obvious analogue of the stopping rule (2) is (cf. (1))

$$T = \inf\left\{n\colon n\left(\frac{1}{s_{1n} f_{1n}}+\frac{1}{s_{2n} f_{2n}}\right) \leqslant \frac{1}{c}\right\}. \tag{7}$$

The main results of this section are Theorems 2 and 3, which correspond to Lemma 1 and Theorem 1, respectively, in the single-population case. Theorem 2 shows that T defined by (7) is uniformly asymptotically efficient and Theorem 3 asserts that it asymptotically provides the correct coverage probability uniformly in p_1, p_2.

THEOREM 2. *Uniformly in* $0 < p_1, p_2 < 1$,

$$\mathrm{E}\,T \sim c\{(p_1 q_1)^{-1}+(p_2 q_2)^{-1}\} \quad (c \to \infty).$$

The inequality in one direction is a consequence of the following trivial lemma:

LEMMA 4. *For all* p_1, p_2 $(0 < p_1, p_2 < 1)$ *and all* c

$$\mathrm{E}\,T \geqslant c\{(p_1 q_1)^{-1}+(p_2 q_2)^{-1}\}.$$

Proof. From (4), Wald's identity and Jensen's inequality one obtains

$$\begin{aligned} c^{-1} &\geqslant \mathrm{E}\{T(1/s_{1T} f_{1T}+1/s_{2T} f_{2T})\} \geqslant \{\mathrm{E}(s_{1T} f_{1T}/T)\}^{-1}+\{\mathrm{E}(s_{2T} f_{2T}/T)\}^{-1} \\ &= \{p_1 q_1 \mathrm{E}T-\mathrm{E}[(s_{1T}-p_1 T)^2/T]\}^{-1}+\{p_2 q_2 \mathrm{E}T-\mathrm{E}[(s_{2T}-p_2 T)^2/T]\}^{-1} \\ &\geqslant (\mathrm{E}T)^{-1}\{(p_1 q_1)^{-1}+(p_2 q_2)^{-1}\}. \end{aligned}$$

To obtain asymptotic upper bounds on $\mathrm{E}T$ it is useful to define (cf. (2))

$$T_i(c) = \inf\{n\colon n/s_{in} f_{in} \leqslant 1/c\}.$$

Since $s_{in} f_{in}/n$ increases with n, for all $\alpha > 1$ and $\beta > 1$ such that $1/\alpha+1/\beta = 1$ one gets

$$T \leqslant \max(T_1(\alpha c), T_2(\beta c)). \tag{8}$$

In what follows $\alpha = (p_1 q_1+p_2 q_2)/p_2 q_2$ and $\beta = (p_1 q_1+p_2 q_2)/p_1 q_1$, so

$$\alpha/p_1 q_1 = \beta/p_2 q_2 = (p_1 q_1+p_2 q_2)/(p_1 q_1 p_2 q_2) = (p_1 q_1)^{-1}+(p_2 q_2)^{-1}. \tag{9}$$

With these fixed values of α and β there is no ambiguity in writing T_1 for $T_1(\alpha c)$ and T_2 for $T_2(\beta c)$.

It is now possible to complete the proof of Theorem 2. Obviously, by (8),

$$\mathrm{E}T \leqslant \mathrm{E}\{\max(T_1, T_2)\} = \int_{\{T_1 \leqslant T_2\}} T_2\,dP+\int_{\{T_2 < T_1\}} T_1\,dP. \tag{10}$$

Let $\varepsilon > 0$ be arbitrary and put $\gamma = \beta c(1+\varepsilon)/p_2 q_2$. Then

$$\begin{aligned} \int_{\{T_1 \leqslant T_2\}} T_2\,dP &\leqslant \int_{\{T_1 \leqslant T_2, T_2 \leqslant \gamma\}} T_2\,dP+\int_{\{T_2 > \gamma\}} T_2\,dP \\ &\leqslant \gamma P\{T_1 \leqslant T_2\}+(p_2 q_2)^{-1}\beta c P\{T_2 > \gamma\}+ \\ &\quad +\int_{\{T_2 > \gamma\}} |T_2-(p_2 q_2)^{-1}\beta c|\,dP. \end{aligned} \tag{11}$$

By Lemma 2

$$(p_2 q_2)^{-1} \beta c P\{T_2 > \gamma\} \leqslant 7 (p_2 q_2)^{-1} \varepsilon^{-2};$$

and by the Schwarz inequality and Lemma 2 again

$$\int_{\{T_2 > \gamma\}} |T_2 - (p_2 q_2)^{-1} \beta c| \, dP$$
$$\leqslant [(p_2 q_2)^{-2} \mathrm{E} |p_2 q_2 T_2 - \beta c|^2 P\{T_2 > \gamma\}]^{1/2} \leqslant 7 (p_2 q_2)^{-1} \varepsilon^{-1}.$$

Putting these inequalities together with (9), (10), and (11) yields

$$\mathrm{E} T \leqslant c\{(p_1 q_1)^{-1} + (p_2 q_2)^{-1}\}(1 + \varepsilon + 14/\varepsilon^2 c),$$

which completes the proof, as ε is arbitrarily small.

THEOREM 3. *For T defined by (7), uniformly in $0 < p_1, p_2 < 1$*

$$\lim_{c \to \infty} P\{c^{1/2} [\log (s_{1T} f_{2T} / s_{2T} f_{1T}) - \log (p_1 q_2 / p_2 q_1)] \leqslant x\} = \Phi(x).$$

With the help of Lemma 5 below, the proof of Theorem 3 may be carried out along the same lines as the proof of Theorem 1.

LEMMA 5. *Let $\mu = \{(p_1 q_1)^{-1} + (p_2 q_2)^{-1}\}^{-1}$. For all $\varepsilon > 0$ and all large c (not depending on ε)*

$$P\{|\mu T - c| > c\varepsilon\} \leqslant 14/c\varepsilon^2.$$

Proof. The proof of Theorem 2 shows that

$$P\{T > c(1+\varepsilon)\mu^{-1}\} = P\{T_1 < T_2, T_2 > (p_2 q_2)^{-1} \beta c (1+\varepsilon)\} +$$
$$+ P\{T_2 \leqslant T_1, T_1 > (p_1 q_1)^{-1} \alpha c (1+\varepsilon)\} \leqslant 7/c\varepsilon^2.$$

The same upper bound for $P\{T < c(1-\varepsilon)\mu^{-1}\}$ follows by a similar calculation and the observation that $T \geqslant \min(T_1(\alpha c), T_2(\beta c))$.

4. Remarks. (a) Unpublished numerical computations of H. Levene in the one-sample case show that the asymptotic theory of Section 2 provides good approximations for $c \geqslant 10$ and reasonable ones for c as small as 3. It seems likely that similar results hold for two populations.

(b) The heuristic principle which suggests the stopping rules (2) and (7) is quite common in the literature of fixed precision estimation (e.g., [1]), and it leads to reasonable stopping rules for more complicated log linear models. However, the uniform asymptotic theory developed here seems to require new ideas for very simple extensions.

One important generalization is a set of (2×2)-tables with equal odds ratios. Appropriate asymptotic theory might involve a large number of observations from each of a small number of tables or a large number of tables.

Another interesting variation is log linear regression. In this case one might also wish to consider sequential design in selecting values of the independent variable.

REFERENCES

[1] F. J. Anscombe, *Sequential estimation*, J. Roy. Statist. Soc., Ser. B, 15 (1953), p. 1-21.
[2] D. R. Cox, *Analysis of binary data*, Chapman and Hall, London 1970.
[3] J. B. S. Haldane, *The estimation and significance of the logarithm of a ratio of frequencies*, Ann. Human Genetics 20 (1955), p. 309-311.
[4] T. L. Lai and D. Siegmund, *A non-linear renewal theory with applications to sequential analysis, II*, Ann. Statist. 7 (1979), p. 60-76.
[5] M. Pollak and D. Siegmund, *Approximations to the expected sample size of certain sequential tests*, ibidem 3 (1975), p. 1267-1282.
[6] A. Rényi, *Wahrscheinlichkeitsrechnung*, VEB Deutscher Verlag der Wissenschaften, Berlin 1966.
[7] H. Robbins and D. Siegmund, *Sequential estimation of p in Bernoulli trials*, in: Studies in Probability and Statistics, Jerusalem Academic Press, 1974.

Department of Statistics
Stanford University
Stanford, California, U.S.A.

Received on 28. 1. 1980;
revised version on 18. 6. 1980