Journal of the Korean Data & Information Science Society
2009, **20**(5), 887–894

한국데이터정보과학회지

Minimum risk point estimation of two-stage procedure for mean[†]

Kiheon Choi[1]

[1]Department of Statistics, Duksung Women's University

Received 12 February 2009, revised 19 May 2009, accepted 12 September 2009

Abstract

The two-stage minimum risk point estimation of mean, the probability of success in a sequence of Bernoulli trials, is considered for the case where loss is taken to be symmetrized relative squared error of estimation, plus a fixed cost per observation. First order asymptotic expansions are obtained for large sample properties of two-stage procedure. Monte Carlo simulation is carried out to obtain the expected sample size that minimizes the risk and to examine its finite sample behavior.

Keywords: Bernoulli trials, minimum risk, risk efficient, two-stage procedure.

1. Introduction

Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables with $\mathrm{P(X_i = 1)} = \theta, \mathrm{P(X_i = 0)} = 1 - \theta, 0 < \theta < 1$. Given a sample of size n, one wishes to estimate θ, by the sample mean $\theta_n = S_n/n$ where $S_n = \sum_{i=1}^{n} X_i$, subject to the loss function

$$L_n = A\left(\frac{\theta_n - \theta}{\theta(1-\theta)}\right)^2 + cn, \tag{1.1}$$

where $A > 0$ is a known weight and $c > 0$ is a known cost per observation. Note that loss is modeled as the sum of a multiple of the symmetrized relative squared estimation error appropriate when θ close to 0 or 1 and the aggregate cost of observations. Consider the following hypothetical situation in medical trials (Hubert and Pyke, 2000). The probability θ that a drug will cure a particular ailment is to be estimated sequentially with a cost $c > 0$ per observation. If this value of θ is large, the drug will tend to be called a cure and research will shift to other ailments. If θ is small, the drug will be discarded and more money and time will be invested in the problem. However, if θ is close to one-half, no dramatic change will occur, that is, the drug will continue to be administered and research will continue

[†] This work was supported by the Institute of Natural Science Fund, Duksung Women's University, 2009.
[1] Professor, Department of Statistics, Duksung Women's University, Seoul 132-714, Korea.
E-mail: khchoi@duksung.ac.kr

in the same direction. Since in both extreme cases a dramatic decision will result, greater accuracy of estimation is demanded. A loss function satisfies such a requirement.

Developments of Robbins's work have included the study of more elaborate loss function for the normal distribution with unknown variance (Starr and Woodroofe, 1969) and extensions to other distributions such as the exponential (Starr and Woodroofe, 1972) and the uniform (Mukhopadhyay *et al.*, 1983). The following discussion is due to Robbins (1959). For fixed n and θ, the expected loss for (1.1) is

$$E_\theta\left(L_n\right) = A\left(n\theta(1-\theta)\right)^{-1} + cn,$$

which is minimized by using the optimal fixed sample size

$$n_0 \equiv n_0\left(\theta\right) \approx \left(A/\left(c\theta(1-\theta)\right)\right)^{1/2}. \tag{1.2}$$

The corresponding optimal fixed sample size risk is

$$E_\theta\left(L_{n_0}\right) = 2cn_0. \tag{1.3}$$

Since θ is unknown, the required sample size n_0 is indeed unknown, and there is no fixed sample size rule that will achieve the risk $E_\theta\left(L_{n_o}\right)$. In the case of squared relative error loss function. Robbins and Siegmund (1974) proposed a purely sequential estimation procedure for a Bernoulli success probability parameter θ. They proved that for any fixed $0 < \theta < 1$, as $c \to 0$, $N/n_0 \to 1$, in probability and $E_p\left(L_N\right)/E_p\left(L_{n_0}\right) \to 1$, so that the purely sequential procedure is asymptotically as good as the optimal fixed sample size rule n_0. For any fixed $c > 0$, however, this procedure performs badly for θ close to 0 or 1. To overcome this difficulty, Cabilio and Robbins (1975), Cabilio (1977) and Liu (2000) introduced the problem from the Bayesian point of view and obtained some asymptotic properties of the optimal Bayes procedure for the uniform prior and unspecified prior. Zacks and Mukhopadhyay (2007) develop the exact distribution of the stopping variable of a sequential procedure that was originally given by Robbins and Siegmund (1974). The stopping variable was designed for estimating the log-odds in a sequence of Bernoulli trials.

There are two major reasons for using sequential procedures in inference. The first is to decrease expected sample size with regard to hypothesis testing. The second reason is if there is no suitable fixed sample size procedure available. An early estimation example of the latter is Stein's two-stage procedure for estimating the mean of normal distribution with unknown variance, a problem for which no fixed sample size procedure suffices for all possible values of the variance. Usually, purely sequential procedures enjoy many desirable asymptotic properties. However, from the practical point of view, sampling with larger batches can be observed quickly and hence one may be able to cut down the operational time significantly compared with sapling sequentially when time or costs are important design factors. Furthermore, Hall (1981) proposed a triple sampling procedure and showed that the procedure combines the simplicity of Stein (1945) with the efficiency of the fully sequential procedures. In particular, two-stage and three-stage point estimation problems have been studied by Ghosh and Mukhopadhyay (1981). Mukhopadhyay (1985), Hamdy *et al.* (1988), Choi (2008) and Ghosh *et al.* (1997) in reviewing the progress of multistage estimation methods.

The plan of this paper is as follows. Section 2 proposes a two-stage sampling and point estimation procedure and then states the main result of this paper concerning its asymptotic

properties. Section 3 presents some results of the moderate sample size performance of the procedure using the Monte Carlo method.

2. Asymptotic for the two-stage procedure

The two-stage procedure is constructed as follows. Let m be a positive integer and we start the experiment with a sample of size m, say $X_1, \ldots, X_m$. Based on the sample, we assume that the pilot sample size m is chosen such that

$$m = m(c) = \left\langle 2\left(A/c\right)^{1/2} \right\rangle + 1 \tag{2.1}$$

where $< a >$ is the largest integer smaller than a. We note that the initial sample size m is going to increase as $n_0(\theta)$ increases and for all practical purposes m is always less than $n_0(\theta) + 1$ since $\theta(1-\theta) \leq 1/4$ for all $\theta \in (0,1)$. So, we may proceed to consider two-stage procedure. Let

$$N = (A/c)^{1/2} \left\{ \bar{X}_m(1-\bar{X}_m) + m^{-1} \right\}^{-1/2}.$$

Note that we make sure that $\bar{X}_m(1-\bar{X}_m) + m^{-1}$ remains well defined if it happens at all that $\bar{X}_m = 0$ or $\bar{X}_m = 1$. Next, the final sample size M as

$$M = M(c) = \max\left\{m, \langle N \rangle + 1\right\}. \tag{2.2}$$

When $M = m$, we do not take any more samples in the second stage. If, however, $M > m$ then we obtain more $M - m$ observations, say, $X_{m+1}, \ldots, X_M$. Finally, we estimate θ by θ_M and the corresponding loss is L_M. As usual, the risk efficiency of the two-stage procedure (2.2) is defined as $e(c) = E_\theta(L_M)/E_\theta(L_{n_0})$. In this section, we study the asymptotic properties of the proposed two-stage stopping rule. Justification of the proposed procedure rests primarily upon its good asymptotic behavior for sufficiently small c. Theoretically, we are not able to investigate its small sample behavior of the proposed procedure when sample size is sufficiently large. Therefore, since the random stopping time M is a function of c, we can get large enough M by letting c get small. From the stopping rule (2.2), we note that $M(c) \geq m(c)$ and thus $\lim_{c\to 0} M(c) \geq \lim_{c\to 0} m(c) = \infty$ with probability one. So, we $\lim_{c\to 0} M(c) = \infty$ with probability one. M has the following properties as $c \to 0$ for fixed $0 < \theta < 1$.

Theorem 2.1 We have
(i) $\lim_{c\to 0} M/n_0 = 1$ in probability.
(ii) $\lim_{c\to 0} E\left\{M/n_0\right\} = 1$.

Proof: There is no loss of generality in supposing that $A = 1$ and from (2.2), we have

$$\begin{aligned} c^{-1/2}\left\{\bar{X}_m(1-\bar{X}_m) + m^{-1}\right\}^{-1/2} \leq M < c^{-1/2} \\ \times \left\{\bar{X}_m(1-\bar{X}_m) + m^{-1}\right\}^{-1/2} + mI(M=m) \end{aligned} \tag{2.3}$$

from which it is see that

$$\left\{\bar{X}_m(1-\bar{X}_m)+m^{-1}\right\}^{-1/2}(\theta(1-\theta))^{1/2} \le \frac{M}{n_0}$$
$$< \left\{\bar{X}_m(1-\bar{X}_m)+m^{-1}\right\}^{-1/2}(\theta(1-\theta))^{1/2} + \frac{m}{n_0}I(M=m). \tag{2.4}$$

We know that $\lim_{c\to 0} m = \infty$, but $m = O(n_0)$. Using the weak law of large numbers it follows that the estimator $\bar{X}_m$ of θ converges in probability to θ as $n_0 \to \infty$. Thus, part (i) will follow from (2.4). if we verify that $I(M=m) \to 0$ in probability as $n_0 \to \infty$. It follows that there exists a $\delta > 0$ such that $[M=m] \subset [|\theta_m - \theta| \ge \delta]$ for all large value of m. Using the basic inequality (Loève, 1977, page 160) for a binomial random variable, it follows that

$$\begin{aligned} P(\theta_m \ge \theta+\delta) =& P(X_1+\cdots+X_m \ge m(\theta+\delta)) \\ \le& \exp\left\{-m\delta t + mt^2\right\} \end{aligned} \tag{2.5}$$

for all $0 < t < 1$. In particular, taking $t = \delta/2$, the right hand side of (2.5) is less than or equal to $\exp(-m\eta_\delta)$ where $\eta_\delta = \delta^2/4 > 0$. Similarly, $P(\theta_m \le \theta - \delta) \le \exp(-mn_\delta)$. Putting together yields

$$P(|\theta_m - \theta| \ge \delta) \le 2exp(-m\eta_\delta)$$

for some $\eta_\delta > 0$. So $P(M=m) \le P(|\theta_m - \theta| \ge \delta) = O(e^{-m\eta_\delta})$. Next, with an arbitrary but fixed $\epsilon > 0$,

$$P(I(M=m) > \epsilon) \le \epsilon^{-1}E\left\{I(M=m)\right\} = \epsilon^{-1}P\left\{M=m\right\} \to 0,$$

as $n_0 \to \infty$. Hence $\lim_{c\to 0} M/n_0 = 1$.

The proof of part (ii) is based on showing that the convergence of part (i) is dominated. We verify that $P(M=m) = O(e^{-m\eta_\delta})$. Also, the Fatou's lemma (Ash, 1972, p. 48) and part (i) together will let us conclude that

$$\liminf_{c\to 0} \; E\left\{\frac{M}{n_0}\right\} \ge E\left\{\liminf_{c\to 0} \; \frac{M}{n_0}\right\} = 1.$$

Thus, in view of the upper bound given by (2.4), part (ii) immediately follow if we can verify the follows results

$$\lim_{c\to 0} E\left\{\frac{1}{\bar{X}_m(1-\bar{X}_m)+m^{-1}}\right\}^{1/2} = \left(\frac{1}{\theta(1-\theta)}\right)^{1/2} \tag{2.6}$$

for every fixed $0 < \theta < 1$. Now, by the Taylor's theorem,

$$\begin{aligned} E\left\{\bar{X}_m(1-\bar{X}_m)+m^{-1}\right\}^{-1/2} &= E\left\{\bar{X}_m(1-\bar{X}_m)+m^{-1}\right\}^{-1/2} I_{\{|\theta_m-\theta|<\delta\}} + 0(m^{-1}) \\ &= E\left\{(\bar{X}_m(1-\bar{X}_m))^{-1/2} I_{\{|\theta_m-\theta|<\delta\}}\right\} - (1/2)m^{-1} \\ &\quad + E\left\{(\bar{X}_m(1-\bar{X}_m))^{-3/2} I_{\{|\theta_m-\theta|<\delta\}}\right\} + 0(m^{-1}) \\ &= \theta(1-\theta)^{-1/2}\left[1 + \frac{3(\theta(1-\theta))^{-1}-8}{8m} - \frac{(\theta(1-\theta))^{-1}}{2m}\right] \\ &\quad + 0(m^{-1}). \end{aligned}$$

Combining the above, we conclude the validity of (2.6), which in turn completes the proof of (ii). □

Theorem 2.2 As $c \to 0$,

(i) $(\theta(1-\theta)/c)^{1/2}\left(\dfrac{\bar{X}_M-\theta}{\theta(1-\theta)}\right)^2 \to Z^2$ in law.

(ii) $(\theta(1-\theta)/c)^{1/2}E\left\{\left(\dfrac{\bar{X}_M-\theta}{\theta(1-\theta)}\right)^2\right\} \to 1,$

where Z is a normally distributed random variable with mean 0 and variance 1.

Proof: To prove (i) we note that

$$(\theta(1-\theta)/c)^{1/2}\left(\frac{\bar{X}_M-\theta}{\theta(1-\theta)}\right)^2 = \frac{(\sum_{i=1}^{M} X_M - M\theta)^2}{M\theta(1-\theta)}\frac{1}{M(\theta(1-\theta)c)^{1/2}}.$$

Thus (i) follows from Anscombe's theorem (Woodroofe, 1982, p. 11) and (i) of theorem 2.1. Since $M \geq 2/c^{1/2}$,

$$(\theta(1-\theta)/c)^{1/2}\left(\frac{\bar{X}_M-\theta}{\theta(1-\theta)}\right)^2 \leq (c/\theta(1-\theta))^{1/2}\frac{(\sum_{i=1}^{M} X_M - M\theta)^2}{4\theta(1-\theta)}$$

and thus (ii) follows from (i) if $(c/\theta(1-\theta))^{1/2}(\sum_{i=1}^{M} X_M - M\theta)^2$ is uniformly integrable for $0 < \theta < 1$. To prove the latter we make use of Wald's lemma for second moments (Woodroofe, 1982, p. 8), together with (ii) of theorem 2.1, to obtain as $c \to 0$

$$E\left\{(c/\theta(1-\theta))^{1/2}(\sum_{i=1}^{M} X_M - M\theta)^2\right\} = E(M)(c\theta(1-\theta))^{1/2} \to 1.$$

Further, (i) of theorem 2.1 and Anscombe's theorem yield

$$(\sum_{i=1}^{M} X_M - M\theta)^2(c/\theta(1-\theta))^{1/2} = \frac{(\sum_{i=1}^{M} X_M - M\theta)^2}{M\theta(1-\theta)}M(c\theta(1-\theta))^{1/2} \to Z^2$$

in law.
The uniform integrability follows from the convergence theorem of Woodroofe (1982). □

The performance of the two-stage procedure is usually evaluated by comparing two risks; one is $R_M(c)$, the risk involved in two-stage estimation of θ using the proposed two-stage procedure, and the other is $R_{n_0}(c)$, the risk associated with the optimal fixed-sample size n_0. As a measure of closeness, the risk efficiency under consideration are defined by

$$e(c) = \frac{R_M(c)}{R_{n_0}(c)}.$$

To show that the proposed two-stage procedure is asymptotically risk efficient. i.e., $R_M(c)/R_{n_0}(c) \to 1$ as $c \to 0$. We establish the following theorem.

Theorem 2.3 As $c \to 0$

$$e(c) \to 1.$$

Proof: From (ii) of theorem 2.1 and (ii) of theorem 2.2 it follows that as $c \to 0$

$$\frac{R_M(c)}{R_{n_0}(c)} = \frac{E\left\{(\bar{X}_M - \theta)^2/(\theta^2(1-\theta)^2)\right\} + cEM}{2(c/\theta(1-\theta))^{1/2}} \to 1.$$

□

3. Monte Carlo Studies

In this section, we conducted to illustrate the performance of the stopping rule in the proposed two-stage procedure (2.2) as $c \to 0$. We show several choice of the parameter θ, namely $\theta = 0.01, 0.02, 0.05, 0.1$ under the loss function (1.1) with $c = 0.10, 0.01, 0.001$ and $n_0 = 25, 100, 500, 1000$. Note that for the given values of n_0, θ, c, the starting sample size m can be computed from (1.2) and (2.1). Simulation results are presented in Tables 3.1-3.4. For each selected values of c, every value in the table is based on 100000 independent replications with the initial sample size m. Each table contains the selected value of c, For each row of the tables, the optimal stopping time n_0, the estimates of series of θ, we computed the mean $\bar{\theta}_M$ and the standard deviation $\widehat{s}_{\theta_M}$ of the 10000 simulated values of θ_M, the corresponding average EM, $E_\theta(M/n_0)$ and the risk efficiency $e(c)$.

From Tables 3.1-3.4, we see that the estimate $\bar{\theta}_M$ converge θ. EM is uniformally smaller than the optimal stopping time n_0. That is, the suggested procedure requires smaller sample size than the fixed-sample procedure. We also observe that as the sampling cost per observation c becomes smaller, the average random stopping time EM increases.

However, the values of c plays a role as a sample inflation factor in the two-stage procedure. Analytically, the risk efficiency approaches one as $c \to 0$. The simulation results provide substantial numerical evidence to conclude that the proposed two-stage estimator θ_M performs satisfactorily.

Table 3.1 For $\theta = 0.01$

c	n_0	m	$\bar{\theta}_M$	$\widehat{s}_{\theta_M}$	$E(M)$	$E(M/n_0)$	$e(c)$
0.1	25	5	0.0094	0.0395	5	0.2379	2.0089
0.1	50	10	0.0097	0.0247	15	0.3125	1.6356
0.1	100	20	0.0100	0.0154	42	0.4258	1.3612
0.1	500	100	0.0100	0.0052	386	0.7733	1.043
0.1	1000	199	0.0100	0.0034	892	0.8921	1.0173
0.01	25	5	0.0098	0.0401	5	0.2379	2.068
0.01	50	10	0.0101	0.026	15	0.3121	1.7896
0.01	100	20	0.0101	0.0154	42	0.4254	1.3571
0.01	500	100	0.0100	0.0052	385	0.7715	1.0511
0.01	1000	199	0.0100	0.0034	893	0.8939	1.0177
0.001	25	5	0.0099	0.0411	5	0.2381	2.1699
0.001	50	10	0.0099	0.0251	15	0.3124	1.6884
0.001	100	20	0.0098	0.0153	42	0.4264	1.354
0.001	500	100	0.0100	0.0052	386	0.7732	1.0446
0.001	1000	199	0.0100	0.0034	894	0.8946	1.0221

Table 3.2 For $\theta = 0.02$

c	n_0	m	θ_M	$\widehat{s}_{\theta_M}$	$E(M)$	$E(M/n_0)$	$e(c)$
0.1	25	7	0.0199	0.046	9	0.3843	1.4384
0.1	50	14	0.0201	0.0286	24	0.4981	1.2117
0.1	100	28	0.0201	0.018	64	0.6445	1.0883
0.1	500	141	0.0200	0.0067	466	0.9321	0.9893
0.1	1000	281	0.0200	0.0046	973	0.9738	0.9889
0.01	25	8	0.0198	0.045	9	0.3881	1.3848
0.01	50	15	0.0201	0.0278	25	0.5156	1.1684
0.01	100	29	0.0200	0.0176	65	0.6520	1.0542
0.01	500	141	0.0201	0.0066	468	0.9364	0.9765
0.01	1000	281	0.0200	0.0046	971	0.9712	0.9773
0.001	25	8	0.0194	0.0445	9	0.3876	1.3586
0.001	50	15	0.0199	0.0275	25	0.5168	1.1485
0.001	100	29	0.0197	0.0175	65	0.6528	1.0462
0.001	500	141	0.0200	0.0066	467	0.9341	0.9836
0.001	1000	281	0.0200	0.0046	970	0.9705	0.9768

Table 3.3 For $\theta = 0.05$

c	n_0	m	θ_M	$\widehat{s}_{\theta_M}$	$E(M)$	$E(M/n_0)$	$e(c)$
0.1	25	11	0.0501	0.0545	16	0.6626	0.9686
0.1	50	22	0.0503	0.0359	39	0.7912	0.9484
0.1	100	44	0.0499	0.0238	90	0.9045	0.938
0.1	500	218	0.0501	0.0099	492	0.9853	0.9141
0.1	1000	436	0.0500	0.0069	993	0.9934	0.9017
0.01	25	11	0.0494	0.0539	16	0.6663	0.9548
0.01	50	22	0.0495	0.0352	39	0.7947	0.9287
0.01	100	44	0.0498	0.0235	90	0.9019	0.9262
0.01	500	218	0.0499	0.0098	491	0.9834	0.8998
0.01	1000	436	0.0501	0.007	991	0.9912	0.9152
0.001	25	11	0.0498	0.0541	16	0.6640	0.9584
0.001	50	22	0.0501	0.0354	39	0.7916	0.9331
0.001	100	44	0.0502	0.0239	89	0.8985	0.9383
0.001	500	218	0.0500	0.0099	491	0.9829	0.9101
0.001	1000	436	0.0500	0.0069	991	0.9918	0.8988

Table 3.4 For $\theta = 0.1$

c	n_0	m	θ_M	$\widehat{s}_{\theta_M}$	$E(M)$	$E(M/n_0)$	$e(c)$
0.1	25	16	0.0994	0.0653	21	0.8692	0.8233
0.1	50	31	0.1007	0.0452	46	0.9282	0.8367
0.1	100	61	0.1003	0.0307	97	0.9706	0.8284
0.1	500	301	0.0998	0.0136	496	0.9928	0.8328
0.1	1000	601	0.1001	0.0096	996	0.9969	0.8321
0.01	25	16	0.1015	0.0651	21	0.8682	0.8199
0.01	50	31	0.100	0.0449	46	0.9332	0.8338
0.01	100	61	0.1002	0.0311	96	0.9671	0.8356
0.01	500	301	0.1001	0.0136	496	0.9925	0.8337
0.01	1000	601	0.1001	0.0095	996	0.9963	0.8248
0.001	25	16	0.1001	0.0647	21	0.8691	0.8162
0.001	50	31	0.1007	0.0448	46	0.9362	0.8335
0.001	100	61	0.0998	0.0308	96	0.9693	0.831
0.001	500	301	0.1001	0.0135	496	0.9932	0.8278
0.001	1000	601	0.0999	0.0094	997	0.9970	0.8175

References

Ash, R. (1972). *Real analysis and probability,* Academic Press, New York.

Cabilio, P. (1977). Sequential analysis in Bernoulli Trials. *Annals of Statistics*, **5**, 342-356.

Cabilio, P. and Robbins, H. (1975). Sequential estimation of p with squared relative error loss. *Proceedings of the National Academy of Sciences U.S.A.*, **72**, 191-193.

Choi, K. (2008). Two-stage procedure with lower bound variance based on bootstrap. *Journal of Korean Data & Information Science Society*, **19**, 657-666.

Ghosh, M. and Mukhopadhay, N. (1981). Consistency and asymptotic efficiency of two-stage and sequential estimation procedure. *Sankya* A, **43**, 220-227.

Ghosh, M., Mukhopadhay, N. and Sen, P. K. (1997). *Sequential estimation,* John Wiley, New York.

Hall, P. (1981). Asymptotic theory of triple sampling for estimation of a mean. *Annals of Statistics*, **9**, 1229-1238.

Hamdy, H. I., Mukhopadhyay, N., Costanza, M. C. and Son, M. S. (1988). Triple stage point estimation for exponential location parameter. *Annals of the Institute of Statistical Mathematics,* **40**, 785-797.

Hubert, S. and Pyke, R. (2000). Sequential estimation of functions of p for Bernoulli Trials. *Lecture Notes-Monograph Series*, **35**, Institute of Mathematical Statistics.

Kook, S., Han, H., Kim, G. and Choi, K. (2008). The anti-hepatotoxic effect of ginseng in rats: Meta-analysis. *Journal of Korean Data & Information Science Society*, **19**, 937-949.

Liu, J. (2000). A note on two-stage point estimation in Bernoulli trials. *Sequential Analysis,* **19**, 161-176.

Loève, M. (1977). *Probability theory I,* 4th ed. Springer-Verlag, New York.

Mukhopadhyay, N. (1985). A note on three-stage and sequential point estimation procedures for a normal mean. *Sequential Analysis,* **4**, 311-319.

Mukhopadhyay, N., Hamdy, H. I., Gosh, M. and Wackerly, D. (1983). Sequential and two-stage point estimation for the range in a power family distribution. *Sequential Analysis,* **2**, 259-288.

Robbins, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics, H. Cramér Volume,* (ed. by U. Grenander), 235-245, Uppsala: Almquist and Wiksell.

Robbins, H. and Siegmund, D. (1974). Sequential estimation of p in Bernoulli Trials. In *Studies in Probability and Statistics,* (ed. by E. J. Williams), 103-107. North-Holland, Amsterdam.

Starr, N. and Woodroofe, M. (1969). Remarks on sequential point estimation. *Proceedings of the National Academy of Sciences U.S.A.,* **62**, 285-288.

Starr, N. and Woodroofe, M. (1972). Further remarks on sequential estimation: The exponential case. *The Annals of Mathematical Statistics*, **43**, 1147-1154.

Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *The Annals of Mathematical Statistics,* **16**, 243-258.

Woodroofe, M. (1977). Second-order approximations for sequential point and interval estimation. *Annals of Statistics,* **5**, 984-995.

Woodroofe, M. (1982). *Non linear renewal theory in sequential analysis,* Society for Industrial and Applied Mathematics, Philadelphia.

Zacks, S. and Mukhopadhyay, N. (2007). Distribution of sequential and two-stage stopping times for fixed-width confidence intervals in Bernoulli trials: Application in reliability. *Sequential Analysis*, **26**, 425 -441.