International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 3 Issue 6 || September. 2015 || PP-24-32

A Two-Stage Approach for Estimation of Prevalence of Human Diseases

*****Rahul Gupta and Neha Jain
Department of Statistics
University of Jammu
Jammu, J and k, India

ABSTRACT *In this paper , for the estimation of P, the prevalence of a disease ,results of a study made have been presented and applied to simulated as well as actual data using a Two-Stage approach. Properties of the proposed procedure are also studied to know the significance and importance of the procedures under symmetrized relative squared estimation error appropriate for prevalence of a diseases close to 0 and the aggregate cost of selecting observations. Asymptotic characteristics for the risk and the moments are studied for the two stage sampling procedure.*

KEYWORDS *Prevalence, Regret , Risk efficiency and Two-stage procedure.*

I. INTRODUCTION

The estimation of prevalence of a disease at a certain point of time and/or location has always been a challenging task for researchers. Many studies have and are being carried out to improve on the available methods to estimate the prevalence of a disease. For planning, coordination, and evaluation of control activities, it is essential to process and obtain reliable disease prevalence estimates at any time or place in this fast changing world.

II. THE SET UP OF THE PROBLEM

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables representing sampling units who are selected for studying certain disease in a population with $P(X_i=1) = p$, representing presence of a disease and $P(X_i=0) = q$, representing absence of the disease. Then $0<p<1$, $p+q=1$.

Given a sample of size n, one wishes to estimate p, by the sample mean $p_n = {S_n}/{n}$ where $S_n = \sum_{i=1}^{n} X_I$ subject to the loss function

$$L_n = A\left(\frac{p_n - p}{pq}\right)^2 + cn \qquad (2.2.1)$$

where $A > 0$ is a known weight and $c > 0$ is the known cost per observation. Note that loss is modelled as the sum of a multiple of the symmetrised relative squared estimation error approximate when p close to 0 and the aggregate cost of observations. The case where A = 1 has been considered by Robbins and Siegmund [1]. Cabilo and Robbins [2] and Cabilo [3], among others. Baran and Magiera[4]have also studied and proposed an estimation procedure for p under Linex loss function. For any value of A , Liu[5] under loss function (2.2.1) has proposed a two stage and a Bayesian procedure for estimation of p and this paper has been prepared along his lines with added applications to study of disease prevalence.

For fixed n and p, the risk for (2.2.1) is

$$E_p(L_n) = A(npq)^{-1} + cn$$

which is minimised by using the optimal fixed sample size

$$n_0 \equiv n_0(p) \approx \left({A}/{(cpq)}\right)^{1/2}$$

(2.2.2) The corresponding optimal risk for fixed sample size is

$$E_p\left(L_{n_0}\right) = 2cn_0 \qquad (2.2.3)$$

Since p is unknown, the required sample size n_0 is indeed unknown, and there is no fixed sample size rule that will achieve the risk $E_p(L_{n_0})$. In the case of A=1, Robbins and Siegmund [1] proposed a purely sequential procedure for the problem of approximating the optimal risk $E_p(L_{n_0})$. The stopping rule for the proposed procedure is

$$\text{N= inf } \{n \geq 1 | n \geq (cp_n q_n)^{-1/2}\}, \tag{2.2.4}$$

where $q_n = 1 - p_n$. , Robbins and Siegmund [1] showed that for any fixed 0<p<1, as $c \to 0$, $E_p(N/n_0)^k \to 1 \; for \; k = 1,2,\ldots$ and $E_p(L_N)/E_p(L_{n_0}) \to 1$, so that the procedure (2.2.4) is asymptotically as

good as the optimal fixed sample size rule n_0

The plan of this paper is as follows. Section 2.3 propose a two-stage sampling and point estimation procedure and then states the main result of this concerning its asymptotic properties. Section 2.3.1 presents some results of the moderate sample size performance of the procedure using the Monte-Carlo method and study prevalence of certain diseases affecting children in age group of 0 to 6 in Jammu District, J and K State.

III. A TWO-STAGE PROCEDURE AND ITS PROPERTIES

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 $X_1, \ldots\ldots\ldots X_m$. Based on the sample, let

$$N_1 = (A/c)^{1/2}(p_m q_m + m^{-\gamma})^{-1/2} \tag{2.3.1}$$

where $\gamma > 0$ is a given constant. Note that the tuning of the term $m^{-\gamma}$ in (2.3.1) is essential so that N_1 becomes finite with probability 1. Sample size is defined as

$$M = \max\{m, [N_1]^{**} + D\}, \tag{2.3.2}$$

where $D = I_{\{U \leq N_1 - [N_1]^{**}\}}$ is the indicator of $\{U \leq N_1 - [N_1]^{**}\}$,$[x]^{**}$ stands for the largest integer smaller than x, and U is uniformly distributed on (0, 1) and independent of $X_1, X_2, \ldots\ldots$ Note that (2.3.2) is a randomised stopping rule first introduced by Woodroofe [6]and, M= m, if $N_1 \leq m$ and M = $[N_1]^{**}$+ D otherwise. 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, $X_{m+1}, \ldots\ldots, X_M$. Finally, we estimate p by p_M and the corresponding loss is L_M. As usual, the regret of the two-stage procedure (2.3.2) is defined as $\omega = E_p(L_M) - E_p(L_{n_0})$.

In the rest of this paper , we assume that the pilot sample size m is chosen such that

$$m = [2(A/C)^{1/2}]^{**} + 1 . \tag{2.3.3}$$

Using (2.2.2), m is always less than $n_0 + 1$ since $pq \leq 1/4$ for all $p \epsilon (0,1)$, but they have the same order of magnitude. This is why we need only two sampling operations.

Before providing the main results we present some Lemmas.

Lemma 2.3.1: For $0 < \delta < 1/2, define \; A_{m,\delta} = \{|p_m - p| < \delta\}$. Then there exists a $n_\delta > 0$ such that as $m \to \infty$,

$$P(\bar{A}_{m,\delta}) = O(e^{-mn_\delta}), \tag{2.3.4}$$

where $\bar{A}_{m,\delta}$ stands for the complement of $A_{m,\delta}$.

Proof : Using the Chernoff bound for a binomial random variable, it follows that

$$P(p_m > p + \delta) = P[X_1 + X_2 + \cdots \ldots + X_m \geq m(p + \delta)]$$

$$\leq exp[-m(p + \delta)t + m(pt + t^2)] \quad (2.3.5)$$

for all $0 < t < 1$.

In particular, taking $t = \delta/2$, the right hand side of (2.3.5) is less than or equal to $\exp(-mn_\delta)$ where $n_\delta = \gamma^2/4 > 0$.

A similar argument works for estimating the probability that $p_m \leq p - \delta$.

Putting the two results together yields

$$P(\bar{A}_{m,\delta}) \leq 2exp(-mn_\delta)$$

for some $n_\delta > 0$. This completes the proof.

Lemma2.3.2: Let q(x) be a real function such that its second derivative, $g''(x)$, is continuous at p. Then, for small $\delta > 0$, we have as $m \to \infty$,

$$E_p\left[q(p_m)I_{A_{m,\delta}}\right] = g(p) + pqg''(p)/(2m) + o(m^{-1}) \quad (2.3.6)$$

In particular for every real number s,

$$E_p\left[(p_m q_m)^s I_{A_{m,\delta}}\right] = (pq)^s\left(1 + \frac{s(s-1)(pq)^{-1} - 2s(2s-1)}{2m}\right) + o(m^{-1}) \quad (2.3.7)$$

Proof: By Taylor's theorem,

$$g(p_m) = g(p) + g'(p)(p_m - p) + \frac{1}{2}g''(\xi_m)(p_m - p)^2,$$

where ξ_m is some random variable between p and p_m. Then, it follows that

$$E_p\left[g(p_m)I_{A_{m,\delta}}\right] = g(p)P(A_{m,\delta}) + g'(p)E_p\left[(p_m - p)I_{A_{m,\delta}}\right] + \frac{1}{2m}E_p(Z_m), \quad (2.3.8)$$

where $Z_m = mg''(\xi_m)(p_m - p)^2 I_{A_{m,\delta}}$. As $m \to \infty$, using (2.3.4),the first term of the right hand side in (2.3.8) is $g(p) + o(m^{-1})$ and the second term is $o(m^{-1})$,so from (2.3.8) we shall complete the proof of (2.3.6) by showing that $E_p(Z_m) = pqg''(p) + o(1)$ as $m \to \infty$. Note that Z_m convergs in distribution to $pqg''(p)\chi_1^2 where$ χ_1^2denotes a chi square random variable with 1 degree of freedom, hence it suffices to show that $\{Z_m\}_{m\geq 1}$ is uniformly integrable.

Observe that, for small $\delta > 0$,

$$E_p(Z_m^2) \leq B^2 E_p[m^2(p_m - p)^4], \quad (2.3.9)$$

where B is a bound for $|g''(x)|$ on $A_{m,\delta}$. It follows that $\{Z_m\}_{m\geq 1}$ is uniformly integrable since the right hand side in (2.3.9) is bounded. This completes the proof.

The main results in this section are given in the following theorems, which separate $p \neq 1/2$ and p=1/2.

Theorem2.3.1. If $p \neq 1/2, \gamma > 1/2$ and (2.3.3) holds, then as $m \to \infty$, we have

1. For every positive integer k,

$$E_p(M/n_0)^k = 1 + \frac{k(k+2)(pq)^{-1} - 4k(k+1)}{8m} - \frac{k(pq)^{-1}}{2m^\gamma} + o(m^{-1}) \quad (2.3.10)$$

In particular,

$$E_p(M) - n_0 = \frac{3(pq)^{-3/2} - 8(pq)^{-1/2}}{16} - \frac{(pq)^{-3/2}}{4} m^{1-\gamma} + o(1) \ , \tag{2.3.11}$$

and the variance of M is

$$V(M) = m[(pq)^{-1/2} - \frac{4(pq)^{-1}}{16} + o(m)$$

2. The regret of the two-stage procedure (2.3.2) is given by

$$w = c[\frac{(pq)^{-3/2}}{8} + (pq)^{-1}.\frac{(pq)^{-1/2}}{2} - 6 + o(o) \tag{2.3.12}$$

Proof: For (2.3.10) using (2.3.2)

$$E_p(M^k) = E_p\{([N_1]^{**} + D)^k I_{[M>m]}\} + E_p(m^k I_{[M=m]})$$

$$= E_p[(N_1 + \beta_m)^k] + E_p\{[m^k - (N_1 + \beta_m)^k I_{[M=m]}]\}, \tag{2.3.13}$$

where $\beta_m = [N_1]^{**} - N_1 + D$. Since $p \neq 1/2$, it follows that there exists a $\delta > 0$ such that [M=m] c $\bar{A}_{m,\delta}$ for all large values of m. Thus , using (2.3.4) and (2.3.1), (2.3.13) becomes

$$E_p(M^k) = \sum_{j=0}^{k} \binom{k}{j} E_p({N_1}^{k-j} {\beta_m}^j) + o(1)$$

$$= (A/c)^{k/2} E_p\left[(p_m q_m + m^{-\gamma})^{-k/2}\right] + k(A/c)^{(k-1)/2} E_p\left[(p_m q_m + m^{-\gamma})^{-(k-1)/2}\beta_m\right] + \quad 0(m^{k-1}) . \tag{2.3.14}$$

Now from (2.3.4) for some $\delta > 0$, (2.3.7) with s=$-k/2$, and Taylor's theorem, we get

$$E_p\left[(p_m q_m + m^{-\gamma})^{-k/2} = E_p\left[(p_m q_m + m^{-\gamma})^{-k/2} I_{A_{m,\delta}}\right]\right] + 0(m^{-1})$$

$$= E_p\left[(p_m q_m)^{-k/2} I_{A_{m,\delta}}\right] - (k/2) m^{-\gamma} E_p\left[(p_m q_m)^{-(k+2)/2} I_{A_{m,\delta}}\right] + 0(m^{-1})$$

$$= (pq)^{-k/2}\left[1 + \frac{k(k+2)(pq)^{-1} - 4k(k+1)}{8m} - \frac{k(pq)^{-1}}{2m^{\gamma}}\right] + 0(m^{-1}) \tag{2.3.15}$$

Using (2.3.14) and (2.3.15), (2.3.10) follows if we show that

$$E_p\left[(p_m q_m + m^{-\gamma})^{-(k-1)/2}\beta_m\right] = o(1),$$

and so it suffices to show that $(p_m q_m + m^{-\gamma})^{-(k-1)/2}$ *and* β_m are asymptotically uncorrelated because $E_p(\beta_m) = 0$. This is easily accomplished since, using (2.3.15),

$$Cov^2((p_m q_m + m^{-\gamma})^{-(k-1)/2}, \beta_m)$$

$$\leq V\left[(p_m q_m + m^{-\gamma})^{-(k-1)/2}\right] V(\beta_m)$$

$$= o(1).$$

This completes the proof of (2.3.10).

For (2.3.12), using (2.2.1) and (2.2.3)

$$E_p(L_M) - E_p(L_{n_o}) = AE_p\left(\frac{p_M - p}{pq}\right)^2 + cE_p(M) - 2cn_o$$

$$= c[(A/c)E_p\left(\frac{p_M - p}{pq}\right)^2 - n_o + E_p(M) - n_o]$$

(2.3.16)

Conditioning on $\mathcal{F}_m$, the $\sigma - field$ generated by U, $X_1, \ldots\ldots, X_m$, gives

$$(A/c)E_p\left(\frac{p_M - p}{pq}\right)^2 = \frac{A}{c(pq)^2}E_p[E_p[\left(\frac{S_M - M_F}{M}\right)^2|\mathcal{F}_m]$$

$$= \frac{A}{c(pq)^2}E_p[\frac{(S_m - mp)^2 + (M-m)pq}{M^2}]$$

$$= \frac{A}{c(pq)^2}E_p\left[\frac{m^2(p_m - p)^2 - mpq}{(N_1 + \beta_m)^2}\right] + \frac{A}{cpq}E_p\left[\frac{1}{N_1 + \beta_m}\right] + o(1)$$

= I+II +o(1), say . (2.3.17)

The fourth equality holds by making use of (2.3.4) since there exists a $\delta > 0$ such that [M=m]c $\bar{A}_{m,\delta}$ for all large values of m. To evaluate the first term I in (2.3.17), we write

$$\text{I} = \frac{A}{c(pq)^2}E_p\left[\frac{m^2(p_m - p)^2 - mpq}{N_1^{\,2}}\right] + \frac{A}{c(pq)^2}E_p\{[m^2(p_m - p)^2 - mpq][(N_1 + \beta_m)^{-2} - N_1^{-2}]\}$$

$= I_a + I_b$, say . (2.3.18)

It follows from (2.3.1) and (2.3.3) that

$$I_b = \frac{-A}{c(pq)^2}E_p\{[m^2(p_m - p)^2 - mpq]\frac{\beta_m(2N_1 + \beta_m)}{N_1^{\,2}(N_1 + \beta_m)^2}\}$$

$$= \frac{-2}{(pq)^2}E_p\{[m(p_m - p)^2\text{-pq}]\,\beta_m\frac{(2 + N_1^{-1}\beta_m)(p_m q_m + m^{-\gamma})^{-3/2}}{(1 + \beta_m N_1^{-1})^2}\} + o(1).$$

Since,

$$\text{V}[\frac{(2 + N_1^{-1}\beta_m)(p_m q_m + m^{-\gamma})^{-3/2}}{(1 + \beta_m N_1^{-1})^2}] = 0(1) and\ E_p[m(p_m - p)^2 - \text{pq}]^2 = O(1),$$

It follows that

$[m(p_m - p)^2\text{-pq}]\,\beta_m$ and $\frac{(2 + N_1^{-1}\beta_m)(p_m q_m + m^{-\gamma})^{-3/2}}{(1 + \beta_m N_1^{-1})^2}$ are asymptotically uncorrelated. Thus,

$$I_b = o(1)\,, \tag{2.3.19}$$

because $E_p\{[m(p_m - p)^2\text{-pq}]\,\beta_m|X_1, \ldots\ldots, X_m\} = 0$. Using (2.3.1), the first term in (2.3.18) becomes

$$I_a = \frac{A}{c(pq)^2}E_p\left[\frac{m^2(p_m - p)^2 - mpq}{N_1^{\,2}}\right]$$

$$= (pq)^{-2}E_p\{[m^2(p_m - p)^2 - mpq](p_m q_m + m^{-\gamma})\}$$

$$= m^2\,(pq)^{-2}E_p[(p_m q_m(p_m - p)^2] - m + 1. \tag{2.3.20}$$

Since $p_m q_m(p_m - p)^2$is a polynomial in p_m, it can be shown , by finding the moments of a binomial distribution with parameters m and p, that

$E_p[(p_m q_m(p_m - p)^2] = (pq)^2/m + (-7p^4 + 14p^3 - 8p^2 + p)/m^2 + (6p^4 - 12p^3 + 7p^2 - p)/m^3.$

This, together with (2.3.20), gives

$$I_a = [(pq)^{-1} - 6)](1 - m^{-1}) . \tag{2.3.21}$$

To evaluate the second term in (2.3.17), we write

$$\text{II} = \frac{A}{cpq} E_p\left(\frac{1}{N_1}\right) + \frac{A}{cpq} E_p\left[(N_1 + \beta_m)^{-1} - N_1^{-1}\right]. \tag{2.3.22}$$

An argument similar to (2.3.19) shows that

$$\frac{A}{cpq} E_p[(N_1 + \beta_m)^{-1} - N_1^{-1}] = -(pq)^{-1} E_p[\beta_m(p_m q_m + m^{-\gamma})(1 + \beta_m N_1^{-1})^{-1}]$$

$$= o(1) \tag{2.3.23}$$

It follows from (2.3.22),(2.3.23) and (2.3.15) with k=-1 that

$$\text{II} = \frac{A}{cpq} E_p\left(\frac{1}{N_1}\right) + o(1)$$

$$= n_0 (pq)^{-2} E_p\left[(p_m q_m + m^{-\gamma})^{1/2} I_{A_{m,\delta}}\right] + o(1)$$

$$= n_0 - (pq)^{-3/2}/16 + (pq)^{-3/2} m^{1-\gamma}/4 + o(1) . \tag{2.3.24}$$

Putting (2.3.16), (2.3.17),(2.3.18),(2.3.19), (2.3.21), (2.3.24) and (2.3.11) together, we get

$$E(L_M) - E(L_{n_0}) = c[\text{I} + \text{II} - n_0 + E_p(M) - n_0 + o(1)]$$

$$= c\left[(pq)^{-3/2}/8 + (pq)^{-1} - (pq)^{-1/2}/2 - 6\right] + o(c)$$

Thus, (2.3.12) follows and completes the proof.

Theorem 2.3.2. If $p = 1/2$ and (2.3.3) holds, then as $A/c \to \infty$ through multiples of 1/4, we have

1. For every positive integer k

$$E_p(M/m)^k = \begin{cases} 1 + k/2m + o(m^{-1}); if\ \gamma > 1 \\ 1 + km^{-1} \int_{|x|>2}^{2} \left(\frac{1}{2}x^2 - 2\right) \phi(x) dx + O m^{-1}; if\ \gamma = 1 \\ 1 + o(m^{-1})\ if\ 0 < \gamma < 1 \end{cases} \tag{2.3.25}$$

where $\phi(x) = \left(1/\sqrt{2\pi}\right) \exp\left(-x^2/2\right)$.

In particular

$$E_p(M) - m = \begin{cases} 1/2 + o(1); if\ \gamma > 1 \\ \int_{|x|>2}^{2} \left(\frac{1}{2}x^2 - 2\right) \phi(x)\, dx + O(1); if\ \gamma = 1 \\ O(1); if\ 0 < \gamma < 1 \end{cases} \tag{2.3.26}$$

and V(M)=O(m)

2. The regret of the two- stage procedure (2.3.2) is given by

$$\text{w} = \begin{cases} -2c + o(c); if\ \gamma > 1 \\ c \int_{|x|>2}^{2} (-5x^2 - x^4 - 4) \phi(x)\, dx + O(c): if\ \gamma = 1 \\ o(c); if 0 < \gamma < 1 \end{cases} \tag{2.3.27}$$

Proof : Let β_m be defined as in the proof of Theorem 2.3.1. Then , since m=$2(A/c)^{1/2}$ and $N_1 > m$ if and only if $|p_m - 1/2| > m^{-\gamma/2}$, so

$$E_p(M^k) = E_p\left(m^k I_{[N_1 \le m]}\right) + E_p\{[(N_1 + \beta_m)^k] I_{[N_1 > m]}\}$$

$$= m^k + E_p[N_1{}^k - m^k) I_{\beta_m}] + o(m^{k-1}), \tag{2.3.28}$$

where $\beta_m = \left\{\left(p_m - 1/2\right)^2 > m^{-\gamma}\right\}$. It follows from (2.3.1)and (2.3.3) and Taylor's theorem that

$$E_p[N_1{}^k - m^k) I_{\beta_m}]$$

$$= (A/c)^{k/2} \frac{-k}{2} (4^{-1})^{-(k+2)/2} E_p[(p_m q_m + m^{-\gamma} - 4^{-1}) I_{\beta_m}] + o(m^{k-1})$$

$$= -2k m^{k-\gamma} P(\beta_m) + (1/2) k m^{k-1} E_p\left[4m(p_m - 1/2)^2 I_{\beta_m}\right] + o(m^{k-1}). \tag{2.3.29}$$

Since p= ½, it follows from vonBahr's [7] extension of the central limit theorem and Markov's inequality that for j=0,1,2,....

$$E_p\left[4m(p_m - 1/2)^2 I_{\beta_m}\right] = \begin{cases} (2j)!\,(2^j j!) + o(1) & if\ \gamma > 1 \\ \int_{|x|>2}^{2} x^{2j} \phi(x) dx + o(1) & if\ \gamma = 1 \\ o\left(m^{-(1-\gamma)}\right) & if\ 0 < \gamma < 1 \end{cases} \tag{2.3.30}$$

Using (2.3.30) with j=0,1,2, (2.3.29) becomes

$$E_p[N_1{}^k - m^k) I_{\beta_m}]$$

$$= \begin{cases} (1/2) k m^{k-1} + o(m^{k-1}) & if\ \gamma > 1 \\ k m^{k-1} \int_{|x|>2}^{2} (\frac{1}{2} x^2 - 2) \phi(x) dx + o(m^{k-1}) & if\ \gamma = 1 \\ o(m^{k-1}) & if\ \ 0 < \gamma < 1 \end{cases} \tag{2.3.31}$$

Putting (2.3.28) and (2.3.31) together, we get (2.3.25).

For (2.3.27), using (2.3.16) with p=1/2,

$$E(L_M) - E\left(L_{n_0}\right) = c[4m^2 E_p\left(p_m - \tfrac{1}{2}\right)^2 - m + E_p(M) - m] \tag{2.3.32}$$

From the third inequality in (2.3.17)

$$4m^2 E_p\left(p_m - \tfrac{1}{2}\right)^2 = 4m^2 E_p\left[\frac{m^2 (p_m - \frac{1}{2})^2 - m/4}{M^2}\right] + m^2 E_p\left(\frac{1}{M}\right)$$

$$= \text{I+II, say.} \tag{2.3.33}$$

It follows from (2.3.22) and (2.3.1) that

$$\text{I} = 4m^2 E_p\left[\frac{m^2\left(p_m - \frac{1}{2}\right)^2 - \frac{m}{4}}{m^2}\right] + 4m^2 E_p\left\{\left[m^2\left(p_m - \tfrac{1}{2}\right)^2 - \frac{m}{4}\right](N_1^{-2} - m^{-2}) I_{\beta_m}\right\} + o(1)$$

$$= 4m^{1-\gamma} E_p\left[4m\left(p_m - \tfrac{1}{2}\right)^2 I_{\beta_m}\right] - 4m^{1-\gamma} P(\beta_m) - E_p\left[16m^2\left(p_m - \tfrac{1}{2}\right)^4 I_{\beta_m}\right] + o(1)$$

Now, using (2.3.30) with j=0,1,2, we get

$$I=\begin{cases} -2+o(1) & if\ \gamma>1 \\ \int_{|x|>2}^{2}(5x^2-x^4-4)\phi(x)\,dx+o(1) & if\ \gamma=1 \\ o(1) & if\ 0<\gamma<1 \end{cases} \quad (2.3.34)$$

To evaluate II, it follows from (2.3.26) and Taylor,s theorem that

$$II = m^2\{m^{-1}+E_p[(N_1^{-1}-m^{-1})I_{\beta_m}]\}+o(1)\}$$

$$= m+2m^{1-\gamma}P(\beta_m)-\frac{1}{2}E_p\left[4m\left(p_m-\frac{1}{2}\right)^2 I_{\beta_m}\right]+o(1),$$

which, together with (2.3.30) for j=0,1, gives

$$II=\begin{cases} m-\frac{1}{2}+o(1) & if\ \gamma>1 \\ m+\int_{|x|>2}^{2}(2-\frac{1}{2}x^2)\phi(x)\,dx+o(1) & if\ \gamma=1 \\ m+o(1) & if\ 0<\gamma<1 \end{cases} \quad (2.3.35)$$

Combining (2.3.32),(2.3.33), (2.3.34) ,(2.3.35) and (2.3.26), we get (2.3.27) and completing the proof.

IV. MODERATE SAMPLE SIZE PERFORMANCE

In this section we conducted a series of Monte Carlo trials to examine the moderate sample size performance of the two-stage procedure (2.3.2). To this end, we specified $n_0 = 25,50,100,150,200,250,,500,1000, p = 0.01, 0.02,0.05,0.1,0.15 \text{ and } 0.5$ under the loss function (2.2.1) with c=1, and then considered the values of $\gamma = 0.75,1,1.25,1.5,2$ for the two-stage procedure (2.3.2). note that for the given values of n_0, p and c, the weight A and the starting sample size m can be computed from (2.2.2) and (2.3.2). Simulation results are presented in Tables 1 and 2.

Each simulation based upon 10000 repetitions. Tables 1 and 2 display results for p =0.02,0.15 and of $\gamma = 0.75,1.0 \text{ and } 1.5$. for each row of the tables, we computed the mean $\bar{p}_M$ and the standard deviation $\hat{s}_{p_M}$ of the 10000 simulated values of p_M, the corresponding mean $\bar{M}$ and standard error $\hat{s}_M$ of $\bar{M}$, and the estimated regret $\bar{\omega}$ of ω. we also give the asymptotic values of $E_p(M), s_M = [V(M)]^{1/2}$ and ω obtained after omitting the remainder terms from (2.3.11) and (2.3.12).

Table 1

Moderate Sample Size Performance of Two- Stage Procedure for **Acute Gastro Enteritis** data with p=0.028
(Data Collected from Hospitals of Jammu District)

γ	n_0	m	$\bar{p}_M$	$\hat{s}_{p_M}$	$\bar{M}(E_p(M))$	$\hat{s}_M(s_M)$	$\bar{\omega}(\omega)$
.75	25	9	0.0280	0.0540	9.311 (-32.711)	0.463(26.014)	22.5(55.55)
	50	17	0.0281	0.0347	22.093 (-24.332)	2.346 (35.752)	28.16 (55.55)
	100	33	0.0281	0.0226	53.704(5.275)	6.646(49.813)	44.30 (55.55)
	150	50	0.0282	0.0173	90.253(40.666)	11.812(61.315)	38.25 (55.55)
	200	66	0.0279	0.0146	128.128(80.025)	17.324(70.446)	106.16 (55.55)
	250	83	0.0281	0.0130	168.349(120.666)	22.934(78.999)	95.55 (55.55)
	500	165	0.0280	0.0085	382.455(339.169)	50.715(111.385)	5.792 (55.55)
	1000	330	0.0280	0.0057	840.017(801.410)	95.727(157.523)	21.74 (55.55)
1.0	25	9	0.0270	0.0488	11.586(8.049)	1.483(26.014)	14.08 (55.55)
	50	17	0.0280	0.0306	30.022(33.049)	5.241(35.752)	25.00(55.55)
	100	33	0.0277	0.0195	75.080(83.049)	16.8351(49.813)	16.25 (55.55)
	150	50	0.0280	0.0151	124.445(133.049)	31.054(61.315)	22.58 (55.55)
	200	66	0.0278	0.0127	176.213(183.049)	45.851(70.446)	17.25 (55.55)
	250	83	0.0282	0.0111	228.048(233.049)	60.031(78.999)	0.96 (55.55)
	500	165	0.0279	0.0076	482.033(483.049)	109.896(111.385)	2.04 (55.55)
	1000	330	0.0279	0.0053	983.900(983.049)	165.069(157.523)	4.45 (55.55)
1.5	25	9	0.0281	0.0398	19.044(45.169)	4.432(26.014)	17.81 (55.55)
	50	17	0.0282	0.0248	53.862(75.225)	19.282(35.752)	28.16 (55.55)
	100	33	0.0280	0.0163	136.996(129.036)	73.261(49.813)	71.05 (55.55)
	150	50	0.0279	0.0310	215.940(180.855)	142.854(61.315)	1.36 (55.55)
	200	66	0.0280	0.0115	281.924(231.875)	209.066(70.44)	11.86 (55.55)
	250	83	0.0280	0.0102	340.829(282.617)	265.438(78.999)	33.09 (55.55)
	500	165	0.0279	0.0071	566.824(534.394)	339.308(111.385)	19.50 (55.55)
	1000	330	0.0279	0.0050	1041.395(1035.664)	228.698(157.523)	1.75 (55.55)

- Results of 10000 simulations with c = 1.
- Quantities in brackets are asymptotic values based on (2.2.9) and (2.2.10).

Table 2
Moderate Sample Size Performance of Two- Stage Procedure for **Anaemia Data** with p=0.0067
(Data Collected from Hospitals of Jammu District)

γ	n_0	m	$\bar{p}_M$	$\hat{s}_{p_M}$	$\bar{M}(E_p(M))$	$\hat{s}_M(s_M)$	$\bar{\omega}(\omega)$
.75	25	5	0.00628	0.0353	5(-324.34)	0(82.872)	80(368.36)
	50	9	0.00669	0.0268	9.27(-408.3)	0.446(111.18)	186.77(368.36)
	100	17	0.00692	0.0173	23.12(495.78)	1.436(152.80)	240.66 (368.36)
	150	25	0.0069	0.0131	39.8(-540.42)	2.604(185.30)	289.78 (368.36)
	200	33	0.0067	0.0107	58.4(-564.43)	4.17(212.90)	326.66 (368.36)
	250	41	0.0065	0.009	78.96(-575.9)	5.85(237.31)	570.06 (368.36)
	500	82	0.0066	0.0058	198.4(-546.4)	17.80(335.60)	536.35 (368.36)
	1000	164	0.0066	0.0036	491.9(-308.6)	474.62(368.36)	361.31 (368.36)
1.0	25	5	0.0068	0.0370	5.0(-96.24)	0(82.87)	80(368.36)
	50	9	0.0062	0.022	12.04(-71.24)	0.85(111.185)	120.33(368.36)
	100	17	0.0065	0.0139	32.63(-21.24)	3.011(152.809)	136.03 (368.36)
	150	25	0.0068	0.0108	58.35(28.75)	6.648(185.308)	129.85 (368.36)
	200	33	0.0066	0.0087	88.217(78.751)	11.295(212.903)	123.10 (368.36)
	250	41	0.0067	0.0075	120.857(128.751)	17.411(237.310)	110.76 (368.36)
	500	82	0.0067	0.0046	317.449(378.751)	61.396(335.60)	46.50 (368.36)
	1000	164	0.0066	0.00302	787.491(878.751)	194.186(474.62)	1.93 (368.36)
1.5	25	5	0.0067	0.031	6.75(158.29)	0.509(82.872)	46.28 (368.36)
	50	9	0.0065	0.018	20.565(235.736)	2.474(111.185)	35.63 (368.36)
	100	17	0.0068	0.0105	64.259(327.54)	11.702(152.809)	15.05 (368.36)
	150	25	0.0068	0.0077	124.78(397.132)	28.84(185.308)	1.44 (368.36)
	200	33	0.0067	0.0063	196.491(459.06)	56.012(212.903)	2.77 (368.36)
	250	41	0.0066	0.0053	278.63(517.313)	91.318(237.31)	277.81 (368.36)
	500	82	0.0066	0.0035	781.095(788.37)	388.54(335.60)	336.82 (368.36)
	1000	164	0.0066	0.0024	1812.053(1303.2)	1362.83(474.62)	100.74 (368.36)

- Results of 10000 simulations with c = 1.
- Quantities in brackets are asymptotic values based on (2.2.9) and (2.2.10).

The results shown in Tables 1 and 2 indicate that the two-stage procedure (2.3.2) tends to oversample if $\gamma \leq 1$and oversample if $\gamma > 1$. for p in the range 0.15 -0.5, varying γ from 0.75 to 1.5 leads to no appreciable change in risk. When p is reduced to 0.02, however γ = 0.75 and γ =1.5 shows a slightly increased risk when the optimal sample size, n_0, is less than or equal to 500. This is due to a tendency toward poor estimate for N_1 when m is relatively small. In this situation, $\gamma = 1$ is a reasonable choice for practical implementation.

V. CONCLUSION

The proposed two-stage methodology is better than the existing sequential or three stage procedures, especially when time and/or cost are important factor designs in point estimation of prevalence of diseases affecting human beings in our society besides in case fixed sample size procedures fail because of their dependence on nuisance parameters this method is the only way by which some type of robust estimates are possible .Moreover this proposed method can be applied widely in different areas globally to synthesise available information and provide better estimates of prevalence of diseases affecting human population.

REFERENCES

[1] [1] Robbins. H. and Siegmund,D. , Sequential estimation of *p* in Bernoulli trials . *In Studies in Probability and Statistics (ed. by E.J. Williams),103-107, Jerusalem Acad Press, Jerusalem, 1974.*

[2] Cabilio, P. and Robbins, H., Sequantial estimation of p with squared relative error loss. *Proc. Nat. Acad. Sci. U.S.A. 72, 1975,191-193*

[3] Cabilio.P., Sequential estimation in Bernoulli trials. *Ann. Statist., 5, 1977,342-356.*

[4] Baran, J and Magiera, R., Optimal Sequential Estimation procedure of a probability of success under LINEX loss. *Stat.Papers, 51, 2010, 511-529.*

[5] Liu, J F., A note on two-stage point estimation in bernoulitrials. *Sequential Analysis,19(4),2000, 161-176.*

[6] Woodroofe, M., Asymptotically sequential point estimation in three stages. *In New Persepectives in Theoretical and Applied Statistics (ed.by M. L Puri), 397-411, Wiley, New York,1975.*

[7] von Bhar B.,. Convergence of moments in Central Limit Theorem..*Ann.Math. Statist. 36, 1965, 808-818*