Kernel Sampling Based Parameter Estimation in Detected Community in Weighted Graph in Big Data

Ram Milan, Diwakar Shukla

•

Department of Computer Science and Applications, Dr. Harisingh Gour Vishwavidyalaya (A Central University), Sagar M.P[1, 2.]
rammilan.in@gmail.com[1], diwakarshukla@rediffmail.com[2]

Abstract

The social media platforms are such examples of big-data where the volume, velocity, and variety are visualized over time domain. Registered users of such platforms bear frequent communication with others and that could be identified as a community. Many methods (algorithms) exist in literature to detect such likely groups of frequent communication. This paper presents contribution to estimate parameters of detected communities using sampling procedure. A Kernel sampling procedure is suggested in the setup of detected community environment. A method is suggested whose efficiency has been estimated using calculations of confidence interval. Simulation procedure is used to obtain the lower and upper limits of confidence intervals with the help of multiple samples.

Keywords: Community Detection, Weighted Graph, Big Data, Internet Technology, 4G, Sampling, Simulations Confidence interval.

I. Introduction

With the expansion of social media platforms and technologies, large numbers of users are interacting with each other by forming groups, based on commonness of characters. Some most popular social networking sites are Face-book, Twitter, Instagram and Whatsapp etc. Where users register them self and communicate with the likeminded peoples. This motivates to think over for the identification of phenomena of community formation and community detection. The formation is usually on commonness but detection needs scientific methodologies.

One can assume that each registered user, on social networking platforms, is a vertex of a graph and his social communication with other people represents an edge of a graph. The quantum of connectivity with each other varies exponentially over time which generates voluminous data in a small span of time. The communication defers in modes like text, voice, image, videos, and many other similar which reveal variety in data. Moreover, in a fraction of time, growth of data on social networking platforms is immensely high which reveal velocity characteristics.

The community size and type detection is one such aspect which generates information in terms of popularity and security. Dongsheng Duan Li et al.[1] suggested algorithms for community mining assuming each user a vertex and density of connecting edges a community. An approach to community discovery based on evaluation of partition matrix has also been considered along with detection of change points. Pizzuti et al. [2] used Genetic algorithms approach for detecting communities in social media platform with mathematical approach using concept of graph theory. The Nan Du, et al. [3] detected community development in large scale social networks. An efficient approach based on faster algorithm for obtaining close community structure was suggested due to Newman et al.[4]. A community may be subdivided into small sub communities whose formation and analysis performed by Ferrara E. [5] .The graph theoretical application for community designing and analysis was attempted by Fortunato S. [6].

Communication at the social networking platform when become highly frequent, close and intense then it reaches up to sentimental level. Deitrick et al. [7] suggested sentiment analysis approach on data obtained through social media platform. Leskovec et al.[8] considered several algorithms for network community detection. A methodological survey based contributions over community detection procedures are due to Plantie et al. [9] and Uthayasankar et al. [10]. This paper focuses on developing parameter estimation approach as a posterior application to the detected community.

II. Graph Based Rules

The methodology of community detection targets to the detection of groups of vertices within which connections are dense. Consider a graph G which is set of vertices V (G), and set of Edges E (G). One can construct rules for cliques and kernel formation based on collection of vertices and corresponding edges as under.

III. Community detection in weighted graph

The clique is referring to a kind of cohesive sub structure whose maxima provide a tool for community detection. The overlapping maximal clique is kernel. In view to N.Du, et al.[3] some of rule are as under:

Rule 1. $S \subseteq V(G)$, $\forall u,v \varepsilon S, u \neq v$, such that $(u,v) \varepsilon E$, then S is a clique in G. if any other S′ is a clique and $S' \supseteq S$ iff $S' = S$, S is a maximal clique of G.

Rule 2. For a given vertex v, $N(v) = \{u | (v,u) \varepsilon E(G)\}$, we call N (v) is the set of all neighbors of v. Given set $S \subseteq V(G)$, $N|_S = \cup N(v_i) - S$, $V_i \varepsilon S$, $N|_S$ is the set of all neighbors of S.

Rule 3. Let Com (G) be the set of all components in G. the giant component is denoted by C_g and M (C_g) is the set of all the maximal cliques of C_g. We use $V_m \subseteq V(G)$ to represent the set of all vertices covered by M (C_g).

Rule 4. Let P_0, P_1,--,-- P_{n-1} be the sub graph of G such that $\forall P_i, P_j$, $V(p_i) \cap V(P_j) = \phi$, and $V(P_0) \cup$,---,$V(P_{n-1}) = V(G)$. For any pair of P_i and P_j, if $|E(P_i)| > |(N| P_i \cap P_j)|$, P_i is defined as a community of G.

Rule 5. Given vertex $v_i \varepsilon V_m$, define $C_i = \{S | S \varepsilon M(C_g), V_i \varepsilon S\}$ to be the set of all maximal cliques containing V_i and C the set of all C_i's. $\forall C_i, C_j \varepsilon C$, if $\frac{|Ci \cap cj|}{cj} >= f$ which is a threshold to describe the extent to which C_i overlaps with C_j, we call C_j is contained in C_i, denoted by $C_j < C_i$. If c_i is not contained by any other element in C, C_i is called the kernel of G and Vi is the center of C_i.

Rule 6. Let K be the set of all kernels in G. $V_k = \{ V_i | V_i \varepsilon K, K_j \varepsilon K \}$ is the set of all vertices covered by K. and $I_k = \cup (K_i \cap k_j)$, $k_i, k_j \varepsilon K$, $i \neq j$ is the union of all the vertices that any pair of element in K has in common.

IV. Problem undertaken

Assume, using any of existing algorithms several communities have been detected. One may be interested to estimate unknown parameter of characteristics associated with edge between any pair of vertices, within the community formed in graphical population structure of a social media platform in the setup of big data. For example, large numbers of registered users are on social networking platform then the average time consumed between any pair of users within a community is a problem to work out. Being a large data setup, growing fast over time and space, the estimation of such is time and cost consuming. This paper considers a solution approach for a problem described herein using sampling procedure.

V.A Graphical Structure:

Assume a fig. 1 where enumeration of cliques is taken into consideration. Among constituted cliques, there exist maximal clique which is a complete sub graph which can represent closed relationship for single entity in a given network.

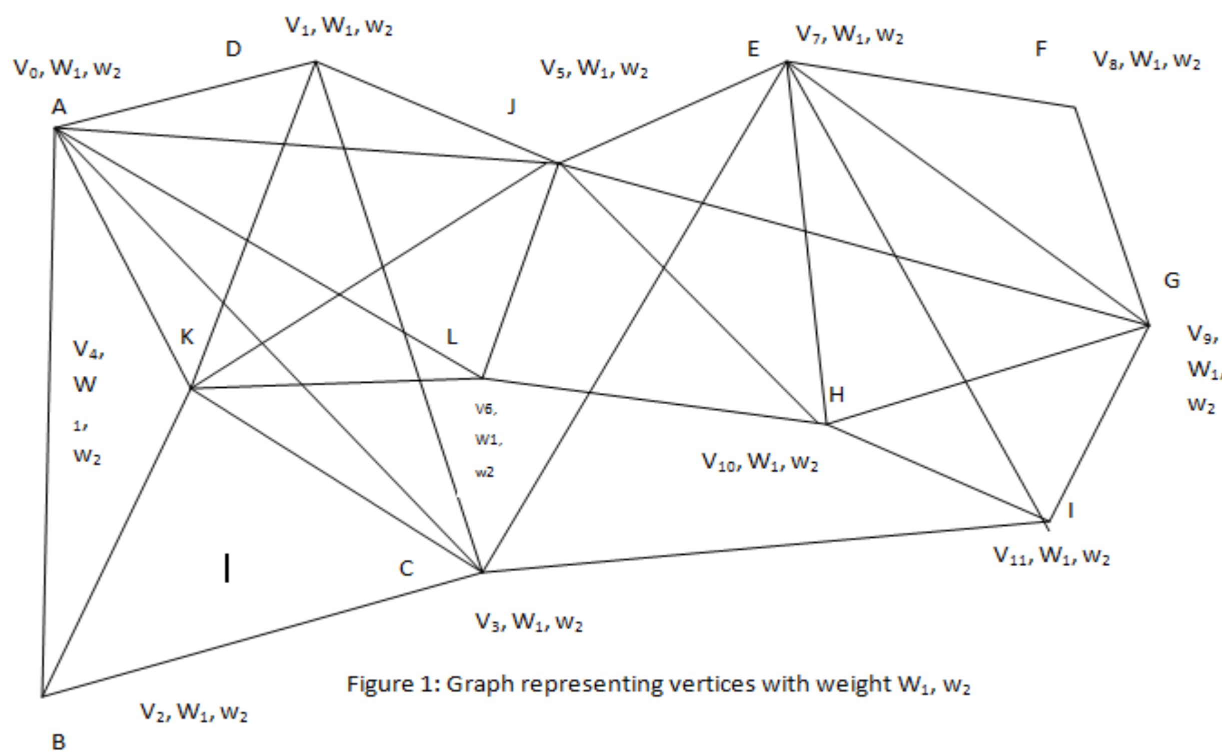

Figure 1: Graph representing vertices with weight W_1, w_2

For enumerate the cliques of a graph using rules 1-5: one can get:

$\mathbf{C_0}$ = {(V_0, W_1,w_2), (V_1,W_1,w_2), (V_4,W_1,w_2), (V_5,W_1,w_2)}, {(V_0, W_1,w_2), (V_1,W_1,w_2), (V_3,W_1,w_2), (V_4,W_1,w_2)}, {(V_0, W_1,w_2), (V_2,W_1,w_2), (V_3,W_1,w_2), (V_4,W_1,w_2)}, {(V_0, W_1,w_2), (V_6,W_1,w_2), (V_4,W_1,w_2), (V_5,W_1,w_2)} V_0 being as the center.

$\mathbf{C_1}$={(V_0, W_1,w_2), (V_1,W_1,w_2), (V_4,W_1,w_2), (V_5,W_1,w_2)}, {(V_0, W_1,w_2), (V_1,W_1,w_2), (V_3, W_1, w_2),(V_4,W_1,w_2)}

$\mathbf{C_2}$ = {(V_0, W_1, w_2), (V_2, W_1, w_2), (V_4, W_1, w_2), (V_3, W_1, w_2)}

$\mathbf{C_3}$ = {(V_0, W_1, w_2), (V_2, W_1, w_2), (V_4, W_1, w_2), (V_3, W_1, w_2)},

$\mathbf{C_4}$={(V_0, W_1,w_2), (V_1,W_1,w_2), (V_4,W_1,w_2), (V_6,W_1,w_2)}, {(V_0, W_1,w_2), (V_2,W_1,w_2), (V_3,W_1,w_2), (V_4, W_1, w_2)}, {(V_1, W_1, w_2), (V_4, W_1, w_2), (V_5, W_1, w_2), (V_6, W_1, w_2)}

$\mathbf{C_5}$= {(V_5, W_1, w_2), (V_1, W_1, w_2), (V_4, W_1, w_2), (V_6, W_1, w_2)}

$\mathbf{C_6}$= {(V_5, W_1, w_2), (V_1, W_1, w_2), (V_4, W_1, w_2), (V_6, W_1, w_2)}

$\mathbf{C_7}$={(V_7, W_1,w_2), (V_8,W_1,w_2), (V_9,W_1,w_2), (V_{10},W_1,w_2)}, {(V_7, W_1,w_2), (V_9,W_1,w_2), (V_{11},W_1,w_2), (V_{10}, W_1, w_2)},

$\mathbf{C_8}$ = {(V_7, W_1, w_2), (V_8, W_1, w_2), (V_9, W_1, w_2), (V_{10}, W_1, w_2)}

$\mathbf{C_9}$= {(V_7, W_1, w_2), (V_8, W_1, w_2), (V_9, W_1, w_2), (V_{10}, W_1, w_2)}

$\mathbf{C_{10}}$= {(V_7, W_1, w_2), (V_8, W_1, w_2), (V_9, W_1, w_2), (V_{10}, W_1, w_2)}

$\mathbf{C_{11}}$= {(V_7, W_1, w_2), (V_8, W_1, w_2), (V_9, W_1, w_2), (V_{10}, W_1, w_2)}

C_8, C_9 $C_{10,}$ C_{11} are contained by C_7.

Therefore C_0 and C_7 are two different kernels respectively with weight associated with vertices.

VI. Parameter estimation

Consider the following graph in figure 3 where first weight the age of the users registered in the social networking sites and the other weight is the number of hours of the social networking sites used. In figure 2, social media communities detected through algorithms and unknown parameters existence are given from which one can extract sample based implementation.

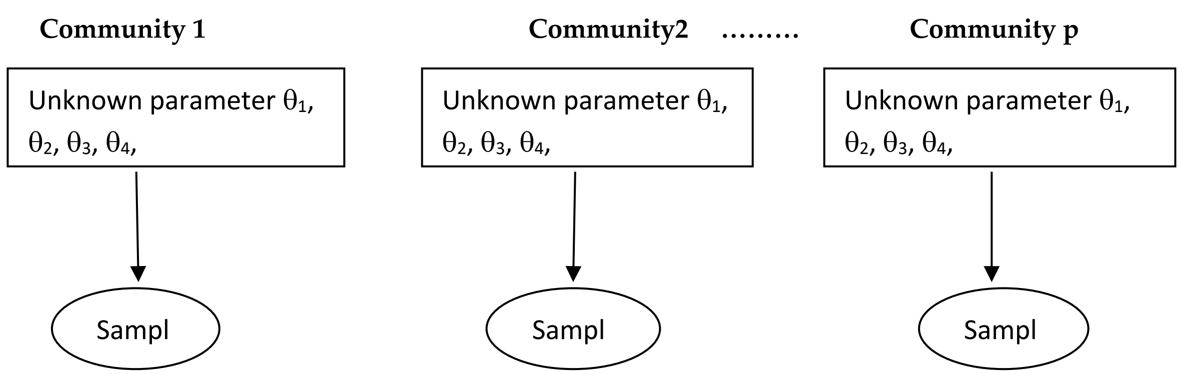

Figure 2: Social media communities & unknown parameters

Consider the graph as population having kernel based k groups classification likes below:-

Table 1: Kernel based groups

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	…..K^{th}
K_{e1}	K_{e2}	K_{e3}	K_{e4}	K_{e5}	K_{e6}	K_{e7}	K_{e8}	K_{e9}	K_{e10}	K_{e11}	K_{e12}	…… K_{en}

VII. Kernel Sampling:

One can consider the graphical population of vertices (node) and edges G= (V, E) divided into k Kernel based groups, derived from given a graphical population (see table 1). This constitutes setup of Kernel Sampling. Assume the strata sizes are N_1, N_2, N_3…………N_k such that $\sum_{i=1}^{k} N_i = N$
Let the total size of population is N from which a sample of size of population n (n< N) is drawn which is divided into Kernel based group wise as n_1, n_2, n_3,….n_k.. Such that $\sum_{i=1}^{k} n_i$=n. Let the sample means are $\bar{m}_1$,$\bar{m}_2$,$\bar{m}_3$,…,…. $\bar{m}_k$of the k strata respectively P.V. Sukhatme[11] and Cochran [12].
Consider vertices of graph G= (V, E) having two variables W_2: number of hours the user is consuming social media website is used in a month (auxiliary variable) and W_1: the age of user (in completer years) as main variable. The unknown parameter is average number of hours consumed by a user W_2. It may assume that mean age of users $\bar{W}_1$ in population is known (due to registration data while creating account on social networking sites). The i[th] Kernel based group has size N_i and pair of values (W_{1ij}, W_{2ij}) where W_{1ij}, W_{2ij} are j[th] value i[th] Kernel based group relating to number of hours consumed by users and ages fo users.

$\bar{W}_1$=$\frac{1}{N}$=$\sum_{i=1}^{K}\sum_{I=1}^{NI} W_{1ij}$ (Known parameter) (4.1)

$\bar{W}_2$=$\frac{1}{N}$=$\sum_{i=1}^{K}\sum_{j=1}^{NI} W_{2ij}$ (Unknown parameter and to be estimated) (4.2)

Moreover some other symbols are as under:

$\bar{W}_{1i}$: Population mean of i[th] strata of variable W_1

$\bar{W}_{2i}$: Population mean of i[th] strata of variable W_2

Estimation method under Kernel Sampling:

To estimate unknown$\bar{W}_2$, the random samples of sizes n_i are drawn from i[th] group N_i paired values (w_{1ij} , w_{2ij}) such that

$$\bar{w}_{1i} = \frac{1}{n_i}\sum_{j=1}^{ni} w_{1ij} \quad (4.3)$$

$$\bar{w}_{2i} = \frac{1}{n_i}\sum_{j=1}^{ni} w_{2ij} \quad (4.4)$$

and (w_{1ij} , w_{2ij}) are pair of sample observations from i[th] group

Method to use for estimation of $\bar{W}_2$is

M = $\sum_{i=1}^{k} \phi(z_i, \acute{z}_i)\bar{W}_{2i}$, where ϕ(z_i,z'_i)= (z_i.z'_i) and $z_i = \bar{w}_{1i}$, $\acute{z}_i = \frac{1}{\bar{W}_{2i}}$ and $\bar{W}_{2i}$ assumed known. (4.5)

The Mean Square Error of method M is

$$\text{MSE (M)}= \sum_{i=1}^{k} Z_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right)(S_i^*)^2 \quad (4.6)$$

R=$\bar{W}_1/\bar{W}_2$; Z$_i$=$\frac{N_i}{N}$; (4.7)

Where $(S_i^*)^2 = [S_{iW1}^2 + R^2 S_{iW2}^2 - 2RS_{iw1w2}]$

$$S_{iW1}^2 = \frac{1}{N_i - 1}\sum_{i=1}^{K}\left(W_{1ij} - \bar{W}_{1i}\right)^2, \qquad S_{iW2}^2 = \frac{1}{N_i - 1}\sum_{i=1}^{K}\left(W_{2ij} - \bar{W}_{i2}\right)^2; \quad (4.8)$$

$$\bar{W}_{1i} = \frac{1}{N_1}\sum_{j=1}^{N1} W_{1ij}\ ;\bar{W}_{2i} = \frac{1}{N_2}\sum_{j=1}^{N2} W_{2ij};\ S_{iw1w2} = \frac{1}{N_{i-1}}\ \sum_{i=1}^{K}\left(W_{1ij} - \bar{W}_{1i}\right).\left(W_{2ij} - \bar{W}_{i2}\right) \quad (4.9)$$

The estimate of $(S_i^*)^2$is est(MSE) = $\sum_{i=1}^{k} Z_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)(S_i^*)^2$

Where $(S_i^*)^2 = [S_{iW1}^2 + R2S_{iW2}^2$-2RSiw1w2] , the S_{iW1}^2, S_{iW2}^2, Siw1w2 are estimated from sample and r=$\frac{\bar{w}_1}{\bar{w}_2}$ exist in sample.

The 95% Confidence interval for estimating $\bar{W}_1$ is:

P [M- 1.96 $\sqrt{MSE(M)}$< M+1.96 $\sqrt{MSE(M)}$]=0.95 (4.10)

VIII. Simulation procedure for confidence interval

Step I: Draw a random sample of size n

Step II: Compute the lower limit and upper limit of confidence interval

Step III: Repeat step I and II for k times (K=200)

Step IV: Compute the less than type and more than type cumulative frequency over all k samples for lover limit and upper limit of confidence interval.

Step V: Plot data of step IV on graph. The perpendicular from point of intersection on the x-axis is the simulated value of lover limit and upper limit of confidence interval for parameter to be estimated.

IX. Numerical illustration:

Consider figure 2 having 11 vertices and consisting of data in the tuple (Vi, W_{1i}, W_{2i}**)**. The relationship of vertices is in the form of edges which is used to constitute form clique and kernel.

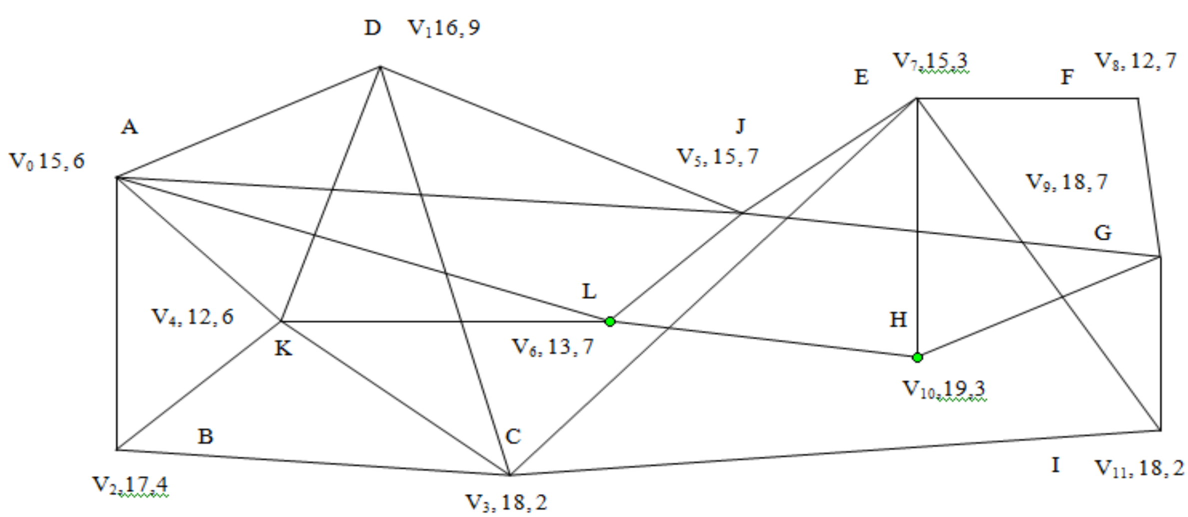

Figure 3: Graph with weight representing Age and hours of use.

The figure has 2 kernels C_0 and C_7.The Kernel constituted based group structure of graphical population is as under. From figure 4 we are extracting samples from group 1(C_0) and group 2(C_7).

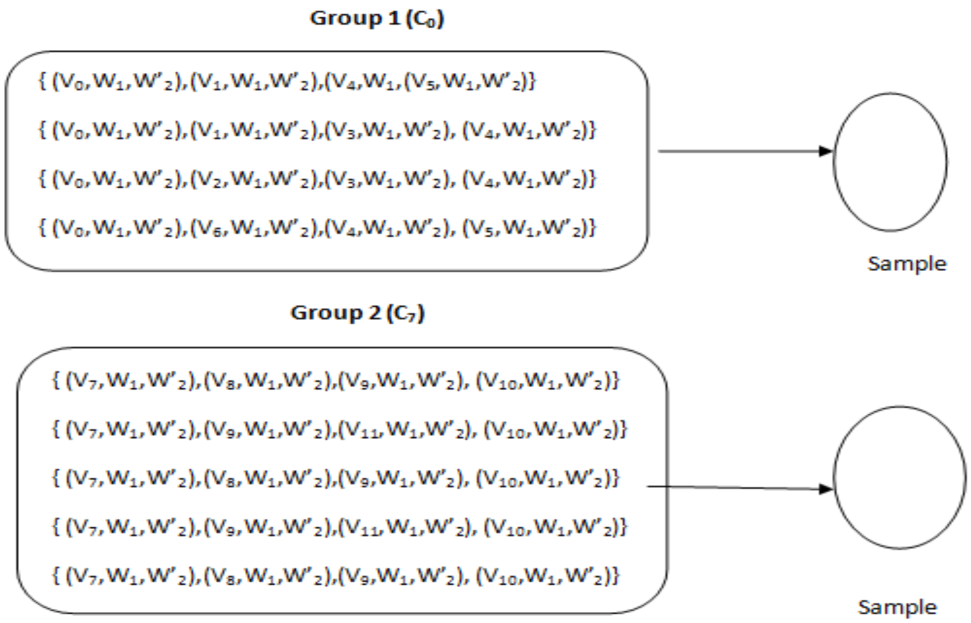

Figure 4: Representing Kernels and Samples

As per figure 3, the representation of the vertices with weight (W1: ages of users) and (W2: time consumed by users) are given below in terms of V_i=(W_{1i} ,W_{2i})

V_0=(15,6) , V_1=(16,9), V_2=(17,4) V_3=(18,2) V_4=(12,6) V_5=(15,7)

V_6=(13,7), V_7=(15,3), V_8=(12,7), V_9=(18,7) V_{10}=(19,3), V_{11}=(18,2)

The group 1 Kernels contains 16 tuple (N_1=16) and group 2 contains 20 tuple (N_2=20).

A random sample of size n_1=6 is drawn from N_1=16. Similarly, random sample of size n_2 is drawn from N_2=20 (n1< N_1, n_2 < N_2). Using these sample values, the objective is to estimate unknown population mean$\overline{W}_1$.

Table 2: Description of population parameters

	Group size N_i	Group mean W_1	Group mean W_2
Group I	N_1=16 Z_1=N_1/N =0.44	$\overline{W}_{1G1}$= 14.75 S^2_{W1G1}=4.2	$\overline{W}_{2G1}$=5.93 S^2_{W2G1}=3.02
Group II	N_2=20 Z_2=N_2/N =0.55	$\overline{W}_{1G2}$=16.6 S^2_{W1G2}=6.25	$\overline{W}_{2G2}$=4.5 S^2_{W2G2}=4.47
	N=N_1+N_2=36	R= $\frac{\overline{W}_1}{\overline{W}_2}$=14.75/16.6=0.88	

Table 3: Sample based computation (First Sample)

	Sample size	Sample values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_0,15,6),(V_1,16,9),(V_4,12,6),(V_5,15,7),(V_3,18,2),(V_6,13,7)	$\overline{w}_{1G1}$=14.83 S^2_{W1G1} = 4.51	$\overline{w}_{2G1}$=6.16 S^2_{W2G1}=5.47	r_1=2.4	$(S_1^*)^2$=12.92	[8.70-19.67]
Group II	n_2=4	(V_7,15,3),(V_8,12,7) (V_9,18,7),(V_{10},19,3)	$\overline{w}_{1G2}$=16.0 S^2_{W1G2} = 10	$\overline{w}_{2G2}$=5.0 S^2_{W2G2} = 5.33	r_2=3.2	$(S_2^*)^2$=131.2	
M=14.19			Est.(MSE)=7.89				

Table 4: Sample based computation (Second sample)

	Sample size	Sample values(V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est. .$(S_i^*)^2$	
Group I	n_1=6	(V_0,15,6),(V_4,12,6), (V_5,15,7) (V_3,18,2) ,(V_6,13,7), (V_1,16,9)	$\overline{w}_{1G1}$=14.83 S^2_{W1G1}=4.51	$\overline{w}_{2G1}$=6.16 S^2_{W2G1}=5.47	r_1=2.40	$(S_1^*)^2$=12.82	[10.98-17.98]
Group II	n_2=4	(V_7,15,3) ,(V_{11},18,2) (V_8,12,7), (V_9,18,7)	$\overline{w}_{1G2}$=15.75 S^2_{W1G2}=8.25	$\overline{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.31	$(S_2^*)^2$=52.31	
M=14.48			Est.(MSE)=3.37				

Table 5: Sample based computation (Third sample)

	Sample size	Sample Values(V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_3,18,2) (V_4,12,6) (V_5,15,7) (V_6,13,7) (V_1,16,9) (V_0,15,6)	$\overline{w}_{1G1}$=14.83 S^2_{W1G1} = 4.51	$\overline{w}_{2G1}$=6.16 S^2_{W2G1}=5.47	r_1=2.40	$(S_1^*)^2$=12.77	[13.19-22.47]
Group II	n_2=4	(V_7,15,3) (V_{11},18,2) (V_9,18,7) (V_{10},19,3)	$\overline{w}_{1G2}$=17.5 S^2_{W1G2}=3.0	$\overline{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.66	$(S_1^*)^2$=52.57	
M=17.831			Est.(MSE)=5.56				

Table 6: Sample based computation (Fourth Sample)

	Sample size	Sample Values(V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_5,15,7) (V_4,12,6) (V_6,13,7) (V_0,15,6) (V_1,16,9) (V_2,17,4)	$\bar{w}_{1G1}$=14.66 S^2_{W1G1} = 3.46	$\bar{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=22.75	[8.74-18.84]
Group II	n_2=4	(V_{10},19,3) (V_9,18,7) (V_8,12,7) (V_7,15,3)	$\bar{w}_{1G2}$=16 S^2_{W1G2} = 10	$\bar{w}_{2G2}$=5 S^2_{W2G2}5.33	r_2=3.2	$(S_1^*)^2$=103.9	
M=13.79			Est.(MSE)=6.66				

Table 7: Sample based computation (Fifth Sample)

			Mean				95% C.I.
	Sample Size	Sample Values(V_0,w_1,w_2)	w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6), (V_3,18,2), (V_6,13,7) (V_2,17,4) (V_1,16,9) (V_0,15,6)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}5.86	r_1=2.67	$(S_1^*)^2$=69.31	[11.48-19.9]
Group II	n_2=4	(V_8,12,7) ,(V_9,18,7) (V_{10},19,3) ,(V_{11},18,2)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$=55.57	
M=15.69			Est.(MSE)=4.64				

Table 8: Sample based computation (Sixth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6) (V_0,15,6) (V_2,17,4) (V_5,15,7) (V_1,16,9) (V_3,18,2)	$\bar{w}_{1G1}$=15.5 S^2_{W1G1}=4.3	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.73	$(S_1^*)^2$=49.05	[7.91-23.81]
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$=260.11	
M=15.86			Est.(MSE)= 16.52				

Table 9: Sample based computation (Seventh Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_0,15,6) (V_1,16,9), (V_4,12,6) (V_5,15,7) (V_2,17,4) (V_3,18,2)	$\bar{w}_{1G1}$=15.5 S^2_{W1G1}=4.3	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.73	$(S_1^*)^2$=49.43	[9.48-21.46]
Group II	n_2=4	(V_7,15,3) (V_8,12,7) (V_9,18,7) (V_{10},19,3)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=5.41	r_2=3.36	$(S_1^*)^2$=140.67	
M=15.47			Est.(MSE)= 9.37				

Table 10: Sample based computation (Eighth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_3,18,2) (V_4,12,6), (V_5,15,7) (V_6,13,7) (V_2,17,4) (V_1,16,9)	$\bar{w}_{1G1}$=14.83 S^2_{W1G1}=5.44	$\bar{w}_{2G1}$=5.83 S^2_{W2G1}=6.15	r_1=2.54	$(S_1^*)^2$=55.55	[10.1,19.54]
Group II	n_2=4	V_7,15,3) (V_8,12,7) (V_9,18,7) (V_{11},18,2)	$\bar{w}_{1G2}$=15.75 S^2_{W1G2}=8.25	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.31	$(S_1^*)^2$=79.59	
M=14.82			Est.(MSE)=5.82				

Table 11: Sample based computation (Ninth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6) (V_3,18,2), (V_6,13,7) (V_0,15,6) (V_2,17,4) (V_1,16,9)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.67	$(S_1^*)^2$=59.16	[8.93-21.59]
Group II	n_2=4	(V_8,12,7) (V_{10},19,3), (V_9,18,7) (V_{11},18,2)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=5.00 S^2_{W2G2}=8	r_1=3.35	$(S_1^*)^2$=155.78	
M=15.26				Est(MSE)=10.45			

Table 12: Sample based computation (Tenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_6,13,7) (V_5,15,7), (V_4,12,6) (V_0,15,6) (V_2,17,4) (V_1,16,9)	$\overline{w}_{1G1}$=14.66 S^2_{W1G1}=3.46	$\overline{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=21.55	[8.86-23.98]
Group II	n_2=4	(V_7,15,3) (V_{10},19,3) (V_{11},18,2) (V_8,12,7)	$\overline{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\overline{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.26	$(S_1^*)^2$=242.52	
M=16.42			Est.(MSE)=14.95				

Table 13: Sample based computation (Eleventh Sample)

		Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_6,13,7) (V_5,15,7), (V_4,12,6) (V_3,18,2) (V_2,17,4) (V_1,16,9)	$\overline{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\overline{w}_{2G1}$=5.83 S^2_{W2G1}=6.15	r_1=2.60	$(S_1^*)^2$ =56.98	[8.87-22.11]
Group II	n_2=4	(V_8,12,7) (V_{10},19,3) (V_{11},18,2) (V_9,18,7)	$\overline{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\overline{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$ =172.80	
M=15.49			Est.(MSE)=11.43				

Table 14: Sample based computation (Twelfth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2		est .$(S_i^*)^2$	
Group I	n_1=6	(V_2,17,4) (V_1,16,9), (V_0,15,6) (V_4,12,6) (V_5,15,7) (V_6,13,7)	$\overline{w}_{1G1}$=14.66 S^2_{W1G1}=3.46	$\overline{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=22.76	[8.19-19.39]
Group II	n_2=4	(V_7,15,3) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\overline{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\overline{w}_{2G2}$=5.00 S^2_{W2G2}=5.33	r_2=3.2	$(S_1^*)^2$=129.42	
M=13.79				Est.(MSE)=8.19			

Table 15: Sample based computation (Thirteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_2,17,4) (V_1,16,9), (V_3,18,2) (V_5,15,7) (V_6,13,7) (V_0,15,6)	$\bar{w}_{1G1}$=15.66 S^2_{W1G1}=3.06	$\bar{w}_{2G1}$=5.83 S^2_{W2G1}=6.15	r_1=2.68	$(S_1^*)^2$= 39.85	[9.23-22.19]
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$= 172.80	
M=15.71			Est.(MSE=11.10				

Table 16: Sample based computation (Fourteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_5,15,7) (V_4,12,6), (V_6,13,7) (V_3,18,2) (V_2,17,4) (V_1,16,9)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.83 S^2_{W2G1}=6.15	r_1=2.60	$(S_1^*)^2$=56.98	[9.82-24.82]
Group II	n_2=4	(V_{10},19,3) (V_{11},18,2) (V_7,15,3) (V_8,12,7)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.26	$(S_1^*)^2$=229.74	
M=17.32			Est.(MSE)=14.85				

Table 17: Sample based computation (Fifteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_1,16,9) (V_0,15,6), (V_4,12,6) (V_5,15,7) (V_2,17,4) (V_3,18,2)	$\bar{w}_{1G1}$=15.5 S^2_{W1G1}=4.3	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.73	$(S_1^*)^2$ =49.43	[9.27-22.43]
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$ =172.8	
M=15.85			Est.(MSE)=11.29				

Table 18: Sample based computation (Sixteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	$(V_6,13,7)$ $(V_5,15,7)$, $(V_4,12,6)$ $(V_0,15,6)$ $(V_2,17,4)$ $(V_3,18,2)$	$\bar{w}_{1G1}$=15.0 S^2_{W1G1}=5.2	$\bar{w}_{2G1}$=5.33 S^2_{W2G1}=3.85	r_1=2.81	$(S_1^*)^2$=66.45	[14.31-23.43]
Group II	n_2=4	$(V_{11},18,2)$ $(V_{10},19,3)$ $(V_9,18,7)$ $(V_7,15,3)$	$\bar{w}_{1G2}$=17.50 S^2_{W1G2}=3.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.66	$(S_1^*)^2$=70.39	
M=18.87			Est.(MSE)=5.46				

Table 19: Sample based computation (Seventeenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2		r_i	
Group I	n_1=6	$(V_5,15,6)$ $(V_4,12,6)$, $(V_6,13,7)$ $(V_2,17,4)$ $(V_1,16,9)$ $(V_0,15,6)$	$\bar{w}_{1G1}$=14.66 S^2_{W1G1}=3.46	$\bar{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=26.08	[8.84-24]
Group II	n_2=4	$(V_{11},18,2)$ $(V_{10},19,3)$ $(V_7,15,3)$ $(V_8,12,7)$	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.26	$(S_1^*)^2$=242.52	
M=16.42			Est.(MSE)= 15.04				

Table 20: Sample based computation (Eighteenth Sample)

	Sampl e size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	$(V_4,12,6)$ $(V_3,18,2)$, $(V_2,17,4)$ $(V_1,16,9)$ $(V_0,15,6)$ $(V_5,15,7)$	$\bar{w}_{1G1}$=15.5 S^2_{W1G1}=4.3	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.73	$(S_1^*)^2$=-0.476	[9.35-22.35
Group II	n_2=4	$(V_{11},18,2)$ $(V_{10},19,3)$ $(V_9,18,7)$ $(V_8,12,7)$	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$=172.80	
M=15.85			Est.(MSE)=10.37				

Table 21: Sample based computation (Nineteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6) (V_3,18,2), (V_2,17,4) (V_1,16,9) (V_0,15,6) (V_6,13,7)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.67	$(S_1^*)^2$=60.92	[9.05-20.73]
Group II	n_2=4	(V_7,15,3) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=5.00 S^2_{W2G2}=5.33	r_2=3.2	$(S_1^*)^2$=129.42	
M=14.89			Est.(MSE)=8.90				

Table 22: Sample based computation (Twenty Samples)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_2,17,4) (V_1,16,9), (V_4,12,6) (V_5,15,7) (V_6,13,7) (V_0,15,6)	$\bar{w}_{1G1}$=14.66 S^2_{W1G1}=3.46	$\bar{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=22.96	[13.4-21.42]
Group II	n_2=4	(V_9,18,7) (V_7,15,3) (V_{10},19,3) (V_{11},18,2)	$\bar{w}_{1G2}$=17.50 S^2_{W1G2}=3.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.9	r_2=4.66	$(S_1^*)^2$=63.47	
M=17.41			Est.(MSE)=4.23				

Table 23: For Confidence interval calculations

For lower limit of Confidence Interval				For upper limit of Confidence Interval			
Class Interval	Probability over 200 samples	LTT	MTT	Class Interval	Probability over 200 samples	LTT	MTT
Below 8.0	0.08	0.08	1.00	Below 17.0	0.05	0.05	1.00
8.0-9.0	0.38	0.46	0.92	17.0-18.0	0.13	0.18	0.95
9.0-10.0	0.29	0.75	0.54	18.0-19.0	0.16	0.34	0.82
10.0-11.0	0.14	0.89	0.25	19.0-20.0	0.18	0.52	0.66
11.0-12.0	0.04	0.93	0.11	20.0-21.0	0.23	0.75	0.48
12.0-13.0	0.05	0.98	0.07	21.0-22.0	0.14	0.89	0.25
13.0-14.0	0.01	0.99	0.02	22.0-23.0	0.09	0.98	0.11
Above 14.0	0.01	1	0.01	23.0-24.0	0.01	0.99	0.02
LTT: Less Than Type; MTT: More Than Type				Above 24.0	0.01	1.00	0.01

Probability = $\left(f_i / \sum f_i\right)$; f_i : frequency of i^{th} class interval

$\sum f_i$: total frequency; P[A]= probability of event A.

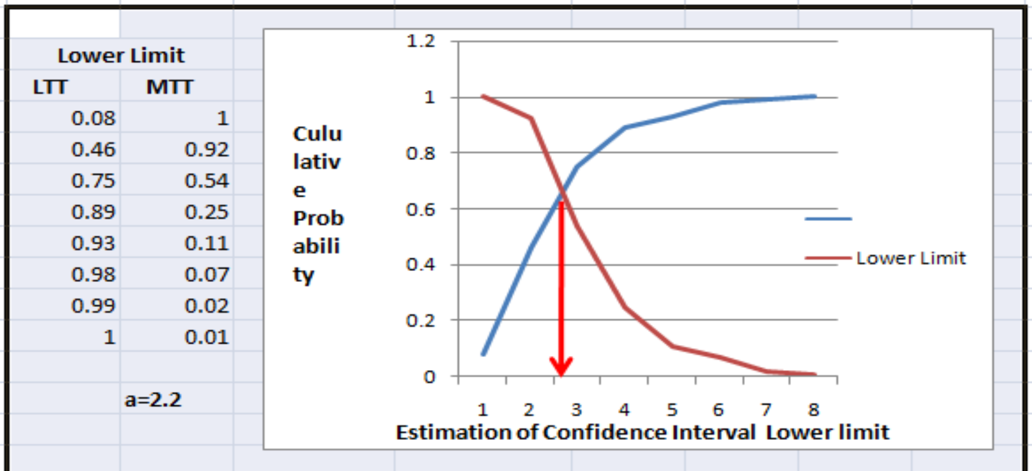

Figure 4: Intersecting point of LTT & MTT for Lower limt

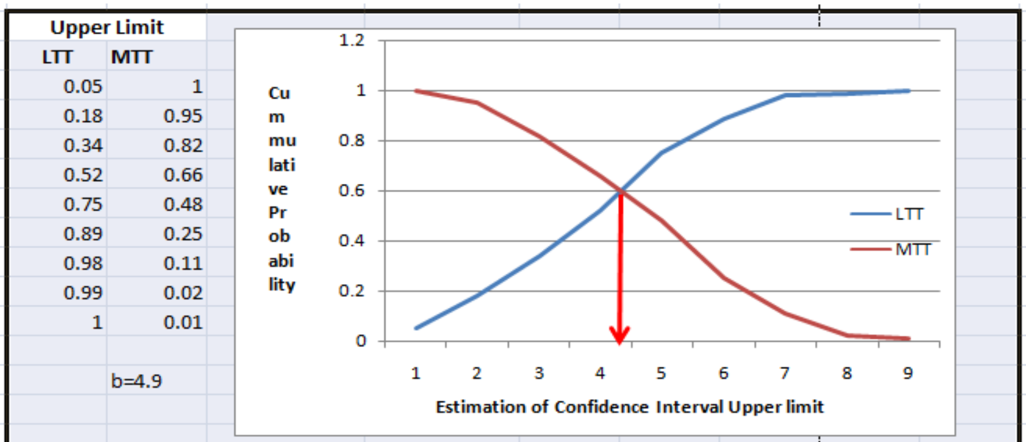

Figure5 : Intersecting point of LTT & MTT for Upper limt

a=2.2, b=4.9

Confidence Interval = P[a<2.3<b]=0.95 ; where P[A] is probability of event A.

Other Computations: - $(S_1^*)^2$ =17.64, $(S_2^*)^2$= 39.08

MSE (M) = $Z_1^2 \left(\frac{1}{n_1} - \frac{1}{N_1}\right)(S_1^*)^2 + Z_2^2 \left(\frac{1}{n_2} - \frac{1}{N_2}\right)(S_2^*)^2 = 4.70$

X. Conclusion

In this paper, a graphical structure of population has been considered and using the Kernel creation procedure rules and closed communities have been detected. The closeness is based on criteria of click formation. In order to estimate the unknown population parameter (average hours used) a scheme named after as Kernel Sampling estimation method is used. The 95% confidence intervals have been computed. It has been found that 95% confidence intervals are catching the true values. The simulation procedure suggested herein provides the well predicted estimated interval. This contribution opens up avenues and opportunities to think for mixing of community detection and parameter estimation.

References:

[1]. Dongsheng Duan YUhua Li et al. "Community Mining on Dynamic Weighted Directed Graphs", CNIKM 09 November 6, 2009, Hong Kong, China.

[2]. C. Pizzuti, "Community detection in social networks with genetic algorithms" Annual conference on Genetic and Evolutionary Computation, pages1137-1138, 2008.

[3]. Nan. Du, B.Wu, Xin Pei et al. "Community detection in large scale social networks" , SNA-KDD, pages 16-25, August 12, 2007, California, USA.

[4]. M.E.J. Newman. "Fast algorithm for detecting community structure in networks" Phys Rev E Stat Nonilin Soft Matter Phys, 2004.

[5]. Ferrara E. "A large-scale community structure analysis in Face-book. EPJ Data Sci ;1(9),2012

[6]. Fortunato S. Community detection in graphs. Phys Rep 2010; 4863(3-5):75-174.

[7]. Deitrick W, Valyou B, et al. "Enhancing sentiment analysis on twitter using community detection". Communications and Network 2013;5(3):192-7.

[8]. Leskovec J, Lang KJ, Mahoney MW. Empirical comparison of algorithms for network community detection". In: International conference on World Wide Web (WWW); 2010.p. 631-40.

[9]. Plantie M, Crampes M. "Survey on social community detection", In: Social media retrieval, computer Communications and Networks. Springer: 2013. P 65-85.

[10]. Uthayasankar Sivarajah et al. "Critical analysis of Big data challenges and analytical methods", Journal of Business research, (2017), P 263-286.

[11]. P.V. Sukhatme, B.V.Sukhatme et al. "Sampling Theory of Surveys with Applications", IOWA State University Press and Indian Society of Agricultural Statistics (New Delhi), 1984.

[12]. Cochran W.G (2005) , Sampling Techniques, John Willey and Sons, New York.