An efficient implementation of the German Tank Problem Statistical Algorithm in JAVA

[1]Joyjit Chatterjee
Amity School of Engineering and Technology
Amity University, U.P, Noida-201313, India
joyjitece@gmail.com

Abstract— The German Tank Problem is one of the most interesting problems in the history of Mathematics and Statistics. During the Second World war, the allies used a simple, but highly efficient statistical formula to identify the total number of German Tanks which were produced based on the serial numbers obtained from the sample of tanks which were captured. But today, this algorithm is not limited for solving only the German Tank problem but can be used for estimation of the total number of objects of any day to day articles based on the serial numbers of some of the articles. The same algorithm was used for estimation of the number of IPhones sold in 2008 which amounted to a huge 9 million. In the paper, the Statistical Algorithm is implemented using a simple, yet efficient JAVA Program which provides for an immediate result based on the user's input.

Keywords— German Tank Problem, Statistical Algorithm, Bayers Theorem, Conditional Probablity, JAVA

I. INTRODUCTION

The German Tank Problem was used by the allies during the Second World War for estimation of number of tanks being produced by Germans based on a simple, yet highly accurate statistical approach. The approach mainly uses the population maximum to estimate the total sample size from a given sample of data. Though, it is quite simple to calculate, but it can be difficult to understand and interpret the conclusion from a larger data set for a layman. So, in the paper, the same algorithm based on the Minimum Variance Unbiased Calculator has been implemented as a quick and simple JAVA Program. JAVA has been preferred for this purpose as it is a platform independent language and is quick to produce the desired result.

The entire paper is structured as follows: Section II describes the Methodology adopted in implementing the program, Section III gives a gist of the Minimum Variance Unbiased Statistical Estimator, Section IV elucidates a Demonstration of Executed Program, Section V gives a glance into the Results of the paper and Section VI concludes the paper.

II. METHODOLOGY

The methodology adopted in the paper is described in the flow chart given below and each step has been briefly discussed in the following paragraph.

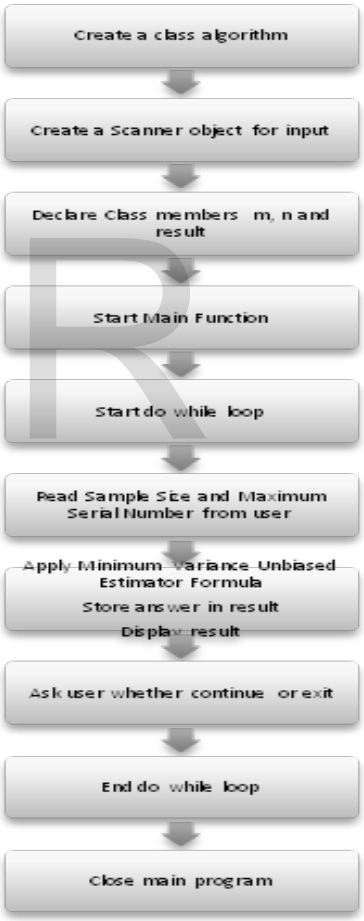

Figure 1: Flow chart showing methodlogy adopted

Firstly, the class algorithm is created for the specific purpose of solving the German Tank Problem. Then, a scanner object making use of java.util.scanner library is created for input

purposes. The integer type data members of algorithm class, m and n are declared. Next, the public static void main(String args[]) funciton is started. A do while loop is initiated and the sample size and maximum serial nunmber observed are inputted from user. Then, the Minimum Variance Unbiased Statistical Estimator formula is applied to calculate the result, which is the total number of estimated objects. Finally, the result is displayed and user is asked whether to continue or not and main program is closed.

III. MINIMUM VARIANCE UNBIASED STATISTICAL ESTIMATOR

The Minimum Variance Unbiased Estimator mainly aims to minimize the variance, which indicates the variation of the sample values from its mean. It is based on the Bayesian Conditional probability estimator. If we can minimize the variance, then we will revolve around the mean, which is the main parameter to be calculated. Without going into the Mathematical complexity, the final formula used for the purpose is:-

$$N= m+ (m/n) -1 \quad \text{……..(1)}$$

Where,

N: Total Number of Estimated Objects

m: Highest Observed Serial Number

n: Total Sample Size taken

IV. A DEMONSTRATION OF THE EXECUTED PROGRAM

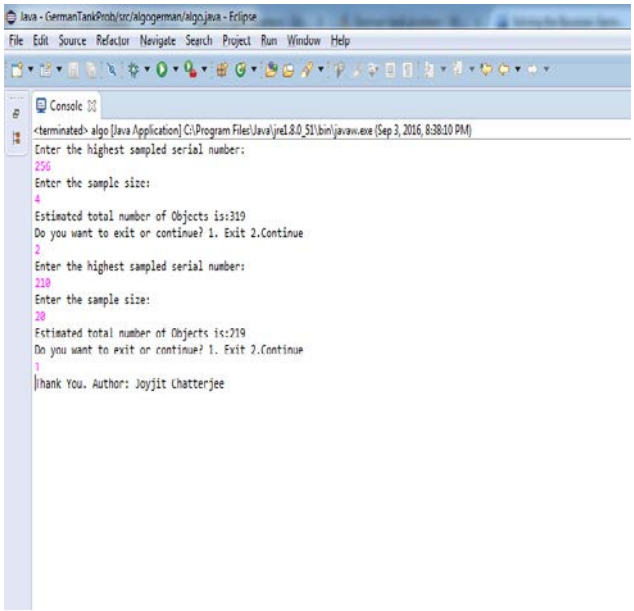

Figure 2: Sample Execution of Program in Eclipse IDE

Figure 2 shows the execution of the German Tank Problem for 2 sample runs. In the first run, highest order sample number is 256 with 4 sample size while there are 20 samples with maximum number 210 in the second sample. The same approach was used by the allies in Second World War. The algorithm yielded the total number of German Tanks produced to be 256 during the Second World War and after the war, when the official German documents were assessed, it was found that the total number of tanks produced was 255, thus only 1 short of the estimated value. Therefore, this methodology is highly accurate for statistical estimations.

Highest Sampled Serial Number	**Sample Size**	**Estimated Total Number of Objects**
256	7	319
210	20	219

Table 1: A glance into the test runs of the program

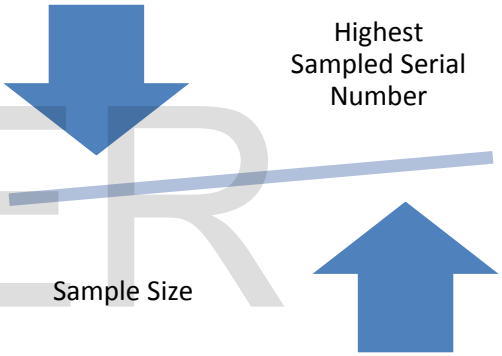

Figure 3: Correlation between Highest Sampled Serial Number and Sample Size

Figure 3 shows the gist of this paper and highlights that the correlation between Highest Sampled Serial Number and Sample Size forms the basis for using the German Tank Problem Algorithm and that's how it is executed in any programming language, in this case, JAVA.

V. RESULT

The German Tank problem algorithm is a highly reliable, accurate and efficient measure for estimation of the algorithm. By the end of 2008, the total number of IPhones sold was estimated at 9.1 million and was estimated from the same algorithm based on the Serial Numbers obtained by users from online discussion forums. The simple and highly efficient

algorithm has been successfully implemented on JAVA and the object code created can be executed on any operating system by create a suitable executable file for that particular OS.

The program has been successfully executed for 2 trial runs based on highest observed serial number and total sample size entered by the end user. Though, this algorithm is simple and efficient, till date, there is minimal Research Work based on implementing it in a programming language and this work can serve as a good initiative for future research in this area.

VI. CONCLUSION

The German Tank Problem is undoubtedly one of the most amazing problems which one can possibly imagine in mathematics. Though, it was first used at the time of Second World War for the estimation of tanks produced by Germans, this problem can now be used in the 21st Century for many applications which can benefit the mankind. The estimation of the total number of objects produced can serve the Industry as well as be used by statisticians and mathematicians in their day to day research work. It can be used by companies for estimating the total number of goods being produced by their rivals, thus creating a healthy competition.

REFERENCES

[1] K. Adenauer, "The German Problem, a world problem," Foreign Affairs, vol. 41, no. 1, p. 59, 1962.

[2] G. Blom, "A property of minimum variance estimates," *Biometrika*, vol. 65, no. 3, p. 642, Dec. 1978.

[3] Wolfe, W. J. (1999). A fuzzy hopfield-tank traveling salesman problem model. *INFORMS Journal on Computing*, *11*(4), 329–344.

[4] Y. Nakamura, "Mean-variance utility," *SSRN Electronic Journal*.

[5] B. Scherer, "A new look at minimum variance investing," *SSRN Electronic Journal*.

[6] Wilson, G. V., & Pawley, G. S. (1988). On the stability of the Travelling Salesman Problem algorithm of Hopfield and Tank. *Biological Cybernetics*, *58*(1), 63–70.

[7] Fortin, M., & DeBlois, J. (2010). A statistical estimator to propagate height prediction errors into a general volume model. *Canadian Journal of Forest Research*, *40*, 1930–1939.

[8] Worton, B. J. (1995). A Convex Hull-Based Estimator of Home-Range Size. *Biometrics*, *51*(4), 1206–1215.