

# Visualization of Performance of Binary Search in Worst Case in Personal Computer using Semi Parametric Technique

Dipankar Das

Assistant Professor

The Heritage Academy, Kolkata, West Bengal, India

**Abstract**— Binary search algorithm is a well-known divide and conquer searching algorithm. In this present work, the researchers have generated the experimental data set by running java code for binary search algorithm in the worst case on Linux operating system in a personal computer (laptop) and have calculated mean values of the observed execution time i.e. one thousand observations for each data size where the data sizes are five thousand to twenty thousand with an interval of five hundred to eliminate and/or decrease the noises from the observations (for each data size). The data points i.e. Data size versus Mean execution time in nano seconds are analyzed using semi parametric approach with varying degrees, residual diagnostics have been carried out for each of these models (fits) and the results of semi parametric fits are graphically represented to visualize the performance of the Binary search algorithm in the worst case in a personal computer (laptop).

**Key words:** Binary Search Algorithm, Binary Search

## I. INTRODUCTION

Empirical analysis of algorithms or experimental algorithmics is one of the finest ways to experience and experiment the different algorithms on specific computers. Visualization of algorithms is really fascinating [15]. Visualization is a process of creating a picture not present. Performance visualization is used to gain insight into the execution behavior [16]. The present study focuses on visualizing the performance of Binary Search algorithm in the worst case in a personal computer (laptop) using Semi Parametric technique. In the semi parametric technique, statistical models or fits are developed which has both non parametric component as well as parametric component [17].

## II. RELATED WORK

Binary search has been studied by the researchers in different context and different point of views e.g. comparative analysis of binary search algorithm with linear search algorithm [7][8][9][10], proposing modified binary search algorithm [11] and performance visualization of binary search algorithm in the worst case [12][13][14].

## III. OBJECTIVES OF THE STUDY

The objective of this study is to visualize the performance (Data Size versus Mean Execution time in nano seconds) of the Binary Search in the worst case in a personal computer (laptop) using Semi Parametric technique with varying degrees.

## IV. RESEARCH METHODOLOGY

The Binary Search algorithm has been executed in a personal computer (laptop) to generate the dataset. The

hardware and software configurations for the purpose are listed below:

- CPU: Intel(R) Core(TM)2 Duo CPU T5870 @ 2.00GHz
- Memory: 3 GB
- Operating system: Ubuntu 12.04.4 LTS
- Language: Java (OpenJDK runtime environment)

We have executed the java code of Binary Search in the worst case from data size five thousand (5000) to twenty thousand (20000) with an interval of five hundred (500) and for each data size one thousand (1000) execution times have been recorded. To eliminate and/or decrease the consequences of any noise present in the readings of the execution time for each data size, we have calculated the mean execution time (Mean Execution Time) for each data size (Data Size). It is to be noted that all the observations are recorded in nano seconds and we have done the experiment on integer data type.

The dataset (Data Size *versus* Mean Execution Time) have been analyzed using semi parametric regression technique. The software configurations for the purpose are given below:

- Language: R
- Package: SemiPar [1][2][3]

The radial basis function has been used to perform the semi parametric regression where different degrees for fitting have been used (degrees equal to 3, 5 and 7). In all the cases, residual diagnostics have been carried out. The models which have passed through the residual diagnostics are graphically represented and their respective model summaries are tabulated.

The following residual diagnostics have been carried out to check whether – (i) the model fits the data well or not (Residual *versus* Predictor plot) [4], (ii) the errors are independent or not (Residual lag plot) [5], (iii) the residuals are normally distributed or not (Histogram of the residuals) [6].

## V. DATA ANALYSIS & FINDINGS

The Residual diagnostics for the three (3) models under consideration have been given below to judge the quality of fit.

### A. Residual versus Predictor Plot

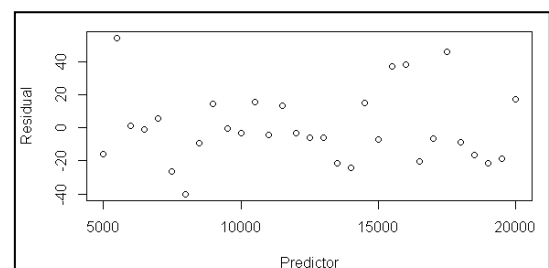


Fig. 1: Residual versus Predictor Plot for degree = 3

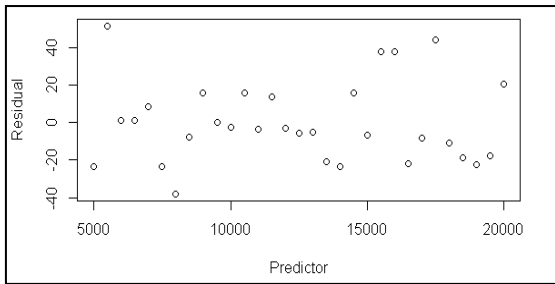


Fig. 2: Residual versus Predictor Plot for degree = 5

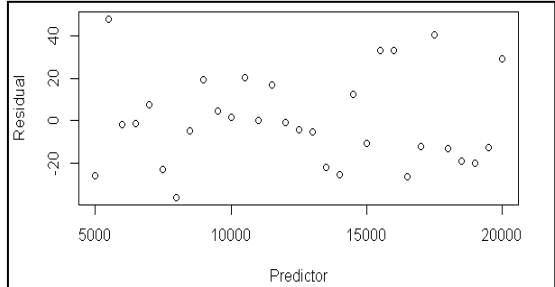


Fig. 3: Residual versus Predictor Plot for degree = 7

From the above figures (Fig. 1, 2 & 3) we observe that in all the three (3) cases the residuals are random and hence it may be concluded that all the three (3) models fit the data well.

### B. Residual Lag Plot

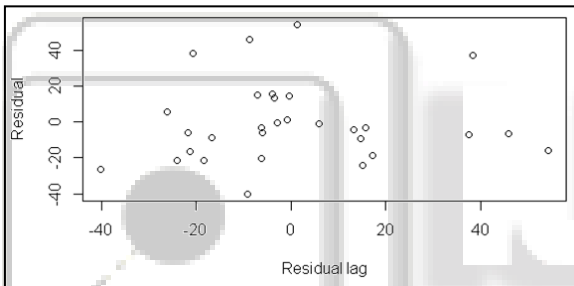


Fig. 4: Residual lag plot for degree = 3

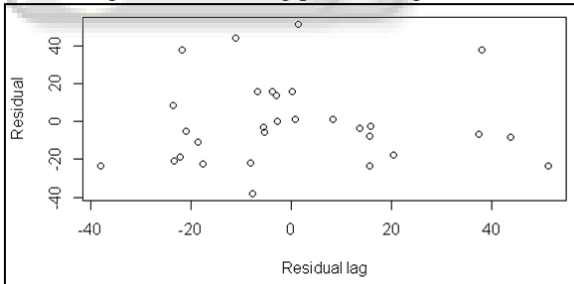


Fig. 5: Residual lag plot for degree = 5

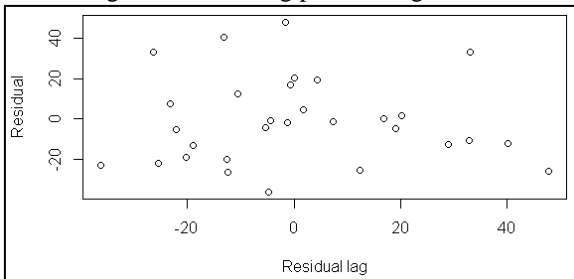


Fig. 6: Residual lag plot for degree = 7

From the above figures (Fig. 4, 5 & 6) we have observed that the data points are randomly distributed in all the three (3) cases and hence it may be concluded that the errors are independent in all the three (3) cases.

### C. Histogram of the Residuals

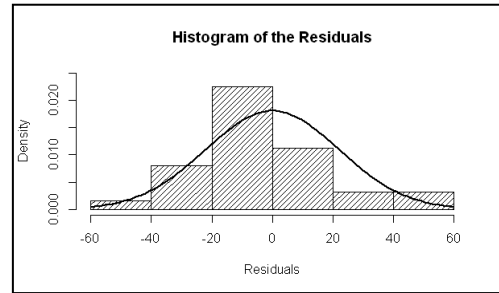


Fig. 7: Histogram of the Residuals for degree = 3

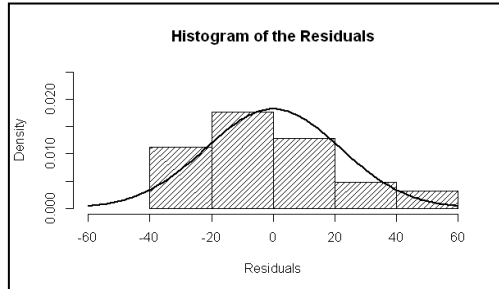


Fig. 8: Histogram of the Residuals for degree = 5

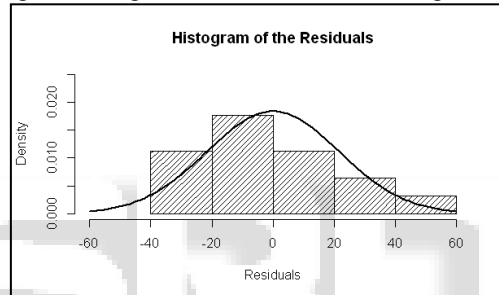


Fig. 9: Histogram of the Residuals for degree = 7

From the above figures (Figure 7, 8 & 9) we have observed that in all the three (3) cases the symmetric bell shaped curves are present which are approximately evenly distributed around zero. Therefore, it may be concluded that in all the three (3) cases the residuals are approximately normally distributed.

### D. Visualization of the Semi Parametric Fits

Therefore, from the above residual diagnostics we may arrive at an opinion that all the three (3) tested models are quite good fit for the experimental data set. The visualizations or the graphical representations of all the three (3) models along with their respective model summaries are given below. The red line represents the semi parametric fit, the purple dots represent the experimental data points and the blue shade represents the standard error bands.

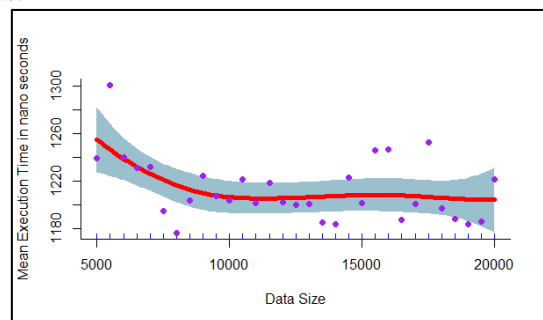


Fig. 10: Result of Semi Parametric fit for degree equal to three

The model summary of the semi parametric fit for degree = 3 are given in the following table (TABLE 1):

	df (degrees of freedom)	spar (smoothing parameter)	knots
f(Data Size)	3.695	9257	6

Table 1: Summary for Non-Linear Components for Model 1 (Degree = 3)

Note this includes 1 df for the intercept.

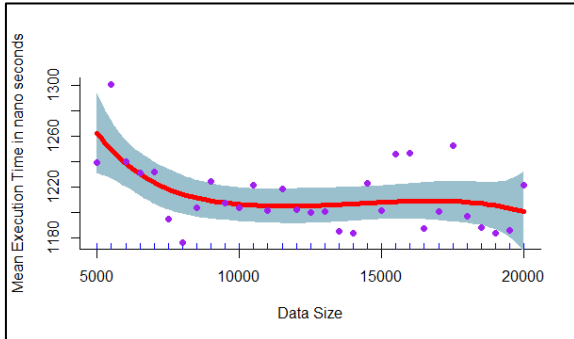


Fig. 11: Result of Semi Parametric fit for degree equal to five

The model summary of the semi parametric fit for degree = 5 are given in the following table (TABLE 2):

	df (degrees of freedom)	spar (smoothing parameter)	knots
f(Data Size)	4.202	11960	6

Table 2: Summary for Non-Linear Components for Model 2 (Degree = 5)

Note this includes 1 df for the intercept.

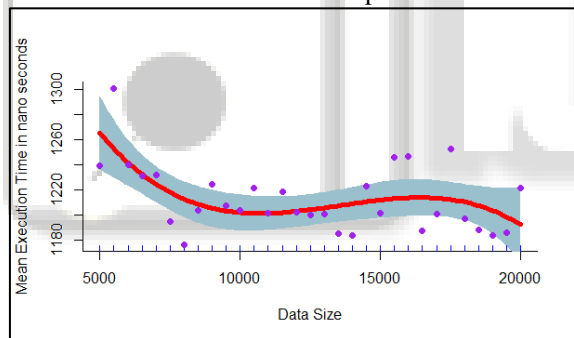


Fig. 12: Result of Semi Parametric fit for degree equal to seven

The model summary of the semi parametric fit for degree = 7 are given in the following table (TABLE 3):

	df (degrees of freedom)	spar (smoothing parameter)	knots
f(Data Size)	4	159200	6

Table 3: Summary for Non-Linear Components for Model 3 (Degree = 7)

Note this includes 1 df for the intercept.

## VI. CONCLUSION

The main objective of this paper is to visualize the performance of Binary Search algorithm in the worst case using semi parametric technique with varying degrees. From the summary tables of the fits we observe that in all the three cases the degrees of freedom and the smoothing parameters are different but the number of knots remain the same. At the same time, some very interesting observation has come to our notice that, as we increase the degrees from three (3) to seven (7) the smoothing parameter increases

considerably. Here, we have not calculated any error measures. In the present work, we have not tried to find the best or optimum semi parametric fit for the experimental data set which will surely be our future scope of study.

## REFERENCES

- [1] Wand, M.P., Coull, B.A., French, J.L., Ganguli, B., Kammann, E.E., Staudenmayer, J. and Zanobetti, A. (2005). SemiPar 1.0. R package. <http://cran.r-project.org>
- [2] Wand, M. (2015, February 19). Package 'SemiPar' [PDF].
- [3] The SemiPar Project. (n.d.). Retrieved July 1, 2016, from <http://matt-wand.utsacademics.info/SemiPar.html>
- [4] 4.4.4.1. How can I assess the sufficiency of the functional part of the model? (n.d.). Retrieved July 1, 2016, from <http://www.itl.nist.gov/div898/handbook/pmd/section4/pmd441.htm>
- [5] 4.4.4.4. How can I assess whether the random errors are independent from one to the next? (n.d.). Retrieved July 1, 2016, from <http://www.itl.nist.gov/div898/handbook/pmd/section4/pmd444.htm>
- [6] 5.2.4. Are the model residuals well-behaved? (n.d.). Retrieved July 1, 2016, from [http://www.itl.nist.gov/div898/handbook/pri/section2/pr\\_i24.htm](http://www.itl.nist.gov/div898/handbook/pri/section2/pr_i24.htm)
- [7] Kumari, A., Tripathi, R., Pal, M., & Chakraborty, S. (2012). Linear Search versus Binary Search: A Statistical Comparison for Binomial Inputs. *International Journal of Computer Science, Engineering and Applications (IJCSSEA)*, 2(2).
- [8] Sapinder, Ritu, Singh, A., & Singh, H.L. (2012). Analysis of Linear and Binary Search Algorithms. *International Journal of Computers & Distributed Systems*, 1(1).
- [9] Parmar, V. P., & Kumbharana, C. (2015). Comparing Linear Search and Binary Search Algorithms to Search an Element from a Linear List Implemented through Static Array, Dynamic Array and Linked List. *International Journal of Computer Applications*, 121(3).
- [10] Pathak, A. (2015). Analysis and Comparative Study of Searching Techniques. *International Journal of Engineering Sciences & Research Technology*, 4(3), 235-237.
- [11] Chadha, A. R., Misal, R., & Mokashi, T. (2014). Modified Binary Search Algorithm.
- [12] Das, D., Kole, A., & Chakrabarti, P. (2015). Sample Based Visualization and Analysis of Binary Search in Worst Case Using Two-Step Clustering and Curve Estimation Techniques on Personal Computer. *International Research Journal of Engineering and Technology*, 2(8), 1508 – 1516.
- [13] Das, D. (2016). Polynomial Curve Fitting of Execution Time of Binary Search in Worst Case in Personal Computer. *International Journal of Recent Trends in Engineering & Research*, 2(6), 245 – 250.
- [14] Das, D. (2016). Visualization of Binary Search in Worst Case Using Spline Interpolation Curve Fitting in Personal Computer. *International Journal of Recent Trends in Engineering & Research*, 2(7), 38 – 42.

- [15] Bostock, M. (2014, June 26). Visualizing Algorithms. Retrieved July 01, 2016, from <https://bost.ocks.org/mike/algorithms/>
- [16] Rover, D. T. (n.d.). Performance Visualization Usability and Reusability [PDF]. Department of Electrical & Computer Engineering; Michigan State University; [www.egr.msu.edu/~rover](http://www.egr.msu.edu/~rover); US/Venezuela Workshop on HPC 2000
- [17] Semiparametric model. (n.d.). Retrieved July 11, 2016, from [https://en.wikipedia.org/wiki/Semiparametric\\_model](https://en.wikipedia.org/wiki/Semiparametric_model)

