DA Assignment-2 Submission Report

Bhushan D Deo 18224

Part-1 : 20 parameter estimation problem: $Z(u,w) = Z_0(w)[1 - \exp\{-b(w)u\}]$

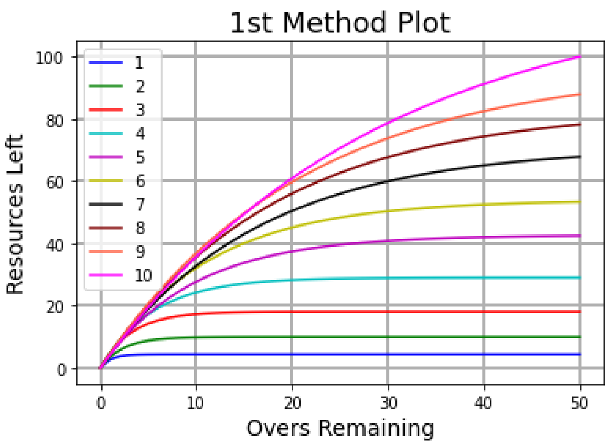
a.) Methodology:

First innings data was used for the computations. The data was cleaned up to avoid negative entries in the Z column of the data-frame. Also some basic sanity restrictions as Z=0|(u=0 or w=0) were imposed. Z column was computed using 'Innings.Total.Runs' – 'Total.Runs'. Wickets left represented w and u = 'Total.Overs' – 'Over'. Mean was used as the statistic to get data points from the samples.

The 'Match' column was used to get the unique innings totals which represented Z for u=50,w=10. The mean squared error function minimization is done using scipy's limited-BFGS iterative solver method. Suitable initial conditions were provided for fast convergence. Error statistics and plots have been generated.

b.) Results:

The plots for different values of w are as shown below:



The plots are normalized accordingly by dividing them by Z(50,10). The cumulative SSE for different values of w is tabulated as follows:

Value of W	SSE for all datapoints
1	238.6
2	2378.
3	4791.47
4	7366.1
5	608.13
6	2920.63
7	1590.23
8	428.86
9	240.17
10	775.19
Total :	21337.38

SSE: Sum of squared errors taken over non-zero valued data points.

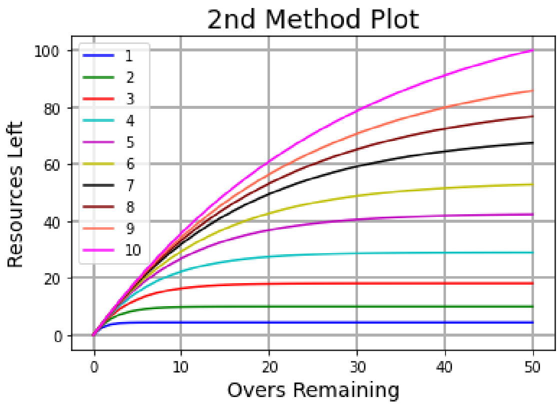
Parameters: $Z_0(w)$ = [10.61 24.02 43.85 70.27 103.37 130.58 172. 201.77 233.86 293.14]
$b(w)$ = [0.96 0.43 0.31 0.18 0.1 0.09 0.06 0.06 0.05 0.04]

Part-2 : 11 Parameter estimation problem : $Z(u,w) = Z_0(w)[1 - \exp\{-Lu/Z_0(w)\}]$

a.) Methodology:

The methodology is nearly similar to the above problem. It is more stable and runs fast as the number of parameters involved is less. As compared to the first method lesser numerical overflow errors were observed.

b.) Results: Parameters: $Z_0(w) = [10.61; 24.02; 43.85; 70.27; 103.37; 130.58; 172; 201.76; 233.86; 293.14]$ $L=10.26$



The plots are as shown below. Following that we tabulate the error functions:

Value of W	SSE for all datapoints
1	238.6
2	2378.05
3	4977.35
4	7782.86
5	679.23
6	3944.62
7	1754.48
8	1889.83
9	2374.12
10	775.2
Total:	26794.34

SSE: Sum of squared errors taken over non-zero valued data points.

Comparison of slopes: The initial slopes for different values of weights in part-1 are as shown. The constant value of the initial slope as obtained in part-2 is also indicated:

Value of W	Initial Slope
1	10.17
2	10.38
3	13.56
4	12.71
5	11.83
6	10.66
7	11.26
8	10.81
9	11.26
10	10.26

Slope as computed by method-2: 10.26

Conclusions: The computation method led to some interesting observations. Firstly, not all values in the 50x10 (overs left x wickets left) matrix were non-zero, meaning some of the combinations never appeared across 1400+ ODI matches. Secondly, the average score for the first innings for u = 50 and w = 10 was found to be less than that of u = 49 and w = 10. This may be explained as follows: u=50 and w=10 includes the case when one or more wickets falls in the 1st over. This inevitably leads to lesser cumulative score. The number of samples for u = 49 and w = 10 are less as compared to the former case (as the latter is contained in the former). So, the data is explainable given that we are working with 1st innings data and teams invariable start slow initially to gauge the conditions and later up the ante.