Data Analytics Assignment 1

Yash Punjabi (SR: 20870)

August 21, 2023

Duckworth-Lewis-Stern Method

Task: Using the first innings data alone in the above data set, find the best fit 'run production functions' in terms of wickets-in-hand w and overs-to-go u. Assume the model

$$Z(u,w) = Z0(w)[1 - exp\{-Lu/Z0(w)\}]$$

Use the sum of squared errors loss function, summed across overs, wickets, and data points for those overs and wickets.

Steps Performed:

1. The dataset consist of ODI matches from 1999 to 2011. For the 'Date' column first I tried to apply python datetime function, but there were some few entries having invalid date formats which were formatted then using custom function
2. Subset of data was fetched, since we require only a few of features, first innings only and only for complete matches (having overs completed as 50 or minimum wickets 0)
3. We define a loss function as $min \sum_{i=1}^{k} l(Z(u_i; Z0, b), y)$
   where l is least squares error for i data points , y is the actual output (runs) , k = no of data points and Z(u,w, Z0, L) predicted value
4. In order to minimize and get the values for Z0, b python library scipy.optimize.minimize was used. Since it contains many methods (L-BFGS-B, BFGS, COBLYA) for minimization all were used and comparison of total loss was made

Results:

Minimization Method	Total Normalized Squared Loss
L-BFGS-B	1383.60

Below are the parameters values for minimum method loss (L-BFGS-B)

Z0(1)	Z0(2)	Z0(3)	Z0(4)	Z0(5)	Z0(6)	Z0(7)	Z0(8)	Z0(9)	Z0(10)	L
14.53	30.06	57.86	91.22	117.18	153.14	184.57	230	261.63	306.06	10.38

Where Z0(i) is score for ith wicket

Below is the plot for 10 functions