Author: Sayan Chaudhuri

SR No.: 05-01-00-10-42-21-1-19743

Implementation Summary

Objective

Using the First Innings data alone, from the dataset provided, find the best fit ‘Run Production Functions’ in terms of Wickets-in-Hand(*w*) and Overs-to-go(*u*).

Model: $Z(u,w)=Z_0(w)[1-\exp\{-Lu/Z_0(w)\}$

Libraries Used

Pandas, Numpy, Scipy, Matplotlib

Procedure

- Loaded the data using Pandas. Then abstracted the data relevant to the problem at hand viz. ‘Match’, ‘Innings’, ‘Over’, ‘Runs.Remaining’, ‘Wicket.in.hand’.
- Filtered out the data to obtain the values of the 1st Innings alone.
- After that, I wrote a short script that will enable me to filter out the matches which were not complete. I assumed all the matches that were played “full innings”, were at least 40 overs long. I did so by creating an empty data frame and iterating a loop over the unique match-ids, adding the matches which were longer than 40 overs, to the new data frame.
- In this process, I also made sure to enforce the condition that at 50 overs remaining, runs remaining equals the total runs, and at 0 overs remaining, the runs remaining equals 0.
- Then I wrote a function ‘pred_score’ that predicts the score based on the model provided.
- I initialized the variables by curve fitting the given data, iterated over each wicket. I obtained 10 different Z_0 and 10 different L for each wicket in hand. Initialized the shared L as the mean of the different L previously obtained.
- Then, I wrote a ‘squared_loss’ function that finds the total squared error summed across overs, wickets and data points for those overs and wickets, normalized by the total number of data points across all overs and wickets for the “full innings”.
- I optimized this squared_loss function by using scipy.optimize.minimize function by passing in the required arguments and the initial variables previously obtained. I used the method as ‘Nelder-Mead’ as the optimization problem is well conditioned, and the dataset might also include some noise, since cricket is a game of ‘glorious uncertainties’
- From the optimized values of the shared parameter L and $Z_0(w)$ obtained, I plotted the curves using the function pred_score, by iterating it over the number of wickets.

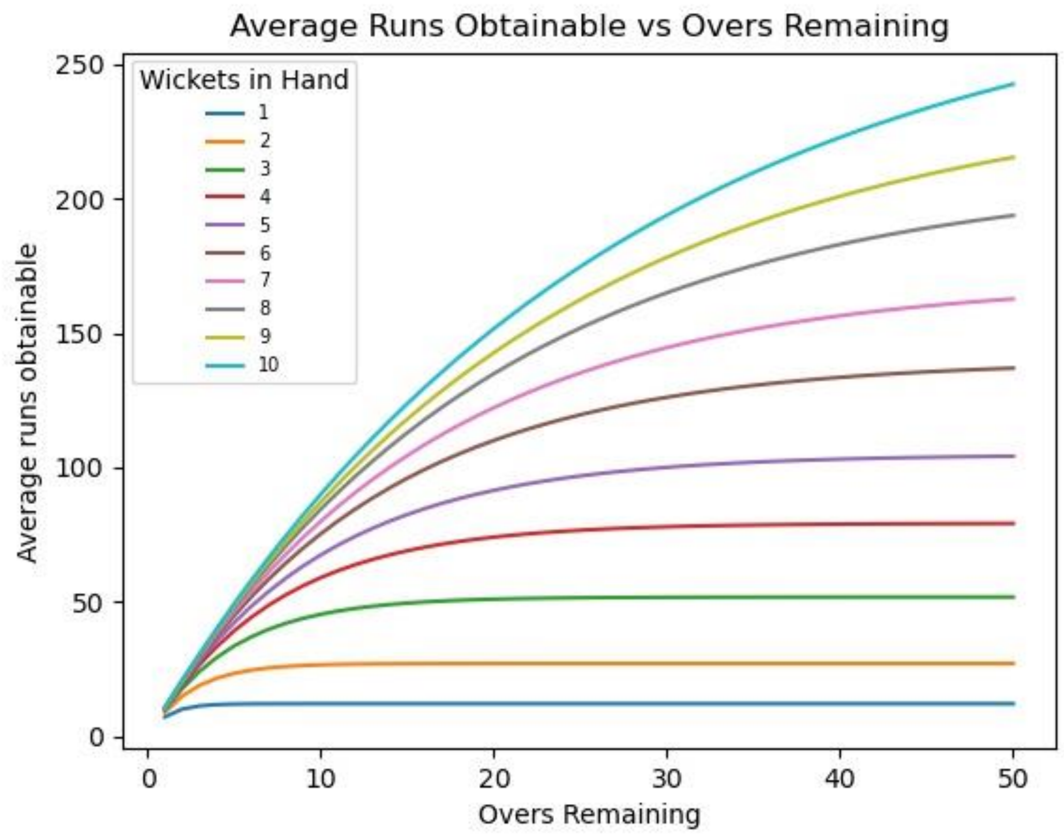
Results

Parameters of the Production Function:

Shared L = 10.111335046786692

Wickets in Hand(w)	**Z_0(w)**
1	**12.05164193**
2	**27.0026869**
3	**51.73998027**
4	**79.20668443**
5	**104.74694589**
6	**139.84887852**
7	**169.71273169**
8	**209.66620091**
9	**240.93756254**
10	**285.76919372**

Normalized Square Error = 1285.947941



Average Runs Obtainable vs Overs Remaining
Average runs obtainable
Overs Remaining
Wickets in Hand
1
2
3
4
5
6
7
8
9
10
0
50
100
150
200
250
0
10
20
30
40
50