

**Mohit Kumar**  
**SR No. 04-01-03-10-51-21-1-19825**  
**MTech Artificial Intelligence**

## Assignment 1: Duckworth Lewis Method

---

### python packages used

1. pandas
2. numpy
3. scipy
4. matplotlib

**D/L method** In D/L method, the Run Production Function is modelled as follows:

$$Z(u, w) = Z_0(w)[1 - e^{-b(w)u}] \quad (1)$$

where  $Z(u, w)$  denotes the runs scored with  $w$  wickets in hand with  $u$  overs remaining.  $Z_0$  and  $b$  are parameters of the model. If only one ball remains, regardless of the number of wickets in hand, we anticipate that the (average) increment to the score is the same. This assumption is OK if a good batsman is on strike. The Run Production Function with this assumption can be modelled as follows:

$$Z(u, w) = Z_0(w) \left[ 1 - \exp \left( -\frac{Lu}{Z_0(w)} \right) \right] \quad (2)$$

where the terms have their usual meaning and  $Z_0$  and  $L$  are parameters of the model.  $L$  which is the slope of the run production functions at  $u=0$  for the  $2^{nd}$  can be computed from 1 as follows :

$$L := \partial_u Z(u, 10)|_{u=0} = b(10)Z_0(10) \quad (3)$$

### Data Cleanup

- (a) Removed match data if data not present from 1<sup>st</sup> over.
- (b) Removed match data if 'Error.In.Data' column was 1.
- (c) Corrected the data in 'Innings.Total.Runs' column by using data from 'Runs' column.
- (d) Created a list of matches which were possibly interrupted due to rain i.e. if the max no of overs bowled were less than 50 and 'Innings.Total.Out' was less than 10. I have experimented with both including and not including the data of the interrupted matches.
- (e) Added data points corresponding to 'Wickets.in.Hand' = 10 and 'Overs.Remaining' = 'Total.Overs'.

**Optimization for computing model parameters** I have used the sum of squared errors loss function, summed across overs, wickets, and data points for those overs and wickets for the optimization of the model parameters. For this I used *optimize.curve\_fit()* function from the *scipy* module of python.

First I found the value of L by using 3 by computing  $Z_0(10)$  and  $b(10)$  from 1. Then I found  $Z_0(1) \dots Z_0(10)$  by using 2.

## Results

1. Including data of matches interrupted by rain

Value of L = 10.67

$w$	$Z_0(w)$
1	11.71
2	26.93
3	50.92
4	79.18
5	104.53
6	138.95
7	170.69
8	209.67
9	241.76
10	286

Mean Square Error = 1612.59

2. Without including data of matches interrupted by rain

Value of L = 10.56

$w$	$Z_0(w)$
1	11.85
2	27.42
3	52.28
4	81.59
5	107.25
6	142.33
7	174.50
8	214.37
9	247.10
10	290.80

Mean Square Error = 1560.79

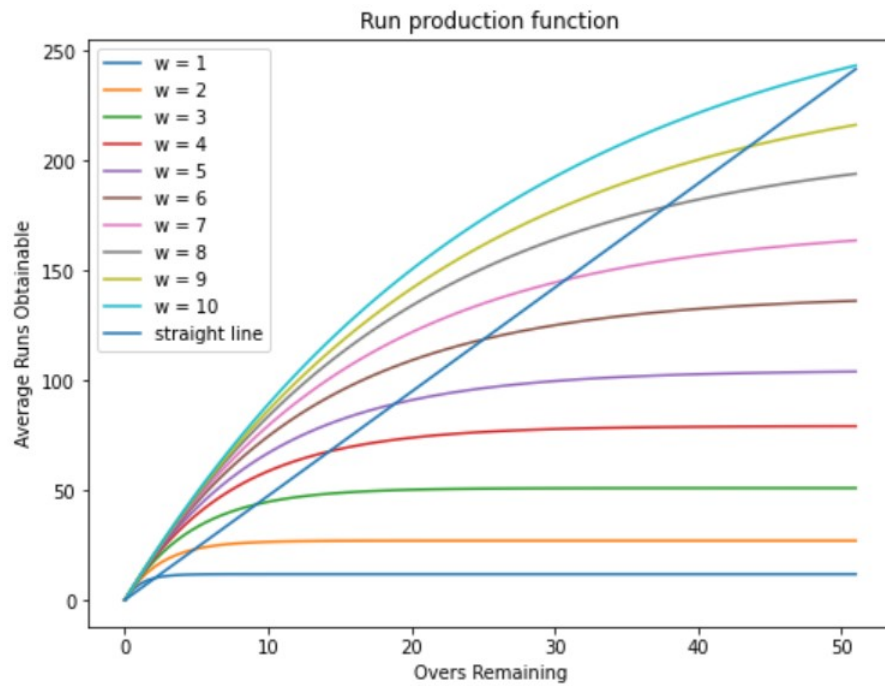


Figure 1: Run Production Function(with interrupted match data)

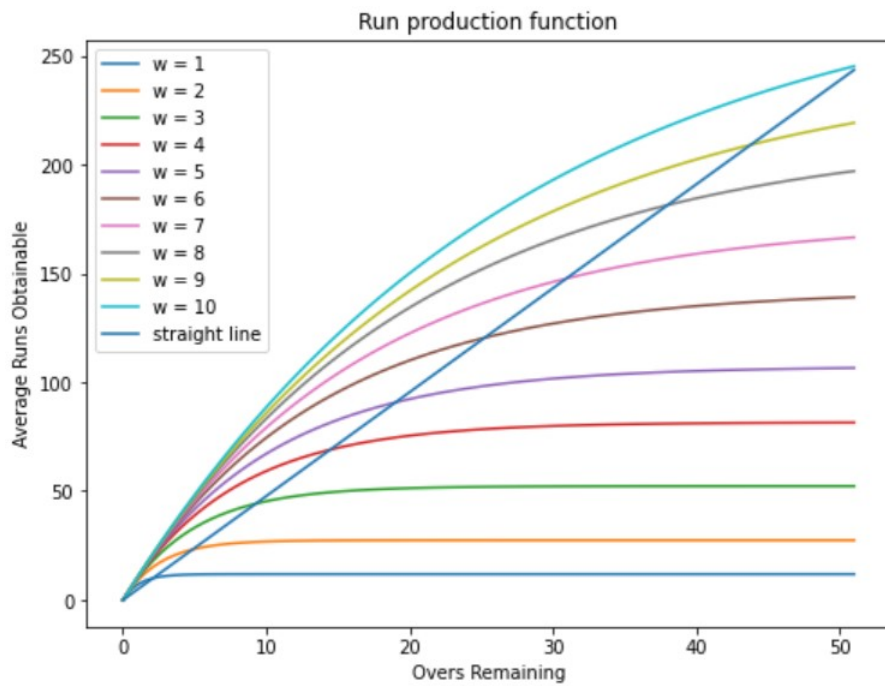


Figure 2: Run Production Function(without interrupted match data)