The Duckworth-Lewis-Stern Method

Data Analytics - DLS Lecture 3
The story thus far

- Revision of targets in shortened matches.

- D/L method:
  Quantify the resources of overs-to-go and wickets-in-hand.
  Resources = run scoring potential.

- Identify quantity of resources available with both teams.
  Set target by equating runs per resource.

- In this lecture: A discussion on the D/L method and new fixes.
Fix high first innings scores: D/L Professional Edition

Introduce dependence of curves on the first innings score.
The fix for high first innings scores

- Higher the first innings score, closer to a straight line.

- Resources remaining after 25 overs is lesser, $R_2$ is higher, so $T = SR_2$ is higher.

- Again, the choice of parameters based on data and is in the Professional Edition.
A discussion on the “relative positions” criterion

- The relative positions of the two teams before and after interruption should be the same.

- Think about the hypothetical IND-AUS game. What was your strategy? What did you try to optimise?

- One appealing criterion is *isoprobability*: The probability of winning before and after the interruption must be the same.

- Does D/L satisfy this?
An example to bring home the point


- Cambridge scored 190 off their 50 overs.

- Oxford were 162/1 off 31 overs when rain interrupted play. (29 to win off 19 overs)

- When rain stopped, 12 overs remained. But Oxford had already exceeded any target that D/L would set. Oxford was declared winner by D/L method.

- Before the rain, Oxford had a huge advantage, but cricket is a game of ‘glorious uncertainties’. The probability of Oxford winning was not 1. Yet, after the interruption, Oxford was declared winner at resumption.

- Isoprobability criterion would have given Cambridge a positive chance to bowl Oxford out. Low probability, but still positive. Would spectators have preferred that? Players?
An insightful pair of comparable games

- Two adjacent grounds A and B hosting two matches. Both Teams 1 scored 250 off 50 overs. Both Teams 2 played 20 overs, lost 3 wickets, when it rained. 10 overs lost due to rain, and now 20 overs remain.

- Team 2A: 120/3
  Team 2B: 50/3.

- Before the break, Teams 2A and 2B needed 131 and 201 (resp.) off 30 overs with 7 wickets in hand.

- But since both teams used up the same amount of resources, and get the same (reduced) resources at resumption, their D/L targets are identical: 221.
  - Team 2A must score 101 off 20
  - Team 2B must score 171 off 20.

- More difficult for Team 2B. D/L improved the advantage for the team that was ahead before the interruption.
Isoprobability criterion

- Isoprobability targets: Team 2A - 228, Team 2B - 216.
  - Team 2A must score 108 off 20. (7 runs more than D/L).
  - Team 2B must score 166 off 20. (5 runs less than D/L).

- A case of “from each according to his ability”? 

Adding a third match (Carter and Guthrie)

Three adjacent grounds A, B, C hosting three matches. Teams 1A and 1B scored 250 off 50 overs. Team 1C scored 180 off 50 overs.

All Teams 2 played 20 overs, lost 3 wickets, when it rained. 10 overs lost due to rain. 20 overs remain.

<table>
<thead>
<tr>
<th>Team 1 score</th>
<th>Team 2 at int</th>
<th>Target at int</th>
<th>D/L target</th>
<th>D/L to-go</th>
<th>IsoP target</th>
<th>IsoP to-go</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 250</td>
<td>120/3</td>
<td>131 (in 30)</td>
<td>221</td>
<td>101 (20)</td>
<td>228</td>
<td>108 (20)</td>
</tr>
<tr>
<td>B 250</td>
<td>50/3</td>
<td>201 (in 30)</td>
<td>221</td>
<td>171 (20)</td>
<td>216</td>
<td>166 (20)</td>
</tr>
<tr>
<td>C 180</td>
<td>50/3</td>
<td>131 (in 30)</td>
<td>159</td>
<td>109 (20)</td>
<td>158</td>
<td>108 (20)</td>
</tr>
</tbody>
</table>

- Isoprobability: Teams 2A and 2C must go the same distance. D/L: Team 2A has it easier.

- Team 2A was the poorer bowling team. D/L gives it a discount. Again a case of 'socialism'?

- Which is the better criterion?
Incentive to alter strategy under D/L rule

- Team 1 is at 160/4 in 39 overs. Rain is expected in the next over. Predicted duration of rain is such that their innings will be terminated, and Team 2 will have about 40 overs.

- Consider two options for Team 1:
  (i) bat carefully and lose no further wickets (Team 2 target 206)
  (ii) bat to maximise score but lose two wickets (Team 2 target 194).

- D/L assume that Team 1 will bat normally as in an uninterrupted match, and maximise their expected score.

- Will Team 1 do this?
  Team 1’s sole objective is also maximising its chances of winning. D/L method is likely to distort their strategy. Is it ok if Team 1 plays (at that stage) contrary to maximising runs? What about ARR method?
Applying D/L to T20

Rain delays play and reduces game to 20 overs.
Applying D/L to T20

- Curves are a lot flatter - D/L is now much closer to ARR method.


- ENG 191/5 off 20. WI 30/0 in 2.2 overs. Rain stops play. At resumption, WI target reduced to 60 (in 6 overs).

- If rain had come before start of WI’s play, RR method target = 58 (in 6 overs).
  But D/L target is only 66 (in 6 overs).
  Too much of an advantage for WI.

- But WI did much better. They consumed very little, and lost quite a bit of resources to rain.
  Revised target was much smaller.

- Is there a fix? ... Shrink the curves. Revised targets for shrink-the-curves method:
  Rain at start: 87 off 6 overs.
  Rain as in game: 69 (in 6 overs).
  VJD method apparently fares better. (A possible project).
Another issue with the D/L method

- Given $u$ overs to go and $w$ wickets in hand, D/L method assumes that both Team 1 and Team 2 will have the same scoring pattern.

- Run scoring potential at any stage mapped to remaining resources. For this Team 1 data alone suffices.

- But Team 2 is maximising probability of winning. It’s pattern of scoring is perhaps different.

- Stern (2009) analysed the pattern of play in the second innings.
The typical pattern in successful chases ...

- If the first innings score is low, use lesser amount of resources.
- If the first innings score is average or high, use the full quota of overs and resources.
- Picture ...
- Also, a fast start, then slow middle overs, followed by acceleration near the end.
Stern’s correction to resources used

\[ R'_2(u, w) = F(R_2(u, w)) \]: D/L Standard edition resources used \( R_2 \) transformed to Stern’s resources used \( R'_2 \)
Stern’s correction to resources remaining

Lose overs where curve is steep $\rightarrow$ lose more resources $\rightarrow$ lower target.
Lose overs either initially or at the end $\rightarrow$ lower target.
Cricket as an Operations Research (OR) problem

- What is the ‘optimal’ batting strategy for Team 1 at any stage of the game?
  What is the ‘optimal’ batting strategy for Team 2?

- Suppose that there are $n$ balls to go and $w$ wickets in hand.

- The batsman can play a risky shot that fetches more runs or a safe shot with fewer runs.
  $p_d = \text{probability of losing wicket.}$
  $p_x = \text{probability of getting } x \text{ runs, } x = 0, 1, 2, 3, 4, 5, 6.$
  Through his choice, the batsman can shape this probability vector.

- What is his best strategy at any stage of Team 1’s innings?
Clarke’s simplified model - Team 1

- Ignore bowler quality or varying quality of later batsmen.
- Also, let us assume that if batsman’s run rate (per ball) is \( r \) then the probability of losing wicket is \( p_d(r) \). Assume known (from historical data).

\[ r = \sum_{x=0}^{6} xp_x. \]

- Let \( Z_b(n, w) \) = expected number of runs in \( n \) balls, with \( w \) wickets.
- Goal: Maximise \( Z_b(300 \, \text{balls}, 10 \, \text{wickets}) \).

- A recursive formula:

\[
Z_b(n, w) = \max_{p_0, \ldots, p_1} \left[ p_d Z_b(n - 1, w - 1) + (1 - p_d) Z_b(n - 1, w) + \sum_x xp_x \right]
\]

This is called the dynamic programming equation or Bellman equation.

- Boundary conditions: \( Z_b(0, w) = 0 \) for all \( w \), \( Z_b(n, 0) = 0 \) for all \( n \). Enough information to determine \( Z_b(n, w) \) for all relevant \( n, w \).
Clarke’s simplified model - Team 2

- Maximise probability of reaching the target.

- \( q(s, n, w) = \) probability of scoring \( s \) with \( n \) balls to go and \( w \) wickets in hand.

- Again a dynamic programming equation:

\[
q(s, n, w) = \max_{p_d, p_0, \ldots, p_6} \left[ p_d q(s, n - 1, w - 1) + \sum_x p_x q(s - x, n - 1, w) \right]
\]

- Simplify even further.
  Assume batsman can score 0 or \( a \) runs only. He can choose \( p_a \).
  This gives a run rate of \( ap_a \).
  Suppose this fixes the \( p_d \) via \( p_d(r) \). Then \( p_0 = 1 - p_a - p_d \).

- Boundary conditions: \( q(l, 0, w) = 0 \) for all \( l < s \) and all \( w \).
  \( q(l, n, 0) = 0 \) for all \( l < s \) and all \( n \).
  \( q(0, n, w) = 1 \) for all \( n \geq 0 \) and all \( w \geq 0 \).
  Enough to determine \( q(n, w) \) for all relevant \( n, w \).
Your assignment

- Generate the production function (as a function of wickets) on 12 years of ODIs. Nonlinear regression to estimate the parameters.

- Projects you could take up:
  The OR problem for Team 1.
  The OR problem for Team 2.
  Stern’s method to predict the probability of a win by Team 2.
  Jayadevan’s method.
References