The Duckworth-Lewis-Stern Method

Data Analytics - DLS Lecture 1

One-day Cricket

- A game between two teams. Each team plays an innings.
  Both get 50 overs and have 10 wickets.
- The team batting first (Team 1) tries to maximise its score using its overs and wickets. The bowling team (Team 2) tries to restrict this score because that score will be its target.
- The bowling team (Team 2) then gets to bat, and tries to reach the target within 50 overs.
- Shorter version than 'test cricket' with a win/lose outcome.
- Often leads to exciting finishes.

Matches affected by bad weather

- Bad weather often leads to quite interesting twists and turns in test cricket, but is not tolerated in result-oriented one-day games.
- There simply isn't time, unlike in test cricket, for the match to continue another day, though reserve days have been used on occasions.
- "Draw" is not a good outcome in knock-out competitions.
- Continue with a shortened match and revised targets.
  Decide a winner based on the state of the match if play can't continue.
- First, some prior approaches through examples.

The average run-rate method

- Third-final of the 1988/89 Benson and Hedges World Series Cup between Australia and The West Indies.
- AUS scored 226/4 off 38 overs. Two hours delay during Australia's innings. (Dean Jones 93 n.o.)
- WI needed 180 off 31.2 overs when rain again stopped play for 1 hour 25 minutes.
- Target revised to 61 off 11.2 overs.
  Criterion used: Average run-rate.
- Most sides would achieve this target. WI (with Haynes and Richards) won easily with 4.4 overs remaining.
- WI had it too easy. Why?
- Post-match, Border called for a revision of the regulations; Richards was happy the existing regulations.

Another example - hypothetical

- Suppose Team 1 plays 50 overs and scores 250.
  Run-rate is 5 per over.
- Team 2 replies, and is 120/0 off 25 overs when rain stops play.
- Who is the winner?
- Par score under ARR method is $25 \times 5 = 125$, or 126 to win.
  ARR method: Team 2 loses. Is it fair?
  What if 120/2?
  120/9?

The 'most productive overs' or MPO method

- 1992 Cricket World Cup Semifinal: ENG vs. RSA.
- ENG made 252/6 off 45 overs.
- RSA were 231/6 and needed 22 runs off 13 balls when rain stopped play for 12 minutes.
- RSA target revised to 22 runs off 7 balls, then 21 runs off 1 ball. Criterion used: Most productive overs.
- The two good overs that RSA bowled were struck off. RSA was being penalised for bowling those overs well. (Actually, Wessels went slow and denied ENG the five final overs of acceleration because innings was scheduled to end latest by 6:10 pm).
- Christopher Martin-Jenkins on radio immediately after the game: "Surely someone, somewhere could come up with something better."

Enter F.C.Duckworth and A.J.Lewis

- The WI/AUS game during the 1988/89 BH WSC:
  - WI initial target would have been 232 off 38 overs due to the 2 hour interruption during AUS innings.
  - WI revised target would have been 139 off 11.2 overs after the second interruption.
  - AUS could have been more aggressive if they had known that their innings were shorter.
  - WI had many wickets in hand and could afford a risk of a much faster scoring rate.

- The RSA/ENG game during the 1992 WC:
  - RSA would need 2 to tie and 3 to win in 1 ball.
    (Updated version: 3 to tie, 4 to win in 1 ball.)
  - Both are reasonable targets.

The kinds of interruptions

- Before first team's innings ... (shorten the game)
- During first team's innings ...
  Repeated interruptions possible
- In between the two innings ...
- During second team's innings ...
  Repeated interruptions possible ... (shorten + revise target)
- Stoppage with no resumption ... (determine winner)

D/L method since 1997

- D/L method was tried out first on 01 Jan 1997, ZIM vs. ENG.
  ZIM scored 200 in 50 overs.
  Rain during ENG innings reduced the game to 42 overs.
  ARR target 169.
  D/L target 186.
  ENG scored 179 off 42 and lost (D/L method).

- Has been the preferred method, with some modifications, for resetting targets in shortened games.

- Other methods: ARR; MPO;
  Discounted MPO 0.5% for every over lost;
  PARAB provides diminishing returns for overs in terms of runs;
  Adaptation of PARAB in WC1996 (ignores wickets in hand);
  CLARK method applies different rules for different kind of stoppages;
  VJD method, nearest contender to D/L.

What is desired of a good method (from the D/L paper)

- Revision must be fair to both sides.
  "Relative positions of the two teams should be the same after the interruption as they were before it."
- Must provide sensible results in all situations.
  Recall RSA-ENG semifinal of 1992.
- Should be independent of the first team's scoring pattern ... because it is so before the interruption.
- Easy to apply, requiring no more than a table of numbers and a pocket calculator.
- Understandable by all those involved - players, officials, spectators, reporters.

The latest version deviates from the last two principles.

The basis: two resources and their valuation

- The batting side has two resources at their disposal to set a target: *overs to go and wickets in hand*.
- Both matter, and in a combined way.
  Twenty overs when ten wickets are in hand is much more valuable than when only 1 wicket is in hand.
- Team 2's target must be reset based on its *resources* before the interruption and after the interruption, such that *relative positions of the two teams should be the same before and after interruption*.
- View total runs that can be scored as a function of overs to go (u) and wickets in hand (w) as the net value of the resources.
  Call it $Z(u, w)$.
  Come up with a model, and then fit to data.

The run production function

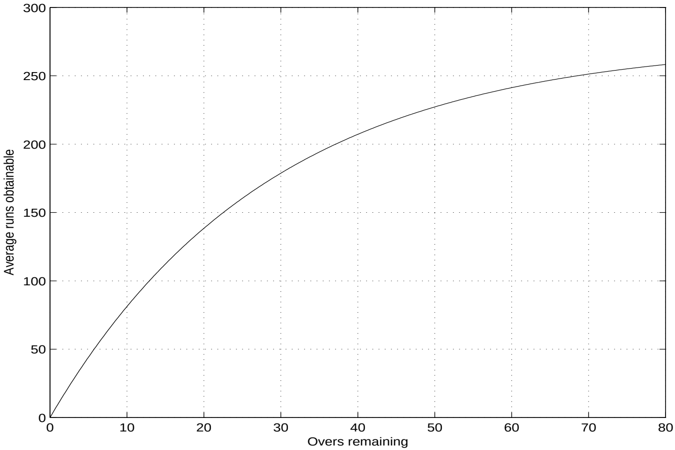
- Suppose a team has all 10 wickets and starts playing.
  They play according to ODI rules, but don't have any over restrictions.
  The total runs scored before they lose all ten wickets is a random quantity.
- Let Z_0 be its average.
- Model for the average score in u overs:

$$Z(u) = Z_0[1 - \exp\{-bu\}].$$

- A plot with $Z_0 = 275, b = 0.035$ ...

Run production function example

- Model for the average score in u overs: $Z(u) = Z_0[1 - \exp\{-bu\}]$. A plot with $Z_0 = 275, b = 0.035$



Getting the curve

- How would you estimate this curve?
- Find all data points with ...
  50 overs remaining, all 10 wickets in hand: get average runs scored in 50 overs;
  49 overs remaining, all 10 wickets in hand: get average runs scored in 49 overs;
  48 overs remaining, all 10 wickets in hand: get average runs scored in 48 overs;
  ...
  Fit a curve, and extrapolate
- You will do this on data from 1999 to 2011.

Resource fraction

- Resources available at the start of an N overs game: $Z(N)$.
  Call this 1 unit.

- Fraction of resources available when only u overs remain is then:

$$\frac{Z(u)}{Z(N)}.$$

What if wickets have fallen?

- A revised relationship:
  If u overs to go and w wickets in hand, then

$$Z(u, w) = Z_0(w)[1 - \exp\{-b(w)u\}].$$

- $Z_0(w)$ depends on the number of wickets in hand.
- Similarly, growth rate too depends number of wickets in hand.
- Anticipate $Z_0(10) > Z_0(9) > \cdots > Z_0(1)$.
- What about the growth rate?

Rate of increase, a connection across curves

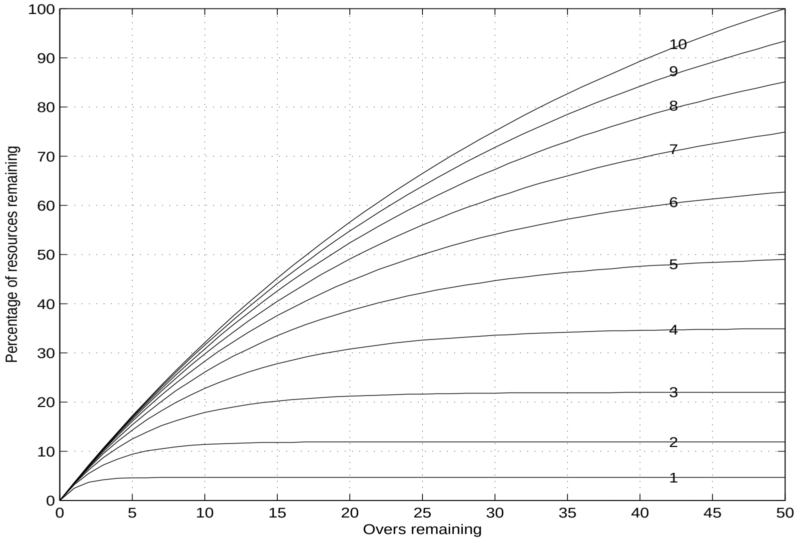
- If only one ball remains, regardless of the number of wickets in hand, anticipate that the increment to the score is the same.
- OK assumption if a good batsman is on strike.
- In mathematical terms, slope when only an incremental number of overs remain, which is $Z_0(w)b(w)$, is a constant independent of the number of wickets in hand.

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

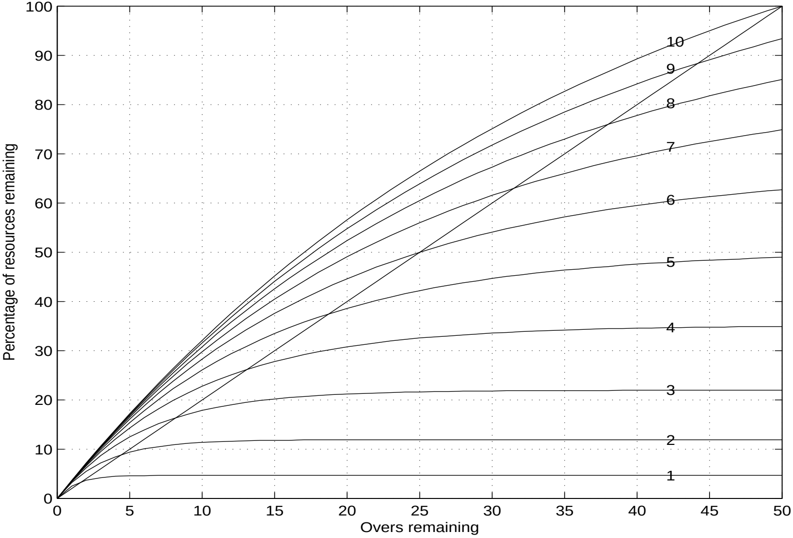
- Fraction of resources remaining

$$P(u, w) := \frac{Z(u, w)}{Z(N, 10)}.$$

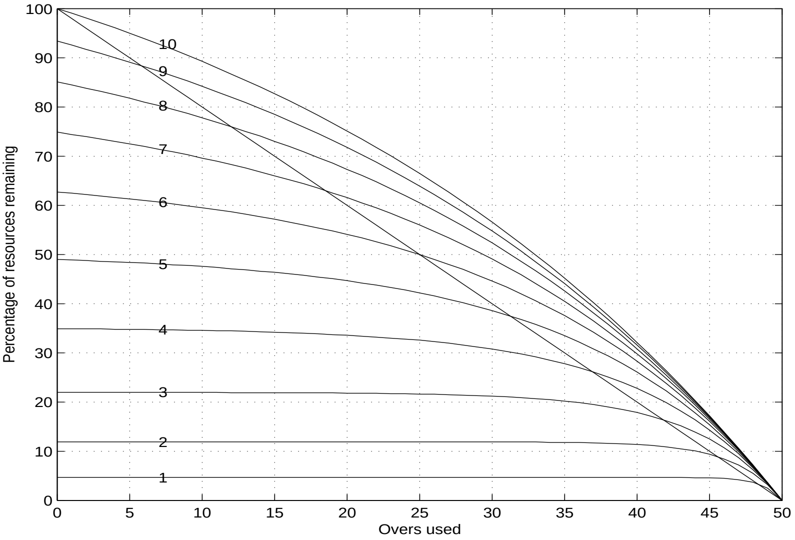
The resources remaining



The resources remaining, and a valuation ignoring wickets



The overs used picture



D/L method: Interruptions in the second team's innings

- Team 1 played 50 overs and scored S runs. They used up 100% of their resources to do this.
- Team 2 replies, and has w wickets in hand and u overs remaining, when rain stops play.
  When play resumes, team 2 has only v overs where $v < u$.
- Team 2 has been deprived of $u - v$ overs. They still have w wickets.
- Proportion of the resources R_2 available/used by team 2:

$$R_2 = \underbrace{1 - P(u, w)}_{\text{fraction used up before interruption}} + \underbrace{P(v, w)}_{\text{fraction remaining}}$$

- Par score $T = SR_2$. Target is the next integer.

Multiple interruptions in the second team's innings

- During team 2's reply:
  First stoppage: u_1 overs to go, w_1 wickets in hand.
  Resume: Reduced overs to v_1, w_1 wickets in hand.
  Second stoppage: u_2 overs to go, w_2 wickets in hand.
  Resume: Reduced overs to v_2, w_2 wickets in hand.

$$R_2 = \underbrace{1 - P(u_1, w_1)}_{\text{used up before int-1}} + \underbrace{P(v_1, w_1) - P(u_2, w_2)}_{\text{used up between ints}} + \underbrace{P(v_2, w_2)}_{\text{remaining}}$$

- Another way to look at it:

$$R_2 = 1 - \underbrace{(P(u_1, w_1) - P(v_1, w_1))}_{\text{lost in int-1}} - \underbrace{(P(u_2, w_2) - P(v_2, w_2))}_{\text{lost in int-2}}$$

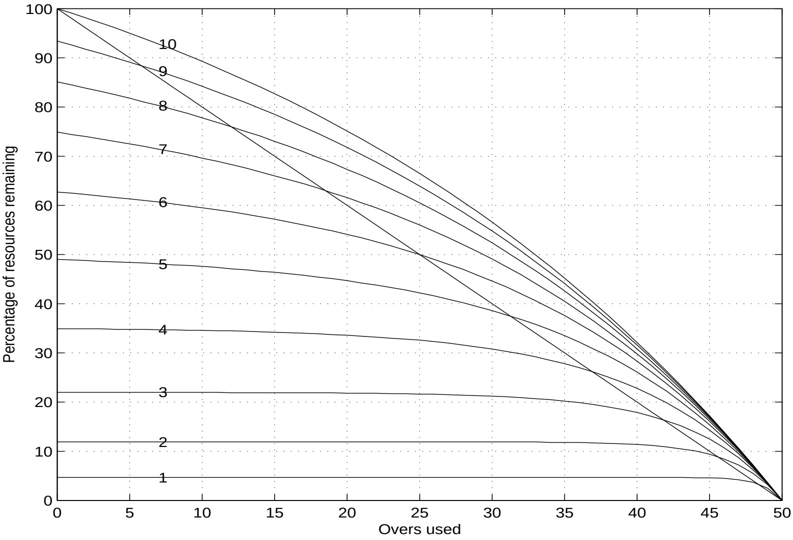
- Par score at resumption after int-2 is $T_2 = SR_2$

Example applications from indicated curves

Team 1 scores 250 off 50 overs.
Team 2's innings stopped at 25 overs.

	I	II	III
Team 2 score at stoppage	120/0	120/5	120/9
Resources lost	0.66	0.42	0.05
R_2	0.34	0.58	0.95
$T = SR_2$	85	145	247.5

The overs used picture



Interruption and resumption

Team 1 scores 250 off 50 overs.
Team 2's innings interrupted at $u = 25$ overs.
When play resumes, team 2 has $v = 10$ overs.

	I	II	III
Team 2 score	120/0	120/5	120/9
Resources rem. at stoppage	0.66	0.42	0.05
Resources rem. at resumption	0.34	0.26	0.05
R_2	1-0.66+0.34	1-0.42+0.26	1-0.05+0.05
	$= 0.68$	$= 0.84$	$= 1.00$
$T = SR_2$	170	210	250

What would be your strategy?

- Australia scored 250 runs off 50 overs.
- India goes in to bat. Forecast says rain in about 1.5 hours.
- What would be your strategy?
- What did you try to optimise?

A famous example: WC2003 IND vs. AUS

- Ganguly won the toss and put AUS into bat.
- AUS went on a rampage and scored 359/2 off 50 overs.
- Tendulkar scored a boundary off McGrath on the fourth delivery, but was out c&b on the fifth.
- Rain briefly stopped play when India were 103/3.
- India were 145/3 off 23 (with Sehwag in the 80s). If India were 159/3 off 25 and had rain terminated the game, India would have won (D/L method).
- No such thing happened. The better team eventually won that day.
- If very high first innings score, the match is already hugely in favour of Team 1, but this wasn't taken into account.
  The newer Professional Edition fixes this to some extent.
  Resources of second team depletes faster, depending on first team's score.

Stoppages during the first innings

- If the interruption occurs before start of play, there is no issue. If 20 overs in total are lost, the game is shortened by 10 overs, and both teams know this before they begin.
- But often, interruptions occur *during* the first innings.
- Match officials still try to arrange that both sides play the same number of overs.
- But Team 1 started out thinking 50 overs, and suddenly, find that their innings is shortened. Team 2 knows, from the start of their innings, that it is shortened.
- Whose loss is greater?
- Mostly Team 1's loss is greater, except ... when they have already lost a lot of wickets.

The D/L method in such interruptions

- Let R_1 be the proportion of the resources of a full uninterrupted innings that was available to Team 1:

$$R_1 = 1 - P(u, w) + P(v, w)$$

  if a stoppage occurs when Team 1 has u overs left and w wickets in hand, and at resumption, had lost $u - v$ overs.

- Similarly compute R_2.
- If $R_2 \leq R_1$, then $T = SR_2/R_1$.
- We will deal with the other case soon enough.

Example: Delayed first innings

- Interruption before the first team's innings resulting in a total loss of 10 overs.
- Game reduced to 45 overs each.
- Team 1 plays 45 overs and scores S.
- Both teams know about the reduction, would have acted accordingly, and so T must be S.
- What does D/L say?
- $R_1 = P(45, 10) = R_2$. Hence $T = S$.

Premature termination of first innings (D/L paper)

- IND vs. PAK, Singer Cup, Singapore, April 1996.
- IND scored 226/8 off 47.1 overs out of 50 overs when rain terminated IND innings.
- PARAB method gave PAK a target of 186/33 overs.
- D/L method:
  - IND used up $R_1 = 0.919$ fraction of their resources and lost 0.081.
  - PAK had $R_2 = 0.815$ fraction of resources available.
  - Since $R_2 < R_1$, D/L par-score is $SR_2/R_1 = 200.42$, or 201 off 33 to win.
- PAK won easily (PARAB target) with 30 balls to spare. (SRT 100, Aamer Sohail and Saeed Anwar 70s).

$R_2 > R_1$ anomaly, example from D/L paper

- Team 1 scores 80/0 off 10 overs. Rain reduces match to 10 overs.
- $R_1 \approx 1 - 0.9 = 0.1$.
- $R_2 \approx 0.34$. $R_2 > R_1$.
- Clearly, team 2 must have a higher target. But what target?
- $SR_2/R_1 \approx 80 \times 0.34/0.1 = 272$ off 10 overs!
- Can the *well above average* scoring rate really be sustained for 50 overs?

An inelegant fix

- If $R_2 > R_1$, then

$$T = S + G(N) \times (R_2 - R_1).$$

- $G(N)$ is the average first innings score in an N over match. (Compare with $G(N)$ replaced by S/R_1.)

- $G(50)$ was
  225 during 1999-2002,
  235 during 2002-2009, and
  245 now.
  Also, it's different for ICC full member nations, associates, under-19.

Interruption in the first innings leading to increased target

- WC2003, AUS vs. NED.
- Rain before play reduced match to 47 overs each. AUS batted first.
- First stoppage: 109/2 off 25, with 22 remaining.
  At resumption, match reduced to 44 overs.
- Second stoppage: 123/2 off 28, with 16 remaining.
  At resumption, match reduced to 36 overs.
- $R_1 = (1 - 0.029) - 0.558 + 0.505 - 0.447 + 0.255 = 0.726.$
- $R_2 = 0.841.$
- $T = 170 + G(50)(R_2 - R_1) = 197.025$ runs (D/L standard method)