A SECOND LOOK AT DUCKWORTH - LEWIS IN TWENTY20 CRICKET

by

Harsha Perera

B.Sc.(Hons.), University of Peradeniya, Sri Lanka, 2008

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2011

APPROVAL

Name: Harsha Perera

Degree: Master of Science

Title of project: A SECOND LOOK AT DUCKWORTH - LEWIS IN TWENTY20 CRICKET

Examining Committee: Dr. Derek Bingham
Chair

Dr. Tim Swartz,
Senior Supervisor
Simon Fraser University

Dr. Tom Loughin,
Supervisor
Simon Fraser University

Dr. Boxin Tang,
Simon Fraser University

Date Approved: 19 July 2011

Abstract

This project investigates the suitability of the Duckworth-Lewis method as an approach to resetting targets in interrupted Twenty20 cricket matches. Whereas the Duckworth-Lewis method has been adopted in both international Twenty20 matches and in the Indian Premier League, there has been growing objections to its use. In this project, we develop methodology for the estimation of a resource table designed for Twenty20 cricket. The approach differs from previous analyses in the literature by considering an enhanced dataset. It is suggested that there exist meaningful differences in the scoring patterns between one-day cricket and Twenty20.

Keywords: Constrained estimation; Duckworth-Lewis method; Gibbs sampling; Twenty20 cricket

To my cherished parents, brother Nalin and sister Yamuna.

Acknowledgments

First and foremost I want to thank my senior supervisor Dr. Tim Swartz and my co-supervisor Dr. Tom Loughin for all their support and guidance in achieving my Masters.

I came to SFU initially because of Dr. Tom Loughin and Graduate Chair Dr. Derek Bingham. I am most indebted to them for reposing their trust and confidence in me which enabled me to begin this great educational experience at SFU.

When trying to decide on a subject for my thesis I gravitated towards a sports related study as sports is an area of great interest to me. In Dr. Tim Swartz I found a supervisor who had done many studies in this area, and this enabled me to combine my interest in sports with my interest in statistics, thus providing me with great satisfaction as I proceeded with my Masters thesis. His expert guidance was always available to me, and he pointed me in the right direction at all times. Needless to say I would not have been able to complete my studies without his valuable input. Going one step further, I am extremely gratified that he has accepted me as a Ph.D student starting in Fall 2011. It was Dr. Tim Swartz and Dr. Derek Bingham who encouraged me to join the Ph.D. program at SFU for which I am most thankful.

To my examining committee consisting of Dr. Tim Swartz, Dr. Tom Loughin and Dr. Boxin Tang - a special word of thanks for a patient hearing and for the valuable inputs which I have incorporated into my thesis.

I also want to thank all the faculty members of the Department of Statistics and Actuarial Science who taught me during these last two years, especially Dr. Boxin Tang, Dr. Rick Routledge, Dr. Brad McNeney, Dr. Jiguo Cao, Dr. Charmaine Dean, Dr.Tom Loughin, and Dr. Tim Swartz. My special thanks to Ian Bercovitz for helping me to improve my statistical writing.

I am grateful for all the financial support provided by the Department of Statistics and Actuarial Science as well as to the donor of the Randy Sitter Annual Graduate Scholarship in Statistics and Actuarial Science. Special thanks to Statistics Workshop Manager Robin Insley for his support and guidance during my time at SFU. My sincere gratitude to Sadika, Kelly, and Charlene for their kind assistance and backup.

I want to thank Rianka Bhattacharya who helped me a lot, especially in the preliminary stages of the project in data collection, which was a real challenge, and Andrew Henrey for his advice and help with coding. Thanks also to my graduate student colleagues for their friendship and camaraderie and for the fun times we had together. To Primal Wijesekera at UBC I want to say thanks for helping me with LaTex and other technical matters.

I am fortunate to have a great family in Sri Lanka. I would not have come this far without their encouragement.

Last but not least I want to thank Dr. Ranil Waliwitiya for introducing me to SFU, and Ranil Bibile for his invaluable support and guidance in so many ways throughout my life in Vancouver.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background, Brief Explanation, and the Formats

Cricket is a sport that originated in England in the 16th century and later spread to her colonies. The first international game however did not feature England but was played between Canada and the United States in 1844 at the grounds of the St George's Cricket Club in New York. In time, in both of these countries, cricket took a back seat to other, faster sports like ice-hockey, basketball, and baseball. International cricket is played today by a number of British Commonwealth countries; the main ones being Australia, Bangladesh, England, India, New Zealand, Pakistan, South Africa, Sri Lanka, West Indies and Zimbabwe. These teams are members of the International Cricket Council (ICC). A second rung of international teams (called affiliates) consist of Afghanistan, Canada, Ireland, Kenya, Netherlands and Scotland.

Cricket is played on an oval-shaped playing field and is the only major international sport that does not define an exact size for the playing field. The main action takes place on a rectangular 22 yard area called the "pitch" in the middle of the large playing field. A diagram showing a cricket field is given in Figure 1.1 and the pitch is magnified in Figure 1.2.

Cricket is a game played between two teams of 11 players each, where the two

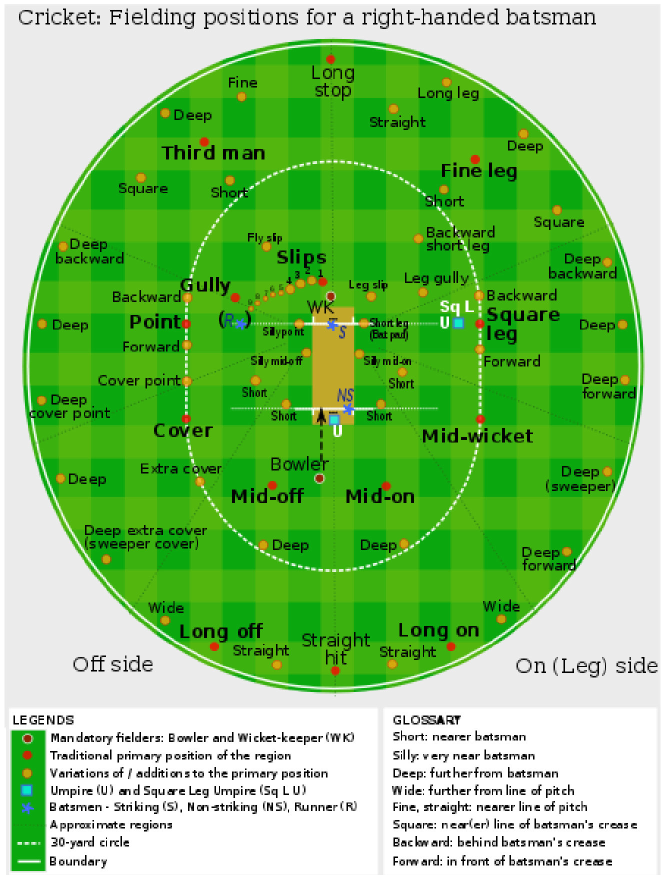


Figure 1.1: A cricket field showing the location of the pitch and some possible positions for players.

teams alternate scoring (batting) and defending (fielding). A player (bowler) from the fielding team delivers a ball to a player (batsman) from the batting team, who should strike it with a bat in order to score while the rest of the fielding team (fielders) defend the scoring. Furthermore, though it is a team sport, the bowler and batsman in particular, and fielders to some extent, act on their own, each carrying out certain solitary actions independently. A similar sport with respect to individual duties is baseball. In the process, batsmen can get dismissed (get "out") due to a variety of lapses on their part. When all the batsmen from the batting team have been dismissed, or faced their allotted number of "overs", (each "over" normally consists

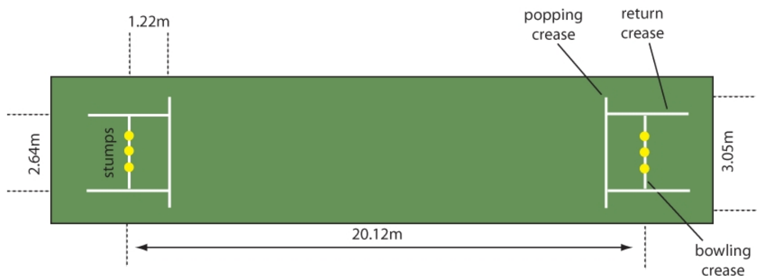


Figure 1.2: The layout of the pitch.

Figure 1.3: The pitch in use, with bowler, batsmen, fielders, and umpire (referee).

of six balls), that team's turn (called "innings") is concluded. Their score (number of runs) is noted down. A photo of some game action is provided in Figure 1.3. The teams now change places and the previously fielding team gets to wield the bat and try to overtake the score of the team that batted first. At the end of one such set of innings (in the shorter versions of cricket) and two such sets of innings (in the longer version of cricket) the winner is selected on the basis of the most runs scored. This is a very simplified explanation of a very complex game, and there are many variables and constraints that come into play. Apart from that, a clear cut winner may not be apparent in all games due to some peculiarities of cricket.

For the purposes of this statistical study it is not necessary to go into the intricacies of how batsmen get out and how the runs are scored. However some information about the various formats of the game is essential to understand why a statistical method is needed to select a winner in certain circumstances.

When international cricket matured, the standard format was a match that could last up to five whole days-called a test match. But even after five days of play the match could end in a draw-meaning there was no winner. This was fine in a more leisurely age when both players and spectators had more time, when playing the game was more important than winning, and when most cricketers were amateur players. But as spectators became ever more reluctant or unable to spend five days watching one match (sometimes with no result) and other, faster sports became crowd pullers and earned much more in the way of ticket sales and TV rights, cricketers became the poor relations in the sports world, earning niggardly sums.

In the nineteen sixties a shorter version of cricket was developed called "one-day cricket" with each batting side given 65 overs, and later 50 overs in which to score as many runs as possible if batting first, or, if batting second, to beat the score of the team batting first. When this format was used in international matches, they became known as One Day Internationals, or ODIs. This version of cricket was much more exciting to watch, and the batsmen had to wield the bat aggressively. It has to be noted that cricket is essentially a fair-weather game requiring good levels of light. The pitch is made of clay and turf and may become a quagmire in wet weather. Dampness affects the bounce of the ball and can endanger the players. The huge field itself has to be reasonably dry. Hence even with limited overs, sometimes bad light or bad weather interrupted play and a match would end with no result. Though artificial lighting helped overcome the bad light constraint, and some covering could be put on the 22 yard pitch itself, it was not feasible to build totally covered stadiums of the size required for cricket. Yet, results had to be obtained in the field of competitive matches even if the game was interrupted due to bad weather. In the early 1990s, various mathematical methods were used to decide on a winner when a limited over game was interrupted and the team batting second did not get to bat their full number of overs. However, all the methods had flaws which could be exploited by one side or the other.

1.2 Limited Overs Cricket and the Duckworth-Lewis method

Frank Duckworth and Tony Lewis, two British statisticians developed their Duckworth-Lewis Method (DL Method, 1998, 2004) to try and overcome the flaws of the previous methods. The DL Method was adopted by ICC and was first used in the ODI World Cup in 1999. The previous methods included factors such as the comparison of run rates and the scores achieved after a given number of overs. However, these methods did not allow for the risk factors inherent in scoring runs, especially when chasing a target. The faster a team tries to score, the more likely it is that they lose batsmen, and possibly lose all batsmen before reaching the target. Conversely, that risk can be reduced by slowing down the rate of scoring, but then the team batting second runs the risk of running out of sufficient overs to reach the target. In an interrupted game the target changes and the number of overs changes, which in turn changes the pacing required by the team batting second. The DL Method split the “assets” of the batting team into two resources, namely “wickets in hand” (i.e. number of batsmen who are still available to bat) and the number of overs remaining, and uses these resources in a table to provide a target for the team batting second in an interrupted game. The DL Method was developed for the 50 over limited game. However, in the newest form of cricket (Twenty20), the aggressiveness and pacing is dramatically different and therefore the relevant resource table needs to be adjusted to compensate for these factors. A brief explanation about Twenty20 is discussed in chapter 2.

1.3 Motivation for the Project

The DL method was developed for one-day cricket based on a large-scale analysis of scoring patterns observed in one-day cricket. It is natural to ask whether the scoring patterns are the same in one-day cricket and Twenty20. Equivalently, one may ask whether the Duckworth-Lewis method is suitable for use in Twenty20 cricket.

Initially, the use of Duckworth-Lewis methodology in T20 did not cause much of a stir. A reason for this may be that inclement weather in Twenty20 cricket tends to lead to the cancellation rather than the shortening of matches. And in the few matches where Duckworth-Lewis was applied, the targets appeared reasonable. However, over time, discontent has grown with various applications of Duckworth-Lewis in Twenty20. For example,

- May 3, 2010: On Duckworth-Lewis applied in the 2010 World Cup match between West Indies and England, captain Paul Collingwood of England complained "Ninety-five percent of the time when you get 191 runs on the board you are going to win the game". He then added, "There is a major problem with Duckworth-Lewis in this form of the game".

- May 5, 2010: Commenting on a variety of issues in cricket, former Pakistani bowler Abdul Qadir expressed the opinion, "One needs to revisit and rethink the D/L method and on top of that its use in T20 format".

- April 18, 2011: Chennai Super Kings coach Stephen Fleming on Duckworth-Lewis after loss to Kochi where Duckworth-Lewis was applied, "it is rubbish for Twenty20".

We suggest that we are now at a point in time where a sufficiently large dataset of T20 matches has accumulated to adequately assess whether there is a difference in scoring patterns between one-day cricket and T20. This project is concerned with the comparison of scoring patterns in the two versions of cricket. Accordingly, we develop a resource table analogous to the Duckworth-Lewis resource table for the resetting of targets in interrupted Twenty20 matches.

1.4 Organization of the project

In chapter 2, we review the Duckworth-Lewis method used in one-day cricket and how it is currently applied in T20 matches. We provide a list of suspicions as to why Duckworth-Lewis may not be appropriate for Twenty20. In chapter 3, we review aspects of Bhattacharya, Gill and Swartz (2010) which was the first attempt at assessing the suitability of Duckworth-Lewis in T20. We modify the Battacharya, Gill and Swartz (2010) methodology, hereafter referred to as BGS, and apply it to a much larger dataset than originally studied by BGS. In chapter 4, we develop new methodology for the estimation of a resource table in T20 which overcomes a prominent weakness of BGS. In particular, we do not aggregrate scoring over all matches but retain the variability due to individual matches. The new approach differs from the Duckworth-Lewis construction in that it does not assume a parametric form on resources. The approach is Bayesian and is implemented using Markov chain Monte Carlo methodology with estimation on a constrained parameter space. The new resource table confirms some of the findings in BGS. We conclude with a short discussion in chapter 5.

Chapter 2

T20 and the Duckworth-Lewis Method

2.1 Twenty20 Cricket

Compared to the five-day long test matches, the advent of the 50-over format was a dramatic improvement vis-a-vis spectator entertainment. However, even a 50-over match lasted about 8 hours and could not compete with the two to three hour match times, TV times, and attention spans of fans of soccer, rugby, ice-hockey, baseball and basketball. As competition increased for the sports fans' dollar, and TV advertising income linked directly to the number of viewers, it was inevitable that a very short format for cricket would emerge.

With declining ticket sales and dwindling sponsorships, the England and Wales Cricket Board (ECB) discussed the options for a shorter and more entertaining game in both 1998 and 2001. But it took until 2003 for the first official games to be played. On 13th June 2003 the exciting new format was launched at the English Counties Twenty20 Cup Tournament with the catchy new marketing slogan, "I don't like cricket, I love it".

Since then it has exploded in popularity with the label "Twenty20" getting a shorter format to read T20. A landmark match occured when Australia played India on 1st February 2008 at the Melbourne Cricket Grounds (MCG) in front of a crowd of 84,000. Later that same year the Indian Premier League (IPL) commenced with the T20 format and dozens of international players, and changed the financial aspect of cricket forever. The Board of Control for Cricket in India (BCCI) was already the richest cricket administration in the world but the T20 format brought the IPL almost up to the NBA level in the USA in terms of team salaries on a pro-rata basis. The IPL brand value is estimated at over $3.5 billion. Cricket had finally found its Eldorado in T20. The IPL is so popular that all 74 matches in 2011 were televised live in Canada, a nation with a limited cricket history.

At an international level, three T20 World Cups have been contested with India, Pakistan and England prevailing in the years 2007, 2009 and 2010 respectively. The next World Cup Tournament is scheduled for 2012 in Sri Lanka.

It should be noted that this is a new and evolving format, and hence the rules are changing quite frequently as more matches are played and experiences are accumulated; resulting in the rules being changed to fine tune the game.

Some salient points about the T20 format according to the International Cricket Council's (ICC's) "Standard Twenty20 International Match Playing Conditions" (October 2010 revision) are:

- Number of overs is 20 per side.
- Each bowler is limited to four overs only.
- During the first six overs of each side's innings, which is called the "power play", there cannot be more than two fielders outside a circle demarcated at 30 yards from the pitch. Refer to Figure 1.1 for the 30 yard circle shown with a dotted white line.
- After the first six overs the number of fielders outside the 30 yard circle cannot be more than five.
- Irrespective of the above two fielding restrictions, there cannot be more than a total of five fielders on the entire "leg-side" at any given time during the match. The "leg-side" is that half of the grounds that is behind the batsman as he stands with his side to the bowler. Refer to Figure 1.1 for leg-side indication.

- The fielding side must complete bowling their 20th over within 75 minutes. If they do not do so the batting side gets six extra runs for every full over that has to be bowled after 75 minutes. To prevent the batting side from deliberately delaying play to gain the benefit of this rule, the umpire can extend the 75 minute period if he feels it is justified. This rule is important as a T20 game is supposed to be completed in 3 hours which is a big selling point for this short format.

- When a “no ball” is called because the bowler oversteps the designated line, the batting side benefits in two ways. Firstly, there has to be an extra ball in that over and an extra run is scored, as in all formats. Secondly, the ball immediately following the no ball is termed a “free hit” which means the batsman can be extremely aggressive as he cannot be made out due to any of the normal rules except a run out.

- For a game to be considered as having been played, each of the two sides should have faced a minimum of five overs.

- In the event the scores of both teams are equal at the end of the designated 20 overs each, i.e. the result is a “tie”, the two teams will compete in a One Over Per Side Eliminator (OOPSE) to decide the winner.

As mentioned in Chapter 1, the Duckworth-Lewis (DL) Method is a carryover from the 50 over one day game and needs to be adjusted for the T20 format.

2.2 Review of Duckworth-Lewis

For resetting targets in interrupted one-day cricket matches, the Duckworth-Lewis method (Duckworth and Lewis, 1998, 2004) supplanted the method of run-rates in the late 1990s and has since been adopted by all major cricketing boards.

As far as I am aware, no other sport uses a statistical method to select the winning target for a match. But the peculiarities of cricket and its susceptibility to bad weather have made it imperative to find such a solution for matches where a result is mandatory.

Cricket is the second most popular game in the world, after soccer. With billions of followers around the world the DL method is probably the greatest contribution to the sporting world from a mathematical, statistical and operational research perspective. It is remarkable that the DL Method has been accepted by the ICC even though the fans hardly understand it. A number of competitors have tried to unseat the DL method (Clarke 1988, Christos 1998, Jayadevan 2002, Carter and Guthrie 2004), but it has withstood the test of time and has been updated every four years to accommodate the changes to the limited over games.

As mentioned in section 1.2, the resources are wickets in hand and overs remaining. Availability of these resources enables the scoring of runs. In 50-over cricket the team batting first starts off with their 50 overs and all 10 wickets in hand. They continue batting and scoring until their 50 overs are completed or all their wickets are lost. The DL method amalgamates these two resources into one quantifiable percentage as a bivariate exponential decay function. In this scenario, 50 overs and 10 wickets in hand at the beginning of the innings is equivalent to 100% of its resources. Depleting either all the 50 overs or all 10 wickets results in a 0% resource balance. When innings are reduced due to a rain-interruption and the team batting second does not get to face 50 overs, the winning target has to be reset to a "fair" value based on the resources remaining.

Table 2.1 shows an abbreviated version of the DL Resource Table (Standard Edition) taken from the ICC Playing Handbook (Section on Duckworth-Lewis Method of Re-calculating the Target Score in an Interrupted Match).

As an illustration of the use of Table 2.1, suppose that in a 50 over game the team batting first (Team 1) scores 285 runs in 50 overs. Due to rain interruption between the innings the team batting second (Team 2) only has time to face 40 overs. In this case it is obvious that Team 2 cannot have a target of 286 runs. If we use the old run rate method then the winning target is (285) (40/50) = 228. Therefore the winning target is set at 229. Since Team 2 has all 10 wickets in hand but has to face only 40 overs, they can be more aggressive, resulting in a faster scoring rate. Therefore 229 is a comparatively easy target and unfair for Team 1. The DL method removes this unfair advantage. In the DL Resource Table shown above, the calculation is based on 40 overs available and 0 wickets lost. According to Table 2.1 the corresponding resources available to Team 2 is 89.3%. Therefore the reset target for Team 2 to win is obtained via 285(0.893)= 254.5 which is rounded up to 255 runs. This is much more equitable target.

To construct the DL table, Duckworth and Lewis used a two-variable mathematical expression for the relationship between the average total score and the two resources. The mathematical function is of the form:

$$Z(u, w) = Z_0(w)[1 - \exp\{-b(w)u\}] \tag{2.1}$$

This was the original development in 1998; however more parameters were added later. Here $Z_0(w)$ is the asymptotic average total score from the last $10 - w$ wickets in unlimited overs and $b(w)$ is the exponential decay constant, both of which depend on the number of wickets already lost. However, the mathematical definitions of these functions have not been disclosed due to commercial confidentiality. Figure 2.1 shows the family of curves described by equation 2.1 using parameters estimated from hundreds of one day internationals. Even though the DL method is of great benefit there are some puzzling anomalies:

- Equation 2.1 uses an asymptotic average total score in unlimited overs cricket. We have to ask the question why a formula of unlimited overs was used for a limited overs game.

- Are there any other parametric curves available which could be more appropriate to obtain a resource table for a limited over game?

- Is there any advantage in using a non-parametric approach to obtain the resource table?

- In Table 2.1 the columns for wickets lost 7, 8 and 9 shows a static number in some cells even though the overs available are decreasing. This pattern is more pronounced as more wickets are lost. We know that the occurrences where this part of the table is applicable are very rare. Nevertheless the question is how the amalgamated resource remains constant when one of its components is decreasing. We cannot decipher the reasons for this anomaly as the method of estimation has not been divulged due to commercial confidentiality.

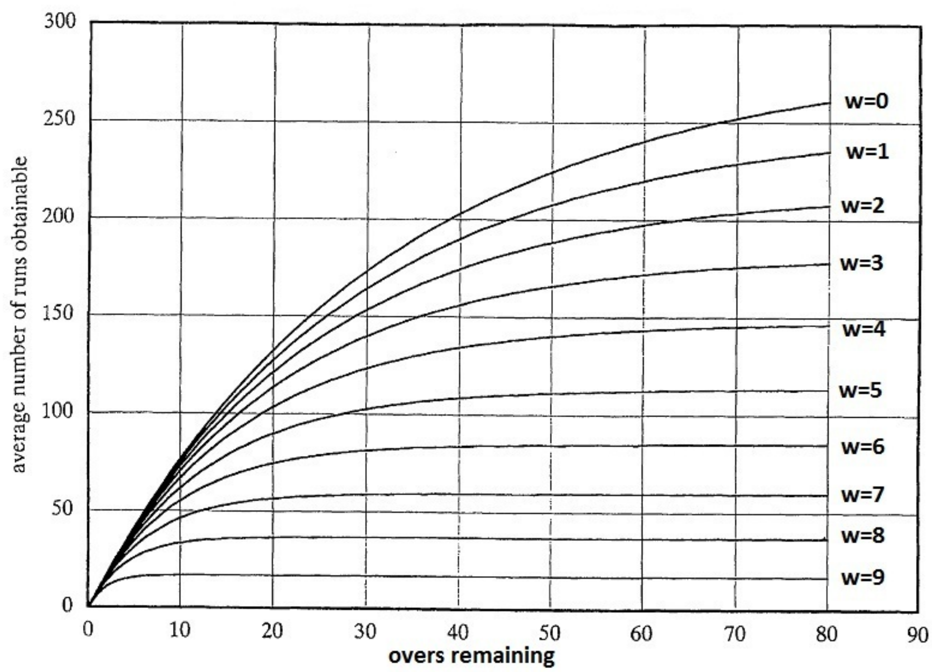


Figure 2.1: Average number of runs $Z(u, w)$ from u overs remaining with w wickets lost: Duckworth and Lewis (1998).

The application of the Duckworth-Lewis method to Twenty20 cricket considers the status of a one-day cricket match when 20 overs and 10 wickets are available. At this state of a match, according to the Standard Edition of the Duckworth-Lewis table, 56.6% of the batting teams resources remain. Therefore, the modified Duckworth-Lewis table for Twenty20 is obtained from the original Duckworth-Lewis table by dividing each cell entry by 0.566. In Table 2.2, we provide the Duckworth-Lewis table (Standard Edition) scaled for Twenty20. The table is monotonic decreasing in both the columns and rows indicating that resources diminish as overs are utilized and wickets are taken.

Although the scaling of resources from one-day cricket to Twenty20 is an intuitive and seemingly sensible procedure, there are a number of suspicions that the scoring patterns in one-day cricket and Twenty20 may not be identical due to some of the reasons already discussed. Therefore it seems prudent to investigate the suitability of the DL method in the context of Twenty20 cricket.

Table 2.1: Abbreviated version of the Duckworth-Lewis resource table (Standard Edition). The table entries indicate the percentage of resources remaining in a match with the specified number of wickets lost and overs available.

	Wickets Lost									
Overs Available	0	1	2	3	4	5	6	7	8	9
50	100.0	93.4	85.1	74.9	62.7	49.0	34.9	22.0	11.9	4.7
45	95.0	89.1	81.8	72.5	61.3	48.4	34.8	22.0	11.9	4.7
40	89.3	84.2	77.8	69.6	59.5	47.6	34.6	22.0	11.9	4.7
35	82.7	78.5	73.0	66.0	57.2	46.4	34.2	21.9	11.9	4.7
30	75.1	71.8	67.3	61.6	54.1	44.7	33.6	21.8	11.9	4.7
25	66.5	63.9	60.5	56.0	50.0	42.2	32.6	21.6	11.9	4.7
20	56.6	54.8	52.4	49.1	44.6	38.6	30.8	21.2	11.9	4.7
15	45.2	44.1	42.6	40.5	37.6	33.5	27.8	20.2	11.8	4.7
10	32.1	31.6	30.8	29.8	28.3	26.1	22.8	17.9	11.4	4.7
5	17.2	17.0	16.8	16.5	16.1	15.4	14.3	12.5	9.4	4.6
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 2.2: The Duckworth-Lewis resource table (Standard Edition) scaled for T20. The table entries indicate the percentage of resources remaining in a match with the specified number of overs available and wickets lost.

	Wickets Lost									
Overs Available	0	1	2	3	4	5	6	7	8	9
20	100.0	96.8	92.6	86.7	78.8	68.2	54.4	37.5	21.3	8.3
19	96.1	93.3	89.2	83.9	76.7	66.6	53.5	37.3	21.0	8.3
18	92.2	89.6	85.9	81.1	74.2	65.0	52.7	36.9	21.0	8.3
17	88.2	85.7	82.5	77.9	71.7	63.3	51.6	36.6	21.0	8.3
16	84.1	81.8	79.0	74.7	69.1	61.3	50.4	36.2	20.8	8.3
15	79.9	77.9	75.3	71.6	66.4	59.2	49.1	35.7	20.8	8.3
14	75.4	73.7	71.4	68.0	63.4	56.9	47.7	35.2	20.8	8.3
13	71.0	69.4	67.3	64.5	60.4	54.4	46.1	34.5	20.7	8.3
12	66.4	65.0	63.3	60.6	57.1	51.9	44.3	33.6	20.5	8.3
11	61.7	60.4	59.0	56.7	53.7	49.1	42.4	32.7	20.3	8.3
10	56.7	55.8	54.4	52.7	50.0	46.1	40.3	31.6	20.1	8.3
9	51.8	51.1	49.8	48.4	46.1	42.8	37.8	30.2	19.8	8.3
8	46.6	45.9	45.1	43.8	42.0	39.4	35.2	28.6	19.3	8.3
7	41.3	40.8	40.1	39.2	37.8	35.5	32.2	26.9	18.6	8.3
6	35.9	35.5	35.0	34.3	33.2	31.4	29.0	24.6	17.8	8.1
5	30.4	30.0	29.7	29.2	28.4	27.2	25.3	22.1	16.6	8.1
4	24.6	24.4	24.2	23.9	23.3	22.4	21.2	18.9	14.8	8.0
3	18.7	18.6	18.4	18.2	18.0	17.5	16.8	15.4	12.7	7.4
2	12.7	12.5	12.5	12.4	12.4	12.0	11.7	11.0	9.7	6.5
1	6.4	6.4	6.4	6.4	6.4	6.2	6.2	6.0	5.7	4.4

Chapter 3

Review and Implementation of BGS

3.1 Data Collection and a Suggested Improvement of the BGS Resource Table

In their initial analysis to assess the appropriateness of the DL method for T20, BGS collected the first innings scores from $n = 85$ international ICC recognized T20 matches which were played over the period 17th February 2005 to 9th November 2009. The second innings data were not collected as the side batting second may change their scoring patterns depending on the target score, the state of play, and the resources at hand. This results in second innings data being distorted and misleading.

For each match, BGS defined $x(u, w(u))$ as the runs scored from the stage in the first innings where u overs are available and $w(u)$ wickets have been taken until the end of the first innings. The variable u took the values $u = 1, \ldots, 20$ and $w(u)$ took the values $w = 0, \ldots, 9$. BGS defined $r_{u,w}$ as the estimated percentage of resources remaining when u overs are available and $w(u)$ wickets have been taken. They calculated $(100\%)r_{u,w}$ by averaging $x(u, w(u))$ over all matches where $w(u) = w$ wickets lost, and dividing by the average of $x(20, 0)$ over all matches. The sample standard deviation corresponding to $r_{u,w}$ was denoted by $\sigma_{u,w}$. In the case of missing values, BGS set $r_{u,w}$ equal to the corresponding Duckworth-Lewis table entry and $\sigma_{u,w} = 5.0$.

BGS constructed a resource table consisting of the posterior means of the $y_{u,w}$ which were estimated via Gibbs sampling from the full conditional distributions

$$[y_{u,w} \mid \cdot] \sim \text{Normal}[r_{u,w},\ \sigma^2_{u,w}] \tag{3.1}$$

subject to $y_{20,0} = 100.0$, $y_{0,w} = y_{u,10} = 0.0$ and the table monotonicity constraints $y_{u,w} \geq y_{u,w+1}$ and $y_{u,w} \geq y_{u-1,w}$ for $u = 1, \ldots, 20$ and $w = 0, \ldots, 9$.

BGS carried out the sampling using a normal generator and rejection sampling according to the above constraints. The DL method assumes a double exponential decay form on resources. However, BGS used a nonparametric procedure where no functional relationship is imposed on the y's and moreover, they considered the limited over nature of T20 instead of an unlimited over formulation which involves an asymptote in the DL construction.

Having collected data over a longer period than BGS, namely from February 17, 2005 to May 28, 2011, we now have a substantially larger dataset. This dataset has first innings data from all ICC recognized international matches including three T20 World Cups, and in addition to that, the first innings data from four Indian Premier League (IPL) seasons. This gives a total of $n = 388$ matches.

In our study we made one improvement to the estimation procedure used by BGS. Instead of the estimate $r_{u,w}$ we used $(100\%)r_{u,w}$ equal to the average of $x(u, w(u))$ over all matches where $w(u) = w$ wickets lost, divided by the average of $x(20, 0)$ in the corresponding matches. The corresponding $r_{u,w}$ resource matrix is shown in Table 3.1 where missing entries correspond to situations where the relevant stages never existed in any of the 388 matches.

The suggested estimates have advantages in reducing variability. This is shown in the Appendix. It should be noted that we used survey sampling theories when comparing the variances between two estimates. Furthermore these estimates prevent some discrepancies which have been noted in the values shown in the BGS estimated resource matrix.

Table 3.2 shows the re-worked BGS resource table. As described above, this incorporates the improved estimates arrived at by estimating posterior means of the

Table 3.1: The matrix $R = (r_{u,w})$ of estimated resources for Twenty20 (calculated by taking the average of the ratio of $x(u, w(u))$ over all corresponding matches where u overs are available to bat and $w(u)$=w wickets lost). Missing entries correspond to match situations where data are unavailable.

	Wickets Lost									
Overs Available	0	1	2	3	4	5	6	7	8	9
20	100.0									
19	95.9	96.8	98.3							
18	91.2	92.0	92.3	93.9						
17	86.1	87.2	86.4	87.1	88.9					
16	80.9	81.8	82.2	82.6	82.2					
15	75.6	76.2	76.9	77.6	77.6	82.7				
14	69.7	71.1	71.8	72.2	72.8	73.4				
13	65.0	67.8	66.9	68.5	67.9	61.8				
12	60.4	63.4	63.0	63.1	64.7	58.4				
11	56.4	58.4	58.5	59.6	59.7	53.8	42.4			
10	51.3	53.8	53.8	55.6	55.8	46.0	46.7	57.4		
9	46.9	48.8	49.0	50.6	50.9	45.2	41.4	44.8		
8	41.8	44.0	43.9	45.5	46.2	41.8	36.9	41.3		
7	37.8	39.7	39.7	40.1	41.1	38.7	35.3	32.3		11.9
6	35.0	34.2	34.6	34.9	35.5	36.3	31.8	26.5	18.9	4.8
5	19.8	30.5	29.7	29.8	30.1	31.0	29.8	24.5	16.8	4.8
4		23.2	25.0	24.4	24.8	23.9	25.6	21.3	13.0	8.9
3		17.1	19.2	18.9	19.5	18.4	19.7	16.4	13.6	6.2
2		11.1	12.1	13.6	12.8	13.1	13.2	13.6	9.4	7.2
1		5.4	6.2	6.7	6.8	6.5	6.6	6.8	6.7	4.6

y's obtained through Gibbs sampling. The estimates stabilized after 50,000 iterations using R.

3.2 Comparison of the Improved Resource Table with the Existing BGS and DL Tables.

In this section we are going to compare both the existing BGS table and the DL table with our Improved Resource Table (IRT) which is shown in Table 3.2. We use two separate heat maps to effectively display the comparisons, firstly between the IRT and the existing BGS table, and secondly between the IRT and the DL table. To obtain these heat maps we used absolute values of the differences between the candidate tables. These heat maps are shown in Figure 3.1 and Figure 3.2.

As apparent in the heat map shown in Figure 3.1, there are significant differences in the matrix when overs available are between 14 and 20 and wickets lost are between 1 and 5.

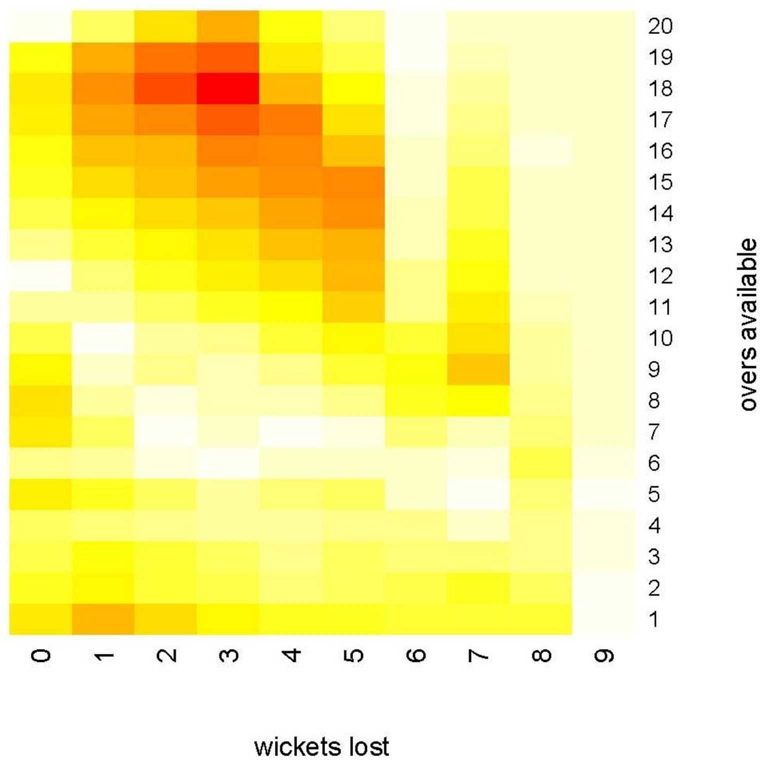
Within this, the biggest differences occur when overs available are 18 and 19 and wickets lost are 2 and 3, with IRT >BGS by approximately 12%. This difference drops to about 8% (IRT >BGS) when overs available is 15 and wickets lost are between 3 and 5. It can be assumed that these differences are due to the IRT having access to data from 388 matches whereas BGS had access to data from only 85 matches. In working out the IRT we had more match situations clustered around the areas which have significant differences. In the rest of the matrix, the differences are not that significant.

As seen in the heat map shown in Figure 3.2, the biggest differences are at the top right corner and in the top middle section (overs available 14-20 and wickets lost 2-5). The differences at the top right corner are due to IRT allocating more resources (18.3%) to the last batsman whereas DL allocates only 8.3%. The IRT allocation is not realistic as the last batsman is usually the weakest batsman. However, this part of the table is not of great relevance as it is extremely unlikely to be ever used; a situation where the last batman is in with 18-20 overs available is an improbable

Table 3.2: The Improved Resource Table (IRT) for Twenty20 based on a slight modification of the methods of BGS as described in section 3.1. The table entries indicate the percentage of resources remaining in a match with the specified number of overs available and wickets lost.

	Wickets Lost									
Overs Available	0	1	2	3	4	5	6	7	8	9
20	100.0	98.6	96.9	93.7	83.9	73.8	60.1	45.4	29.1	18.1
19	98.2	96.7	95.4	91.6	80.6	70.4	56.7	42.9	26.6	15.8
18	95.3	93.4	91.9	90.4	78.5	68.3	54.6	41.3	25.2	14.5
17	91.1	88.3	85.7	82.4	77.1	66.6	52.6	40.1	24.0	13.4
16	86.3	83.3	80.6	77.6	73.3	65.3	50.8	39.0	23.1	12.5
15	81.6	78.5	75.9	73.0	69.3	64.4	49.0	38.1	22.1	11.7
14	77.0	73.9	71.3	68.5	65.3	61.2	47.2	37.1	21.3	11.0
13	72.8	69.6	66.8	64.2	61.2	56.9	45.2	36.2	20.4	10.3
12	68.4	65.3	62.7	60.1	57.1	52.8	43.1	35.4	19.5	9.7
11	64.0	60.9	58.3	55.8	52.7	48.5	40.7	34.6	18.5	9.1
10	59.2	56.1	53.6	51.3	48.2	43.9	38.3	33.9	17.6	8.5
9	54.7	51.7	49.2	46.9	43.8	40.0	35.9	33.2	16.6	7.8
8	50.1	47.2	44.7	42.5	39.7	36.2	32.7	29.1	15.5	7.1
7	45.6	42.5	40.2	38.0	35.5	32.2	28.8	24.2	14.2	6.4
6	40.5	37.5	35.3	33.2	31.1	28.4	24.8	20.4	12.4	5.5
5	32.8	31.0	29.3	27.6	25.8	23.7	20.8	16.9	10.7	4.6
4	29.1	26.5	24.8	23.1	21.3	19.3	16.9	13.5	8.8	3.8
3	23.3	20.1	18.9	17.6	16.4	14.8	12.9	10.1	6.7	2.8
2	17.3	14.1	13.2	12.2	11.2	10.1	8.7	6.9	4.4	1.8
1	10.0	6.1	5.7	5.3	4.8	4.3	3.7	2.9	1.9	0.8

Figure 3.1: Heat map of the absolute differences between the BGS table and the IRT. Darker shades indicate larger differences.



scenario.

In addressing the differences in the top middle section it should be noted that the IRT has allocated 3%-8% more resources than DL. A possible reason for this is the use of the one-day (50 over) DL table as a basis for T20 matches. We feel that some distortion is inherent due to comparing two somewhat non-comparable situations. In one day cricket, by the time 20 overs are left the batsmen have already faced 30 overs and have reached a comfort level with the match conditions (the bowlers, the pitch, the ball etc.) and this is further augmented if only a few wickets have been lost as is the case here. In T20 however, when 20 overs are left, the match is just starting and the batsmen are only beginning to adjust to the match conditions and are therefore more cautious.

Figure 3.2: Heat map of the absolute differences between the DL table and the IRT. Darker shades indicate larger differences

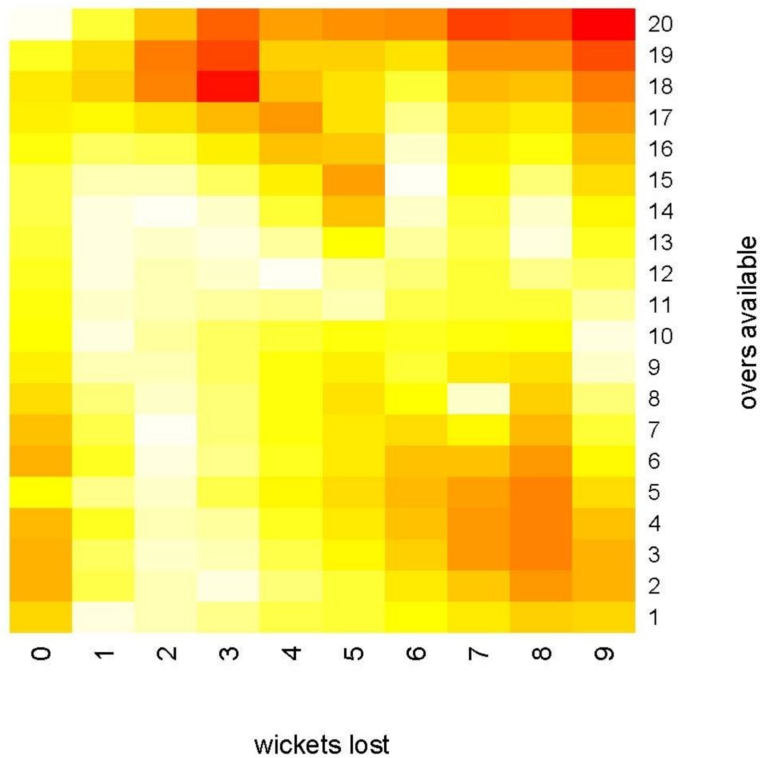


Figure 3.2 also shows minor differences at the bottom of the left column (overs remaining 1-7) where IRT >DL by approximately 3%, and in the bottom right area (overs remaining 2-6 and wickets lost 7-9) where IRT <DL by approximately 5%. The resource allocation given by the IRT appears to be more consistent with the format of the T20 game. In the rest of the matrix, the differences are not that significant.

Chapter 4

A New Resource Table for T20 Cricket

4.1 Data Collection

Actual first innings data from 388 T20 matches were collected from www.cricinfo.com for the period February 17, 2005 to May 28, 2011. These included 146 ICC recognized international T20 matches and 242 IPL matches over four seasons. We excluded data from matches where the first innings was shortened to less than 20 overs as this data would be inappropriate to develop the resource tables. The second innings data were not collected as the side batting second can change their scoring patterns depending on the target score, and this causes the second innings data to be distorted and misleading.

The match summaries at cricinfo provide over by over information and this was collected and transferred into a tabulated form.

4.2 Development of the New Resource Table

After collecting the match results on an over by over basis for each match, we defined $x(u, w(u))$ as the runs scored from the stage in the first innings where u overs are available and $w(u)$ wickets have been taken until the end of the first innings. The variable u took the values $u = 1, \ldots, 20$ and $w(u)$ took the values $w = 0, \ldots, 9$. This was also the initial starting point used by BGS.

A weakness of the BGS procedure is that the observed resource percentages $r_{u,w}$ are calculated as aggregates over all matches. For the estimation of resources, it is desirable to avoid the unnecessary summarization of data, and instead account for the variability of scoring patterns with respect to individual matches. We propose a more refined statistical model which utilizes the individual match data instead of aggregate data using substantially more matches (n=388).

Let $r_i = (r_{i,20}, \ldots, r_{i,n_i})'$ for match $i = 1, \ldots, n$ where $r_{i,u}$ is the percentage of runs scored in match i with u overs available until the end of the first innings, $u = n_i, \ldots, 20$. Note that $n_i > 1$ implies that the batting team used up all of its wickets. For example, $r_{i,20} = 100\%$ and $r_{i,19}$ is the percentage of runs scored in the first innings since the end of the first over. The covariate $w_{i,u}$ is the number of wickets taken at the stage of the ith match where u overs are available.

Let $[A \mid B]$ generically denote the density function or probability mass function corresponding to A given B. Then using conditional probability, the likelihood of the first innings data are given by

$$\prod_{i=1}^{n} [r_{i,20}, \ldots, r_{i,n_i}] = \prod_{i=1}^{n} [r_{i,n_i} \mid r_{i,n_i+1}] \cdots [r_{i,19} \mid r_{i,20}]. \quad (4.1)$$

It should be noted that the likelihood in the above formula is based on the Markov assumption and does not take into account the momentum. Our goal in this project is to obtain a resource table whose entries $\theta_{u,w}$ are the expected percentage of resources remaining when u overs are available and w wickets have been taken. The percentages satisfy the constraints $\theta_{20,0} = 100\%$, $\theta_{0,w} = \theta_{u,10} = 0\%$, $\theta_{u,w} \geq \theta_{u-1,w}$ and $\theta_{u,w} \geq \theta_{u,w+1}$, for $u = 1, \ldots, 20$, $w = 1, \ldots, 9$. Our key modelling assumption is that

$$[r_{i,u} \mid r_{i,u+1}] \sim \text{Normal}[r_{i,u+1} + \theta_{u,w_{i,u}} - \theta_{u+1,w_{i,u+1}},\ \sigma^2] \quad (4.2)$$

which states that the observed change in resources $r_{i,u} - r_{i,u+1}$ is centred about the expected change. As with BGS, the proposed model does not imply a functional relationship on resources; the goal is to allow the data to determine the resource percentages $\theta_{u,w}$, subject to the monotonicity constraints.

4.3 Markov Chain Monte Carlo and Gibbs Sampling

Markov Chain Monte Carlo (MCMC) methods use the values of previous samples to generate the next sample values on a random basis to generate a chain. These methods are a set of algorithms for sampling from probability distributions. The two most popular algorithms for MCMC are:

- Gibbs sampling
- Metropolis-Hastings

The Gibbs Sampling algorithm is the simplest of MCMC algorithms which is used when the full conditional distributions are known and are tractable.

For the estimation of $\theta_{u,w}$, we consider a Bayesian approach where we impose a flat prior $[\theta] \propto 1$ and a standard reference prior $[\sigma^2] \sim$ Inverse Gamma$[1.0, 1.0]$. Using the likelihood (4.1), the modelling assumption (4.2) and the prior specification, we obtain the full conditional distribution

$$[\sigma^2 \mid \cdot] \sim \text{Inverse Gamma}[m_{u,w} + 1,\ (m_{u,w} + 2)/2] \tag{4.3}$$

The remaining full conditional distributions are given by

$$[\theta_{u,w} \mid \cdot] \sim \text{Normal}[\tau_{u,w},\ \sigma^2/(2m_{u,w})] \tag{4.4}$$

subject to the constraints $\max(\theta_{u-1,w}, \theta_{u,w+1}) \leq \theta_{u,w} \leq \min(\theta_{u+1,w}, \theta_{u,w-1})$ for $u = 1, \ldots, 19$ and $w = 1, \ldots, 9$, where $m_{u,w}$ is the number of matches that pass through (u, w),

$$\tau_{u,w} = \frac{1}{2m_{u,w}} \sum r_{i,u} - r_{i,u+1} + \theta_{u+1,w_{i,u+1}} + r_{i,u} - r_{i,u-1} + \theta_{u-1,w_{i,u-1}}$$

and the sum is taken over all matches i that pass through (u, w).

The corresponding $\tau_{u,w}$ resource matrix is shown in Table 4.1 where missing entries correspond to situations where the relevant stages never existed in any of the 388 matches. To obtain the complete table we took entries from the DL Table to replace the missing entries. We have coded a Gibbs sampling algorithm which iteratively simulates from the full conditional distributions in equations 4.3 and 4.4. Using 50,000 iterations, we estimate the posterior means of the $\theta_{u,w}$ leading to the resource table shown in Table 4.2.

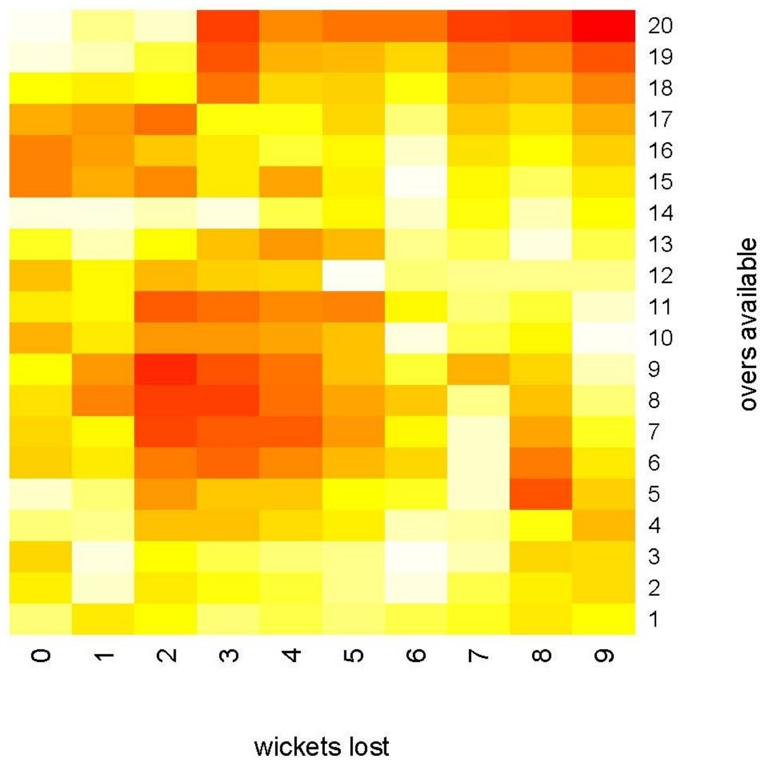
4.4 Comparison of New Resource Table with the IRT and DL Tables

In this section we are going to compare our new resource table (NRT) with the IRT and DL tables. For this exercise we use two separate heat maps where the absolute values of the differences between the relevant tables are compared. These heat maps are shown in Figures 4.1 and 4.2 respectively.

As visible in Figure 4.1 there are four areas where we can see significant differences:

- Overs available 15-20 and wickets lost 0-3, i.e. towards the top left of the map, which represents the very early part of the innings. Here NRT allocates roughly 6% more resources than IRT.
- Overs available 5-11 and wickets lost 2-5, i.e. towards the left lower middle of the map, which is approximately half way through the innings. Here NRT allocates roughly 4%-8% less resources than IRT.
- Overs available 16-20 and wickets lost 7-9, which is represented at the top right corner of the map. However, this is not problematic as it is highly unlikely for match situations to occur here.
- Overs available 4-9 and wickets lost 8, i.e. towards the bottom right corner of the map which is near the end of the innings. Here NRT allocates roughly 3%-7% more resources than IRT.

Figure 4.1: Heat map of the absolute differences between the NRT and the IRT. Darker shades indicate larger differences.



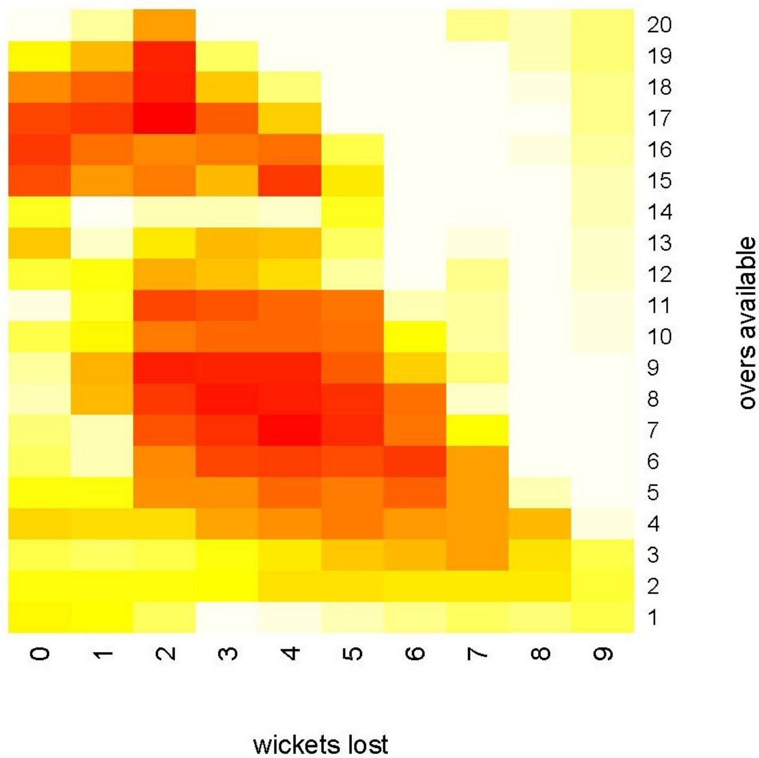
It can be seen therefore that NRT better captures the characteristics of a T20 innings as the batsmen tend to be more cautious at the beginning of the innings and more aggressive towards the middle and the end of the innings. This characteristic is better captured in the NRT due to its enhanced data set.

As can be seen in Figure 4.2 there are two major areas showing significant differences along the main diagonal.

- Overs available 15-20 and wickets lost 0-4, towards the top left of the map, which represents the very early part of the innings. Here NRT allocates roughly 4%-8% more resources than DL.

- Overs available 4-11 and wickets lost 2-6, towards the middle of the map, which

Figure 4.2: Heat map of the absolute differences between the NRT and the DL. Darker shades indicate larger differences.



represents the middle part of the innings. Here NRT allocates roughly 3%-8% less resources than DL.

The reasoning here is the same as the reasoning we expounded when we compared DL vs IRT in Figure 3.2. The batsmen are more cautious in the early part of the innings and become progressively more aggressive towards the middle and the end of the innings.

Although the shading in the lower left corner of the heat map is not as intense as in the diagonal section, the differences in this corner are nevertheless meaningful and are worthy of comment. In the next section we discuss matches where this section of the table is relevant.

4.5 Discussion

In this section we are going to compare the actual target scores set using the DL table with target scores which would have been set using our NRT. These have been applied to two matches where the targets set by the DL method were controversial.

- Match 1: England vs West Indies, 15th June 2009.
  England made 161 for 6 wickets in 20 overs. After rain interruption the West Indies innings was reduced to 9 overs. The DL method was used to reset the target to 80 runs which is 8.9 runs per over. The West Indies achieved the target +2 runs (82 runs) in 8.2 overs (8 overs and 2 balls) i.e. with 4 balls to spare. Their scoring rate was 9.84 runs per over. Applying the NRT, the target would have been 86 runs (9.55 runs per over) which is closer to what was actually achieved, which would have been a more competitive target.

- Match 2: England vs West Indies, 3rd May 2010.
  England batted first and scored 191 for the loss of 5 wickets. After rain interruption West Indies were limited to 6 overs and were set a target of 60 runs by the DL method (10 runs per over). The West Indies reached this target in 5.5 overs (5 overs and 5 balls) with one ball to spare a scoring rate of 10.29 runs per over. Applying the NRT the target would have been 73 runs (12.16 runs per over). Since West Indies had 10 wickets in hand facing only 6 overs the chances of being all out was minimal and so should have been set a more challenging target with greater risk.

In both these examples the targets set using the DL table were controversial. Since our table is based on the actual data from 388 T20 matches, we feel it is more representative of the T20 format than the DL table which was designed for 50 over games. As more T20 matches are played and more data are available we feel that the DL table must be re-visited and checked against actual results.

Finally, we must state that our intention is not to supplant the DL method but to highlight its shortcomings so that it can be improved upon to become less controversial and more acceptable to players, officials, and fans alike.

Table 4.1: The matrix $\tau_{u,w}$ of estimated resources for T20. Missing entries correspond to match situations where data are unavailable.

	Wickets Lost									
Overs Available	0	1	2	3	4	5	6	7	8	9
20	100.0									
19	98.2	98.5	98.0							
18	94.9	94.7	91.8	90.7						
17	91.7	90.0	89.1	87.5	78.7					
16	84.5	83.9	82.0	79.4	75.4					
15	81.6	78.4	76.5	73.8	72.7	65.2				
14	73.7	70.8	68.8	67.8	65.1	62.4				
13	71.1	67.8	65.9	63.2	61.4	57.8				
12	67.2	61.0	57.9	55.4	54.7	53.9				
11	62.1	56.8	53.2	51.8	51.2	48.3	48.6			
10	56.3	50.5	47.2	44.5	43.3	41.2	39.6	43.3		
9	51.9	46.5	42.8	40.6	38.0	36.8	34.2	30.1		
8	46.7	39.9	37.4	34.8	32.3	30.7	28.6	28.4		
7	38.6	35.6	32.9	31.4	28.8	26.8	25.8	24.3		8.0
6	37.2	31.4	28.6	26.6	24.5	22.4	20.4	18.4	16.3	8.9
5	26.0	27.5	24.4	22.4	21.2	19.9	18.0	15.8	16.5	9.4
4		23.6	20.9	18.6	17.0	14.8	14.7	12.7	9.2	9.6
3		20.1	17.5	16.0	14.9	13.0	11.9	9.3	7.6	5.5
2		14.9	12.1	11.7	9.9	9.1	8.7	7.7	6.0	4.0
1		6.3	6.0	4.9	5.5	4.9	4.6	4.2	3.9	2.2

Table 4.2: The New Resource Table (NRT) for Twenty20 based on the method as described in sections 4.2 and 4.3. The table entries indicate the percentage of resources remaining in a match with the specified number of overs available and wickets lost.

	Wickets Lost									
Overs Available	0	1	2	3	4	5	6	7	8	9
20	100.0	97.6	97.3	86.7	78.8	68.2	54.5	38.4	22.0	9.5
19	98.5	97.3	97.1	85.2	76.6	66.6	53.5	37.4	21.6	9.4
18	97.5	95.9	94.0	84.8	75.3	65.0	52.6	37.0	21.3	9.3
17	95.3	93.0	91.4	84.4	75.1	63.4	51.5	36.5	21.1	9.2
16	91.5	87.8	84.2	80.3	75.0	62.9	50.4	36.1	21.0	9.1
15	86.8	82.8	80.9	75.6	73.7	61.9	49.1	35.7	20.8	9.0
14	77.2	73.6	70.7	68.7	63.9	58.8	47.7	35.2	20.7	8.9
13	74.6	69.0	64.6	60.5	56.5	53.0	46.1	34.8	20.6	8.8
12	64.7	63.0	58.9	56.8	53.9	52.7	44.3	34.5	20.4	8.7
11	61.4	58.5	52.0	50.0	47.6	43.3	43.1	33.5	20.2	8.6
10	55.1	53.4	48.8	46.5	43.8	40.2	38.1	32.4	20.0	8.5
9	52.6	47.0	41.7	40.5	38.2	36.3	34.3	29.1	19.7	8.4
8	47.3	41.9	37.8	35.6	33.9	31.3	29.2	28.2	19.2	8.3
7	42.4	40.2	33.4	31.7	29.2	27.7	26.4	24.6	18.5	8.2
6	37.2	34.9	29.8	27.2	26.0	24.5	21.7	19.9	17.8	8.1
5	32.4	32.1	24.6	24.1	22.3	21.6	19.0	17.4	17.2	8.0
4	27.9	27.5	21.1	19.4	18.3	16.8	16.3	14.2	10.8	7.7
3	20.2	19.9	16.8	16.2	15.2	13.9	12.8	10.7	9.8	5.8
2	14.8	14.6	10.5	10.2	9.5	9.1	8.9	8.3	6.9	4.8
1	8.8	8.7	7.8	6.4	6.2	5.5	5.2	4.7	4.5	2.9

Chapter 5

Conclusions

This topic was selected as the basis for my Masters thesis as I felt that the application of the Duckworth-Lewis method to T20 matches has anomalies which need to be addressed. Whilst Battacharya, Gill and Swartz (2010) had already carried out a related study in 2010, the data set available to them was limited to 85 matches. At the present moment we have access to data from 388 matches which vastly improves the base on which to carry out a good statistical study.

The resource table that Duckworth-Lewis (DL) developed in 1998 was aimed solely at solving the problem of resetting targets for interrupted 50 over one-day matches. When T20 cricket came into being, the same DL table was applied with modifications to the number of overs. However no allowance was made for the considerably different scoring patterns and match strategies adopted by T20 batsmen. Though each team still has 10 wickets in hand, their exposure to the risk of getting out was limited to just 20 overs (120 balls) vs the 50 overs (300 balls) of the one day format. This enables the T20 batsmen to take more risks and be much more aggressive, and this in turn impinges on the scoring patterns.

We considered a Bayesian model as it has the capability to impute missing values to deal with monotonic constraints in the resource table. In Chapter 4, in Table 4.2, we introduced a new resource table of posterior mean estimates obtained from constrained Gibbs sampling and argued that it captures the characteristics of the T20 scoring patterns. This suggested that there are indeed differences in the scoring patterns between 50 over one-day matches and T20 matches. Therefore we feel that the DL method needs to be revisited to take into account these patterns.

Finally we want to re-state that our intention is not to supplant the DL method but to highlight its shortcomings for the T20 game so that the DL table can be adjusted as necessary. This will enable the setting of less controversial target scores which will be more acceptable to players, officials, and fans alike.

Bibliography

[1] Bhattacharya, R., Gill, P.S. and Swartz, T.B. (2010). Duckworth-Lewis and twenty20 cricket. *Journal of the Operational Research Society*, doi: 10.1057/jors.2010.175.

[2] Carter, M. and Guthrie, G. (2004). Cricket interruptus: fairness and incentive in limited overs cricket matches. *Journal of the Operational Research Society*, **55**: 822-829.

[3] Christos, G.A. (1998). It's just not cricket. *Mathematics and Computers in Sport* (N. de Mestre and K. Kumar, editors), Bond University, Queensland, Australia, pp 181-188.

[4] Clarke, S.R. (1988). Dynamic programming in one-day cricket - optimal scoring rates. *Journal of the Operational Research Society*, **39**: 331-337.

[5] Duckworth, F.C. and Lewis, A.J. (1998). A fair method for resetting targets in one-day cricket matches. *Journal of the Operational Research Society*, **49**: 220-227.

[6] Duckworth, F.C. and Lewis, A.J. (2004). A successful operational research intervention in one-day cricket. *Journal of the Operational Research Society*, **55**: 749-759.

[7] Jayadevan, D. (2002). A new method for the computation of target scores in interrupted, limited over cricket matches. *Current Science*, **83**, 577-586.

[8] Thompson, S.K. (2002). *Sampling, second edition,* Wiley : Interscience.

Appendix A

Variance Comparison

We consider the estimation of $r_{u,w}$ in chapter 3. For ease of notation, we drop the subscript u, w and let the index $i = 1, \ldots, m$ denote the matches which pass through the stage of the first innings where u overs are available and w wickets have been taken. Further, let x_i be the runs scored from that juncture in the ith match until the end of the first innings, let y_i be the total first innings runs in the ith match, and let z_j be the total first innings runs in the jth match, $j = 1, \ldots, n-m$ where the index j corresponds to the remaining matches in the dataset. The data are of the form

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \ldots, \begin{pmatrix} x_m \\ y_m \end{pmatrix} \text{iid}$$

with $\mathrm{E}(x_i) = \mu_1$, $\mathrm{Var}(x_i) = \sigma_1^2$, $\mathrm{E}(y_i) = \mu_2$, $\mathrm{Var}(y_i) = \sigma_2^2$ and $\mathrm{Cov}(x_i, y_i) = \sigma_{12}$. We assume that the correlation

$$\frac{\sigma_{12}}{\sigma_1 \sigma_2} > 1/2 \tag{A.1}$$

reflecting the tendency that high (low) scoring matches are typically high (low) scoring throughout the match. In addition, there are independent data

$$z_1, \ldots, z_{n-m} \text{ iid}$$

with $\mathrm{E}(z_j) = \mu_2$ and $\mathrm{Var}(z_j) = \sigma_2^2$.

Using the above notation, the BGS estimator takes the form

$$\hat{r}_{\text{BGS}}/100\% = \frac{\sum_{i=1}^{m} x_i/m}{\left(\sum_{i=1}^{m} y_i + \sum_{j=1}^{n-m} z_j\right)/n} = \frac{\bar{x}}{\left(\frac{m}{n}\right)\bar{y} + \left(\frac{n-m}{n}\right)\bar{z}} \tag{A.2}$$

and the "improved" estimator takes the form

$$\hat{r}_{\text{IMP}}/100\% = \frac{\sum_{i=1}^{m} x_i/m}{\sum_{i=1}^{m} y_i/m} = \frac{\bar{x}}{\bar{y}} \ . \tag{A.3}$$

An application of the delta method to the expressions A.2 and A.3 then gives

$$\text{Var}(\hat{r}_{\text{BGS}}/100\%) \approx \frac{\sigma_1^2}{m\mu_2^2} - \frac{2\mu_1\sigma_{12}}{n\mu_2^3} + \frac{\mu_1^2\sigma_2^2}{n\mu_2^4}$$

and

$$\text{Var}(\hat{r}_{\text{IMP}}/100\%) \approx \frac{\sigma_1^2}{m\mu_2^2} - \frac{2\mu_1\sigma_{12}}{m\mu_2^3} + \frac{\mu_1^2\sigma_2^2}{m\mu_2^4}.$$

Now suppose that the variability in runs scored is roughly proportional to the number of expected runs scored (i.e. $\sigma_1/\sigma_2 \approx \mu_1/\mu_2$). Then

$$\begin{aligned} \text{Var}(\hat{r}_{\text{BGS}}/100\%) - \text{Var}(\hat{r}_{\text{IMP}}/100\%) &\approx \tfrac{(n-m)\mu_1}{mn\mu_2^3}\left(2\sigma_{12} - \tfrac{\mu_1\sigma_2^2}{\mu_2}\right) \\ &\approx \tfrac{(n-m)\mu_1}{mn\mu_2^3}\left(2\sigma_{12} - \sigma_1\sigma_2\right) \\ &> 0 \end{aligned}$$

using the correlation assumption A.1. Hence the estimator $\hat{r}_{\text{IMP}}$ is preferred over $\hat{r}_{\text{BGS}}$ in terms of variability.