THE DUCKWORTH-LEWIS METHOD AND TWENTY20 CRICKET

by

Rianka Bhattacharya

M.Sc., Statistics, University Of Calcutta, India, 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

SIMON FRASER UNIVERSITY
Spring 2010

APPROVAL

Name: Rianka Bhattacharya

Degree: Master of Science

Title of project: THE DUCKWORTH-LEWIS METHOD AND TWENTY20 CRICKET

Examining Committee: Dr. Derek Bingham
Chair

Dr. Tim Swartz,
Senior Supervisor
Simon Fraser University

Dr. Paramjit Gill,
Supervisor
Simon Fraser University

Dr. Gary Parker,
External Examiner
Simon Fraser University

Date Approved:

Abstract

Cricket has been a very popular sport around the world. But since most versions of cricket take longer to play than other popular sports, cricket is likely to be affected by unfavourable weather conditions. In 1998, Duckworth and Lewis developed a method for resetting the target scores for the team batting second in interrupted one-day cricket.

Twenty20 is the latest form of cricket. Currently, the Duckworth-Lewis method is also used for resetting targets in interrupted Twenty20 matches. However, this may be less than ideal since the scoring pattern in Twenty20 is much more aggressive than that in one-day cricket.

In this project, we consider the use of the Duckworth-Lewis method as an approach to reset target scores in interrupted Twenty20 matches. The construction of the Duckworth-Lewis table is reviewed and alternate resource tables are presented for Twenty20. Alternative resource tables are constructed in a nonparametric fashion using two different approaches.

Keywords: Duckworth-Lewis method; Gibbs sampling; Isotonic regression; Twenty20 cricket.

To my dear parents, Mr. Arun Prokas Bhattacharya and Mrs. Bani Bhattacharya.

"T20 is a fun game. It is like practising fours and sixes."

— *Kamran Akmal*, 2009 TWENTY20 WORLD CUP *(after being judged the "Man Of The Match" against Netherlands)*, 2009

Acknowledgments

I would like to begin by thanking the faculty in the Department of Statistics and Actuarial Science for their great commitment to their students. Their kind attention to all my queries encouraged me a lot in continuing my studies.

I am truly indebted to my senior supervisor, Dr. Tim Swartz. Without his expert guidance it would not be possible to for me to continue with my studies in Statistics. His support on academic, professional and personal issues have been undoubtably priceless. He was always there to help me to get on to the right direction and to correct my chapters. He taught me how to write scientific papers and how to be persistent in accomplishing the project. It is impossible to thank him enough for his support throughout my studies at SFU. It has really been a great honour for me to pursue my Masters degree under his supervision.

I would like to extend my heartiest thanks to my co-supervisor, Dr. Paramjit Gill who made some valuable suggestions on my project. He helped me a lot, especially in the preliminary stages of the project with the data collection as that has been a real challenge. I would also like to thank my examining committee member, Dr. Gary Parker, for commenting on the project. My special thanks to Ian Bercovitz for his guidance in helping me to learn how to write a paper.

I would also like to thank Dr. Brad McNeney for giving me an oppurtunity to pursue my studies at SFU. The prompt actions of both Dr. Brad McNeney and Dr. Tim Swartz on my visa issues were really appreciated.

I am grateful for the financial support provided by the Department of Statistics and Acturial Science. My sincere gratitude to Sadika, Kelly and Charlene for their kind assistance.

Thanks to all the graduate students in the Department for making my studies at SFU enjoyable. Special thanks to Saman Muthukumarana for helping me to settle down on my arrival and also for his valuable suggestions throughout my studies, especially while working on this project. My thanks to Caroll Co, Aruni Tennakoon, Harsha Perera, Joslin Goh, Jervyn Ang, Ryan Lekivetz, Mathew Pratola, Jia Xu and Monica Lu for your friendship.

Thank you Suddeshna for helping me learn Latex and providing such a friendly atmosphere at home. Thank you to my family for encouraging and standing by my side at all times. Last, but not the least, I would like to thank you Niladri for your understanding and endless support in so many ways.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Different Forms of Cricket and a Brief Background

Cricket is a very popular sport that has been played around the world for a long time. It is a team sport that originated in southern England in the 16th century. By the end of the 18th century, cricket had become the national sport of England. The expansion of the British Empire introduced the game overseas and by the mid-19th century the first international matches were being held. Today, the game's international governing body is the International Cricket Council (ICC). This body is responsible for the organization of major tounaments around the world. As of now, there are 10 full-member ICC countries: Australia, Bangladesh, England, India, New Zealand, Pakistan, South Africa, Sri Lanka, West Indies and Zimbabwe.

One of the features of cricket is that it involves strategy and this is one of the main reasons why cricket fans love the game so much. From a statistician's point of view cricket is a great game since the simplicity of the rules provides a vast oppurtunity for research. A great amount of data have also been collected on cricket matches (www.cricinfo.com).

There exists at least three versions of cricket that are being popularly played at the present time: 1) Test cricket, 2) One-Day cricket, and 3) Twenty20 cricket.

Test cricket is the original form of the sport. It is widely recognized as the ultimate test of playing ability. A test cricket match typically lasts for three to five days. Test matches often end in draws which is unsatisfying to many fans. As a natural response to the intrinsic weakness of first class cricket, "one-day" cricket evolved. Though one-day cricket began between English county teams in 1962, the first one-day international match was played in

Melbourne in 1971. As the name suggests, a one-day cricket match is completed in one day. In this form, each team bats for a maximum of 50 overs. For more information on cricket see, http://en.wikipedia.org/wiki/Cricket.

The popularity of one-day cricket was at its peak when Twenty20 was introduced in 2001 in the Twenty20 Cup between the English counties. The Twenty20 game, also commonly known as T20 cricket, involves two teams, each having a single innings, batting for a maximum of 20 overs. This version of the game is completed in about three and a half hours, with each innings lasting around 75 minutes, thus bringing the game closer to the timespan of other popular team sports. With the introduction of the biennial World Twenty20 tournament in 2007 and the Indian Premier League in 2008, Twenty20 cricket has gained widespread popularity.

1.2 The Duckworth-Lewis Method and One-day Cricket

When considering the substantial amount of research that has been directed towards the sporting world from a mathematical, statistical and operational research perspective, the Duckworth-Lewis method (Duckworth and Lewis, 1998, 2004) perhaps stands as the most significant contribution to sport. This method was first used during the 1999 World Cup and since then it has been adopted by every major cricketing board and competition.

Limited overs cricket (both one-day and Twenty20) is intended to be completed within the stipulated timespans. Thus a major problem in limited overs cricket is that it is intolerant of interruptions due to weather. As a ‘draw’ is contrary to the purpose of limited overs cricket and a definite result is demanded, the Duckworth-Lewis method was introduced as a solution to interrupted matches.

The method used prior to the adoption of the Duckworth-Lewis method in interrupted one-day cricket matches was to award victory to the team with the highest average run rate. However, a difficulty with run rates is that the approach does not take wickets into account. Consequently, the method of run rates is generally seen as unfair.

Alternatively, the Duckworth-Lewis method recognizes the fact that the batting side has two resources at its disposal, the number of overs left and the wickets in hand. A team with ten overs to bat with ten wickets in hand tends to score runs more aggressively than a team with ten overs to bat but say, only two wickets in hand. In one-day cricket, the Duckworth-Lewis method considers resources lost (wickets and overs) when a match is interrupted. In general, people do not understand the method and how targets are set, but they do agree that the targets are sensible and are preferable to the target scores set by the approach based on run rates.

1.3 Motivation of the Project

The Duckworth-Lewis Resource table was designed for one-day cricket matches, but it has also been applied to Twenty20 matches. Although Twenty20 cricket is similar to one-day cricket in many senses, there still exists some subtle variations in the rules (e.g. fielding restrictions, limits on bowling, etc.) between the two forms of cricket. The key difference between one-day and Twenty20 is the reduction of overs from 50 to 20, and this suggests that the scoring patterns in Twenty20 may differ from that in one-day matches. In general, Twenty20 is a more lively form of the game where the ability to score 4's and 6's is more highly valued.

Since the Duckworth-Lewis method and its associated resource table are based on the scoring patterns in one-day cricket matches, one may doubt the use of it in Twenty20. The investigation of the use of Duckworth-Lewis in Twenty20 is the prime focus of this project.

Until now, it might not have been possible to investigate the application of the Duckworth-Lewis method to Twenty20 due to the lack of Twenty20 match results. But now, we have nearly 100 international match results at our disposal, and by the use of efficient estimation procedures the question may be at least partially addressed. Also, since Twenty20 matches are of shorter duration, till now very few matches have been interrupted and resumed according to the Duckworth-Lewis method. Consequently, if there has been a problem with the method being applied to Twenty20, it may not have yet manifested itself.

1.4 Organization of the Project

In chapter 2, we briefly discuss the game of cricket, more precisely Twenty20 cricket. The laws of cricket that apply to this form of the sport are also discussed. Then the standard version of the Duckworth-Lewis table is introduced with a simple demonstration of its use. The construction of the table is also reviewed. Then, we obtain the scaled Duckworth-Lewis resource table so as to make it easily interpretable for Twenty20 cricket. In chapter 3, the construction of a new resource table for Twenty20 cricket is presented. Data from all the international Twenty20 matches involving ICC teams that have taken place from February 17/2005 through November 9/2009 is used in the construction of the table. The methodology used to obtain table entries is based on isotonic regression and Markov chain simulations. The resultant table is compared with the Duckworth-Lewis table. We conclude with a discussion in chapter 4.

This project is an expansion of the companion paper by Bhattacharya, Gill and Swartz (2010).

Chapter 2

Twenty20 and Duckworth-Lewis

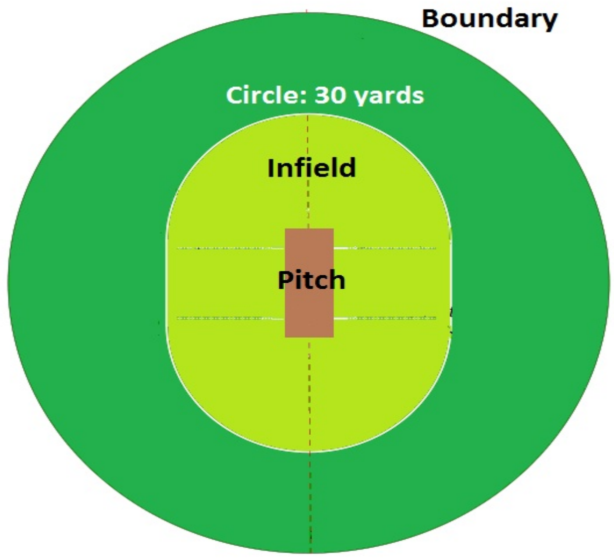
2.1 Twenty20 Cricket: The Game

We begin by describing Twenty20 cricket in greater detail. Most of the rules of Twenty20 cricket are the same as one-day cricket. In Twenty20 cricket, there are two teams (called "sides") each with eleven players. The first team bats its "innings", followed by the second team batting its innings, each innings lasting about 75 minutes. So, each team gets a chance to bat only once. A coin flip prior to the game decides which team has the option to bat first or second. Whoever scores the most runs wins the game.

Cricket is played on a large oval-shaped field with a rope marking the outer edge of the cricket field. A diagram of a typical cricket field is given in Figure 2.1. A rectangular area called the "pitch", with three wooden stakes known as the stumps at each end is located in the middle of the field. These stumps are, in fact, located outside the pitch, though not far away from it. The pitch is ideally 22 yards long. The stumps are usually two feet tall, with two crosspieces called the "bails" atop them. Each set of three stakes with the two bails are collectively known as a wicket. More details concerning the cricket field are provided in http://en.wikipedia.org/wiki/Cricket_field.

At the commencement of an innings, the first two players in the batting order of the batting side come to bat in the field. This is called a "partnership". They bat until one of the batsmen gets out and then the third player in the order comes in to replace the batsman who just made out. This partnership continues until one of these two batsmen gets out. Then the fourth player in the order comes in, and so on.

Figure 2.1: A typical cricket field.



Now, let us describe how the batsmen bat. The two batsmen stand at opposite ends of the pitch. For each ball that is bowled, only one of these two batsmen is the "striker". There are eleven players on each side. Among the players of the fielding side, one player is the bowler and another is the wicketkeeper. The bowler bowls from the opposite end of the pitch to the striker. When a wicket is hit by a ball, and one or both the bails fall off, a batsman gets "stumped out". This is one of the most common methods to get a batsman out. If a ball is hit by the striker, the two batsmen may attempt to change places while the fielders chase the ball and try to throw it back to one of the wickets. A run is scored every time the batsmen change places. However, the batsmen should be careful while changing places as they might get "run out", if the fielder succeeds in hitting the wicket with the ball while they are still on the pitch. This is another way of a batsman getting out. Runs are also scored when the batsman hits the ball and no fielder succeeds in stopping it before the ball touches the boundary line. Four runs are scored if the ball touches the ground at least once before touching the boundary and six runs are scored if it goes over the boundary without touching the ground in between. The wicketkeeper's job is to stop the ball when the striker fails to hit the ball.

There are other ways in which a batsman may be put out by the fielding team. If the ball is hit and a fielder catches it before it touches the ground, then a batsman is out. It is called a "catch out". Another form of getting out occurs when the ball hits the striker's leg and the umpire rules that the ball would have hit the wicket if the leg had not been there. This is called "lbw" (leg before wicket). When a batsman gets out, we say that a wicket has been taken or lost. This plays an important role towards ending an innings.

Basically, there are two ways for an innings to end: when all 10 wickets are lost or when 20 overs have been played. An over consists of six balls bowled. There also exists a third way to end an innings. This happens when the team batting second surpasses the total runs of the team batting first.

In Twenty20, an innings does not go beyond the 20 overs or, in other words, 120 balls. However, sometimes more than 120 balls may be bowled in an innings because of "no balls" and "wides". The former occurs if the bowler performs any illegal action while bowling and the latter is called by the umpire if he feels that the ball is unreachable to the striker. In both cases, the batting side is awarded one run and the ball is also not counted as a part of the over, thus explaining why more than 120 balls are sometimes bowled.

All of the above rules are common to limited overs cricket. The "laws of cricket" which differentiate Twenty20 from one-day cricket are given below.

- The number of overs are reduced from 50 to 20.
- Each bowler may bowl for a maximum of 4 out of 20 overs.
- If the bowler delivers a no ball by overstepping in the pitch, a run is awarded to the batting team and the next delivery is designated as a "free hit". A batsman can be made out from a free hit only by run out and some other uncommon methods.
- No more than five fielders can be on the leg-side at any time during a match. From the point of view of a right-handed batsman facing the bowler, it is the left hand side of the cricket field. With a left-handed batsman the leg-side is the right-handed side of the batsman.
- During the first six overs only two fielders can be outside the 30 yards circle which surrounds the pitch. This period of the match is known as the "powerplay". After the first six overs a maximum of five fielders can be outside the circle.

- Umpires may award five penalty-runs at their discretion if they feel that either team is wasting time.
- If a fielding team does not start to bowl the 20th over within 75 minutes, then the batting team is awarded an extra six runs for every over bowled after the 75th minute.

Thus we can say that Twenty20 is a faster and even more exciting format than one-day cricket. The batting team tends to bat more aggressively as only 20 overs are available to score runs.

2.2 The Duckworth-Lewis Resource Table

The Duckworth-Lewis method is one of the most important statistical contributions to the sporting world. It was developed by two British academics, Frank Duckworth and Tony Lewis in 1998 for resetting targets in interrupted one-day matches. The associated Duckworth-Lewis table has been updated regularly, most recently in 2004, as it is clear that one-day international matches are achieving significantly higher scores than in previous decades. The essence of the table for resetting targets in interrupted limited overs cricket is the concept of resource, namely wickets in hand and overs available. In Table 2.1, an abbreviated version of the Duckworth-lewis resource table (Standard Edition) taken from the 2008-2009 ICC Playing Handbook found at www.icc-cricket.com is provided.

Referring to Table 2.1, it is noted that in a full innings of one-day cricket, a team begins batting with 100% of its resources available, that is 50 overs and ten wickets in hand. A simple demonstration of the use of the Duckworth-Lewis resource table is given below.

Suppose that in a one-day match, the team batting first scores 256 runs upon completion of its innings. It then rains heavily resulting in a delay to the start of the next innings. To make up for the lost time, the second team is told that they can bat for only 30 overs during its innings. In such a situation, it is not fair to make the second team score 257 runs to win. Here, the Duckworth-Lewis resource table is used to reset a target for the second team. According to the resource table, the team batting second has only 75.1% of its resources available. So the target for winning the match is set at $256(0.751) \rightarrow 193$ runs. But if we use the approach based on run rates, then we would set an unreasonably low target of $256(30/50) \rightarrow 154$ total runs. Thus we can see that the Duckworth-Lewis resource table gives a better target in one-day matches.

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
40	89.3	84.2	77.8	69.6	59.5	47.6	34.6	22.0	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
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
1	3.6	3.6	3.6	3.6	3.6	3.5	3.5	3.4	3.2	2.5
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

In constructing the table, Duckworth and Lewis (1998) provide only partial information citing reasons of commercial confidentiality. In the paper it is stated that 20 parameters $Z_0(w)$ and $b(w)$, w = 0,1,...,9 are estimated corresponding to the function

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

where $Z(u, w)$ is the average total score obtained in u overs in an unlimited overs match where w wickets have already been taken. The parameter $Z_0(w)$ 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 lost. However, the estimation procedure is not disclosed. Although we admit the utility of the Duckworth-Lewis resource table in one-day cricket, some questions naturally arise based on equation (2.1) and the estimated values in Table 2.1:

- There are many parametric curves available that could be fit to obtain a resource table. Is (2.1) the best curve to fit? Is there any advantage of using a nonparametric approach?
- The function (2.1) uses an asymptote (i.e. unlimited overs cricket). Is there any advantage in taking the limited overs nature of the one-day game into account?

- What is the estimation procedure? If the 10 curves corresponding to w = 0,1,...,9 are fit separately, sparse data is available beyond $u = 30$ for fitting the curve with $w = 9$. Also, the asymptotes of the curves with w = 0,1,2 (Figure 2.2) fall beyond the range of the data. [Figure 2.2 shows the family of curves described by (2.1) using parameters estimated from hundreds of one-day internationals. Thus we see that Figure 2.2 is the fitted version of average number of runs plotted against the overs remaining for w=0,1,...,9 wickets lost. The figure is taken from Duckworth and Lewis (1998)].

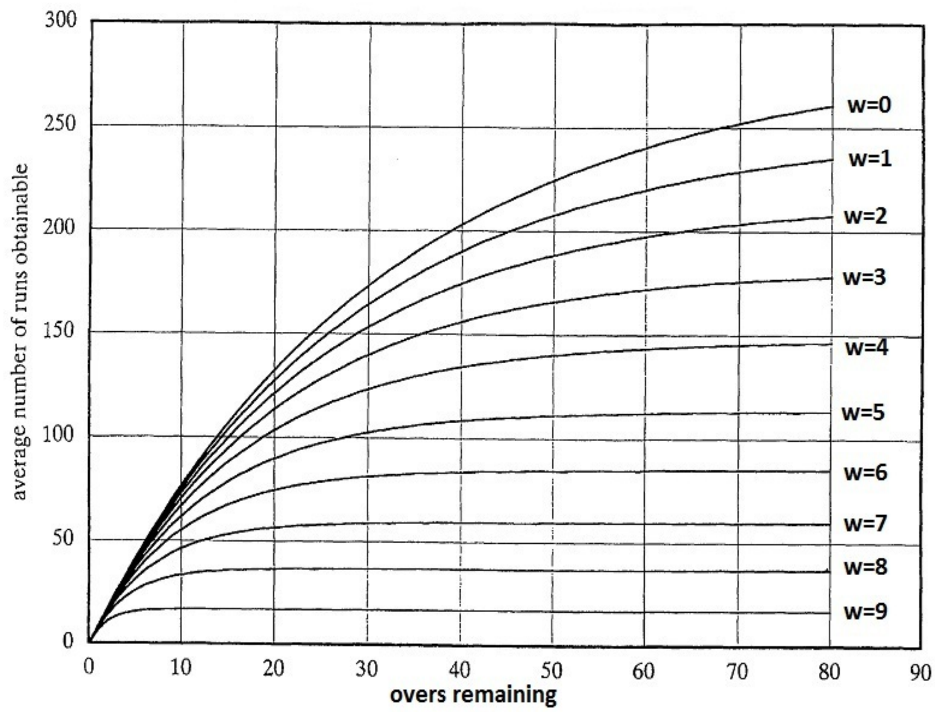
- In Table 2.1, the last two columns have some identical entries going down the columns. Though very few matches occur under these conditions, the question remains whether it is sensible for resources to remain constant as the number of available overs are decreasing. This may be a consequence of the asymptote imposed by (2.1).

For ease of discussion, we convert the Duckworth-Lewis resource table to the context of Twenty20 cricket. That is, we scale the entries in the table so that the resources available to a team with 20 overs and 10 wickets in hand corresponds to 100% resources. In Table 2.2, the full Duckworth-Lewis resource table (Standard Edition) for Twenty20 is given. In this table the entries are obtained by dividing the corresponding entries in Table 2.1 by 0.566 (which is the resources remaining in a one-day match where 20 overs are still available and 10 wickets are available).

Table 2.2: The Duckworth-Lewis resource table (Standard Edition) scaled for Twenty20. 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
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

Figure 2.2: Average number of runs $Z(u, w)$ from u overs remaining with w wickets lost: Duckworth and Lewis (1998).



Chapter 3

New Resource Tables for Twenty20

3.1 Data Collection

Given the questions posed in the previous chapter concerning the Duckworth-Lewis resource table, we consider the construction of new resource tables.

In constructing a new table, it is important to consider the scoring pattern of a Twenty20 match. We therefore consider all international matches played between ICC teams from February 17/2005 through November 9/2009. These matches are of a consistently high standard. There are in total 85 matches used in this project. Here, we have excluded the four shortened matches where the Duckworth-Lewis method was applied. The data were taken from www.cricinfo.com.

In this analysis, we only consider the first innings data. This is because the scoring pattern of the second innings are highly influenced by the number of runs scored in the first innings. Swartz, Gill and Muthukumarana (2009) study the batting behaviour in second innings while developing a one-day cricket simulator.

Our investigation in this project requires ball-by-ball data for the Twenty20 matches. From the Cricinfo website we can see that though the match summary results are readily available, the ball-by-ball information is given in a commentary form. Thus, a Java script was used to parse the associated commentary log for each match to provide a more convenient data structure. This script extracts the relevant details on a ball-by-ball basis, and stores the data in a tabular form for easy access.

3.2 Development of the Resource Table

3.2.1 Construction of R matrix and its Flaws

After running the Java script we get the match results on a ball-by-ball basis. For each match, define $x(u, w(u))$ 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. We calculate $x(u, w(u))$ for all values of u and $w(u)$ that occur in the first innings of a Twenty20 match. The variable u takes the values 0,1,...,20 and $w(u)$ takes the values 0,1,...,10 with $w(20)$=0.

We now define r_{uw} as the estimated percentage of resources remaining when u overs are available and w wickets have been taken. Thus the matrix $R = (r_{uw})$ becomes our first attempt as a new resource table for Twenty20. We calculate (100%)r_{uw} by averaging $x(u, w(u))$ over all the matches where $w(u) = w$ wickets have been taken and dividing by the average of $x(20, 0)$ over all the matches considered. Note that the denominator is simply the average number of first innings runs over all matches. In the case of u=0, we set r_{uw}=r_{0w}=0.0%. In this matrix, $r_{20,0}$=100% is also desired. The calculated matrix is given in Table 3.1.

The new estimated resource matrix R is calculated in a non-parametric fashion and it does not make any assumptions concerning the scoring patterns in Twenty20. In spite of these facts, the estimated matrix is less than ideal. In Table 3.1 there are many missing entries as the data for those situations are missing. There are other flaws associated with this table. For example, the table entries do not exhibit the monotonicity that we expect. Logically, we require a resource table that is decreasing as we go from left to right along the rows and also as we go down the columns. This property of the resource table can be explained by a simple example. Consider the situation in a match where 16 overs are left to bat with 10 wickets in hand and the other situation is that 16 overs are left in a match with only 2 wickets in hand. In the former case the batsmen bat aggressively since a lost wicket is not a big problem. In the latter case, the batsmen bat less aggressively as they try to retain the two wickets in hand. Thus, in the former case the team has more resources available. This explains why the resources are decreasing from left to right along rows. Similarly, we consider two different situations in a match to explain the decreasing property of the table down the columns. Suppose the match situation is such that 16 overs are left to bat with 10 wickets in hand and the other situation is 5 overs left with 10 wickets in hand. Clearly in the former case, the extra overs provide extra opportunity to score runs.

There are also some dubious entries in Table 3.1. One such entry is 110.2% resources corresponding to 19 overs available and 2 wickets taken. This entry is clearly misleading as all the entries in the table should be less than 100%. This particular entry arises due to the fact that the sample size corresponding to the given situation is very small, only 2 matches.

We emphasize that the investigation in this project is one of discovery rather than an attempt to replace the existing Duckworth-Lewis table for Twenty20.

Table 3.1: The matrix $R = (r_{uw})$ of estimated resources for Twenty20 (calculated by taking ratio of the average of $x(u, w(u))$ over all matches where u overs are left to bat and $w(u)$=w wickets have been taken to the average number of first innings runs over all the matches). 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	93.6	83.0	110.2							
18	90.4	85.8	78.3							
17	86.7	80.5	82.8	53.7						
16	81.7	74.5	81.9	70.7	32.8					
15	76.5	71.4	71.5	65.9	59.9					
14	68.3	69.1	67.6	66.2	58.4					
13	63.8	68.2	62.4	62.9	59.0	24.3				
12	62.1	62.3	60.6	57.3	58.8	44.1				
11	60.5	56.3	57.0	53.6	61.0	39.7				
10	57.6	49.6	52.1	52.8	48.1	38.6	41.0	35.2		
9	54.9	52.1	43.6	49.0	44.1	33.8	35.0	29.7		
8	51.0	46.4	41.7	42.2	41.2	36.7	27.5	28.7		
7	48.6	45.8	38.9	35.9	39.1	34.8	24.1	25.5		
6	54.0	37.9	36.6	30.3	36.2	31.3	20.9	21.4	26.7	
5		44.0	32.5	25.4	28.7	29.4	23.9	17.1	14.9	
4			28.2	23.4	22.5	22.2	20.9	14.3	10.6	6.7
3			20.6	19.9	16.9	17.8	15.8	12.4	7.6	1.2
2			21.2	17.6	11.9	13.4	10.6	11.0	7.2	1.4
1				8.7	5.2	7.3	6.0	5.5	6.0	2.6

3.2.2 Isotonic Regression

In numerical analysis, *isotonic regression* in one dimension involves finding a weighted least-square fit $y \in R^n$ to a vector $r \in R^n$ with weight vector $q \in R^n$ that minimizes the function

$$\sum_{i=1}^{n} q_i(r_i - y_i)^2$$

subject to the set of monotonicity constraints $y_i \geq y_j$, $\forall\ i \geq j$.

Isotonic regression is also sometimes referred to as *monotonic regression*. Isotonic regression refers to a trend that is increasing, while monotonic regression implies a trend that is either increasing or decreasing.

Isotonic regression provides an approach to our problem. However, in this problem, there are two independent variables, overs left to bat and number of wickets lost with monotonicity constraints in the rows and columns. For this, we consider the minimization of

$$F = \sum\sum q_{uw}(r_{uw} - y_{uw})^2 \tag{3.1}$$

with respect to the matrix $Y = (y_{uw})$ where the double summation corresponds to $u =$ 1,...,20 and $w =$ 0,...,9. The q_{uw} are the weights and the minimization is subject to the constraints $y_{u,w} \geq y_{u,w+1}$ and $y_{u,w} \geq y_{u-1,w}$. We impose, $y_{20,0} = 100$ and $y_{0,w} = 0$ for $w =$ 0,...,9 and $y_{u,10} = 0$ for $u =$ 1,...,20.

The fitting of Y is completely nonparametric. But still there are some arbitrary choices made in the minimization of (3.1). The "squared error" discrepancy in (3.1) is one of many functions that could be minimized. Also, there are various weights that might be chosen. We note that minimization of the function F in (3.1) with the squared error discrepancy is equivalent to the method of constrained maximum likelihood estimation where the data r_{uw} are independently distributed normal variates with means y_{uw} and variances $1/q_{uw}$.

Again, in this project, we consider the matrix Y : 20 $\times$ 10 based on overs. Instead, we might have considered a larger matrix Y : 120 $\times$ 10 based on balls. But we prefer the former case as it involves less missing data and leads to less computation. Not only that, if we have a matrix Y based on overs, it is always possible to find the Y matrix based on balls by simple interpolation.

Since we are considering a weighted least squares approach, the choice of weights plays an important role and may affect the result. We choose a very simple set of weights. We take $1/q_{uw}$ equal to the sample variance with respect to the calculation of r_{uw}. The reason behind this choice is that when r_{uw} is less variable, y_{uw} should be close to r_{uw}.

In Table 3.2, a nonparametric resource table based on the minimization of (3.1) is given. The algorithm for isotonic regression in two independent variables was first introduced by Dykstra and Robertson (1982). Fortran code was subsequently developed by Bril, Dykstra, Pillers and Robertson (1984). The algorithm for bivariate isotonic regression has recently been updated in R in the "Iso" package on the Cran website (www.cran.r-project.org). In this project, we have used the R code to implement bivariate isotonic regression.

The program requires 27 iterations to achieve convergence. But, this table also has some drawbacks. First, there exist adjacent table entries that are of the same value which is not desirable. Clearly, resources cannot be constant with decreasing overs left or with increasing wickets lost. Secondly, this table also has a large number of missing values corresponding to those match situations where there is no data available. This is natural as it is impossible to have matches, for example where 20 overs are left but at the same time 9 or even 10 wickets have been taken.

3.3 Another Approach in Constructing the R Matrix

In the approach discussed in the previous section, we calculated the R=(r_{uw}) matrix in such a way that (100%)r_{uw} is the ratio of the average of $x(u, w(u))$ over all matches where u overs are left to bat and $w(u)$=w wickets have been taken to the average number of first innings runs over all the matches. In this section, we are going to calculate R differently.

When the numerator and denominator are positively correlated, we know that the standard deviation of the average of ratios is smaller than the standard deviation of the ratio of averages. We use this fact in the alternative calculation of R. We know that $x(u, w(u))$ is the total number of runs scored from that point in a first innings match where u overs are available and w wickets are taken. We therefore define (100%)r_{uw} as the average of the ratios $x(u, w(u))$ to $x(20, 0)$ taken over all the matches where u overs are left and $w(u)$ wickets have been taken. For u=0, we set r_{uw}=r_{0w}=0.0% and also $r_{20,0}$=100% which are desired conditions. The calculated matrix R is given in Table 3.3.

Table 3.2: A nonparametric resource table based on Table 3.1 for Twenty20 using isotonic regression. The table entries indicate the percentage of resources remaining in a match with the specified number of wickets lost and overs available. 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	93.6	85.5	85.5							
18	90.4	85.5	80.8							
17	86.7	80.8	80.8	64.7						
16	81.7	77.4	77.4	64.7	55.9					
15	76.5	71.5	71.5	64.7	55.9					
14	68.8	68.8	67.6	64.7	55.9					
13	66.6	66.6	62.6	62.6	55.9	38.4				
12	62.2	62.2	60.6	57.3	55.9	38.4				
11	60.5	56.8	56.8	54.8	54.8	38.4				
10	57.6	52.1	52.1	52.1	48.1	38.4	34.1	29.3		
9	54.9	52.1	46.5	46.5	44.1	36.3	34.1	29.3		
8	51.0	46.4	42.0	42.0	41.2	36.3	28.6	28.6		
7	48.6	45.8	38.9	37.3	37.3	34.8	25.3	25.3		
6	39.7	39.7	36.6	32.8	32.8	31.3	23.0	21.4	21.4	
5		39.7	32.5	28.0	28.0	28.0	23.0	17.1	15.5	
4			27.9	23.4	22.5	22.2	20.9	14.3	10.7	10.7
3			20.7	19.9	17.4	17.4	15.8	12.4	7.7	7.7
2			20.7	17.6	12.5	12.5	10.8	10.8	7.2	1.8
1				8.7	6.6	6.6	6.0	5.7	5.7	1.8

Table 3.3: The matrix $R = (r_{uw})$ of estimated resources for Twenty20 (calculated by taking the average of the ratio of $x(u, w(u))$ to $x(120, 0)$ over all matches where u overs are left to bat and $w(u)$=w wickets have been taken). 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	95.8	95.9	97.6							
18	91.8	91.1	90.8							
17	86.0	86.1	86.6	81.0						
16	80.2	80.1	83.0	80.7	73					
15	73.6	74.3	76.1	78.4	67.8					
14	67.1	69.3	71.6	73.4	69.2					
13	61.6	66.8	65.6	69.7	67.1	54.1				
12	57.7	61.5	61.0	63.3	66.0	57.6				
11	54.3	55.5	57.0	58.9	63.5	53.1				
10	50.2	50.1	51.2	56.1	53.4	48.6	48.1	57.4		
9	44.6	47.9	45.6	49.9	48.7	42.7	41.3	44.8		
8	40.7	42.2	41.6	42.8	43.7	43.0	35.2	42.5		
7	37.6	41.0	37.0	36.3	41.2	38.1	31.6	37.7		
6	40.8	32.3	34.4	30.5	35.9	34.6	26.2	29.2	38.3	
5		34.5	29.0	26.2	28.1	30.5	28.2	23.7	21.7	
4			23.1	23.9	21.9	22.8	24.0	17.0	15.4	14.9
3			16.7	18.5	17.0	17.2	17.7	14.4	11.4	2.7
2			16.7	14.3	11.6	12.9	11.5	12.4	9.3	2.3
1				7.0	5.3	7.0	6.3	5.9	6.7	3.4

This table also suffers from some of the same problems as Table 3.1. For example, the missing values in the table are a drawback. As before, we ran the isotonic regression program on this estimated resource table. The resultant nonparametric resource table is given in Table 3.4. In this case, the program requires 32 iterations to achieve convergence.

Table 3.4: A nonparametric resource table based on Table 3.3 for Twenty20 using isotonic regression. The table entries indicate the percentage of resources remaining in a match with the specified number of wickets lost and overs available. 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	96.2	96.2	96.2							
18	91.8	91.2	91.2							
17	86.2	86.2	86.2	86.2						
16	81.3	81.3	81.3	81.3	81.3					
15	75.9	75.9	75.9	75.9	75.9					
14	71.3	71.3	71.3	71.3	71.3					
13	67.2	67.2	67.2	67.2	67.2	67.2				
12	62.5	62.5	62.5	62.5	62.5	62.5				
11	58.3	58.3	58.3	58.3	58.3	58.3				
10	53.3	53.3	53.3	53.3	53.3	48.7	48.7	48.7		
9	53.3	48.1	48.1	48.1	48.1	45.3	45.3	45.3		
8	53.3	43.1	43.1	43.1	43.1	43.1	43.1	43.1		
7	53.3	41.0	38.6	38.6	38.6	38.6	38.6	38.6		
6	53.3	39.0	34.4	33.8	33.8	33.8	33.8	33.8	33.8	
5		39.0	29.0	28.5	28.5	28.5	28.5	28.5	28.5	
4			27.1	23.9	22.6	22.6	22.6	19.6	19.6	19.6
3			27.1	18.5	17.2	17.2	17.2	14.9	14.9	14.9
2			27.1	14.3	12.2	12.2	11.9	11.9	9.3	3.8
1				7.0	6.6	6.6	6.3	6.1	6.1	3.8

Table 3.4 also cannot be regarded as ideal. In fact, we hoped that the new approach would give us a more reliable resource table. But in this table we also see equal adjacent values. This table also suffers from missing entries. Thus though we hoped to see a better resource table, it can hardly be regarded as an improvement to the previous one at this stage.

3.4 Gibbs Sampling

Markov Chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The most popular algorithms for MCMC are

- Metropolis-Hastings
- Gibbs sampling

Gibbs sampling is a MCMC algorithm that is applicable when the associated conditional distributions are tractable.

We saw that the nonparametric resource tables generated by isotonic regression have many shortcomings. To address those criticisms, we now use Gibbs sampling as another approach to estimation. Note that the minimization of (3.1) arises from the maximization of the normal likelihood

$$\exp\{-\frac{1}{2}\sum\sum \quad q_{uw}(r_{uw} - y_{uw})^2\} \tag{3.2}$$

Thus, we now consider a Bayesian model where the unknown parameters in (3.2) are the y's. A flat default prior is assigned to the y's subject to the required monotonicity constraints as mentioned in (3.1). Thus the posterior density takes the form (3.2) and Gibbs sampling can be carried out via sampling from the full conditional distributions

$$[y_{uw} \mid \cdot] \sim \text{Normal}(r_{uw}, \frac{1}{q_{uw}}) \tag{3.3}$$

subject to the local constraints on y_{uw} in the given iteration of the algorithm. Sampling from (3.3) is easily carried out using a normal generator and rejection sampling according to the constraints. Here it should be mentioned that though we are taking a parametric form for the random variables in (3.3), we still refer to the approach as nonparametric since we are not assuming any functional relationship that is imposed on y. In Table 3.5 and Table 3.6 we present the resource matrices that we get after carrying out Gibbs sampling using the R matrices from Table 3.1 and Table 3.3 respectively. These are the alternate resource tables for Twenty20 cricket.

In Table 3.5, the table entries are the estimated posterior means of the y's obtained through Gibbs sampling. In this project we have used Fortran code where the estimates stabilize after 50,000 iterations. Note that the R matrix has many missing entries. So to address this problem, we actually impute the missing r_{uw}'s with the corresponding entries in the scaled Duckworth-Lewis table given in Table 2.1. This imputation is in the spirit of a Bayesian approach as we are using prior information. A nice feature about this table is that it is a complete resource table and devoid of any missing values. No adjacent table entries are identical. Thus one major drawback of the previous resource tables are been overcome. We should also mention that the approach is able to take into account expert opinion. For example, if there is an opinion that the table entry y_{ij} should be a particular value k, then to set the table entry to that value we simply take $r_{ij} = k$ and assign a sufficiently small standard deviation.

Table 3.5: A nonparametric resource table for Twenty20 based on Gibbs sampling using the R matrix in Table 3.1. 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
20	100.0	96.9	93.0	87.9	81.3	72.2	59.9	44.8	29.7	17.6
19	95.6	90.9	87.7	83.0	76.9	68.3	56.5	42.0	27.2	15.3
18	91.7	86.7	82.9	78.7	73.2	65.4	54.2	40.2	25.7	13.9
17	87.7	82.3	78.9	73.8	69.7	62.8	52.2	38.7	24.6	12.8
16	83.5	78.2	75.3	70.5	66.4	60.2	50.3	37.4	23.5	12.0
15	79.2	74.3	70.9	66.9	62.6	57.4	48.4	36.2	22.7	11.2
14	75.1	70.7	67.3	63.7	59.3	54.6	46.4	35.0	21.8	10.5
13	71.5	67.4	63.6	60.3	56.2	51.5	44.3	33.8	21.0	9.8
12	68.3	63.7	60.2	56.8	52.9	47.5	41.9	32.6	20.2	9.1
11	65.0	59.9	56.6	53.3	49.7	43.9	39.3	31.3	19.4	8.5
10	61.3	56.0	52.6	50.1	46.0	40.8	36.1	30.0	18.6	7.9
9	57.9	52.3	47.9	46.1	42.5	37.8	33.1	28.3	17.7	7.2
8	54.0	48.3	44.3	41.7	38.9	34.9	30.2	26.1	16.7	6.6
7	49.3	44.2	40.2	37.4	35.4	32.1	27.2	23.4	15.7	5.9
6	41.7	38.5	35.7	33.0	31.7	29.0	24.2	20.0	14.5	5.2
5	36.2	33.4	31.0	28.6	27.3	25.5	21.5	17.0	12.2	4.4
4	30.8	28.0	26.1	24.1	22.4	20.7	18.3	14.2	10.0	3.5
3	25.4	22.8	21.1	19.4	17.7	16.5	14.4	11.6	7.9	2.5
2	19.7	17.2	15.5	14.1	12.7	11.9	10.6	9.3	6.2	1.6
1	13.7	11.3	9.7	8.5	7.3	6.7	6.0	5.2	4.2	0.9

Table 3.6 is the resource table that we obtained from Gibbs sampling using Table 3.3 as the R matrix. Again, we impute the missing values with the corresponding Duckworth-Lewis table entries given in Table 2.2. The resultant resource table has some advantages over the earlier resource tables. Note that the table entries are slightly different from those in Table 3.5. In the next section we compare the two suggested resource tables with the scaled Duckworth-Lewis resource table given in Table 2.2.

Table 3.6: A nonparametric resource table for Twenty20 based on Gibbs sampling using the R matrix in Table 3.3. 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
20	100.0	98.0	95.1	89.8	82.5	72.7	60.1	44.8	29.8	17.8
19	96.0	95.5	92.4	85.9	78.6	69.0	56.8	42.1	27.3	15.5
18	91.9	91.0	89.5	82.9	75.7	66.3	54.5	40.2	25.8	14.1
17	86.7	86.1	85.3	80.5	73.2	63.9	52.6	38.8	24.7	13.1
16	81.6	81.0	80.5	78.2	71.1	61.7	50.7	37.5	23.7	12.2
15	76.2	75.5	75.1	74.3	68.7	59.5	49.0	36.3	22.8	11.5
14	71.9	71.3	70.9	70.0	66.6	57.5	47.2	35.1	22.0	10.8
13	67.8	67.0	66.3	65.8	63.8	55.5	45.4	34.0	21.2	10.1
12	63.8	62.6	61.9	61.3	60.1	53.7	43.4	32.8	20.4	9.5
11	59.7	58.3	57.7	57.1	56.2	51.1	41.4	31.5	19.6	8.9
10	54.8	53.8	53.3	52.9	51.5	47.1	39.1	30.2	18.8	8.3
9	51.3	49.5	48.3	47.8	46.8	42.8	36.9	28.8	18.0	7.7
8	47.5	44.9	43.3	42.5	41.6	39.5	34.1	27.4	17.1	7.1
7	44.5	41.4	39.3	38.3	37.7	36.2	31.1	26.0	16.2	6.5
6	40.9	36.6	34.9	33.6	33.1	32.2	28.0	24.4	15.2	5.8
5	36.1	33.5	30.8	29.0	28.4	27.9	25.4	21.9	14.1	5.2
4	30.5	27.5	25.4	24.0	22.8	22.0	20.9	17.2	12.8	4.4
3	24.9	22.1	20.2	18.8	17.7	16.9	15.9	13.9	10.8	3.4
2	19.7	17.3	15.7	14.1	12.6	12.2	11.5	10.9	8.5	2.4
1	13.4	10.9	9.2	7.8	7.0	6.7	6.3	5.8	5.2	1.6

3.5 Comparison of the Suggested Resource Tables with the Duckworth-Lewis Resource Table

In this section we compare the suggested resource tables with the scaled Duckworth-Lewis resource table for Twenty20 cricket presented in Table 2.2. Since our nonparametric resource tables are based on Twenty20 matches and the Duckworth-Lewis resource table was developed for one-day matches, differences might be expected. Thus to facilitate a meaningful comparison, we take the absolute values of the differences between each of the two suggested nonparametric resource tables with the Duckworth-Lewis resource table and produce two separate heat maps. The heat maps are respectively presented in Figure 3.1 and Figure 3.2.

From Figure 3.1, we see that the greatest absolute differences occur in three different regions. First, large differences occur in the top-right hand corner and bottom left-hand corner of the heat map. These regions are not so important as they correspond to unlikely match situations. However, it should be noted that our nonparametric resource table (Table 3.5) assigns more resources in these two regions than the Duckworth-Lewis resource table.

For example, let us consider a Twenty20 match with a single over remaining and two wickets lost. Assume that the match averages 150 runs. In such a situation, the Duckworth-Lewis resource table suggests that $(0.064)150 = 9.6$ runs is expected in the final over for an average of $9.6/6 = 1.6$ runs per ball. Our resource table (Table 3.5) suggests $(0.097)150 = 14.6$ runs in the final over for an average of $14.6/6 = 2.4$ runs per ball. It may be noted that in such a match situation, a talented batsman might reasonably score 2.4 runs per ball. Thus assigning more resources does not seem to be absurd.

We also observe discrepancies in the middle of the innings, that is when 8 to 13 overs are remaining and 3 to 6 wickets are lost in a match. In these cells of Table 3.5, we can see that there are about 5% fewer resources available to the batting team than is provided by the Duckworth-Lewis resource table. This coincides with our intuition since the scoring pattern in Twenty20 is more aggressive than in one-day cricket. Since there are 50 overs in one-day cricket, the batsmen need to protect their wickets for a longer period while scoring runs. We remark that a difference of 5% resources may be very meaningful as a target of 240 runs diminished by 5% gives a target score of 228 runs.

To demonstrate this point let us consider a Twenty20 match with 11 overs left and five wickets lost. Let the match average be 240 runs. The Duckworth-Lewis resource table suggests that $(0.491)240 \to 118$ runs is expected from the remaining overs while our resource table (Table 3.5) suggests that (0.439)240 $\rightarrow$ 106 runs is expected. It is interesting to note that in spite of the aggressive batting style of Twenty20 cricket the table is suggesting a lower target score. As an explanation, Twenty20 has more aggressive batting than one-day cricket throughout the innings. One-day cricket begins its innings in a less aggressive fashion. Therefore, at the "halfway" point of an innings, Twenty20 has used up more of its resources. Thus the Duckworth-Lewis table (Table 2.2) does not seem to capture this characteristic as it is based on scoring patterns of one-day cricket. So, the target of 106 runs seems to be more sensible in Twenty20 cricket.

Figure 3.1: Heat map of the absolute differences between the Duckworth-Lewis resource table (Table 2.1) and the nonparametric resource table based on Gibbs sampling (Table 3.5). Darker shades indicate larger differences.

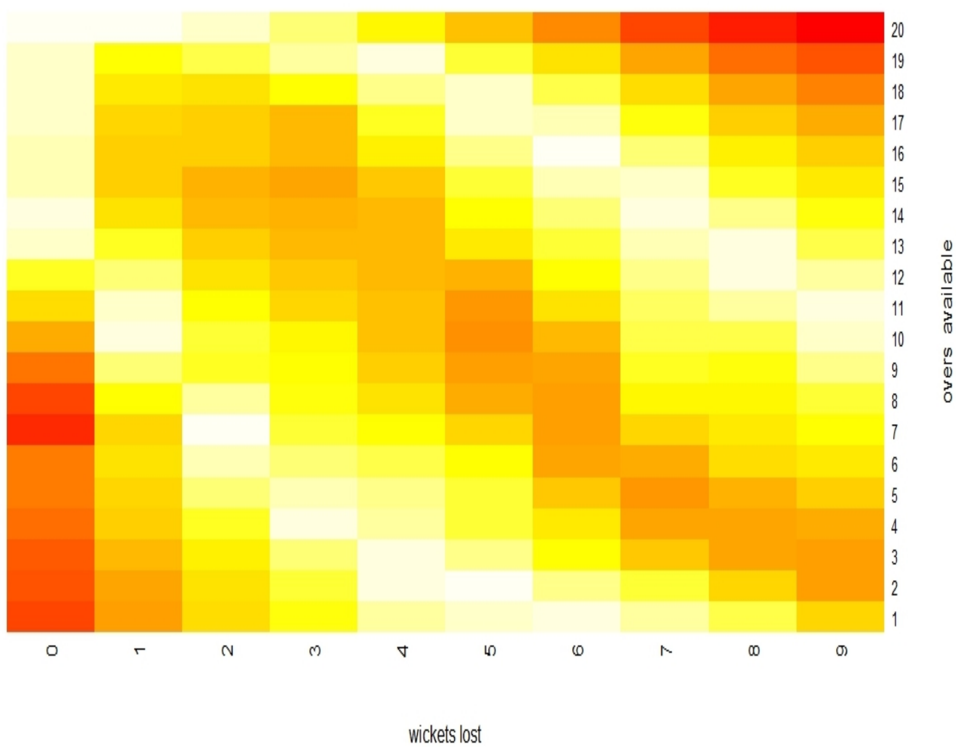


Figure 3.2 shows the heat map of the absolute differences between the resource table (Table 3.6) and the Duckworth-Lewis resource table for Twenty20 (Table 2.2). In this figure, we can

see that the greatest difference lies in the top-right hand corner and the bottom-left hand corner of the heat map. We can also see a little difference when 11 to 16 overs are remaining and 4 wickets are lost. In these situations we see that the new resource table assigns nearly 6% to 8% more resources than the resource table in Table 3.5. Thus it assigns more resources in these situations than the Duckworth-Lewis table in Table 2.2, when our intuitions tells us that fewer resources should be available in these areas. A meagre difference also lies when 2 to 4 overs are left and 9 wickets are lost, though it can be regarded as insignificant.

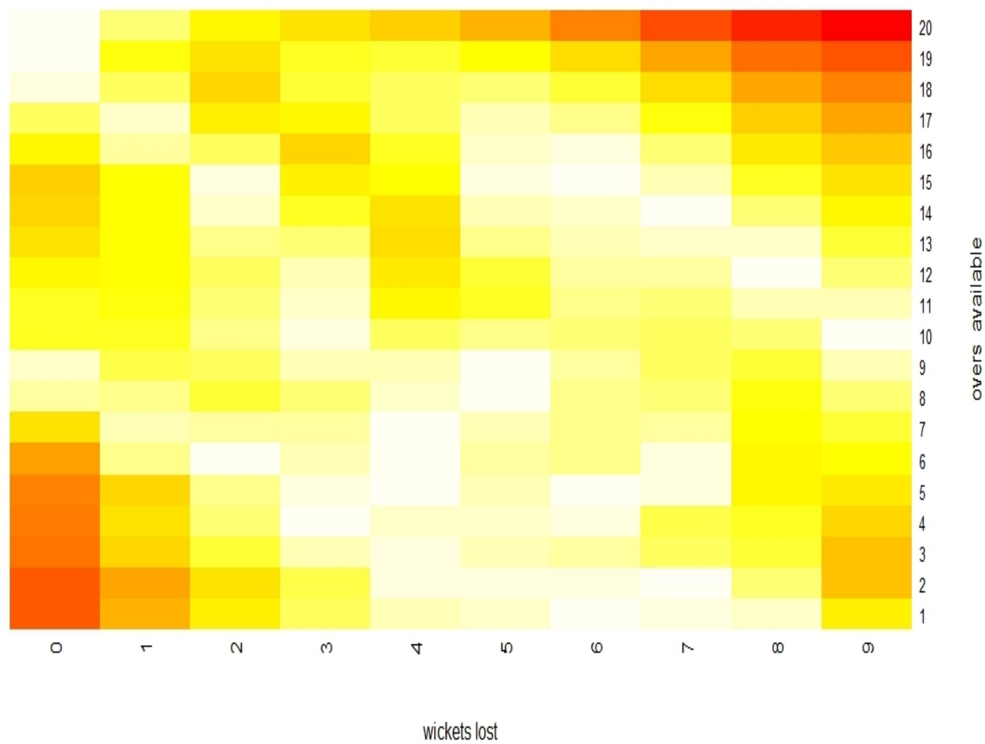
The heatmap given in Figure 3.3 shows the absolute difference of the two new suggested resource tables (Table 3.5 and Table 3.6). These two tables differ in situations where 8 to 15 overs are left with no wickets lost and also where 9 to 19 overs are available with 1 to 5 wickets taken. As mentioned earlier, Table 3.6 assigns more resources than Table 3.5 in these areas whereas that is not desirable. These situations arise often in Twenty20 cricket matches. We thus regard Table 3.5 preferable to Table 3.6 since Table 3.6 does not seem to capture the aggressive scoring pattern of the Twenty20 game.

3.6 Discussion

In this section, our intention is to provide a discussion about the use of alternate resource tables for Twenty20. We recall that we have only considered international Twenty20 matches played between ICC teams up to November 9, 2009. Until that time there were only four such instances where the Duckworth-Lewis method was applied.

To demonstrate our proposed resource table, we revisit the match between England and West Indies during the 2009 World Cup on June 15, 2009. This was a very important game as the winner advanced to the semi-finals. In the first innings, England scored 161 runs at the expense of 6 wickets using 100% of their resources. The second innings was shortened to 9 overs with a target of 80 runs for West Indies. West Indies scored 82 runs in 8.2 overs with 5 wickets still in hand and eliminated England from the tournament. The English fans were upset and a report in the Guardian claimed that the Duckworth-Lewis method will be reviewed to take into account Twenty20 matches (http://www.cricinfo.com/ci-icc/content/story/409482.html). Note that in this match, the Professional Edition of the Duckworth-Lewis Table (which is not available from the 2008-2009 ICC Playing Handbook) was used to set the target score of 80. If we use the scaled

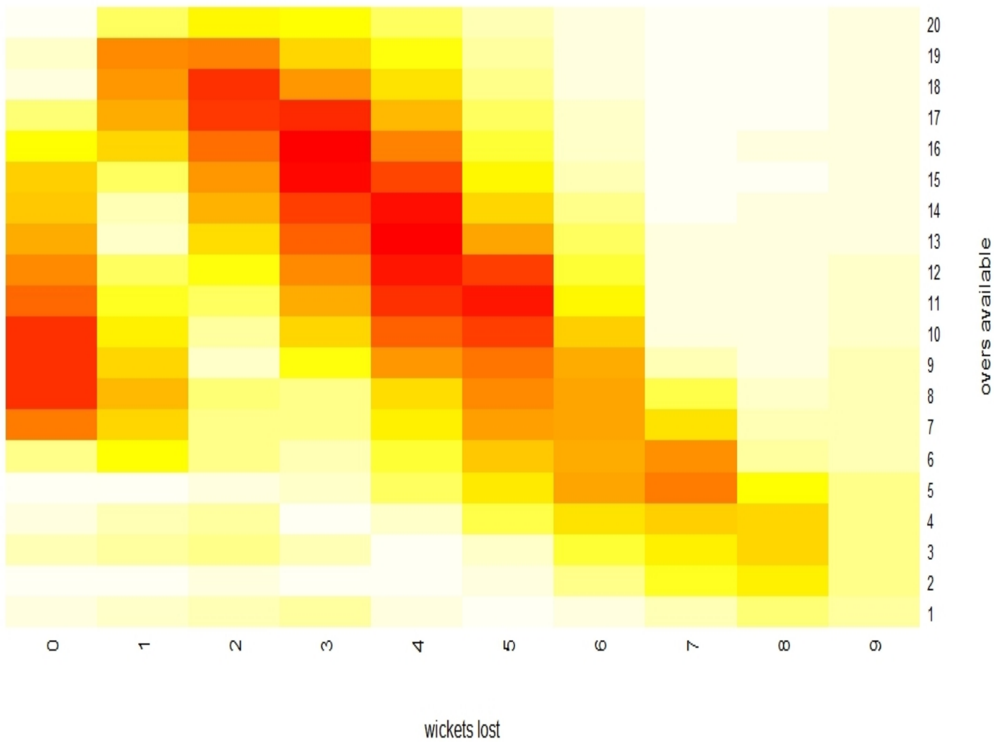
Figure 3.2: Heat map of the absolute differences between the Duckworth-Lewis resource table (Table 2.1) and the nonparametric resource table based on Gibbs sampling (Table 3.6). Darker shades indicate larger differences.



Duckworth-Lewis Table given in Table 2.2 (based on the Standard Edition), the target score would have been (0.518)161 $\rightarrow$ 84 runs. However, the alternative resource table in Table 3.5, sets an even higher target score of (0.579)161 $\rightarrow$ 94 runs. And if we use the alternate resource table in Table 3.6, the target score is set at (0.513)161 $\rightarrow$ 83 runs.

In a Twenty20 match between England and South Africa on November 13, 2009, the Duckworth-Lewis method was again used. In that match, England scored 202 runs losing 6 wickets in the full first innings. But due to rain, the second innings was reduced to 13 overs with a target score of 129 using the Duckworth-Lewis method. But South Africa scored 127 runs losing 3 wickets in 13 overs. Thus England won the match. To set the target score the Professional Edition of the Duckworth-Lewis table was used. Using Table 2.2 (based on the Standard Edition), the target score is (0.71)202 $\rightarrow$ 144 runs. If we use our Table 3.5,

Figure 3.3: Heat map of the absolute differences between the nonparametric resource table based on Gibbs sampling (Table 3.5) and the nonparametric resource table based on Gibbs sampling (Table 3.6). Darker shades indicate larger differences.



the target score is (0.715)202 $\rightarrow$ 145 runs and Table 3.6 gives a target of (0.678)202 $\rightarrow$ 137 runs. We see that all targets are higher than the target score used in the match. Taking into account the aggressive scoring pattern of the Twenty20 game, we can say that 145 runs is a preferable target in 13 overs where the average runs scored in the first innings is 202.

Both of these examples question the validity of the existing Duckworth-Lewis method when applied to Twenty20. It is important to emphasize that we are not claiming our nonparametric resource table (Table 3.5) is a replacement for the Duckworth-Lewis resource table in Twenty20. Our resource table is based on only 85 matches. This is too small a sample to provide confident table entries. However, the table does suggest that there may be a significant difference in the scoring pattern of one-day cricket and Twenty20 cricket. When more Twenty20 matches become available, we endorse a review of the use of the

Duckworth-Lewis method in Twenty20 and the associated estimation techniques.

Chapter 4

Conclusion

The goal of this project was to develop a resource table for the resetting of target scores in interrupted Twenty20 matches. In 1998, Duckworth and Lewis developed a resource table for one-day cricket. However, after the introduction of Twenty20 cricket, the existing Duckworth-Lewis table is used for resetting target scores in Twenty20. But it is a known fact that the scoring pattern in Twenty20 is much more aggressive than it is in one-day cricket. Thus the question arises whether it is sensible to use the Duckworth-Lewis table for Twenty20.

In this project we tried to produce an alternate resource table taking into account the scoring pattern of the Twenty20 game. We tried to frame our resource table in a nonparametric fashion using isotonic regression. But due to the lack of data or no data in some of the match situations, the resultant resource table remains incomplete. We therefore considered a Bayesian model as it has the capability to impute missing values. Posterior mean estimates obtained from Gibbs sampling provide us with a new nonparametric resource table for Twenty20 cricket.

In chapter 3, we argued that Table 3.5 exhibits the characteristics of the scoring patterns of Twenty20 game. This table suggests that there are differences between the scoring rates in the two types of limited overs cricket. We reiterate that we are not claiming that our resource table is a viable replacement of the existing Duckworth-Lewis resource table since our table is based on a very small amount of data. But with the availability of more data, a review of the Duckworth-Lewis approach should be undertaken.

Bibliography

[1] Bhattacharya, R., Gill, P.S. and Swartz, T.B. Duckworth-Lewis and Twenty20 Cricket. *under review*, 2010.

[2] Bril, G., Dykstra, R.L., Pillers, C. and Robertson. Isotonic regression in two independent variables. *Journal of the Royal Statistical Society*, 33:352–357, 1984.

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

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

[5] Dykstra, R.L. and Robertson, T. An algorithm for isotonic regression for two or more independent variables. *Annals of Statistics*, 10:708–716, 1982.

[6] Swartz, T.B., Gill, P.S. and Muthukumarana, S. Modelling and simulation for one-day cricket. *The Canadian Journal of Statistics*, 37:143–160, 2009.