Ranking ODI Cricketers in the Bradley-Terry Framework

Peter Matthews

September 2021

1 Introduction

This report is on work produced for the Undergraduate Research Support Scheme (URSS) at Warwick University under the supervision of Dr Nicholas Tawn and Ian Hamiton.
It tackles the ranking of batsmen and bowlers in ODI cricket, which can be challenging for a number of different reasons. Traditionally, the accepted measure of batting ability is the average, defined as

$$Average = \frac{Runs}{Wickets}$$

where runs are the total number of runs that a batsman has scored or the total number of runs that a bowler has conceeded in their career. A higher average is better for batsmen and a lower average is better for bowlers.

However, this is flawed as it fails to account for batsmen facing different bowlers at different stages of a match, reflecting differing levels in the difficulty of scoring runs. It also fails to account for the speed of scoring runs which is important in limited-over cricket. The later criticism of traditional averages is often addressed by also reporting a batter's strike rate - $\frac{runs\ scored}{balls\ faced} \times 100$

The former criticism is improved on by working within the paradigm of models proposed by Bradley and Terry (1952) and expanded on by Davidson (1970); Firth (2017); Hamilton (2018) to produce two models measuring each batter's and bowler's ability on the dimensions of wicket-taking and protecting, and run scoring and saving. A chief benefit of this approach is that wickets taken or lost and runs scored or conceded are sufficient statistics in each respective model, ensuring that they are in agreement with traditional measures of strike rate under idealised conditions. A more unified ranking of ability is presented based on the intuitive measure: how much would a team, made up of copies of the same batsmen, score in 50 overs against a team made up of "average" bowlers. This is calculated from monte-carlo simulations of the models of wicket-preserving and run-scoring ability.

2 Data

This project uses ball-by-ball data obtained from `https://cricsheet.org/downloads/` for all One Day International (ODI) matches played between ICC Full Member nations over a 4 year period from 1st June 2017 to 1st July 2021. Ball level data includes the names of the bowler, striker and non-striker; the over and ball of the innings; the number of runs scored off the bat; whether a wicket fell; and in the case of a wicket - the outgoing batsman, and the method of dismissal. Match and innings level data is also included such as the date, the venue, the bowling team, the batting team, and the innings of the match. All balls that where there were extras or a run out are excluded from the analysis, as these runs and wickets are not traditionally credited to the batter and bowler respectively.

3 Models

This section explains the formulation of Bradley-Terry based models. In each case we want a model of player ability that is consistent with the total runs scored (or equivalently, strike rate) and wickets lost in the hypothetical scenario that each batsman faces each bowler for the same number of balls in the same match situation. This is ensured if wickets taken or lost and runs scored or conceded are sufficient statistics in their respective models. We also aim to have the simplest possible models so that all factors that a player can control for themselves are absorbed into their ability.

3.1 Wicket-Based Model

In order to measure the wicket-protecting ability of each batter and wicket-taking ability of each bowler, we can consider each ball in an ODI to be a trial. A batter "wins" the trial if they successfully survive the delivery, whereas a bowler "wins" if they take the batters wicket. In the context of pairwise comparison trials the Bradley-Terry model shown below is standard. Here, $\mu_{i,w}$ and $\lambda_{j,w}$ is the positive valued "ability" of the i^{th} batter and j^{th} bowler, respectively. The probability of batter i surviving a given ball delivered by bowler j can be formulated as follows

$$P(i\ survives\ j) = \frac{\nu_w \mu_{i,w}}{\nu_w \mu_{i,w} + \lambda_{j,w}} \tag{1}$$

I include an order effect ν_w, comparable to that used to the home advantage effect often used in Bradley-Terry models for team sports, as most balls do not result in a wicket so in any given ball the batter has a significant advantage. ν can be interpreted as the number of balls the average batter would be expected to survive against the average bowler before getting out.

The model can also be expressed as a logistic regression which allows it to be fitted as a generalised linear model in R.

$$logit(i\ survives\ j) = log(\nu) + log(\mu_{i,w}) - log(\lambda_{j,w}) \tag{2}$$

The current model is not identifiable and requires further constraints on the parameters. To see this consider $\mu^*_{i,w} = \alpha\mu_{i,w}$, $\lambda^*_{j,w} = \beta\lambda_{j,w}$, $\nu^* = \frac{\beta}{\alpha}\nu$, we then have

$$logit(i\ survives\ j) = log(\nu) + log(\mu_{i,w}) - log(\lambda_{j,w}) = log(\nu^*) + log(\mu^*_{i,w}) - log(\lambda^*_{j,w})$$

The parameters can vary over 2 degrees of freedom, which are suppressed by adding 2 constraints to the model shown below

$$\sum_i log(\mu_{i,w}) = 0 \tag{3}$$

$$\sum_j log(\lambda_{j,w}) = 0 \tag{4}$$

These constraints ensure that for both batters and bowlers, their log-abilities sum to 0, or equivalently, the geometric means of $\mu_{i,w}$ and $\lambda_{j,w}$ are both equal to 1.

3.2 Run-Based Model

To measure run-scoring and run-saving ability, the aim is to create a model that has runs scored as a sufficient statistic for reasons explained above. Following the precedent of Firth (2017) and Hamilton (2018), a natural choice would be to construct a likelihood where the ability of each player is raised to the power of the number of runs scored for each ball. We can equivalently think of a batter as scoring runs - raising their ability $\mu_{i,r}$ the positive power - and a bowler as conceding that same amount of runs - thus raising their ability $\lambda_{j,r}$ to the negative of that power. The parameters $\nu_{0,r}, \nu_{1,r} \ldots \nu_{6,r}$ represent the relative frequencies of each run scoring outcome for the average player. We can set $\nu_{0,r} = 1$ for identifiability.

$$P(i\ scores\ 0\ from\ j) \propto \nu_{0,r}$$

$$P(i\ scores\ 1\ from\ j) \propto \nu_{1,r}\left(\frac{\mu_{i,r}}{\lambda_{j,r}}\right)$$

$$P(i\ scores\ 2\ from\ j) \propto \nu_{2,r}\left(\frac{\mu_{i,r}}{\lambda_{j,r}}\right)^2$$

$$\ldots$$

$$P(i\ scores\ k\ from\ j) \propto \nu_{k,r}\left(\frac{\mu_{i,r}}{\lambda_{j,r}}\right)^k$$

Let μ be the vector of batsman ratings, λ the vector of bowlers ratings, and ν the vector of parameters $\nu_{0,r}, \nu_{1,r} \ldots \nu_{6,r}$. Define $N^{\mathbf{0}}$ as the matrix whose $(i,j)^{th}$ entry is the number of times batter i faced a dot ball from bowler j, and similarly define the matrices $N^{\mathbf{1}}$, $N^{\mathbf{2}}$, $N^{\mathbf{3}}$, $N^{\mathbf{4}}$, and $N^{\mathbf{6}}$. Let N_μ be the number of batsman, N_λ, be the number of bowlers, and $N^T = N^{\mathbf{0}} + N^{\mathbf{1}} + N^{\mathbf{2}} + N^{\mathbf{3}} + N^{\mathbf{4}} + N^{\mathbf{6}}$ be the total number balls that batter i faced from bowler j. Finally let $K = \{1, 2, 3, 4, 6\}$ and $N = \{\ N^k : k \in K \cup \{\ T\}\}$

We can use this to formulate a likelihood function $\mathcal{L}_r(\mu, \lambda, \nu; N)$ and show that runs scored r_μ and runs conceded r_λ are a sufficient statistic for the calculating player ratings.

$$\mathcal{L}_r(\mu, \lambda, \nu; N) = \prod_{i=1}^{N_\mu} \prod_{j=1}^{N_\lambda} \frac{\prod_{k \in K} \left(\nu_k \left(\frac{\mu_i}{\lambda_j} \right)^k \right)^{N_{i,j}^{\mathbf{k}}}}{\left(1 + \sum_{k \in K} \nu_k \left(\frac{\mu_i}{\lambda_j} \right)^k \right)^{N_{i,j}^T}} \tag{5}$$

Defining $r_i^s := \sum_{j=1}^{N_\lambda} \sum_{k \in K} k \times N_{i,j}^{\mathbf{k}}$, $\quad r_j^c := \sum_{i=1}^{N_\mu} \sum_{k \in K} k \times N_{i,j}^{\mathbf{k}}$, and

$n_k := \sum_{i=1}^{N_\mu} \sum_{j=1}^{N_\lambda} N_{i,j}^{\mathbf{k}}$. We can re-write

$$\mathcal{L}_r(\mu, \lambda, \nu; N) = \frac{\prod_{i=1}^{N_\mu} \mu_i^{r_i^s} \prod_{j=1}^{N_\lambda} \lambda_j^{-r_j^c} \prod_{k \in K} \nu_k^{n_k}}{\prod_{i=1}^{N_\mu} \prod_{j=1}^{N_\lambda} \left(1 + \sum_{k \in K} \nu_k \left(\frac{\mu_i}{\lambda_j} \right)^k \right)^{N_{i,j}^T}} \tag{6}$$

Thus the statistic (r^s, r^c, n) is a sufficient statistic for (μ, λ, ν).

We can view the vector $[N_{i,j}^{\mathbf{1}}, \ldots, N_{i,j}^{\mathbf{6}}]$ to be multinomially distributed with $N_{i,j}^{\mathbf{T}}$ trials. And from the likelihood we have the probability of falling into the category corresponding to k runs as

$$P(k \mid i, j, \mu, \lambda, \nu) = \frac{\nu_k \left(\frac{\mu_i}{\lambda_j} \right)^k}{\left(1 + \sum_{k \in K} \nu_k \left(\frac{\mu_i}{\lambda_j} \right)^k \right)} \tag{7}$$

By setting $a_{i,j} = log\left(N_{i,j}^T\right) + log\left(1 + \sum_{k \in K} \nu_k \left(\frac{\mu_i}{\lambda_j} \right)^k \right)$ we have the log of the expectation of $N_{i,j}^{\mathbf{k}}$

$$log\left(\mathbf{E}\left[N_{ij}^{\mathbf{k}} \mid i, j, \mu, \lambda, \nu\right]\right) = log(\nu_k) + k\left[log(\mu_i) - log(\lambda_j)\right] + a_{ij} \tag{8}$$

where a_{ij} is a nuisance parameter that accounts for both the denominator of the likelihood and the number of balls faced. This can be fit in the model but is not of particular interest as we already have information on the ratings of players from the numerator. This allows us to find ratings by fitting a log-linear model. Furthermore, for the purposes of maximum likelihood estimation there is no difference between Multinomial and Poisson models as shown by Birch (1963) (A moddified argument is provided in A). This permits the computationally efficient fitting of the model as a log-link Poisson regression using R's gnm package - Turner and Firth (2007).

Similarly to the wicket-based model there needs to be 2 extra constraints on the parameters for identifiability, so we require the mean log-ability of the batters and bowlers to both be zero.

$$\frac{1}{N_\mu}\sum_{i=1}^{N_\mu} log(\mu_i) = 0 \tag{9}$$

$$\frac{1}{N_\lambda}\sum_{j=1}^{N_\lambda} log(\lambda_j) = 0 \tag{10}$$

3.3 Model Extensions

The original models proposed are an improvement in the sense that they account for the differences in the difficulty faced by batters (bowlers) due to the abilities of the different bowlers (batters) they face. However, a fairer model would also account for other challenges faced by each player that they do not have direct control over. This subsection looks at some of these challenges and how they are included as co-variates in the model.

3.3.1 Phase of Match

The clearest example of effects on the ability of players is the stage of an innings. In general, at later stages of the innings there are fewer balls left with which to score runs and so the cost of loosing a wicket is smaller. This results in batters being more aggressive at later stages in the innings and so more runs are scored and more wickets fall than at earlier parts of the innings. This is accounted for by adding a factor variable, phase, which splits the match into 10 stages (5 in each inning) of 10 overs. In each stage the average player would be expected to get out with a different frequency and so we can model this by changing ν_w. We can calculate 10 separate parameters subset-ed by phase l, and the equation becomes

$$logit(i\ survives\ j \mid l, \mu, \lambda, \nu) = log(\nu_{w,l}) + log(\mu_{i,w}) - log(\lambda_{j,w}) \tag{11}$$

For the runs model we consider a more parsimonious solution, we model the phase l as having a multiplicative effect e^{ζ_l} on the ability of each player, relative to the base scenario - at the start of the first innings. This similar to order effect extensions of Bradley-Terry models shown in... It is an attractive option as this treatment introduces 9 new parameters to the model (one for each non-base phase) whereas the alternative of calculating separate $\nu_{k,l}$s for each phase would increase the number of parameters in the model by $6 \times (10-1) = 54$. The log of the expectation of $N^{\mathbf{k}}_{i,j,l}$ is then

$$log\left(\mathbf{E}\left[N^{\mathbf{k}}_{i,j,l} \mid i,j,l,\mu,\lambda,\nu\right]\right) = log\left(\nu_k\right) + k\left[log(\mu_i) - log(\lambda_j) + \zeta_l\right] + a_{ij} \quad (12)$$

3.3.2 Home Advantage

Playing at home has a particular advantage due to a player's familiarity with the pitch and the climate conditions which influence ball trajectories, as well as the common advantages in sports such as having support from home crowds. We consider playing at home to have a similar multiplicative effect $\mathcal{H}$ on playing ability in both the wicket and run based models to be calculated separately. Broadly, major cricket playing nations can be categorised into Asia (Afghanistan, Bangladesh, India, Pakistan, Sri Lanka) and "SENA" (Australia, England, Ireland, New Zealand, South Africa, West Indies, Zimbabwe). There is a broad difference in the climate and pitches between either region and so there is often an advantage to playing in the your home region even if not in your home country. For example, New Zealand are considered to have an advantage over India when playing in England. As such, the models also include a region effect $\mathcal{R}$. Defining $\gamma_{\mathcal{H}}$ and $\gamma_{\mathcal{R}}$ as follows, the log-linear representation of the model is shown in (11)

$$\gamma_{\mathcal{H}} = \begin{cases} log\left(\mathcal{H}\right) & i \text{ at home} \\ 0 & \text{neutral venue} \\ -log\left(\mathcal{H}\right) & j \text{ at home} \end{cases}$$

$$\gamma_{\mathcal{R}} = \begin{cases} log\left(\mathcal{R}\right) & i \text{ in home region, } j \text{ not in home region} \\ 0 & i \text{ and } j \text{ have the same home region} \\ -log\left(\mathcal{R}\right) & i \text{ not in home region, } j \text{ in home region} \end{cases}$$

$$log\left(\mathbf{E}\left[N^{\mathbf{k}}_{i,j,l} \mid i,j,l,\mu,\lambda,\nu\right]\right) = log\left(\nu_k\right) + k\left[log(\mu_i) - log(\lambda_j) + \zeta_l + \gamma_{\mathcal{H}} + \gamma_{\mathcal{R}}\right] + a_{ijl} \quad (13)$$

3.4 DLS Runs Model

A further extension of the model was also considered that incorporate measures from Duckworth and Lewis (1998). These measures are used to frequently to calculate targets in rain-affected matches, and so are widely considered to be fair and accurate. An advantage of this approach is that it allows for a more refined measures of how aggressive a batsman needs to be given the match context. Rather than purely a function of innings phase, how aggressive a batsman should be is a function of both overs remaining and wickets in hand.

While one could derive and incorporate a number of different functions that account for this, the "resources" defined by Duckworth and Lewis (1998) are a fair approximate measure. Importantly for a ranking context, the DLS Method is used frequently, batters are aware of the par score and resources used during an innings, and they form part of a method that is already deemed to be fair. This makes using them in a ranking model more transparent then defining separate, previously unknown, proxies for agression. Let $R(o, w)$ be the $(0, w)^{th}$ entry of the DLS Resource table (Appendix B). we consider the proxy $r_{l,w}$ for how aggressive a batter can afford to be during the l^{th} over, while w wickets down defined as

$$r_{l,w} = R(o, w) - R(o + 1, w) \tag{14}$$

which can be interpreted as the resources used during the l^th over, given that a wicket is not lost. This reflects a batters capacity for aggression, which is an advantage when measuring run-scoring ability. As such, we assume in our model that $r_{l,w}$ has a linear effect on log ability.

In the second innings, a DLS derived co-variate is used as a proxy for the pressure induced by the run chase. The DLS par score is commonly presented during 2nd innings as the score that the batting team would need to be declared the winner in the event of a game-ending interruption at that point in the match. As such, if it over l we define d_l as

$$d_l = runs\ at\ start\ of\ over - par\ score\ at\ start\ of\ over$$

d_l in effect shows how far ahead or behind the chasing team are, and consequently, how much more aggressive the batter has to be. So it is also assumed that d_l will have a linear impact on log-ability, resulting in the log-linear model 15.

$$log\left(\mathbf{E}\left[N^{\mathbf{k}}_{ijlw}\right]\right) = log\left(\nu_k\right) + k\left[log(\mu_i) - log(\lambda_j) + \beta_r r_{lw} + \beta_d d_l + \gamma_{\mathcal{H}} + \gamma_{\mathcal{R}}\right] + a_{ijlw} \tag{15}$$

One disadvantage of this approach is that, as the co-variates are more granular, pooling the data results in a much greater number of pools than in the inital models. This greatly increases the order of the model matrix, making fitting the model more computationally expensive. Due to a lack of computing power available, the DLS-based runs model was therefore only fit on a small subset of matches.

4 Model Fits and Selection

The AIC - Akaike (1974) - of each model is compared with its extensions below. The model with the lowest AIC is shown in bold. The baseline model are referred to as "Wickets" and "Runs" respectively, with "+ [Covariate]" added when the corresponding effect is included.

4.1 Wicket Models

In the wicket-based models, the region effect did not improve the likelihood when the home-advantage effect was also included, so there was a lack of evidence to suggest that there is a particular advantage to playing away in the same region over playing in the opposite region. Incorporating phase effects had the most dramatic improvement in AIC, and the best perforing model (by AIC), "Wickets + Phase + Home", is used in subsequent analysis.

Model	AIC
Wickets	23000.56
Wickets + Home	22996.35
Wickets + Home + Region	22996.35
Wickets + Phase	20911.53
Wickets + Phase + Home	**20909.58**

Table 1: Wicket-Based Models and their AICs

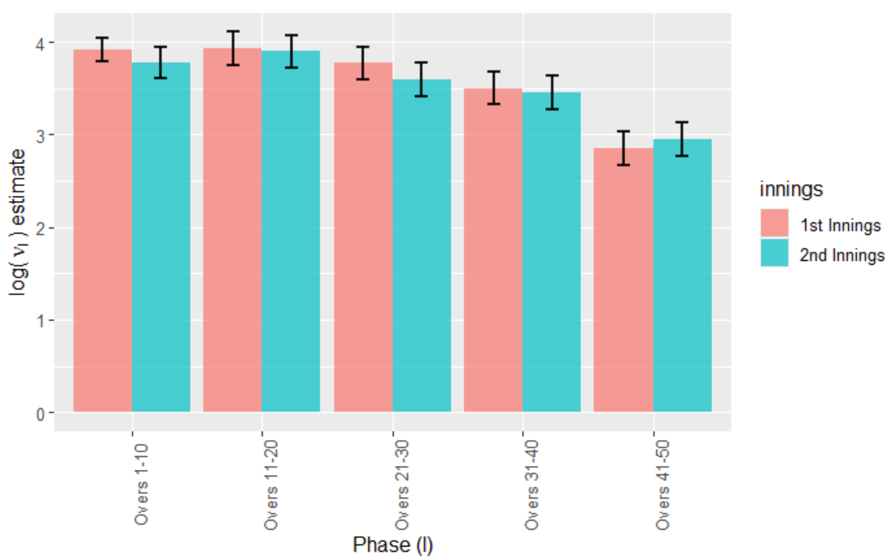

Figure 1: Parameter estimates of $log\left(\nu_l\right)$ for each phase l of the innings

1 shows the estimates of each $log\left(\nu_l\right)$ for each ten over phase in the model "Wickets + Phase + Home". Error bars presented are 2 times the quasi-standard error of the estimate. It shows that, as one would expect, surviving a given ball becomes harder as the innings progresses, and that this effect is magnified towards the end of the innings.

The best batters and bowlers in "Wickets + Phase + Home" are shown in 2 and 3 along with quasi-standard errors.

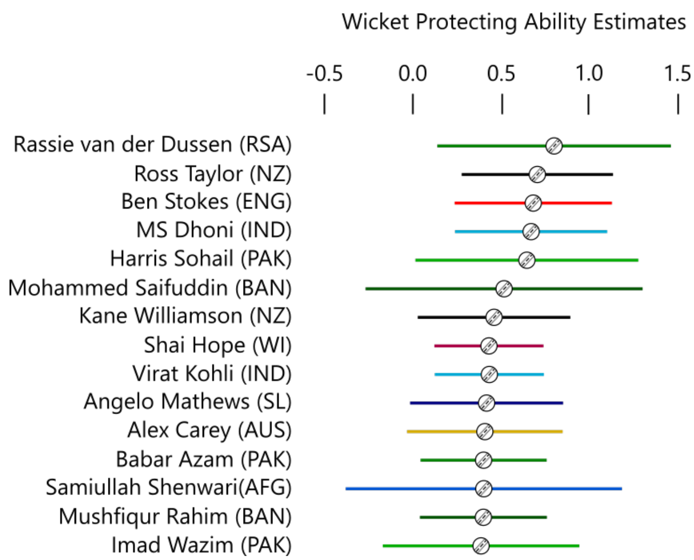

Figure 2: Parameter estimates of $log\left(\mu_{w,i}\right)$ for the best 15 batters i

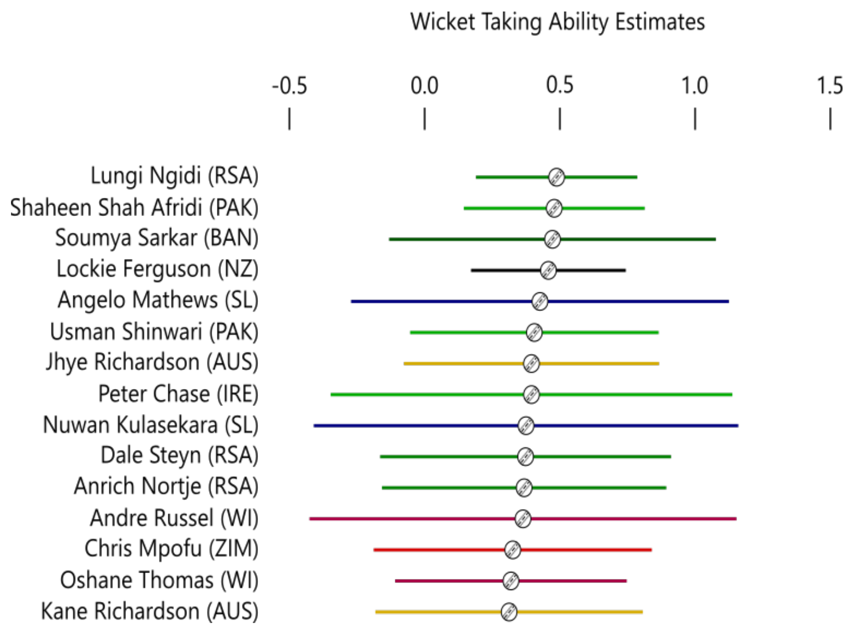

Figure 3: Parameter estimates of $log\left(\lambda_{w,j}\right)$ for the best 15 bowlers j

4.2 Run Models

Model	AIC
Runs	477389.4
Runs + Home	477389.4
Runs + Home + Region	477389.4
Runs + Phase	475257.1
Runs + Home + Phase	475257.1
Runs + Home + Phase + Region	475257.1

Table 2: Runs-Based Models and their AIC

The likelihood was not improved by adding a home or region advantage parameter in any instance, and so they were dropped from the model. There was, however, a significant improvement when phase effects were included. As such, "Runs + Phase" is the model used in subsequent analysis. Further research is needed to establish why home advantage is significant in the prevention of wickets but not in the scoring of runs. One explanation based on anecdotal evidence could be that an effective counter-tactic for batters playing on unfamiliar away pitches is to "swing hard", causing them to get out more frequently as well as scoring runs more quickly, which counter-balances the effect of it being easier to score runs on familiar pitches. There is however insufficient data to properly evaluate this claim.

5 Expected Score

Given results from the wicket-based and runs-based models, there is a problem of how to adequately combine them into one ranking. The rest of this paper considers this problem in light of ranking batters, although similar procedures could be used to rank bowlers. The advantage of working within a probabilistic framework is that model outputs can be used to find probabilities of unobserved events. Given a batter has a wicket-preserving ability μ_w and a run-scoring ability μ_r. For a given phase of the innings l, against the "average" bowler $\lambda_w = \lambda_r = 1$ the model equations show the probability of a wicket, and the probability of each run-scoring outcome given no wicket falls. This is used to simulate a hypothetical 1st innings of a team made up of 11 copies of a given batter against 11 average bowlers. Batters can then be ranked based on how much they would be expected to score if given 50 overs and 10 wickets against the average bowler, presumed to be the first innings of a match at a neutral venue where both teams are from the same region. This was used to define the metric Expected Score (xS). xS was calculated by simulating 1,000 team innings for each batsman, where upon each ball, the probability of a wicket was calculated from the wicket model and whether or not a wicket fell was determined. Then the expected number of runs scored for that ball - given that a wicket did not fall - was calculated from the runs model and was added to the total. Each innings was complete when either all 300 balls has been simulated or when 10 wickets had fallen. The results of the best batters by xS are shown below.

Player	$log(\mu_w)$	$log(\mu_r)$	xS
AB de Villiers	0.10714	0.21815	**371.54**
JJ Roy	-0.16378	0.25390	**370.17**
JM Bairstow	-0.03698	0.23551	**368.75**
S Dhawan	0.24414	0.19724	**363.37**
CH Gayle	-0.08341	0.21993	**352.58**
DA Warner	0.19906	0.16966	**341.49**
MDKJ Perera	0.02472	0.17268	**333.38**
Fakhar Zaman	0.07601	0.16472	**329.37**
RG Sharma	0.46781	0.14030	**328.87**
SO Hetmyer	-0.06195	0.17572	**326.95**

There is a prevalence of attacking batters across the best performers in xScore, as illustrated by the fact that the mean $log(\mu_r)$ amongst the top 10 batters is 0.194781 (and all are positive), whereas the mean $log(\mu_w)$ amongst the group is only 0.077276 - with 4 of the 10 batters having a below average wicket-protecting ability. One could argue that this reflects that attacking batsmen are underrated by the more traditional measures of "Average" as well as the recent trends towards more attacking batters in limited overs matches. However, it could be argued that comparing batsmen on the basis of what each would make as an entire team slightly inflates the value of aggressive players. Amongst the best batters in the world, the batters coming in after them in a real match are likely to not be as able to score runs as easily as themselves. As such, the run scoring potential of the team is hurt more by such batters losing their wickets in a real match, then it would be if that batter was allowed to replace their-self, as in this hypothetical. I do not believe there is a fair way to account for this as doing so would turn the question into "Which batters are the best value for their given team?" Such a question, whilst interesting in its own right, would depend on the quality of a player's teammates, making it separate from the idea of ranking each batsmen on the same scale. Despite this slight flaw, the paradigm presented is still beneficial in that it combines measures of wicket-protecting and run-saving ability in a way that acknowledges that the trade-off between runs and wickets is not constant and varies depending on the match situation. It does so in an intuitively explainable and consistent way, which gives it a strong philosophical basis to be used in ranking.

References

Akaike, H. (1974). "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19(6), pp. 716–723. DOI: `10.1109/TAC.1974.1100705`.

Birch, MW (1963). "Maximum likelihood in three-way contingency tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 25(1), pp. 220–233.

Bradley, Ralph Allan and Milton E Terry (1952). "Rank analysis of incomplete block designs: I. The method of paired comparisons". In: *Biometrika* 39(3/4), pp. 324–345.

Davidson, Roger R (1970). "On extending the Bradley-Terry model to accommodate ties in paired comparison experiments". In: *Journal of the American Statistical Association* 65(329), pp. 317–328.

Duckworth, Frank C and Anthony J Lewis (1998). "A fair method for resetting the target in interrupted one-day cricket matches". In: *Journal of the Operational Research Society* 49(3), pp. 220–227.

Firth, David (2017). *Maths. Football. That's all.* URL: `https://alt-3.uk/` (visited on 09/13/2021).

Hamilton, Ian (2018). *Ranking in Schools Rugby*. URL: `https://warwick.ac.uk/fac/sci/statistics/staff/research_students/ihamilton/msc_dissertation_ian_hamilton_30aug18.pdf` (visited on 09/13/2021).

Turner, Heather and David Firth (2007). "Generalized nonlinear models in R: An overview of the gnm package". In:

A Equivalence between Poisson and Multinomial MLE

Consider 2 experiments, A - with N trials of a Multinomial Distribution with K categories - and B - with K independent Poisson random variables. Assume that in experiment A, the probability of falling in the k^{th} category is given by $p_k(\theta)$ and

$$\sum_{k=1}^{K} p_k(\theta) = 1$$

For experiment B, the k^{th} row is Poisson distributed with parameter $Np_k(\theta)$. Let x_k be the number of times that a result falls in the k^{th} category, as well as the value of the k^{th} row of experiment B. We impose that $\sum_{k=1}^{K} x_k = N$ for the Poisson experiment, although this is trivially true in the Multinomial case, and so must be true if both experiments have the same data.

Consider $\mathcal{L}_A$ and $\mathcal{L}_B$, the likelihoods of A and B respectively.

$$\mathcal{L}_A = \frac{N!}{\prod_{k=1}^{K} x_k!} \prod_{k=1}^{K} (p_k(\theta))^{x_k} = N! \prod_{k=1}^{K} \frac{(p_k(\theta))^{x_k}}{x_k!}$$

$$\mathcal{L}_B = \prod_{k=1}^{K} \frac{(Np_k(\theta))^{x_k}}{x_k!} e^{-Np_k(\theta)}$$

because $\sum_{k=1}^{K} p_k(\theta) = 1$ and $\sum_{k=1}^{K} x_k = N$ we can rewrite

$$\mathcal{L}_B = N^N e^{-N} \prod_{k=1}^{K} \frac{(p_k(\theta))^{x_k}}{x_k!}$$

so we have

$$\frac{\mathcal{L}_A}{\mathcal{L}_B} = \frac{N!}{N^N e^{-N}}$$

which is constant with respect to θ, so the θ that maximises $\mathcal{L}_B$ will also maximise $\mathcal{L}_A$.

B DLS Resource Table

Overs Remaining	Wickets Lost 1	2	3	4	5	6	7	8	9	10
50	100	93.4	85.1	74.9	62.7	49	34.9	22	11.9	4.7
49	99.1	92.6	84.5	74.4	62.5	48.9	34.9	22	11.9	4.7
48	98.1	91.7	83.8	74	62.2	48.8	34.9	22	11.9	4.7
47	97.1	90.9	83.2	73.5	61.9	48.6	34.9	22	11.9	4.7
46	96.1	90	82.5	73	61.6	48.5	34.8	22	11.9	4.7
45	95	89.1	81.8	72.5	61.3	48.4	34.8	22	11.9	4.7
44	93.9	88.2	81	72	61	48.3	34.8	22	11.9	4.7
43	92.8	87.3	80.3	71.4	60.7	48.1	34.7	22	11.9	4.7
42	91.7	86.3	79.5	70.9	60.3	47.9	34.7	22	11.9	4.7
41	90.5	85.3	78.7	70.3	59.9	47.8	34.6	22	11.9	4.7
40	89.3	84.2	77.8	69.6	59.5	47.6	34.6	22	11.9	4.7
39	88	83.1	76.9	69	59.1	47.4	34.5	22	11.9	4.7
38	86.7	82	76	68.3	58.7	47.1	34.5	21.9	11.9	4.7
37	85.4	80.9	75	67.6	58.2	46.9	34.4	21.9	11.9	4.7
36	84.1	79.7	74.1	66.8	57.7	46.6	34.3	21.9	11.9	4.7
35	82.7	78.5	73	66	57.2	46.4	34.2	21.9	11.9	4.7
34	81.3	77.2	72	65.2	56.6	46.1	34.1	21.9	11.9	4.7
33	79.8	75.9	70.9	64.4	56	45.8	34	21.9	11.9	4.7
32	78.3	74.6	69.7	63.5	55.4	45.4	33.9	21.9	11.9	4.7
31	76.7	73.2	68.6	62.5	54.8	45.1	33.7	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
29	73.5	70.3	66.1	60.5	53.4	44.2	33.4	21.8	11.9	4.7
28	71.8	68.8	64.8	59.5	52.6	43.8	33.2	21.8	11.9	4.7
27	70.1	67.2	63.4	58.4	51.8	43.3	33	21.7	11.9	4.7
26	68.3	65.6	62	57.2	50.9	42.8	32.8	21.7	11.9	4.7
25	66.5	63.9	60.5	56	50	42.2	32.6	21.6	11.9	4.7
24	64.6	62.2	59	54.7	49	41.6	32.3	21.6	11.9	4.7
23	62.7	60.4	57.4	53.4	48	40.9	32	21.5	11.9	4.7
22	60.7	58.6	55.8	52	47	40.2	31.6	21.4	11.9	4.7
21	58.7	56.7	54.1	50.6	45.8	39.4	31.2	21.3	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
19	54.4	52.8	50.5	47.5	43.4	37.7	30.3	21.1	11.9	4.7
18	52.2	50.7	48.6	45.9	42	36.8	29.8	20.9	11.9	4.7
17	49.9	48.5	46.7	44.1	40.6	35.8	29.2	20.7	11.9	4.7
16	47.6	46.3	44.7	42.3	39.1	34.7	28.5	20.5	11.8	4.7
15	45.2	44.1	42.6	40.5	37.6	33.5	27.8	20.2	11.8	4.7
14	42.7	41.7	40.4	38.5	35.9	32.2	27	19.9	11.8	4.7
13	40.2	39.3	38.1	36.5	34.2	30.8	26.1	19.5	11.7	4.7
12	37.6	36.8	35.8	34.3	32.3	29.4	25.1	19	11.6	4.7
11	34.9	34.2	33.4	32.1	30.4	27.8	24	18.5	11.5	4.7
10	32.1	31.6	30.8	29.8	28.3	26.1	22.8	17.9	11.4	4.7
9	29.3	28.9	28.2	27.4	26.1	24.2	21.4	17.1	11.2	4.7
8	26.4	26	25.5	24.8	23.8	22.3	19.9	16.2	10.9	4.7
7	23.4	23.1	22.7	22.2	21.4	20.1	18.2	15.2	10.5	4.7
6	20.3	20.1	19.8	19.4	18.8	17.8	16.4	13.9	10.1	4.6
5	17.2	17	16.8	16.5	16.1	15.4	14.3	12.5	9.4	4.6
4	13.9	13.8	13.7	13.5	13.2	12.7	12	10.7	8.4	4.5
3	10.6	10.5	10.4	10.3	10.2	9.9	9.5	8.7	7.2	4.2
2	7.2	7.1	7.1	7	7	6.8	6.6	6.2	5.5	3.7
1	3.6	3.6	3.6	3.6	3.6	3.5	3.5	3.4	3.2	2.5