Modifying Bradley-Terry and other ranking models to allow ties

ROSE BAKER, School of Business
University of Salford
The Crescent, Salford M5 4WT, UK
email rose.baker@cantab.net

PHILIP SCARF, Cardiff Business School
Cardiff University
Aberconway Building, Colum Drive, Cardiff CF10 3EU
email scarfp@cardiff.ac.uk.

December 16, 2020

Abstract

Models derived from distributions of order-statistics are useful for modelling ranked data. The well-known Bradley-Terry and Plackett-Luce models can be derived from the order statistics of the exponential distribution but cannot handle ties. However, ties often occur in sports, and the ability to accommodate them leads to more useful ranking models. In this paper, we use discrete distributions, principally the geometric distribution, to obtain modified Bradley-Terry and Plackett-Luce models and some others that allow tied ranks. Our methodology is introduced for some mathematically tractable and some less tractable distributions and is illustrated using test match cricket.

Keywords

Bradley-Terry model; Plackett-Luce model; order-statistics model; ties; iterative summation; cricket.

1 Introduction

Ranked data arise in many subject areas, such as politics, voting, market research, food preference, medical treatments, etc., and a major area is sport.

Distributions of order-statistics provide a fruitful class of models for ranked data. Here, the ranks are imagined to be ranks of finishing times in a race with n competitors, so that with finishing times $T_1 \cdots T_n$ the probability of the ranking $1 \cdots n$ is $r_{123 \cdots n} = \text{Prob}(T_1 < T_2 < T_3 \cdots < T_n)$. Finishing times are usually assumed independent, so that in this case with probability density of finishing time $f_i(t_i)$ for the ith competitor, we have that

$$r_{123 \cdots n} = \int_0^\infty f_1(t_1)\,\mathrm{d}t_1 \int_{t_1}^\infty f_2(t_2)\,\mathrm{d}t_2 \cdots \int_{t_{n-1}}^\infty f_n(t_n)\,\mathrm{d}t_n. \tag{1}$$

Naturally, these densities are specified from the same family, e.g. the exponential distribution $f(x) = \lambda \exp(-\lambda x)$, but with different 'strength' or rate parameters λ_i (for the ith competitor). Using (1) with this distribution in the simplest case where $n = 2$, we have $r_{12} = \lambda_1/(\lambda_1 + \lambda_2)$. This is the Bradley-Terry (BT) model.

The integral (1) is only analytically solvable for a few distributions, the exponential and some distributions based on it, and for $n > 2$ a concise analytic solution is only available for the exponential distribution itself, giving the Plackett-Luce (PL) model. Baker (2019) gives details of analytic solutions and numerical computations for a variety of distributions.

In sporting data, ties often occur. For example, we could classify a soccer result simply as a home win, an away win, or a tie. Golf scores are also often tied. Ties or draws also occur in other sports, such as rugby (rare), cricket, and chess, where at the highest level over half of games end in a draw. In some sports as in soccer, there is a simple, explicit score. In cricket there is a more complicated score (and margin of victory), and in some games such as chess the idea of a score is purely notional. Outside sport, in preference data, ties are much less common. For example, in a preference study soft drinks might be ranked in order of preference, but in a typical study consumers cannot express indifference. The methodology described here allows such tied data to be analysed, and so preference studies could become less prescriptive.

To generate a model with a theoretical foundation that can cope with ties, we propose using a discrete distribution for the family of survival distributions in (1). With probabilities p_{ij_i} that the finishing time for the ith competitor is j_i, we have

$$r_{123\cdots n} = \sum_{j_1=0}^{\infty} p_{1j_1} \sum_{j_2=j_1+1}^{\infty} p_{2j_2} \cdots \sum_{j_n=j_{n-1}+1}^{\infty} p_{nj_n}. \tag{2}$$

This gives the probability of a ranking with no ties. Using the notation that the indices of tied competitors are enclosed in parentheses, the probability that competitor 1 and competitor 2 are tied, but that none of the others are tied, is

$$r_{(12)3\cdots n} = \sum_{j_1=0}^{\infty} p_{1j_1} p_{2j_1} \sum_{j_3=j_1+1}^{\infty} p_{3j_3} \cdots \sum_{j_n=j_{n-1}+1}^{\infty} p_{nj_n},$$

and so on. Hence using a discrete distribution instead of a continuous one in the order-statistics model yields a class of models for ranked data that can accommodate tied ranks.

Thus, intuitively, if two competitors are ranked using a continuous measure such as time-taken, they cannot tie under an order-statistics model because their respective times-taken cannot be exactly equal. Whereas, if they are ranked using a discrete measure, such as whether they are fast or slow, then a tie is possible (either if both are fast or if both are slow).

Turning to management, order-statistics based models using discrete distributions can model data of interest, such as consumer preference data. Here the competitors are, for example, competing brands of soft drink. Contests then become judgements made by individual consumers. Sometimes items are compared in subsets, for example different consumers might rank different subsets of (say) three different product brands. This avoids giving respondents unduly demanding tasks. Covariates must often be included in the models, as is the case with sport. Vigneau *et al* (1999) give an account.

These various cases can be modelled using discrete distributions, so that 'no preference' becomes an option, and the 'competitor strengths' can be found for example by fitting the discrete model using maximum-likelihood. In general, existing methodologies can be extended to allow indifference between products.

There is quite a lot of previous work on tied-ranks. Glenn and David (1960) modified the Thurstone-Mosteller model to allow small differences to become ties, and Rao and Kupper (1967) corrected the BT model similarly by assuming that small values of $\ln(\lambda_1) - \ln(\lambda_2)$ would be declared ties. This led to formulae $r_{12} = \lambda_1/(\lambda_1 + \phi\lambda_2)$, $r_{(12)} = \lambda_1\lambda_2(\phi^2 - 1)/(\lambda_1 + \phi\lambda_2)(\phi\lambda_1 + \lambda_2)$, where $\phi \geq 1$. Beaver and Rao (1973) generalized the above work to ties when there are three competitors. Kuk (1995) applied the approach of Glenn and David to football.

Davidson (1970) gave an *ad hoc* correction to the BT model for ties, i.e. $r_{12} = \lambda_1/(\lambda_1 + \lambda_2 + \nu\sqrt{\lambda_1\lambda_2})$ where $\nu \geq 0$, so that $r_{(12)} = \nu\sqrt{\lambda_1\lambda_2}/(\lambda_1 + \lambda_2 + \nu\sqrt{\lambda_1\lambda_2})$.

In related work, Lengyel (2009) considered the gambler's ruin problem allowing for ties. DeWart and Gillard (2019) apply the BT model to cricket, where draws occur but do not depend strongly on team strengths.

Other attempts to do inference in the presence of ties do not introduce an extra parameter. Su and Zhou (2006) point out a similarity between the BT model and the Cox proportional hazards model. Here ties can be dealt with using Cox's exact method or be broken in all possible ways with equal probability. Baker and McHale (2015) analyse golf tournaments assuming that ties can be broken in all possible ways. The drawback of this type of approach is that although inference about player or team strengths can be made, the probability of a tie cannot be forecast. For that, a model that accommodates ties is needed.

The next section introduces this type of model, and in particular we derive modified BT and PL models using the geometric distribution. We make some general points about modelling ties, and then more complicated models from distributions related to the geometric distribution are discussed. Finally, computations are described for discrete distributions that should yield realistic models, but for which analytic solutions are impossible. We use test-match cricket, where draws are relatively frequent, to illustrate an application of the model. We finish with some conclusions.

Note, throughout this paper we shall use the terms 'draw' and 'tie' interchangeably, although in some sports they are perceived as distinct outcome types, e.g. test-match cricket. Thus here a draw or tie between two competitors in a match is an outcome such that the two competitors cannot be separated on ability. One cannot be ranked above or below the other.

2 Models derived with the geometric distribution

The geometric distribution has probability mass function (pmf) $p_k = (1-p)p^k$ for $K \in \{0, 1, 2 \cdots\}$, with mean $p/(1-p)$, and is the discrete analogue of the exponential distribution. The survival function $S_k = \sum_{n=k+1}^{\infty} p_n = p^{k+1}$.

In a familiar example, that of tossing a biased coin with probability p of getting 'tails', $K+1$ is the number of tosses before a head is obtained (because the count starts at zero). In a game such as golf, K is the score, and high p is bad, so that the reciprocal of the mean score, $(1-p)/p$, would be a possible measure of player ability. Applying (2) with $n = 2$, we have that

$$r_{12} = \sum_{j>i} p_{1i}p_{2j} = \sum_{i=0}^{\infty} p_{1i}S_{2i} = (1-p_1)p_2 \sum_{i=0}^{\infty} (p_1p_2)^i = (1-p_1)p_2/(1-p_1p_2).$$

The tied probability

$$r_{(12)} = (1-p_1)(1-p_2)\sum_{i=0}^{\infty}(p_1p_2)^i = (1-p_1)(1-p_2)/(1-p_1p_2),$$

and of course $r_{21} = p_1(1-p_2)/(1-p_1p_2)$.

This does not yet look much like the BT model. Nonetheless, the BT model can be obtained as follows. We take strength $\lambda_i = (1-p_i)/p_i$. Note that $(1-p_1p_2) = p_1(1-p_2)+p_2(1-p_1)+(1-p_1)(1-p_2)$. It can be shown that $r_{12} = \lambda_1/(\lambda_1+\lambda_2+\lambda_1\lambda_2)$, $r_{(12)} = \lambda_1\lambda_2/(\lambda_1+\lambda_2+\lambda_1\lambda_2)$. The full derivation is simple but laborious and is omitted for brevity. As $p_i \to 1$, the strengths go to zero with their ratio constant, $r_{(12)} \to 0$, and the BT model is regained.

This is the tied BT model for sports where the lowest score wins, such as a round of golf. For sports where highest score wins, the definition is reversed, so that $\lambda_i = p_i/(1-p_i)$. Using (2) again, we now have

$$r_{12} = \lambda_1/(\lambda_1+\lambda_2+1), \quad r_{(12)} = 1/(\lambda_1+\lambda_2+1). \tag{3}$$

Now as the $p_i \to 1$, the strengths go to infinity, and again the BT model is regained.

When fitting models by likelihood-based methods, one can of course estimate the strengths λ_i or the probabilities p_i. In the simplest case, there are several games between the same two teams over a short period, so that strengths can be assumed constant. We then have estimates $\hat{p}_{12}$, $\hat{p}_t$, $\hat{p}_{21}$, and maximum likelihood or the method of moments gives $\hat{p}_1 = \hat{p}_{21}/(1-\hat{p}_{12})$ etc.

The probabilities for the 3-competitor case are:

$$r_{123} = \frac{(1-p_1)(1-p_2)p_2p_3^2}{(1-p_2p_3)(1-p_1p_2p_3)},$$

$$r_{(12)3} = \frac{(1-p_1)(1-p_2)p_3}{1-p_1p_2p_3},$$

$$r_{1(23)} = \frac{(1-p_1)(1-p_2)(1-p_3)p_2p_3}{(1-p_2p_3)(1-p_1p_2p_3)},$$

$$r_{(123)} = \frac{(1-p_1)(1-p_2)(1-p_3)}{1-p_1p_2p_3}.$$

In general, the probability $\mathcal{L}$ of one of the patterns of score can be found numerically or symbolically as follows:

1. Set up an array giving the number of competitors at each distinct score, e.g. for the 4 patterns just quoted, these arrays would hold 111, 21, 12, 3 respectively. Let there be M elements $s_1 \cdots s_M$.
2. Set the probability $\mathcal{L} = \prod_{i=1}^{n}(1-p_i)$, where n is the number of competitors.
3. For j from 1 to M:
   (a) let $m = 1 + \sum_{k=1}^{j-1} s_k$, where an empty sum with upper suffix less than the lower counts as zero.
   (b) If $j > 1$, $\mathcal{L} \to \mathcal{L} \prod_{t=m}^{n} p_t$.
   (c) $\mathcal{L} \to \mathcal{L}/(1 - \prod_{t=m}^{n} p_t)$.

This would be the best method for numerically computing probabilities of rankings or deriving analytic expressions for more than 3 competitors.

Note that sometimes the bottom m competitors out of n are not ranked, i.e. we are only interested in ranking (say) the top three players. This situation can be dealt with under the exponential model (1) and also under the geometric model (2). Because the survival function p^{k+1} is simple, (2) can be evaluated. For example, when the top 3 competitors are ranked out of $m+3$, the term $1 - p_1p_2p_3$ appearing in the numerators of r_{123} etc. becomes $1 - \prod_{s=1}^{m+3} p_s$.

3 Theoretical considerations

3.1 The number of tied cases

With $n > 2$ competitors there are 2^{n-1} distinct patterns of tie. Thus, representing a tie by a link between competitors, each of the $n-1$ links can be present or absent. For example, for $n = 3$ there are 4 patterns, 123, 1(23), (12)3, (123).

The number of different results possible is 3 for $n = 2$, 13 for $n = 3$, and 75 for $n = 4$. The number for general n is the ordered Bell number or Fubini number. Usually thought of as the number of ways n non-identical balls can be put into identifiable boxes with at least one ball per box, it has also been used to give the possible number of rankings including ties in e.g. a horse race (De Koninck, 2009).

3.2 The different types of sport

As mentioned, sports can be dichotomized into those where the smallest score wins, such as racing and golf, and those where the highest score wins, as in ball-games. For a given order-statistics distribution, ranking models derived from (1) are in general different in the two cases. Hence one must be careful to choose a sensible model, based on a discrete distribution that approximates the reality. In the continuous case, one obtains the BT model using the exponential distribution in (1) under either scoring system, but the PL model with $n > 2$ differs in the two cases. In the discrete case, the probability of a tie is a different function of strengths even in the 2-competitor case. The notation used here is that e.g. r_{12} is the probability that competitor 1 wins, whether through low or high score.

Sports can also be dichotomized in another way important for ranking. In stroke-play in golf or time-trials in bicycle racing for example, a competitor performs as well as he or she can, independently of other competitors. This is an approximation, because competitors may have some knowledge about the performance of others during the event and this may affect performance. Order-statistics models assume this independence.

However, in match play, a competitor's performance (score) depends on the strength of the other competitor. For example, in the case of soccer this interaction is usually modelled by taking mean goals μ_1 for team 1 as the ratio of that team's offensive strength λ_1 to the defensive strength η_2 of the opposing team, team 2, i.e. $\mu_1 = \lambda_1/\eta_2$. If we simplify by taking $\eta_1 \propto \lambda_1$ etc., the formula for mean goals μ_1 scored by team 1 would be

$$\mu_1 = \beta\lambda_1/\lambda_2, \tag{4}$$

where $\beta > 0$, the common ratio of offensive to defensive strength, is to be estimated.

Taking the tied model (3) and replacing the means λ_1, λ_2 by $\mu_1 = \beta\lambda_1/\lambda_2, \mu_2 = \beta\lambda_2/\lambda_1$, we have that

$$r_{12} = \frac{\lambda_1^2}{\lambda_1^2 + \lambda_2^2 + \beta^{-1}\lambda_1\lambda_2}.$$

Since λ_1, λ_2 are just unknown strengths, we can set $\lambda \to \sqrt{\lambda}$ to obtain

$$r_{12} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \beta^{-1}\sqrt{\lambda_1\lambda_2}}.$$

This is precisely the *ad hoc* model of Davidson (1970). Hence this model can be derived more theoretically, as an order-statistics model based on the geometric distribution.

3.3 Inference for discrete and continuous models

A feature of continuous models derived from (1) is that the probability of a ranking depends only on ratios of strength parameters. This is because survival distributions must have a scale parameter, and a change of variables can be made that rescales all n strengths but does not change the value of the integral. However this cannot be done for discrete-distribution models, in fact the only transformation that preserves the lattice nature of the distribution is to change the support of the distribution from zero; this would make the geometric model more applicable to golf, for example. Hence it can be seen that effectively, as an extra parameter has been added to model the probabilities of ties, there are now n strengths instead of $n-1$ ratios. As the means of the discrete distributions go to infinity, the sums in (2) tend to integrals, and the continuous score model with no ties is regained. Hence (2) can model heavily or lightly-tied data. Thus, as all strengths tend to zero, all scores are tied at zero.

Furthermore, statistical inference is easier than for the no-ties continuous case, as now all the strengths or probabilities p_i can be estimated. In the no-ties case, only ratios of strengths matter, so one strength must be held fixed. However, when tracking player or team strengths through time, higher strengths may just mean fewer draws. It would still be necessary then to display team strengths relative to the average strength. Otherwise, increasing strengths could just mean that there are fewer ties for whatever reason.

This situation does not apply to sports where (4) holds. In this case, strength ratios feature as in continuous models, plus an additional parameter β, so that if $\beta \to \infty$ there are no ties.

3.4 Further modelling of ties

Kuk (1995) and DeWart and Gillard (2019) discuss modelling win/lose/tie data in soccer and cricket respectively, where teams can have a 'strength to draw' so that draws occur more or less often than a simple model would predict. In the context of order-statistics models for team sports with offensive and defensive strength, it is merely necessary to allow each team a unique defensive strength to allow highly variable numbers of ties. Clearly, if both teams play a defensive game, few goals are scored and there are many tied matches. Hence the formulae

$$\mu_1 = \lambda_1/\eta_2, \qquad \mu_2 = \lambda_2/\eta_1 \tag{5}$$

are used. Inference can proceed by allowing all the η_i to be free, when the parameter β is not needed.

For the model arising from using the geometric distribution, we now have $p_1/(1-p_1) = \theta\lambda_1/\eta_2$, $p_2/(1-p_2) = \lambda_2/\eta_1$ where θ is the home-advantage parameter, giving the model:

$$\begin{aligned} r_{12} &= \frac{\theta\lambda_1/\eta_2}{1+\theta\lambda_1/\eta_2+\lambda_2/\eta_1}, \qquad (6) \\ r_{21} &= \frac{\lambda_2/\eta_1}{1+\theta\lambda_1/\eta_2+\lambda_2/\eta_1}, \\ r_{(12)} &= \frac{1}{1+\theta\lambda_1/\eta_2+\lambda_2/\eta_1}. \end{aligned}$$

A good measure of overall or total strength is $s_1 = \lambda_1\eta_1$, in that $r_{12} > r_{21}$ if and only if $\lambda_1\eta_1 > \lambda_2\eta_2$. We call $a_1 = \lambda_1/\eta_1$ the team 1 aggressiveness. If both teams have equal total strength s, then $r_{12} = 1/(2+(a_1a_2)^{-1/2})$, so an aggressive team is less likely to draw.

3.5 Relation between continuous and discrete models

It is known that the geometric distribution is a discrete case of the exponential distribution, and the negative binomial model is the discrete case of the gamma distribution, so a negative binomial order-statistic model is a discrete version of the model of Stern (1990). For the negative binomial, the support is $K \in \{0,1,2,3\cdots\}$ and the pmf is

$$p_k = \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)}(1-p)^\alpha p^k, \tag{7}$$

where $\alpha > 0$ need not be an integer. The mean is $\alpha p/(1-p)$. We now look at the transition to the continuous limit. Given a spacing h between lattice points, the Euler-Maclaurin theorem gives the difference between a sum and the corresponding integral as being $O(h)$. The scale factor (and hence the mean μ) of the discrete distribution is allowed to tend to infinity, so that $h\mu$ remains finite. As $h \to 0$ we obtain the continuous case (1). This of course assumes that the mean can tend to infinity, which it cannot for some discrete distributions, e.g. the Salvia-Bollinger distribution discussed in section 4. Hence models based on such distributions do not include the continuous (zero-tie) case.

4 Other tractable models

These are nearly all variants of the geometric distribution, such as the sum of two or more geometric distributions and the exponentiated geometric distribution.

The sum of two geometric distributions is a negative binomial distribution, called here a Pascal(2) distribution. The pmf is $p_k = (1-p)^2(k+1)p^k$, and it is possible to evaluate (2) for simple cases using simple algebra for series summation. The results are:

$$r_{12} = \frac{2(1-p_2)^2(1-p_2)p_2}{(1-p_1p_2)^3} + \frac{(1-p_1)^2p_2^2}{(1-p_1p_2)^2},$$

$$r_{(12)} = \frac{(1-p_1)^2(1-p_2)^2(1+p_1p_2)}{(1-p_1p_2)^3}.$$

These unintuitive formulae would be suitable for sports where a low score is good, and for the more common case where a high score is good, $r_{12} \leftrightarrow r_{21}$. This is the discrete version of the gamma model used by Henery (1983) and Stern (1990). Another way to find these formulae is to remark that the Pascal(r) probabilities are differentials of geometric probabilities:

$$q_k = \binom{k+r-1}{k}(1-p)^r p^k = \frac{(1-p)^r(\,\mathrm{d}/\,\mathrm{d}p)^{r-1}p^{k+r-1}}{(r-1)!},$$

so that, interchanging series summation and differentiation, any BT or PL-type probability R can be converted to the corresponding probabiity Q using

$$Q = \{(r-1)!\}^{-n}\{\prod_{i=1}^{n}(1-p_i)^r(\frac{\partial}{\partial p_i})^{r-1}p_i^{r-1}\}\frac{R}{\prod_{i=1}^{n}(1-p_i)}.$$

This does not seem easier by hand but a computer algebra package would evaluate the probabilities more easily in this form.

The exponentiated geometric distribution has distribution function $F_k = (1-p^{k+1})^\gamma$ for $\gamma > 0$. When $\gamma = 1$ we regain the geometric distribution, and when $\gamma = 2$ we have $F_k = 1-2p^{k+1}+p^{2k+2}$ or $q_k = 2(1-p)p^k - (1-p^2)p^{2k}$, a negative mixture of two geometric distributions. This distribution has a mode. Using (2) the probabilities

$$r_{12} = (1-p_1)\{\frac{4}{1-p_1p_2} + \frac{(1+p_1)p_2^2}{1-p_1^2p_2^2} - 2\frac{(1+p_1)p_2}{1-p_1^2p_2} - 2\frac{p_2^2}{1-p_1p_2^2}\},$$

$$r_{(12)} = (1-p_1)(1-p_2)\{\frac{4}{1-p_1p_2} + \frac{(1+p_1)(1+p_2)}{1-p_1^2p_2^2} - 2\frac{1+p_1}{1-p_1^2p_2} - 2\frac{1+p_2}{1-p_1p_2^2}\}$$

are obtained.

Finally, we briefly mention the Salvia-Bollinger distribution (*ibid*, 1982), because here $r_{12}, r_{(12)}$ can be written simply, albeit in terms of special functions. The pmf is $p_k = c^k\{1-c/(k+1)\}/k!$, and $\text{Prob}(K \geq k) \equiv S_k = c^k/k!$, for $K \in \{0, 1, 2\cdots\}$ and $0 < c < 1$. Here (2) can be evaluated for $n = 2$, giving

$$r_{12} = \sqrt{c_2/c_1}I_1(2\sqrt{c_1c_2}) + 1 - I_0(2\sqrt{c_1c_2}),$$

$$r_{(12)} = 2I_0(2\sqrt{c_1c_2}) - 1 - \frac{c_1+c_2}{\sqrt{c_1c_2}}I_1(2\sqrt{c_1c_2}),$$

where I_0, I_1 are modified Bessel functions of the first kind.

Given that modified Bessel functions are quick to compute, this looks an interesting possibility. However, apart from the lack of a probabilistic basis, it has the problem that for equally matched competitors, the probability of a tie cannot be lower than approximately 0.33, which occurs at $c_1 = c_2 = 0.7725$. This is a little too high to model football matches. However, a left-truncated and left-shifted distribution with $S_k = n!c^k/(k+n)!$ can rectify this problem. The pmf is $p_k = \frac{n!c^k}{(k+n)!}\{1 - c/(k+n+1)\}$, so that $c \leq n+1$. With $n = 1$, the minimum percentage of ties for equally-matched teams is 25.5. The probabilities are

$$r_{(12)} \to \{r_{(12)} - (1-c_1)(1-c_2)\}/c_1c_2,$$

$$r_{12} \to \{r_{12} - (1-c_1)c_2\}/c_1c_2.$$

This distribution is perhaps best regarded as a mathematical curiosity, and the next section deals with more realistic distributions.

5 Computation for intractable models

5.1 The Poisson distribution

The Poisson distribution is a good model for scores in many ball games, where each side independently scores following a Poisson process. This ignores many possible effects, such as competitors playing differently when trailing their opponents. Strengths could be taken as the means μ_1, μ_2, or (better) (4) or (5) can be used. Home advantage θ multiplies the mean for the home team.

In soccer, the Poisson model is indeed a reasonable approximation, but there might be more ties than it would predict. However, this is not a problem for the ranking model, where the proportion of ties is adjustable, but one should be wary of taking $\hat{\mu}_1, \hat{\mu}_2$ as estimates of mean goals scored. These quantities may have been scaled down to accommodate the high proportion of ties.

Let the Poisson means be μ_1, μ_2, and take the higher score as denoting a win. Then

$$r_{(12)} = \exp(-\mu_1 - \mu_2) \sum_{i=0}^{\infty} (\mu_1 \mu_2)^i / i!^2.$$

Also,

$$r_{12} = \exp(-\mu_1 - \mu_2) \sum_{i=0}^{\infty} \mu_2^i \sum_{j=i+1}^{\infty} \mu_1^j / j!. \tag{8}$$

This can be evaluated by truncating the sums at some large number N such as 100, and proceeding as per Baker (2019) for the continuous case. Thus, starting from zero and going up to N, $S_i = \exp(-\mu_1) \sum_{j=i+1}^{N} \mu_1^j / j!$ is tabulated. Then $r_{12} = \exp(-\mu_2) \sum_{i=0}^{N} \mu_2^i S_i / i!$. The Poisson sums can be rescaled to unity using $S_0 = 1$ etc., so that this gives the exact result for a truncated Poisson distribution.

The probabilities can be written in terms of special functions, as

$$r_{(12)} = \exp(-\mu_1 - \mu_2) I_0(2\sqrt{\mu_1 \mu_2}),$$

where I_0 is the modified Bessel function of the first kind. Using the relation

$$\int_0^{\mu} \frac{y^n \exp(-y)\, \mathrm{d}y}{n!} = \sum_{j=n+1}^{\infty} \frac{\mu^j \exp(-\mu)}{j!},$$

we have from (8) that

$$r_{12} = \exp(-\mu_2) \int_0^{\mu_1} I_0(2\sqrt{\mu_2 y}) \exp(-y)\, \mathrm{d}y. \tag{9}$$

Which method of computation is faster and more convenient is moot, and depends on which computing platform is used.

Note that $r_{12}, r_{(12)}$ can be derived from the Skellam distribution (the difference of two Poisson r.v.s X_1, X_2) although the useful result (9) for $\mathrm{Prob}(X_1 > X_2)$ does not seem to appear in the literature. The distribution is applied directly to goal difference in soccer in Karlis and Ntzoufras (2009).

The Poisson distribution with mean μ tends to normality as $\mu \to \infty$, and twice the square root of the random variable tends to standard normality fast. The probability $r_{12} \to \Phi\{2(\mu_1^{1/2} - \mu_2^{1/2})\}$, where Φ is the normal distribution function. It does not seem that this is a ratio of strengths as it should be, but writing $\mu^{1/2} = \ln(s)$, we have that $r_{12} = \Phi(2\ln(s_1/s_2))$, a Thurstonian model, so this is still a ratio of strengths when the strengths are defined as $s = \exp(\mu^{1/2})$.

5.2 General distributions

In general, the computation of probabilities of results must be done numerically. A useful distribution for the general case is the negative binomial, which has one extra parameter α besides the measures of strength for each competitor. The pmf is given in (7). This is a distribution that includes useful special cases such as the geometric ($\alpha = 1$) and the Poisson ($\alpha \to \infty$ and $p \to 0$). Computation would be done using (2), somewhat analogously to the algorithm in section 2. An algorithm is:

1. Set up an array giving the number of competitors at each distinct score as in section 2. Let there be M elements $s_1 \cdots s_M$.
2. Set $S_0 \cdots S_N$ to 1 for N 'large', e.g. 100.
3. for j from M down to 1:
   (a) let $m = 1 + \sum_{k=1}^{j-1} s_k$, where an empty sum with upper suffix less than the lower counts as zero. The probabilities to be used in this step are $p_m \cdots p_{m+s_j-1}$.
   (b) For k from N down to 0, compute $A_k = S_k \prod_{t=m}^{m+s_j-1} p_{tk}$ and set $S_k \to \sum_{i=k+1}^{N} A_i$. These two operations can be done in one loop.
4. The required probability is S_0.
5. It is possible to cumulate all probabilities $T_m = \sum_{k=0}^{N} p_{mk}$ and so make a correction $S_0 \to S_0 / \prod_{m=1}^{n} T_m$.

6 Example

Test match cricket was used to illustrate some of the models presented here. Test match results were collected from the 'Cricinfo' results archive (ESPN, 2020). Data were taken as win/draw/lose. Results of all matches from 1980-2019 were analysed, excluding those involving Ireland or Afghanistan who played very few games. This left 10 teams and 1480 matches in total over the period. The method of modelling draws in detail in section 3.4

was followed. Here each team has an attacking strength λ and a defensive strength η. Defensive strength can be taken as proportional to attacking strength, in which case the model reduces to the BT model with Davidson's correction for ties, or defensive strengths η can be 'floated' (allowed to vary). The models used were the BT model (6) and the Poisson model. It is convenient to reparameterize to 'total strength' $\lambda\eta$ and 'aggressiveness' λ/η. As pointed out in section 3.4, higher total strength is what makes a team more likely to win than its opposing team. Aggressiveness or defensiveness is a playing style that does not by itself change the odds of winning, but aggressiveness does reduce the probability of a draw. Because the model with variable η generalizes the model where $\lambda/\eta = \beta$, it follows that 9 extra parameters are used. The model was fitted by maximum-likelihood, using a NAG (Numerical Algorithms Group) function minimiser called from a fortran program. Standard errors were found from the Hessian matrix, obtained by numerical differentiation. England were taken as the baseline team, with strength unity, so all strengths are relative to England.

It is clear from table 1 that the variable-η model fits the results data significantly better. Taking the BT model, we have $X^2[9] = 33.67$, showing that the variable defensive parameters are needed.

Table 2 shows fitted parameter values for the variable-η model. The home advantage parameter for the minimum AIC model was $\hat{\theta} = 1.84 \pm 0.13$.

Model	$-\ell$	AIC
Bradley-Terry	1480.74	2983.47
BT with η	1463.90	2967.80
Poisson	1479.28	2980.55
Poisson with η	1466.38	2972.75

Table 1: Minus log-likelihood and Akaike Information Criterion for the 4 models, BT and Poisson, with and without the η_i floated.

The BT model fits acceptably: from looking at home-away team combinations with at least 5 identical results, the goodness of fit $X^2[38] = 41.05$, $p = 0.34$. For the Poisson model, $X^2[48] = 70.46, p = 0.02$. It is necessary to give some details of the chi-squared calculation. First, each home/away team pair gives rise to 3 cells, home win, draw, lose, with 2 degrees of freedom. We required the minimum predicted number in any of the 3 cells to be at least 5. Second, the estimation of model parameters poses a problem, because removing the 20 degrees of freedom from the chi-square is excessive. The half-sample method was used, in which only half the data is used in

Country	Games	Strength	se	Agressiveness	se
Australia	422	1.897	0.289	1.636	0.405
South Africa	260	1.786	0.335	1.182	0.355
India	352	1.263	0.222	0.309	0.073
Pakistan	318	1.209	0.216	0.666	0.172
England	458	1	N/A	0.737	0.164
West Indies	351	0.860	0.142	0.859	0.217
Sri Lanka	283	0.708	0.135	0.766	0.212
New Zealand	297	0.704	0.131	0.460	0.120
Zimbabwe	107	0.077	0.029	0.516	0.271
Bangladesh	114	0.072	0.026	2.594	1.552

Table 2: Total strength $\lambda\eta$ and aggressiveness λ/η by country 1980 to July 2019, in order of total strength, with standard errors.

fitting the model, and then the chi-squared is calculated on the full dataset with no correction for fitted parameters.

The ranking by total strength is much as expected. Both the BT model and the Poisson-distribution model are special cases of order-statistics models derived using the negative binomial distribution (7) with $\alpha = 1$ and $\alpha \to \infty$ respectively. On using this model with the parameter α floated, $\hat{\alpha} = 0.84 \pm 0.52$, and the log-likelihood only increased by 0.1. Clearly, the fit is not very sensitive to the model used, but the tied BT model is about the best in the wide class of negative binomial order-statistics models, as well as the simplest.

What is new in the model (6) is that draws are handled by giving teams an offensive strength λ and defensive strength η. More defensive play leads to lower notional scores in (2), and hence to more draws. Australia has few drawn matches in this model because of the team's aggressive style of play.

Forecasting matches in-play is of great interest for bettors and others: see e.g. Akhtar and Scarf (2012), Asif and McHale (2016). Here models like (6) could be fitted to recent matches using for example a time-discounted likelihood function and with covariates added. That is not however the focus here, where the point of the analysis is simply to illustrate the usefulness of the models.

7 Conclusions

Using discrete distributions in order-statistics based models of ranking allows ties to be included in a natural way. The geometric distribution allows the popular BT and PL models to be given tied analogues, and more realistic models can be devised, using e.g. the Poisson distribution for team sports where high score wins. Computation of likelihoods for these models is still fast, using the iterative method of evaluating the sums in (2).

A caveat is that in this methodology, ties have no particular significance: they arise because of the discrete nature of an underlying score (which may be notional), and one extra parameter is all that is assigned. Sometimes this is not adequate, and ties or draws must be modelled in more detail, as in Kuk (1995) and DeWart and Gillard (2019). Such modelling is also possible with the order-statistics approach, by using models with both attacking and defensive strengths to give a model that is less *ad hoc*. The reason that more or fewer ties arise is that the teams are playing more or less defensively. This gives sensible results when applied to test-match cricket.

What is novel in this work is a new approach to ties that yields many new models. In particular, the geometric distribution specification allows the Plackett-Luce model to be extended to include tied rankings. This allows multi-player sports where tied rankings occur (such as golf) to be modelled for any number of players, and likelihoods to be quickly computed. Also, the ties model of Davidson (1970) is now given a more theoretical basis as an order-statistics model for team sports.

In future work it would be interesting to use some of the models developed here to forecast results, e.g. in-play.

References

[1] AKHTAR S., SCARF P. (2012) Forecasting test cricket outcomes in play, *International journal of forecasting* **28**, 632-643.

[2] ASIF, M., MCHALE I. G. (2016) In-play forecasting of win probability in one-day international cricket: a dynamic logistic regression model. *International journal of forecasting* **32**, 34-43.

[3] BAKER, R. (2019) New order-statistics-based ranking models and faster computation of outcome probabilities, *IMA Journal of Management Mathematics*, 10.1093/imaman/dpz001

[4] BAKER, R. D. & MCHALE, I. G. (2015) Deterministic evolution of strength in multiple comparisons models: Who is the greatest golfer? *Scand. J. Statist.* 42 (2015), pp. 180196.

[5] BEAVER, R. J. & RAO, P. V. (1973). On ties in triple comparisons, *Trabajos de Estadistica y de Investigacion Operativa* **24**, 77-92.

[6] DAVIDSON, R. R. (1970) On extending the Bradley-Terry model to accommodate ties in paired comparison experiments, *Journal of the American Statistical Association* **65**, 317-328.

[7] DE KONINCK, J-M, Those Fascinating numbers (2009), American Mathematical Society, USA.

[8] DEWART, N. & GILLARD, J. (2019) Using Bradley-Terry models to analyse test match cricket. *IMA Journal of Management Mathematics*, **30**, 187-207. and Corrigendum, **30**, 265.

[9] ESPN (2020) List of cricket match results (by year). https://stats.espncricinfo.com/ci/content/records/307847.html, retrieved 2.4.2020

[10] GLENN, W. A. & DAVID, H. A. (1960) Ties in paired-comparison experiments using a modified Thurstone-Mosteller model, *Biometrics*, **16**, 86-109.

[11] HENERY, R. J. (1983) Permutation probabilities for gamma random variables. *J. Appl. Probab.*, **20**, 822-834.

[12] KARLIS, D. & NTZOUFRAS, I. (2009) Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference, *IMA Journal of Management Mathematics*, **20**,133-145.

[13] KUK, Y. C. (1995) Modelling paired comparison data with large numbers of draws and large variability of draw percentages among players, *The Statistician*, **44**, 523-528.

[14] LENGYEL, T. (2009). The conditional gambler's ruin problem with ties allowed, *Applied Mathematics Letters* **22**, 351-355.

[15] RAO, P. V. & KUPPER L. L. (1967) Ties in paired-comparison experiments: a generalization of the Bradley-Terry model, *Journal of the American Statistical Association* **62**, 194-204.

[16] SALVIA, A. A. & BOLLINGER, R. C. (1982). On discrete hazard functions. *IEE Trans. on Reliab.*, **R-31**, 458-459.

[17] STERN, H. (1990) Models for distributions on permutations, *J. Amer. Statist. Assoc.*, **85**, 558564.

[18] SU, Y. & ZHOU, M. (2006) On a connection between the Bradley-Terry model and the Cox proportional hazards model, *Statistics and Probability Letters* **76**, 698-702.

[19] VIGNEAU, E., COURCOUX, P. & SÉMÉNOU, M. (1999). Analysis of ranked preference data using latent class models, *Food quality and Preference*, **10**, 201-207.