

OPINION

Football scores, the Poisson distribution and 30 years of final year projects in Mathematics, Statistics and Operational Research

Phil Scarf, Salford Business School, University of Salford, Manchester, UK. Email: p.a.scarf@salford.ac.uk

Abstract

The development of the Poisson match as a model used in the prediction of the outcome of football matches is described. In this context, many interesting modelling projects arise that are suitable for undergraduate, final year students. In a narrative that discusses the author's engagement with this model and other related models, the paper presents a number of these projects, their attractions and their pitfalls, and poses a number of questions that are suitable for investigation. The answers to some of these questions would be worthy of the attention of the administrators of their respective sports.

Keywords: Poisson distribution; sport; competitive balance; tournament design.

1. Background

This paper is a personal story as well as a description of some mathematics. I first met the "Poisson match" in 1983 as a Mathematics undergraduate at Sheffield University. On a project-based module, students read and summarised papers on sports statistics, among which was the now seminal work of Mike Maher on modelling football scores (Maher, 1982). This was the start of more than thirty years of interaction with this model. A few years later, in a final year project that I set as a young lecturer in the early 1990s, two final year mathematics students forecast football match scores using this model in order to play the *Football Pools*. The 'Pools' was a forerunner to the National Lottery. Millionaires who had correctly forecast eight score draws were front page celebrities. In the simplest stake in this game, a player selected eight matches from among all the matches to be played in the English Football League on a given weekend, earning 3 points if the result of a selected match was a score-draw, 2 for a no-score-draw, $1\frac{1}{2}$ for an away win, and 1 point for a home win. Those who scored 24 points shared the jackpot. So, the object was to forecast match scores, and then make the selection on the basis of these forecasts.

Selecting 8 matches at random achieved 14 points on average, and using Maher's model we could improve this to 17, reducing the expected waiting time for a jackpot from 10^7 plays to 10^6 approximately. There are three problems. Firstly, football match scores are inherently difficult to forecast because games are often finely balanced, and arguably more so than any other game. This is one reason why the game is so popular (Forrest and Simmons, 2002; Buraimo and Simmons, 2015), and I will return to this point later. The second problem is shortcomings in the model itself, and indeed Maher's model has been developed and refined by many. The third problem is sparsity of data, so that parameters are not sufficiently well estimated.

To explain this last point, we must describe the Maher model in a little detail. Doing so also suggests many open questions that are suitable for student projects. It is immediately apparent that the score of a team A in a match M is a random variable taking values 0,1,2,..., and therefore the Poisson distribution is a candidate model for this score. To use this model, the mean score (the

parameter of the Poisson distribution) must be estimated. Now obviously, when A plays team B at venue C at time t say, this is a *random trial* that has never been observed before. Even if A played B at venue C in the previous season, conditions would be different, the teams would be different, a different referee, weather, etc. So what data should we use? Ignoring the issues just raised and assuming all matches AvB at C are statistically identical (and independent), we could use the scores from all the matches AvB at C in the recent seasons. However, this would still give only a small number of datapoints, bearing in mind that the further one goes back in time the more dubious are the assumptions, not least because the abilities or strengths of A and B evolve over time. In fact if one proceeds in this naïve way, for the English Premier League (EPL) now with twenty teams, this approach implies 380 parameters to be estimated. Further, one season's data is itself 380 matches. Maher's trick was to propose that each team possesses an attacking strength (or tendency to score goals) and a defensive weakness (or tendency to concede goals), and the mean score of A when A plays B is the product of A's attacking strength, α_1 , and B's defensive weakness, β_2 , and a home advantage parameter, δ say, which does not depend on the venue. Thus, if X_1 is the random variable that denotes the score of A when A plays B at home, and X_2 is the score of B in this match, then $E(X_1) = \alpha_1 \beta_2 \delta$ (and $E(X_2) = \alpha_2 \beta_1$), and X_1 and X_2 have independent Poisson distributions. With this model, there are two parameters per team, plus the home advantage, making 41 in total for the EPL (actually only 40 because strength is relative). Nonetheless, estimation remains a problem because teams' strengths evolve over time—some teams improve, others decline. Owen (2011) and Koopman and Lit (2015) model this strength evolution. McHale and Kharrat (BBC Sport, 2017) take a different approach, using player line-ups to determine the strengths. The independence of scores is questioned by Dixon and Coles (1998), although McHale and Scarf (2007, 2011) found only slight negative dependence. Now I may be digressing slightly here, but it is important to depict the modelling landscape in order to develop ideas for further related projects.

Returning to my own relationship with the Maher model, a decisive moment was the opportunity to develop the statistical model that underlies the EA Sports Player Performance Index, which is described in McHale et al. (2012). This fortunate event rekindled my work in this area, and further student projects and some publications followed, connected with tournament design (Scarf et al., 2009), cricket (Scarf et al., 2011) and my own passion for mountain running and orienteering (Scarf, 2007). My latest work concerns the question of competitive balance in rugby union, which also uses the Poisson distribution as its basis.

So where is this narrative leading and what has it to do with final year projects for undergraduates? Well, my first point is that the sports industry wants analysts (Brady et al., 2017), not least to repeat the 'Moneyball' success (Lewis, 2003). My second is that sport provides interesting projects, for which data are widely available and easily collected. The third is that sport gives students the opportunity for the application of modelling to the real world (Porter and Bartholomew, 2016). The fourth is that while developments of Maher's model are too difficult for undergraduate projects, many simpler, related questions remain, and arise contemporaneously. Indeed, in the next section, in which I describe open questions that would make suitable projects, I begin with the recent decision of FIFA to extend the soccer World Cup finals. The open questions are presented as something of a list, organised around broad headings, with some discussion of how they might be tackled, and what technical issues may arise and how they can be avoided. The paper finishes with the rugby union question. I hope this list will provide useful inspiration for teachers of undergraduates. I am not aware of a work that has set out to classify projects in this way.

2. Open questions and projects

2.1. Tournament design

FIFA has announced that the World Cup finals in 2026 will comprise 48 teams rather than 32. The immediate question is what effect this will have on the tournament. Obviously it will be larger (more matches), but what will be the effect on its competitiveness? A rather simple argument is that there will be none because the strength of the teams ranked, during the qualification stage, between 33rd and 48th in the world are only very marginally weaker than those ranked say 17th to 32nd. Actually, one can speculate that a greater proportion of matches in the tournament as a whole will be more balanced. What will be the effect on the tournament outcome? The simplest way to study this is to calculate the probability that the best team wins using simulation. A final year project might take the form: *'Use simulation to study the effect on the probability that the best team wins of increasing the tournament size from 32 to 48'*. Another interesting question is: *'How does the number of unimportant matches vary with the size of the tournament?'* A match is deemed unimportant if its outcome has no effect on the tournament outcome, e.g. in a group stage a match between teams who are already eliminated. To tackle these questions, the Maher model can be used to simulate individual matches, using either team strengths that are estimated by the investigator using data collected by the investigator (a more difficult variation of the project) or team strengths that have been reported by others in published work and that are assumed typical of the tournament studied (a less difficult variation of the project). In the less difficult project, the investigator might assume that team strengths of additional teams are the same as those ranked 17 to 32, for example. Or existing ranking lists might be used to modify strengths slightly. In both variations, it is necessary to code: assignment of teams to groups and matches, match outcomes between known teams; progression of teams to knockout stages; match outcomes decided following a tie at full-time (a 'coin-toss' simplest); and repetitions of the complete tournament.

Many variations on this project are then possible: *'Investigate different forms of tournament for the soccer World Cup finals'*. Here, different combinations of group sizes, number of group rounds, and number of knockout rounds can be investigated. Other projects can look at variation in rules for seeding, *'Investigate the effect of different seeding procedures on the probability that the best team wins'*. Interesting questions relate to the UEFA Champions League: its pre-tournament qualification rounds; and the appropriate number of qualifying teams from each national league. These are timely questions because sports administrators continue to fiddle with tournament design.

Reproduction of the results of Maher (1982) would make a starting point for a more demanding project that attempts to investigate developments of the Maher model and its estimation. Other sports tournaments, in e.g. cricket, rugby union, tennis, can be investigated using the model of Bradley and Terry (1952). Within this theme softer projects are possible, for example: *'Carry out a comparative study of tournament design across European football leagues'*; *'Discuss the relative merits of the different designs that have been used in the cricket World Cup'*.

2.2. Rule Changes

Studies of tournament design changes are similar in scope to studies of changes to scoring systems. Thus one might set a project to: *'Study the effect of changes in the scoring system in badminton'*, focusing on the scoring rules before and after 2006, when scoring only on serve changed to scoring regardless of serve, and games from first to 15 to first to 21. It turns out that this question has been studied by (Percy, 2009). However, a large number of variations on this theme in any one of a number of sports are possible. The investigator might even propose a favourite sport to investigate in what-if analysis. Such studies require a simple model for winning a

point, on serve and against serve. A *Bernoulli trial* with a different win probability in each case will do. Tennis, with its archaic scoring system, could be investigated: *'Is serve dominance the determining factor in game length?'* This study could contrast tennis, where games are short (to 4) and many (>12), with badminton, where games are long (to 21) and few (3).

2.3. Score distributions and dependency

In football, scores are almost independent. *'In basketball, what is the nature of score dependency?'* Possession changes and high scoring rates suggest strong dependency, but this may not be true. One could *'Classify team ball sports by score dependency'*, by collecting scores in high profile tournaments for each sport, calculating correlations between scores, and then attempting to relate these to the nature of the sports and perhaps even their popularity. In sports with many different means of scoring points (e.g. variations of football), one has to determine whether to focus on points or numbers of scoring events. One might *'Investigate dependency between scoring types in sports with multiple scoring modes'*. For example, in rugby union, is the number of tries correlated with the number of penalties? The "Poissonness" of scores might be investigated: *'Investigate in sport S the nature of the distribution of scores'*.

2.4. Competitive balance

The final questions that I consider are presented in a little more detail, and investigate the relationship between outcome uncertainty and scoring rate. They originate from three questions that may themselves be posed in projects: *'What are the essential characteristics of a popular sport?'*; *'Why is football (soccer) so popular worldwide?'*; and *'Is rugby union becoming increasingly uncompetitive?'*. The first two questions here are rather broad. Consequently, they offer possibilities for more or less technical solutions. A softer study might carry out a survey using a questionnaire. A more technical project might relate competitiveness to measures that are surrogates for popularity. In the latter, one requires a measure of competitiveness. Many have been proposed in the sports economics literature (e.g. Utt and Fort, 2002). Here the terms competitiveness, competitive balance, and uncertainty of outcome are used interchangeably.

It is my own conjecture that rugby union is becoming less competitive. If this is true, it is natural to ask *'Why?'* Stepping away from a real sport for a moment, let us construct a mathematical sport in which scores follow independent Poisson distributions, and investigate the relation between competitive balance and scoring rate. Let us call such a game a "pure Poisson match" and investigate the question: *'In a pure Poisson match, what is the relationship between team strengths, scoring rates and uncertainty of outcome?'*

To develop this idea a little further, we need some preliminaries. Let Y be a random variable. Then for any $t \neq 0$, $\Pr(Y \geq 0) = \Pr(e^{tY} \geq 1) \leq E(e^{tY})$ by Markov's inequality (since e^{tY} is a non-negative random variable). Now let $Y = X_1 - X_2$ where X_1 and X_2 are independent. Then

$$\Pr(X_1 - X_2 \geq 0) \leq E(e^{t(X_1 - X_2)}) = E(e^{tX_1})E(e^{-tX_2}). \quad (1)$$

Now suppose $X_1 \sim \text{Po}(\lambda_1)$ and $X_2 \sim \text{Po}(\lambda_2)$ independent, so that (X_1, X_2) is a pure Poisson match. The moment generating function of X_i is given by $E(e^{sX_i}) = \exp\{\lambda_i(e^s - 1)\}$. Therefore from (1) we obtain

$$\Pr(X_1 - X_2 \geq 0) \leq \exp\{\lambda_1(e^t - 1) + \lambda_2(e^{-t} - 1)\}.$$

Now setting $t = \log(\sqrt{\lambda_2 / \lambda_1})$, we obtain

$$\Pr(X_1 - X_2 \geq 0) \leq \exp\{\lambda_1(\sqrt{\lambda_2 / \lambda_1} - 1) + \lambda_2(\sqrt{\lambda_1 / \lambda_2} - 1)\} = \exp\{-(\sqrt{\lambda_1} - \sqrt{\lambda_2})^2\}.$$

Now let $\lambda_1 = \lambda$ and $\lambda_2 = \varepsilon\lambda$ for some $\varepsilon > 1$, so that team 2 is slightly stronger than team 1. Then we have that $\Pr(X_1 \geq X_2) \leq \exp\{-\lambda(1 - \sqrt{\varepsilon})^2\} \rightarrow 0$ as $\lambda \rightarrow \infty$ for fixed ε . Therefore $\Pr(X_1 < X_2) \rightarrow 1$ as $\lambda \rightarrow \infty$ for fixed ε . Therefore, in a pure Poisson match, no matter how close are the strengths of the two teams, in the limit (for a very large scoring rate) the stronger team will always win and the match is perfectly competitively unbalanced. Further, if $\lambda_1 = \lambda_2 = \lambda$ then $\Pr(X_1 = X_2) \rightarrow 0$ as $\lambda \rightarrow \infty$ (proof omitted) and $\Pr(X_1 < X_2) = \Pr(X_1 > X_2) \rightarrow 1/2$ as $\lambda \rightarrow \infty$ (by symmetry). A technical project might ask for proofs of these results.

The above deals with the asymptotic behaviour of $\Pr(X_1 \geq X_2)$. For the exact calculation of $\Pr(X_1 \geq X_2)$ for $X_1 \sim \text{Po}(\lambda)$ and $X_2 \sim \text{Po}(\varepsilon\lambda)$ we can use

$$\Pr(X_1 < X_2) = \sum_{y=1}^{\infty} \sum_{x=0}^{y-1} \{e^{-\lambda} \lambda^x / x!\} \{e^{-\varepsilon\lambda} (\varepsilon\lambda)^y / y!\}, \quad (2)$$

and

$$\Pr(X_1 = X_2) = \sum_{x=0}^{\infty} \{e^{-\lambda} \lambda^x / x!\} \{e^{-\varepsilon\lambda} (\varepsilon\lambda)^x / x!\} = \sum_{x=0}^{\infty} e^{-\lambda(1+\varepsilon)} \lambda^{2x} \varepsilon^x / (x!)^2. \quad (3)$$

A less mathematical, more empirical project might use these exact formulae to illustrate the asymptotic results graphically. The probabilities (2) and (3) could also be evaluated by simulation.

Now the question of dependence of scores in particular sports is a pertinent one. An '*Investigation of the relationship between scoring rates and outcome uncertainty in a double Poisson match*' would make an interesting, empirical study. In the double Poisson match (Karlis and Ntzoufras, 2003) scores are not independent. One might speculate that scores in rugby union show some dependence, but in spite of this as scoring rates increase outcomes become less certain. Finally, the project to '*Investigate the evolution of scoring rates in rugby union over time*' can shed some light on the state of this sport and what if anything its administrators should do about it.

3. Conclusion

Sport offers many opportunities for projects. This is because data are relatively abundant and easy to obtain, models are relatively intuitive, and the context is often evident to the investigator. There is also a sport for everyone, and arguably a mathematical project exists within every sport. This paper has considered a number of projects that are suitable for final year undergraduates. These projects are unified within the notion of a Poisson match. The projects are by no means trivial, and some pose questions of which administrators of the respective sports should take note. One wonders indeed if administrators use modelling at all to consider proposals for change. Important decisions should be based on evidence and such evidence should be scrutinised by good modellers. Thus as the business of sport grows, there is a growing need for trained modellers to work in sport, and projects on sport are a stepping stone to employment.

4. References

- BBC Sport, 2017. Premier League: Chelsea to win by seven points? Arsenal outside top four? Available at: <http://www.bbc.co.uk/sport/football/38524102>. [Accessed 12 January 2017].
- Bradley, R.A. and Terry, M.E., 1952. The rank analysis of incomplete block designs 1. The method of paired comparisons. *Biometrika*, 39, pp.324-345.
- Brady, C., Forde, M. and Chadwick, C., 2017. Why your company needs data translators. *MIT Sloan Management Review*, Winter 2017 issue.

- Buraimo, B. and Simmons, R., 2015. Uncertainty of outcome or star quality? Television audience demand for English premier league football. *International Journal of the Economics of Business*, 22, pp. 449-469.
- Dixon, M. and Coles, S., 1997. Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, 46, pp. 265–280.
- Forrest, D. and Simmons, R., 2002. Outcome uncertainty and attendance demand in sport: The case of English soccer. *Journal of the Royal Statistical Society: Series D*, 51, pp.229–241.
- Karlis, D. and Ntzoufras, I., 2003. Analysis of sports data using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D*, 52, pp.381–393.
- Koopman, S. and Lit, R., 2015. A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society: Series A*, 178, pp.167–186.
- Maher, M., 1982. Modelling association football scores. *Statistica Neerlandica*, 36, pp.109–118.
- Lewis, M., 2003. *Moneyball: The Art of Winning an Unfair Game*. New York: Norton.
- McHale, I. and Scarf, P., 2007. Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Statistica Neerlandica*, 61, pp.432-445.
- McHale, I. and Scarf, P., 2011. Modelling the dependence of goals scored by opposing teams in international soccer matches. *Statistical Modelling*, 11, pp.219–236.
- McHale, I., Scarf, P. and Folker, D., 2012. On the development of a soccer player performance rating system for the English Premier League. *Interfaces*, 42, pp.339-351.
- Owen, A., 2011. Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics*, 22, pp.99–113.
- Percy, D., 2009. A mathematical analysis of badminton scoring systems. *Journal of the Operational Research Society*, 60, pp.63-71.
- Porter, R. and Bartholomew, H., 2016. When will I ever use that? Giving students opportunity to see the direct application of modelling techniques in the real world. *MSOR Connections*, 14, pp.45-49.
- Scarf, P., 2007. Route choice in mountain navigation, Naismith's rule and the equivalence of distance and climb. *Journal of Sports Sciences*, 25, pp.719-726.
- Scarf, P., Mat Yusof, M. and Bilbao, M., 2009. A numerical study of designs for sporting contests. *European Journal of Operational Research*, 198, pp.190-198.
- Scarf, P., Shi, X. and Akhtar, S., 2011. The distribution of runs scored and batting strategy in test cricket. *Journal of the Royal Statistical Society, Series A*, 174, pp.471-497.
- Utt, J. and Fort, R., 2002. Pitfalls to measuring competitive balance with Gini coefficients. *Journal of Sports Economics*, 3, pp.367-373.