CHANGING THE ASSOCIATION CROQUET WORLD RANKING SYSTEM

A REPORT BY THE WCF RANKING REVIEW COMMITTEE:
IAN BURRIDGE, CHRIS DENT, JONATHAN KIRBY,
DAVID MAUGHAM, LOUIS NEL AND CHRIS WILLIAMS

9TH MAY 2011

Contents

1. Introduction

The WCF Ranking Review Committee was asked to look at the algorithm used to calculate the Association Croquet World Rankings. There were two particular concerns with the existing system (the Continuous Grading System, or CGS): the lag effect, and whether or not it treats rapid improvers well.

After long deliberations, the committee has come up with a system called Dynamic Grading (DG) which it is recommending as a replacement for the CGS as the official Association Croquet World Rankings. The DG system does not have the lag effect at all, and it incorporates a feature called a dynamic modulator which, amongst other things, accommodates rapid improvers better than the CGS.

The DG system is described in detail in the article *Introduction to Dynamic Grading*, which is available at `http://oxfordcroquet.com/tech/nel-dg/index.asp`. In this report we give a brief comparison of it with the CGS, focusing on the major differences in practice: the lag effect and the dynamic modulator.

Rankings calculated by the DG system have been available in the same place as the official world rankings for the past year or so: `http://butedock.demon.co.uk/cgs/rank.php`

2. The CGS Index

Every world ranking system we have considered for croquet is really a *grading system*, which means each player has a grade (a number representing how good their results have been) and the ranking list is produced by listing the players in grade order. The simplest grading system we considered was the Index of the current CGS. When each game is played, 50 points are at stake. If the two players start with the same index, the winner gains 25 points and the loser loses 25, so the difference between winning and losing is 50 points. If the winner had a higher index to start with then fewer points are transferred, likewise more points change hands for an upset result, but for each player the difference between winning and losing is always 50 points. With each game worth 50 points, the index produces very volatile rankings, so this system has never been used on its own.

3. Smoothing, the CGS, and lag

The approach to reducing the volatility taken in the CGS is that the CGS grade is a smoothed version of the index. In effect, the points won or lost from a game are not added or subtracted from the grade immediately, but instead gradually as more games are played. (This is not exactly how the grade is calculated, but is equivalent.) For a player who has a good balance of wins and losses, the gains and losses of points mostly cancel out, so the grade is very steady. However, if a player has a run of wins, their index will get well ahead of their grade, so the grade then lags behind and may not reflect their true performance level. This is a particular problem for rapid improvers who often have an extended run of good results, and their grade is constantly well below their observed form largely because of this time lag. However, it is a problem to a lesser extent for all players. For example, it is very common for a player to win a game and yet for their grade to go down or to lose and for their grade to go up. This is because the grade change depends to a large extent on earlier games, and not just the most recent game. Due to an inevitable (but surprising) feature of the mathematics of the smoothing mechanism, the most recent game result has only a very small bearing on the current grade (about 2.5 grade points), with games around 12 to 30 prior to the most recent each have a seven times larger effect (about 18 grade points) on average. The practical effect is that CGS grades are always out of date – they

give very small weighting to the most recent games a player has played compared with older games.

4. Modulators and volatility

A different way to reduce the volatility is to replace the number 50 in the calculation of the index by a different number, but then use the resulting index directly, without further smoothing. We called this number the *modulator*. The higher the modulator, the more volatile the system, and the more a player's index jumps around depending on the most recent games. The lower the modulator, the more stable the index and the more it reflects a long-term average performance. Our tests showed that setting the modulator to about 24 gave the best results overall – not nearly as volatile as the original index but still sufficiently volatile for players' genuine changes in form to be recognised.

5. Dynamic modulators and DG

However, just using the index with a modulator of 24 is a compromise: for the more steady players, a modulator of about 16 would be ideal, but for rapid improvers, something as high as 35 would be better. The Dynamic Grading system allows the best of both worlds with a dynamic (variable) modulator. Each player has their own personal modulator which can vary between 16 and 35.2. For players whose results over the last 37 games have been closely in accordance with their grade, their modulator will be at the low end of the scale. For those whose results have been very much better or worse than might have been expected, the modulator will be high.

In fact the published DG rankings list two numbers: the grade itself and the *performance deviation trend* or *pdt*, which measures approximately how many points above or below grade the player has played over the last 37 games. The modulator M is calculated in terms of the pdt by the formula:

$$M = 16 + 19.2 \times \mathrm{PDT}^2/(1 + \mathrm{PDT}^2)$$

where $\mathrm{PDT} = \mathrm{pdt}/92$.

Note, the pdt is how well you played compared to your grade at the time of each game, not how well you played compared with your current grade. The current grade is still the system's best guess at how good you are now. The other thing to bear in mind is that pdt only very approximately measures how many points ahead or behind your grade you have played – that is not exactly how it is calculated, but just a rule of thumb.

One difference with the CGS is that, since the two players of a game may have different modulators, the winner may gain a different number of points from the number the loser loses. It is not a *zero sum* system. This does not seem to cause any problems in practice.

6. Other systems

The committee came up with various other ranking systems during its deliberations, the most notable being the Bayes Ranking system and the Adaptive Bayes Ranking system, both documented on the OxfordCroquet website. However, none of the other systems is close to being as good as DG. Indeed, the second best system we came up with was the simplest system, just using a fixed modulator of 24.

7. Class Factors

The existing CGS uses *class factors* with games in the most important events (World Championships, International matches, and some other major tournaments) called class 1, most plate (consolation) events being class 3, and everything else being class 2. In fact the CGS uses a modulator of 50 only for class 2 events, with class 1 events having a modulator of 60 and class 3 events having a modulator of 40. The DG system (or indeed any other system) can incorporate class factors by increasing the modulator by 20% for class 1 games and reducing it by 20% for class 3 games.

The main reason for using class factors is to make the weight of a game for ranking purposes depend on its class, which it achieves. However, there is an significant side effect: any player who plays a significant number of class 1 games will have a higher modulator and therefore a more volatile grade, which means more based on the most recent results and less based on longer-term form. Since it is mostly the top players who play most class 1 games, the result of using class factors is that the rankings of the top players tend to become more volatile and dependent on recent results. In around 1993, it was decided that the CGS was too volatile for the top players and so a variable smoothing factor was introduced specifically to reduce the volatility of the top players' grades. A side-effect of that is that the lag effect in CGS is worse for the top players. Removing the class factors would have been another way to reduce volatility at the top. In this regard, it is interesting to note that Chess ratings use a smaller modulator for higher-graded players – essentially the opposite of class factors. While we found no advantage to such a move for croquet rankings, introducing class factors into DG did produce a marked and measurable degradation of its performance, so we recommend the system without class factors.

8. Numerical comparison of the different systems

An important test of whether the rankings and grades produced are accurate is to measure how well the grades predict the results of games. More accurately, the difference in grades between two players gives a prediction of the probability of who will win a game between them, and one can measure how good these predictions are, using a statistic called Grade Deviation (GDev). It is important

that this prediction is as good as possible, since if one player's grade is wrong, then anyone who plays him will get the wrong reward for beating him, or the wrong penalty for losing to him, and so his grade will become inaccurate and the effect will propagate through the system. The modulator in DG has been set to optimise the prediction.

We tabulate GDev below, and also the Percentage of Wild Performance Games (PWPG) and the Average Rank Variation (ARV). The former measures the percentage of games in which at least one player has a grade which has been at least 200 points different from their performance over the previous 37 games. It effectively measures how many players have grades which appear to have been substantially wrong over a recent period. ARV measures the volatility of the ranking lists (not the grades), by giving the average number of places a player moves up or down per month under that system.

System	GDev	PWPG	ARV
DG	0.895	6.78	8.96
DGcf	0.964	6.85	8.91
I_{24}	0.986	6.91	9.20
CGS	2.642	10.13	8.62

DG = Dynamic Grade
DGcf = Dynamic Grade with class factors
I_{24} = Simple Index system with fixed modulator of 24
CGS = Continuous Grading System (the current system)

In the case of GDev and PWPG, a lower number is better. The lag effect of CGS which means that the grades are always out of date results in a much worse overall performance than the other systems.

The optimal value of ARV is more subjective, but the values allow comparison between the systems. In fact, all are similar, so the actual ranking lists produced by each method change about the same amount from month to month.