

Glicko Explained

Introduction

There has been some interest and discussion in recent months on the ACF rating system. Mostly, this has concerned the relative virtues and shortcomings of the current Glicko system and the previously employed Elo system. In this article I shall try to explain the workings of both of these systems, but with particular emphasis on the Glicko system. The reason for this emphasis is that the Glicko system is newer and, it seems, the less widely understood of the two.

The Elo System

The Elo system takes its name from Arpad Elo who proposed the system in the 1950's and published his well-known monograph *The Rating of Chessplayers, past and present* in 1978. The Elo system was employed by the ACF until the mid-1990s. As far as I am aware, it is also still in use by FIDE, the Correspondence Chess League of Australia (CCLA) and many national chess federations throughout the world.

Under the Elo system each player has a rating, which is a number that nominally varies from 0 to 3000. Whenever a player participates in a game with a rated opponent, the player's ratings is modified depending on the actual result of that game compared to the expected result of the game.

The formula used to calculate the new rating is

$$r_{post} = r_{pre} + K (S - E)$$

r_{post} is the new rating after the adjustment has been made

r_{pre} is the published rating at the time of the game

K is a measurement of how much the rating will change as a result of each game. Generally this number will be fixed to one or just a couple of values for an entire rating system. For example the ACF's old system used a K factor of 15. The FIDE system uses K -factors of 25, 15 or 10 depending on the player's rating.

S is the result of the game. 1 if the player won, 0.5 in the case of a draw, or else 0 if the player lost.

E is the expected result of the game. This varies from 0 to 1 depending on the difference in ratings of the player and their opponent. Where the ratings are equal, the expected result is 0.5. Where the ratings vary, the expected score is calculated under the assumption that the results between players are distributed normally with a standard deviation of $200 \times \sqrt{2}$ (that is, approximately 282.84) rating points. This can be quickly calculated using a spreadsheet, calculator or looked up in a Statistics reference book. (Some systems use other values for this standard deviation or even approximations based on other formulae.)

For example, where the player has a rating of 1500 plays and defeats a player with a rating of 1780 assuming a K value of 15 the new rating (r_{post}) would be

$$r_{post} = 1500 + 15 (1 - E)$$

where, in this case, $E \approx 0.1611$

$$r_{post} \approx 1500 + 15 * 0.8389 \approx 1500 + 13 = 1513$$

So the player would effectively gain 13 rating points from the game.

The Elo system worked well for many years however there were some problems. Firstly, there was rating drift, where the ACF rating would fall out of kilter with the FIDE ratings, even though a similar system was being used to calculate both ratings. This has caused the ACF to take steps like adding 150 rating points to most ratings so that an ACF rating can be more closely equated to a FIDE rating.

Another problem was that ratings officers would like to adjust some players' ratings more than others. Particularly newly rated players and quickly improving juniors may find themselves grossly underrated and it may take an long time before their Elo rating was a good approximation of their actual playing strength. The other scenario is the player who stops playing chess in Australia for a number of years and then makes a return. In this scenario the player's strength may have changed (up or down) substantially and it may take some time for the player's rating to reflect this change.

Many systems have been employed to address this second problem in the context of the Elo system. One of this is to vary the K value depending on rating, age, or number of rating games played. However, Professor Mark Glickman of Boston University had another idea. His concept was to not only record a rating for every player but also record a measurement of the level of confidence in that rating.

The Glicko system

The name *Glicko* is derived from the names Glickman and Elo. The Glicko system records a rating which looks very much like a Elo rating, again nominally varying from 0 to 3000. However, a Glicko system also records a Rating Deviation (R.D.), which is a measurement of the confidence that is held in any rating. The R.D. varies from 30 to 350.

When thinking about the Glicko system it is important to remember that a rating is nothing more than an approximation of an actual playing strength. Since playing strength can never be known exactly, the rating is always just an approximation. The R.D. is a measurement of how good this approximation is held to be.

For example, a rating of 1500 with an R.D. of 60 means there is around a 70% confidence that the actual playing strength is between 1440 and 1560 (that is, within 1 R.D. of the rating). Another way of expressing this is a 95% confidence that the actual playing strength is between 1380 and 1620 (that is, within 2 R.D.s of the rating).

The Glicko system works very much like the Elo system where adjustments are made to ratings, depending on actual compared to the expected results to derive a new rating. However, the R.D. comes into play, especially where an unexpected result occurs against an opponent with a high R.D. where the effect of that result is de-emphasised. Likewise, when a player has a high R.D. their rating will change quickly compared to someone with an established rating (that is, lower R.D.).

The calculations used in the Glicko system are an order of magnitude more complex than those used in the Elo system. I will cover the formulae and include a worked example but will not dwell on them. At the end of the article I will include a web reference including link to the more detailed work by Prof. Glickman as well as a Microsoft Excel® spreadsheet developed by yours truly which performs Elo and Glicko calculations.

With Glicko there is one system parameter, *c* which is a measurement of how quickly the R.D. increase over time. The ACF ratings officer informs me that they are currently using a *c* of around 42.4264 (that is, $c^2 = 1800$). There is also a constant, *q*, which is equal to $\log_e(10)/400$ or around 0.0058565.

The first step in a Glicko calculation is to calculate the *Onset R.D.* This is done for the player and the opponent.

$$RD = \sqrt{(RD_{old})^2 + c^2 t}$$

RD_{old} is the original RD.

c is the aforementioned parameter.

t is the number of rating periods since the player last played.

Step two is to calculate a *g* based on the opponent's onset RD where

$$g = 1 / \sqrt{(1 + 3q^2 (RD)^2 / \pi^2)}$$

The expected result can then be calculated by

$$E = 1 / (1 + 10^{-g(pr - or)/400})$$

pr is the player's rating

or is the opponent's rating

Before the new rating is calculated, the new RD for the player is calculated. This factor is used in the rating calculation. As previously mentioned, the RD is inflated to an onset RD. This is then adjusted according to the following formulae.

$$d^2 = 1 / (q^2 g^2 E (1 - E))$$

$$RD_{new} = 1 / \sqrt{(1/RD^2 + 1/d^2)}$$

The new rating is then calculated as follows

$$r_{post} = r_{pre} + q RD_{new}^2 g (S - E)$$

r_{post} is the new rating after the adjustment has been made

r_{pre} is the published rating at the time of the game

S is the result of the game. 1 if the player won, 0.5 in the case of a draw, or else 0 if the player lost.

For example, where the player has a rating of 1500 plays and defeats a player with a rating of 1780 assuming a RD of 60 for both ratings then the new rating would be calculated as follows.

$$RD = \sqrt{(60^2 + 1800 \times 1)} = \sqrt{(3600 + 1800)} \approx 73.4847$$

$$g \approx 1 / \sqrt{(1 + (3 \times 0.0000331396 \times 5400)/9.8696)} \approx 1 / \sqrt{1.0544} = 0.97386$$

$$E \approx 1 / (1 + 10^{-0.97386(1500 - 1780)/400}) \approx 1 / (1 + 10^{0.68170}) \approx 0.17226$$

$$d^2 \approx 1 / (0.0000331396 \times 0.97386^2 \times 0.17226 (1 - 0.17226)) \approx 1 / 0.0000044815 \approx 223139$$

$$RD_{new} \approx 1 / \sqrt{(1/5400 + 0.0000044815)} \approx 1 / 0.013772 \approx 72.611$$

$$r_{post} \approx 1500 + 0.0058565 \times 72.611^2 \times 0.97386 \times (1 - 0.17226) \approx 1500 + 24.891 \approx 1525$$

So the new rating would be 1525 with a new RD of 73. The R.D. increased because the player did not play enough games in this rating period (one) to maintain the relatively low R.D.

To calculate a rating adjustment when more than one rated game has been played in a single rating period, the $g^2 E (I - E)$ for each opponent is totalled and a single d^2 is calculated for the whole period. Likewise the $g (S - E)$ factor is totalled for each game and a single r_{post} calculation is performed. To get a better idea of how these calculations work please review the spreadsheet or the Prof Glickman article, which contains a worked example for a player with three games in a single period.

OK, so these calculations can appear daunting and, if you are still reading, I promise there will be no more. As mentioned previously I'll include a link to a Microsoft Excel® spreadsheet that you can use to perform Elo and Glicko calculations without headaches.

Pros and Cons

One benefit of the Glicko system is that every rating has a measurement of confidence in the accuracy of the rating. This can be useful in itself. However, Glicko's main benefit is that ratings are adjusted by an amount depending on the confidence held in the ratings involved. Confidence decreases as time passes which is reasonable since a player's ability is likely to have changed during an extended period of non-rated play.

On the downside, the calculations are more complex. While it may take specialist statistics knowledge (which I do not possess) to understand the derivation of the formulae, these calculations could be performed by anyone with a scientific calculator and the patience to crank the handle. Alternatively, once the calculations have been programmed into an application (like a spreadsheet) all one need do is enter the ratings and results in the appropriate cells and let their computer do the rest.

Another criticism that has been aimed at the Glicko system is that it encourages people to not play for some time so to increase their RD. Then, when they make their triumphant return to rated play, their rating increases will be magnified. While this sounds good (or bad) in theory, one should remember that it is unlikely that the rating will exceed that player's current performance strength. So you won't get a 2500 rating unless you are having those sort of results anyway. Secondly, if on returning the performance in the first few tournaments is not so good, then the rating will reduce more quickly too. As it is easier (in my experience) to have a bad tournament than to have a good one, it seems to me that it is more likely that this tactic would backfire.

There does exist the phenomenon where a player's rating may increase beyond current playing strength, which I call "rating overrun". I don't want to dwell on this problems too much but briefly there has been some criticism of the Glicko system based on some scenarios where a player plays many game above their currently rating and the Glicko adjusted rating overshoots their performance rating by a substantial margin. There is also a related phenomenon where the rating is decreasing below performance rating, which I call "rating underrun" - which is basically the same thing happening in reverse. What must be understood is that both of these problems exist in the Elo and the Glicko system. However, with Glicko's quick adjustment of ratings with high RD's it seems to be more noticeably a problem with the Glicko system. I understand that the ACF have implemented a system where they look for cases of these phenomena and intervene to ensure their effect is reduced.

Unrated Players

When reading the articles by Professor Glickman material one should note that the standard way of handling unrated player (start on 1500 with RD of 350) is not the system that the ACF uses. The ACF looks a performance of the first games played and uses this to estimate the starting rating. Therefore, when a player has played a game against an unrated player, there is no way for that player to determine what affect that game may have on their rating, if any at all. The starting rating of a player depends on their performance in the period before their first published rating.

Conclusion

I hope this article has gone some way to clear up the workings of the Glicko system and how it compares to the Elo system. As promised, there is a spreadsheet available on the world-wide web at http://www.geocities.com/b_j_cox/glicko.htm. On this page you will also find links to the original work by Prof Glickman as well as the ACF own rating lists.