Tiddlywinks World Ratings (Version 3.1)

Patrick Barrie

"An elegant example of error analysis using a Bayesian statistical framework"

Note: a description of Version 3.0 of the tiddlywinks world ratings calculation method has been published: P. J. Barrie, *Journal of Applied Statistics*, **30**, 361-372 (2003).

1. Introduction

The ratings program calculates an estimate of a player's tiddlywinks ability based on their scores in national tournaments and official club matches. Ratings are calculated on a tournament-by-tournament basis. For each player the ratings change takes into account the game scores, the ability of partners and opponents, and the uncertainties in the ability of the players involved.

For the ratings to be reliable, they need a reasonably firm mathematical basis that does not make too many arbitrary judgements over the parameters that affect the calculations.

The program has to handle tournaments that have

- Many players or very few players
- Many games or very few games
- Singles game or pairs games (or a combination of the two)
- Fixed partnerships or varying partnerships

In addition, some participants may be playing in their first ever tournament, while others may be returning to the game after a long absence.

The ratings of "active" players – those who have played in a formal tournament within the last calendar year – are reported after every tournament. The ratings of "inactive" players are stored so that they can be used should they return to tournament winks.

It should be noted that the program is a ratings system with predictive power of game scores. It is not an arbitrary rankings system, even though rankings are reported because people seem interested in them.

2. Predicted game score function

The ratings programme predicts the score of a player in game number i by the following function(s)

$$\tilde{y}_i = 3.5 + 3.55 \operatorname{erf}\left(\frac{x + p_i - q_{1i} - q_{2i}}{1600}\right) \quad \text{for pair-pair games}$$

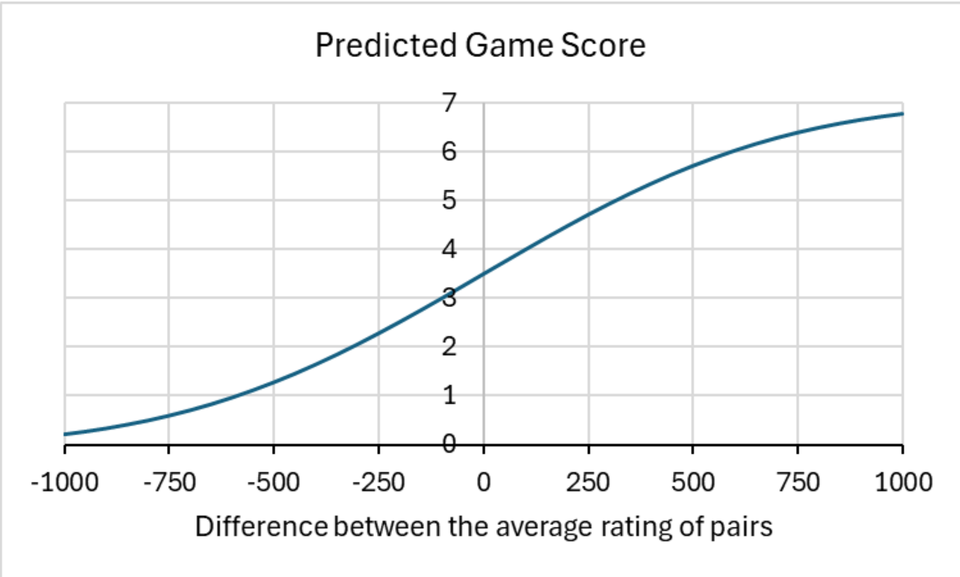
where x denotes the player's rating, p_i denotes the partner's rating, and q_i denotes an opponent's rating. The function erf is the error function, $\operatorname{erf}(z) = (2/\sqrt{\pi}) \int_0^z \exp(-t^2)\, dt$. Analogous equations describe the predicted score in other game situations:

$$\tilde{y}_i = 3.5 + 3.55 \operatorname{erf}\left(\frac{2x - 2q_i}{1600}\right) \quad \text{for single-single games}$$

$$\tilde{y}_i = 3.5 + 3.55 \operatorname{erf}\left(\frac{x + p_i - 2q_i}{1600}\right) \quad \text{for pair-single games}$$

$$\tilde{y}_i = 3.5 + 3.55 \operatorname{erf}\left(\frac{2x - q_{1i} - q_{2i}}{1600}\right) \quad \text{for single-pair games}$$

The predicted game score function is illustrated in the graph below.



It can be seen that an average ratings points difference between pairs of:

- 100 points predicts a 4-3 win
- 205 points predicts a 4.5-2.5 win
- 315 points predicts a 5-2 win
- 440 points predicts a 5.5-1.5 win
- 590 points predicts a 6-1 win
- 805 points predicts a 6.5-0.5 win

3. Prior and posterior distributions

At the start of each tournament, each player involved has a current rating (μ_0) and a current uncertainty or error bar (σ_0) in that rating. The program assumes that the ability of each player is described by a normal distribution with a mean of μ_0 and a standard deviation of σ_0. This forms what is known as the prior distribution for that player.

In Bayesian statistics, the posterior probability distribution for a random variable is found by adjusting the known prior distribution using a likelihood function that is determined from measurements. The game scores in a tournament are used, together with previous ratings and uncertainties, to calculate the likelihood function. The likelihood function is assumed to have a normal distribution characterised by μ_X and σ_X where

- μ_X is the tournament rating for that player – a measurement of the player's ability from that tournament alone
- σ_X is the uncertainty in the tournament rating for that player and is calculated by error analysis.

For a prior distribution that is normal characterised by μ_0 and σ_0, and a likelihood function that is normal characterised by μ_X and σ_X, the resulting posterior distribution is also normal and is characterised by

$$\mu_n = \frac{\sigma_0^2 \mu_X + \sigma_X^2 \mu_0}{\sigma_0^2 + \sigma_X^2}$$

$$\sigma_n^2 = \frac{\sigma_0^2 \sigma_X^2}{\sigma_0^2 + \sigma_X^2}$$

These equations are how a player's new rating μ_n and associated uncertainty σ_n are calculated.

The ratings program constrains the standard deviation σ_n in a player's rating to lie between 70 ratings points (for established players) and 250 ratings points (for tournament newcomers). Rather than report the standard deviation σ_n directly, the ratings program instead reports a "Rating Reliability Factor" (RRF) given by

$$\mathrm{RRF} = \frac{250 - \sigma_n}{1.8}$$

This means that RRF values lie between 0 (for tournament newcomers) and 100 (for established players).

4. Calculating the tournament rating μ_X of a particular player

The total number of points scored in a tournament by a particular player is a random variable; it will differ if the tournament is repeated, even if the participants play to the same standard. Let this random variable be Y. It will be the sum of the individual games scores (each of which are random variables):

$$Y = \sum_{i=1}^{n} Y_i$$

where n is the number of games played by the player. The tournament rating of the player – a measure of "how well the player performed" is described by random variable X.

In a tournament, the value of the total number of points $y = \sum y_i$ gained by a player enables the tournament rating x for that player to be calculated from

$$f(x) = y - \sum_{i=1}^{n} \tilde{y}_i = 0$$

where $\tilde{y}_i$ is the predicted score in game number i.

This means that the tournament rating is the value of x for the player that would predict exactly the total number of points gained by that player (assuming at this point that the ratings of partner and opponents are known exactly). Because there is no direct analytical formula for x, the tournament rating for each player is found by iteration using the fsolve function within the SciPy Python library. The calculated value of x gives the value of μ_X in the likelihood function for Bayesian statistics.

5. Error analysis: calculating the uncertainty $\boldsymbol{\sigma_X}$

We now need to estimate σ_X, the standard deviation of the random variable X, which quantifies the uncertainty or error bar in the calculated tournament rating of a particular player. The error propagation equation gives

$$(\sigma_X)^2 \approx \left(\frac{\partial x}{\partial y}\sigma_Y\right)^2 + \sum_{j=1}^{m}\left(\frac{\partial x}{\partial r_j}\sigma_j\right)^2$$

where

m is the number of other players in the tournament

σ_Y is the intrinsic standard deviation of the random variable Y

σ_j is the standard deviation quantifying the uncertainty, or error bar, in the rating of participant j

This equation has assumed that any covariance terms are zero because each game score is an independent measurement.

We shall consider the two terms on the right-hand side separately.

5.1 First term in error propagation equation

For the first term:

$$\frac{\partial x}{\partial y} = \frac{1600}{3.55}\frac{\sqrt{\pi}}{2}\frac{1}{\sum_i^n \alpha_i g_i} \approx \frac{400}{\sum_i^n \alpha_i g_i}$$

where

$$g_i = \exp\left[-\left(\frac{x + p_i - q_{1i} - q_{2i}}{1600}\right)^2\right] \qquad \text{for pair-pair games}$$

and related expressions for other types of games, and

$\alpha_i = 1$ if the player was in a partnership in game i

$\alpha_i = 2$ if the player played singles in game i

Even if the ratings of players are known precisely, the population of game scores has an intrinsic distribution. σ_Y quantifies the intrinsic uncertainty in the total points gained from n games. Because each game score is an independent measurement

$$\sigma_Y = \sigma_{\text{game}}\sqrt{n}$$

where σ_{game} is the intrinsic standard deviation of scoreline for a single game.

Hence

$$\text{term 1 contribution to } (\sigma_X)^2 = \left(\frac{\partial x}{\partial y}\sigma_Y\right)^2 = n\left(\frac{400\,\sigma_{\text{game}}}{\sum_i^n \alpha_i g_i}\right)^2$$

An analysis of tournament game scores from 1985-2024 shows that the standard deviation of the population of game scores is 2.19. However, with a predictive model (that used in a previous variant of the rating program) the variation of past tournament game scores from predicted scores gave a standard deviation 1.76. Recognising that there is some uncertainty in the ratings of other players, the rating program uses a value of σ_{game} = 1.70

5.2 Second term in error propagation equation

The second term in the error propagation equation considers the fact that the ratings of partners and opponents in games are not known exactly but have associated uncertainties.

For the second term:

$$\frac{\partial x}{\partial r_j} = \frac{\partial x}{\partial y}\frac{\partial y}{\partial r_j}$$

We can derive

$$\frac{\partial y}{\partial r_j} = \frac{3.55}{1600}\frac{\sqrt{2}}{\pi}\left\{\sum_{i=1}^{n} \beta_{ij} g_i\right\}$$

where

$\beta_{ij} = 0$ if participant j was not involved in game i

$\beta_{ij} = 1$ if participant j partnered the relevant player in game i

$\beta_{ij} = -1$ if participant j was in a partnership opposing the relevant player in game i

$\beta_{ij} = -2$ if participant j played singles against the relevant player in game i

and g_i is as previously defined.

This leads to

$$\frac{\partial x}{\partial r_j} = \frac{\sum_i^n \beta_{ij} g_i}{\sum_i^n \alpha_i g_i}$$

Hence

$$\text{term 2 contribution to } (\sigma_X)^2 = \sum_{j=1}^{m}\left(\frac{\partial x}{\partial r_j}\sigma_j\right)^2 = \sum_{j=1}^{m}\left(\frac{\sum_i^n \beta_{ij} g_i}{\sum_i^n \alpha_i g_i}\sigma_j\right)^2$$

5.3 Combining the two terms

We now combine the two factors contributing to the uncertainty in the tournament rating of a players

$$(\sigma_X)^2 = n\left(\frac{400\,\sigma_{\text{game}}}{\sum_i^n \alpha_i g_i}\right)^2 + \sum_{j=1}^{m}\left(\frac{\sum_i^n \beta_{ij} g_i}{\sum_i^n \alpha_i g_i}\sigma_j\right)^2$$

To see the functionality within this expression, it is helpful to consider a limiting case. In the special (but fairly common) case that the values of g_i are all close to unity we get

$$(\sigma_X)^2 \approx n\left(\frac{400\,\sigma_{\text{game}}}{\sum_i^n \alpha_i}\right)^2 + \sum_{j=1}^{m}\left(\frac{\sum_i^n \beta_{ij}}{\sum_i^n \alpha_i}\sigma_j\right)^2$$

$$(\sigma_X)^2 \approx n\left(\frac{400\,\sigma_{\text{game}}}{n_p + 2n_s}\right)^2 + \sum_{j=1}^{m}\left(\frac{n_{j\text{ is partner}} - n_{j\text{ is opp in a pair}} - 2n_{j\text{ is opp in singles}}}{n_p + 2n_s}\sigma_j\right)^2$$

where n_p and n_s denote the number of pairs and singles games played by that participant.

The first term depends on the number of games played in the tournament, becoming smaller as more games are played.

The second term does not decrease with the number of games played. For instance, it shows the following behaviour (which are all “expected” from simple error analysis):

- For a series of single-single games against the same opponent (e.g. World Singles)

$$\text{term 2 contribution to } (\sigma_X)^2 = \sigma_{\text{opp}}^2$$

- For a series of pairs-pairs games with the same partner and the same opponents (e.g. World Pairs)

$$\text{term 2 contribution to } (\sigma_X)^2 = \sigma_{\text{part}}^2 + \sigma_{\text{opp1}}^2 + \sigma_{\text{opp2}}^2$$

- For a series of single-single games against different opponents with no repeats (e.g. National Singles)

$$\text{term 2 contribution to } (\sigma_X)^2 = \sum_{i=1}^{n}\frac{\sigma_{\text{opp},i}^2}{n}$$

- For a series of pairs-pairs games with same partner against different opponents with no repeats (e.g. National Pairs)

$$\text{term 2 contribution to } (\sigma_X)^2 = \sigma_{part}^2 + \sum_{i=1}^{n}\frac{\sigma_{\text{opp1},i}^2 + \sigma_{\text{opp2},i}^2}{n}$$

Some formats involve both partnering and opposing the same player (e.g. Cambridge Open, Individual Pairs). In that case, β_{ij} values for that participant j will contain some positive and some negative values, reducing the value of term 2. This is because that may be systematically “underrated” or “overrated” whether they are partnering or opposing – and so the effects of the uncertainty in that player’s rating may be cancelled out.

6. Calculating new ratings and uncertainties

We have now calculated μ_X and σ_X of the likelihood function and so can now calculate the new rating and uncertainty for the player using

$$\mu_n = \frac{\sigma_0^2 \mu_X + \sigma_X^2 \mu_0}{\sigma_0^2 + \sigma_X^2}$$

$$\sigma_n^2 = \frac{\sigma_0^2 \sigma_X^2}{\sigma_0^2 + \sigma_X^2}$$

However, some adjustments are needed in certain cases.

6.1 Established players

As a player plays more games and more tournaments, the uncertainty in their rating gets smaller.

A minimum value of 70 is imposed by the ratings programme.

This value controls the volatility of the rating of established players. Setting this particular value as the minimum for sigma means that an established player's rating will rarely change by more than 70 ratings points in a tournament.

6.2 Tournament newcomers and players with very low ratings

Tournament newcomers are assigned an initial rating of 1500 with an uncertainty of 250.

The high uncertainty reflects the fact that the rating program does not know whether the newcomer is a complete beginner or someone who is fairly accomplished having played many games at club level before competing in a national tournament.

It is possible that a calculated rating will be below 1500. In these cases, the calculated rating is adjusted so that it does not fall greatly below the value for a nominal beginner.

If the new rating μ_n < 1500 then

$$\mu_{n,\text{adjusted}} = 1400 + 100 \exp\left(\frac{\mu_n - 1500}{200}\right)$$

This means that ratings cannot go below a lower limit of 1400. The uncertainty σ_n in rating is increased by half the difference between 1500 and $\mu_{n,\text{adjusted}}$ (subject to its maximum value being 250).

6.3 Inactive players

Inactive players are not published in the "live" ratings. However, if they return to winks, their rating and its associated uncertainty are adjusted from their last values based on how long since they last played a tournament.

The ratings program has to deal with cases where there are many tournaments in a short space of time and cases when there may be no tournament for a significant period (almost 2 years when the COVID-19 pandemic happened). It also has to recognise that there have been different schedules of tournaments in the USA compared to UK. For that reason, the program assigns time increments. With the exception of the COVID-19 pandemic period, there are 2 time increments in a calendar year, one in spring and one in autumn. These almost always correspond to the ETwA National Singles and ETwA National Pairs tournaments. The program calculates the time increment difference t_{inc} between the current tournament and the last tournament played for each participant.

The rating and uncertainty of any player who has not played in a tournament in the last 367 days (provided that at least 2 time increments have elapsed) is adjusted. The adjusted rating is

$$\mu_{0,\text{adjusted}} = \frac{\left(\frac{340^2}{t_{\text{inc}} - 1}\right)\mu_0 + \sigma_0^2(1400)}{\left(\frac{340^2}{t_{\text{inc}} - 1}\right) + \sigma_0^2}$$

This corresponds to the rating adjustment that would take place if a participant played $t_{\text{inc}} - 1$ games of singles with a tournament rating of 1400.

The ratings uncertainty for the player is increased by

$$\sigma_{0,\text{adjusted}} = \sigma_0 + 18\sqrt{t_{\text{inc}} - 1}$$

subject to this not exceeding 250. The rationale for this functionality is that, for a diffusional process such as a random walk, the standard deviation of a distribution is expected to increase by the square root of the number of steps taken. The value of 18 is an arbitrary decision, but was chosen because it gives a plausible increase in the uncertainty in the test cases investigated.