# Understanding the Elo rating system

**Cédric Beaume**

*Comments and questions are welcome to* `ced.beaume@gmail.com`.

In chess, as in many other competitive games, governing bodies make extensive use of metrics to quantify the strength of competitors. These metrics are beneficial to the game for many reasons: (i) they allow players to track their progress, (ii) they allow to set up games against opponents of a similar strength, (iii) they enable tournament organizers to generate fair pairings and (iv) they provide a way to generate accurate predictions on game results. The Elo rating system is the most popular such metric. We delve below in some basic material in order to explain it.

Disclaimer: I will not explain the history of the Elo system. Historical information is easily available on the Internet, for example: `https://en.wikipedia.org/wiki/Elo_rating_system`. Instead, you will find below an explanation aiming to provide the reader with the necessary elements to understand Elo ratings. I have made an effort to write this document in a comprehensive way so that everyone can benefit from it, irrespective of the extent of their mathematical knowledge.

## 1 Where does the Elo system come from?

### 1.1 Relative strength

Before introducing the Elo system, it is important to understand how points are scored in chess. In most chess competitions, the following point scoring system is applied:

- a win is worth $1$ point,
- a draw is worth $0.5$ point,
- a loss is worth $0$ point.

The expected score of Player A against Player B, $S_{A/B}$ ("$S$" for score; simply read "$S$ of $A$ over $B$"), is the average score the former would obtain over a large number of games against the latter. Since there is exactly one point distributed per game:

$$S_{A/B} + S_{B/A} = 1. \tag{1}$$

Example: Expected score calculation

Player A plays 10 games against Player B: they win 6 of them, draw 3 of them and lose 1 game. Player A scored 7.5 points out of a total of 10 possible points. Their expected[a] score for one game against Player B is thus $S_{A/B} = 0.75$ point. Conversely, Player B scored 1 win, 3 draws and 6 losses, for a total number of points of 2.5 out of 10. Player B's expected score is thus $S_{B/A} = 0.25$ point and $S_{A/B} + S_{B/A} = 1$ follows.

[a] we assume, for the sake of simplicity, that 10 games are sufficient for an accurate calculation of the expected score

To predict the outcome of a game, the quantity that matters is the relative strength of the competitors. This quantity, $R_{A/B}$ ("$R$" for relative strength; simply read "$R$ of $A$ over $B$"), is intuitively defined as the ratio of the players expected scores:

$$R_{A/B} = \frac{S_{A/B}}{S_{B/A}}, \tag{2}$$

$$R_{B/A} = \frac{S_{B/A}}{S_{A/B}}, \tag{3}$$

which, upon making use of equation (1) yields:

$$R_{A/B} = \frac{S_{A/B}}{1 - S_{A/B}}, \tag{4}$$

$$R_{B/A} = \frac{S_{B/A}}{1 - S_{B/A}}. \tag{5}$$

Example: Relative strength calculation

Player A played 10 games against Player B, won 6 of them, drew 3 of them and lost 1 of them. Their expected[a] score is thus $S_{A/B} = 0.75$ and:

$$R_{A/B} = \frac{S_{A/B}}{1 - S_{A/B}} \tag{6}$$

$$= \frac{0.75}{1 - 0.75} \tag{7}$$

$$= 3, \tag{8}$$

implying that Player A is 3 times stronger than Player B. Conversely:

$$R_{B/A} = \frac{S_{B/A}}{1 - S_{B/A}} \tag{9}$$

$$= \frac{0.25}{1 - 0.25} \tag{10}$$

$$\approx 0.33, \tag{11}$$

so that Player B is $0.33$ times stronger than Player A or, equivalently put, $1/0.33 = 3$ times weaker than Player A.

[a] we assume, for the sake of simplicity, that 10 games are sufficient for an accurate calculation of the expected score

## 1.2 The Elo hypothesis

A rating system should be applicable over a community of players, in such a way that the rating of two players who never faced each other should allow an accurate representation of their respective strengths. For example: imagine that we know how Player A would fare against Player B (we know $R_{A/B}$ or, equivalently, $S_{A/B}$) and how Player B would fare against Player C (we know $R_{B/C}$ or, equivalently, $S_{B/C}$), a rating system should be able to inform us how Player A would fare against Player C (it should allow us to calculate $R_{A/C}$ or, equivalently, $S_{A/C}$).

This problem is not an easy one and an assumption is necessary to continue building a rating system. The hypothesis underlying the foundations for the Elo system is quite natural:

$$R_{A/C} = R_{A/B}\, R_{B/C}. \tag{12}$$

Intuition: Elo hypothesis

If Player B is twice stronger than Player C and if Player A is 3 times stronger than Player B, then Player A is 6 times stronger than Player C. This intuition justifies hypothesis (12).

Recalling the definition of the relative strength in equation (4), we also have:

$$R_{A/C} = \frac{S_{A/C}}{1 - S_{A/C}}. \tag{13}$$

This can be inverted to find the expected score as a function of the relative strength of the competitors:

$$S_{A/C} = \frac{R_{A/C}}{1 + R_{A/C}}. \tag{14}$$

We can now make use of hypothesis (12) so as to replace all unknown quantities in the right-hand-side of equation (14) by known quantities:

$$S_{A/C} = \frac{R_{A/B}\, R_{B/C}}{1 + R_{A/B}\, R_{B/C}}. \tag{15}$$

Example: Expected score contraction

We assume that $R_{A/B} = 3$ and $R_{B/C} = 2$. The Elo hypothesis (12) gives:

$$\begin{aligned} R_{A/C} &= R_{A/B}\, R_{B/C} && (16)\\ &= 6. && (17) \end{aligned}$$

Furthermore, equation (15) yields:

$$\begin{aligned} S_{A/C} &= \frac{R_{A/B}\, R_{B/C}}{1 + R_{A/B}\, R_{B/C}} && (18)\\ &= \frac{6}{1+6} && (19)\\ &\approx 0.86. && (20) \end{aligned}$$

So, if Player A is three times stronger than Player B and if Player B is twice stronger than Player C, under the Elo hypothesis, Player A is expected to score on average $0.86$ point per game against Player C.

Note that, although the Elo hypothesis (12) can be thought of as the primary weakness of the Elo system, it provides very good predictions. Let us not worry about it here.

## 1.3 Toward the Elo system

So far, all the calculations we made rely on the relative strength of the players, *e.g.*, $R_{A/B}$. Unfortunately, a rating system built over this quantity would be confusing. To illustrate this, let us imagine that we build a rating system on the relative strength of players. The rating of Player A is simply their strength and, to predict the outcome of a game between Player A and Player B, we would divide the rating of Player A by that of Player B to obtain the relative strength of Player A against Player B: $R_{A/B}$. This quantity is directly related to the expected score through equation (14). The calculation is straightforward but the numerical values that represent the player ratings are confusing to interpret. A player rated $214$ would be expected to score twice as many points against a player rated $107$ due to the fact that their relative strength is equal to 2 in this matchup. A similar result is expected from a player rated $0.18$ against a player rated $0.09$ despite the fact that their ratings are much closer. Additionally, gaining one rating point is insignificant to the player rated $214$ but it represents an outstanding jump in performance to the one rated $0.18$. These oddities are due to the fact that it is the *rating ratio* and not the *rating difference* that matters[1]. While this rating system is not mathematically flawed, this example shows how much confusion it can produce.

To provide ratings that are amenable to straightforward comparison, we seek a system based on *rating differences*. Mathematically, we want to turn equation (12), involving a product of strength ratios, into:

$$D_{A/C} = D_{A/B} + D_{B/C}, \tag{21}$$

where $D_{A/B}$ ("$D$" for difference), is the rating difference between Player A and Player B. To map $R_{A/B}$ into $D_{A/B}$, we introduce a function $f$:

$$D_{A/B} = f(R_{A/B}), \tag{22}$$

which is just a way to mathematically say that we want $D$ to be a function of $R$. Upon replacing $D$ into equation (21), we get:

$$f(R_{A/C}) = f(R_{A/B}) + f(R_{B/C}) \tag{23}$$

$$\implies f(R_{A/B}\, R_{B/C}) = f(R_{A/B}) + f(R_{B/C}), \tag{24}$$

where we have, again, used hypothesis (12). Equation (24) is a characteristic property of the logarithm[2], so we can write:

$$D_{A/B} = \alpha \log_{10}\left(R_{A/B}\right), \tag{25}$$

where $\alpha$ is a constant to be determined and $\log_{10}$ is the base 10 logarithm[3][4].

Equation (25) allows us to recover the strength ratio as a function of the rating difference:

$$R_{A/B} = 10^{\left(\frac{D_{A/B}}{\alpha}\right)}. \tag{26}$$

The coefficient $\alpha$ is determined at this point: it represents the rating difference in a game where a player is 10 times stronger than their opponent. Federations have historically used $\alpha = 400$, which has become a standard in the chess world.

Knowing the relative strength of Player A compared to Player B, we can calculate their expected score using equation (4):

$$\frac{S_{A/B}}{1 - S_{A/B}} = 10^{\left(\frac{D_{A/B}}{\alpha}\right)} \tag{27}$$

$$\implies S_{A/B} = \frac{1}{1 + 10^{-\left(\frac{D_{A/B}}{\alpha}\right)}}. \tag{28}$$

This formula is generally the starting point of many documents on the Elo system. Now, you know where it comes from!

[1] in fact, ratings generated by a scaling system such as the one introduced here scale exponentially

[2] $\log(ab) = \log(a) + \log(b)$

[3] any base for the logarithm can be used; the transformation from one to another only changes the coefficient $\alpha$

[4] I chose the base 10 logarithm as it is the one that yields the most straightforward explanation

### 1.4 Some important relationships

The relative strength between Player A and Player B as a function of their rating difference can be obtained using:

$$R_{A/B} = 10^{\left(\frac{D_{A/B}}{400}\right)}. \tag{29}$$

Example: Relative strength from rating difference

We assume that Player A is rated 200 Elo higher than Player B. Equation (29) gives:

$$\begin{aligned} R_{A/B} &= 10^{200/400} && (30) \\ &\approx 3.16. && (31) \end{aligned}$$

Player A is thus 3.16 times stronger than Player B.

The expected score between Player A and Player B can be deduced from their rating difference as follows:

$$S_{A/B} = \frac{1}{1 + 10^{-\left(\frac{D_{A/B}}{400}\right)}}. \tag{32}$$

Example: Expected score from rating difference

We assume that Player A is rated 200 Elo higher than Player B. Equation (32) gives:

$$\begin{aligned} S_{A/B} &= \frac{1}{1 + 10^{-200/400}} && (33) \\ &\approx 0.76. && (34) \end{aligned}$$

Player A is thus expected to score $0.76$ point per game against Player B.

## 2 How does an Elo rating change?

In order for a rating to track the performance level of a player, it needs to be updated after each playing session. There is no unique way to do this but the principle is simple. A win should increase the player's rating in proportion to the opponent's rating: the stronger the opponent, the more gain. Conversely, a loss should decrease the player's rating in a symmetric way. Lastly, a draw should narrow the rating gap between the players.

### 2.1 One-game calculation

After a game, the Elo rating of both competitors evolves in the following fashion:

$$E'_A = E_A + K(G_{A/B} - S_{A/B}), \tag{35}$$
$$E'_B = E_B + K(G_{B/A} - S_{B/A}), \tag{36}$$

where $E_A$ (resp. $E_B$) is the Elo rating of Player A (resp. Player B) before the game, $E'_A$ (resp. $E'_B$) is the Elo rating of the same player after the game, $G_{A/B}$ is the score obtained in the game by Player A against Player B, $S_{A/B}$ is the expected score of Player A against Player B and $K$ is a constant, often referred to as the $K$-factor. Note that the rating change of Player B is the opposite of that of Player A, since $G_{A/B} = 1 - G_{B/A}$ and $S_{A/B} = 1 - S_{B/A}$.

There are two simple, complementary ways to understand the meaning of the $K$-factor:

- If two players of equal strength and rating face each other, their expected score is even: $S_{A/B} = S_{B/A} = 0.5$. Since both players are of equal rating, a draw between them would not change their ratings. Indeed, $G_{A/B} = G_{B/A} = 0.5$ leads to $E'_A = E_A$ and $E'_B = E_B$. If Player A wins, $G_{A/B} = 1$ and $E'_A = E_A + K/2$, so that the rating of player A goes up by $K/2$. The rating of Player B goes down by the same amount so that, after the game, the player's ratings are separated by $K$.
- If Player A's rating is much[5] larger than Player B's: $S_{A/B} = 1$ and $S_{B/A} = 0$. Reproducing the above calculation with these new variables, we discover that a draw between these players will make Player A's rating go down by $K/2$ and that of Player B up by $K/2$ so that their difference has decreased by $K$. If Player A wins, the ratings remain unchanged. If Player B wins, Player A's rating goes down by $K$ and Player B's rating goes up by $K$, thereby decreasing the difference in rating by $2K$.

Consequently, the constant $K$ controls the speed at which the rating evolves: when a player wins against another player of the same rating, their rating difference after the game is $K$; and the maximum rating gain/loss after a game is $K$. Standard values for $K$ are found between 10, for a slowly evolving rating system, and 50, for a quickly evolving one.

Example: Rating change after a game

We assume that Player A is rated 1704 Elo, Player B is rated 1623 Elo and $K = 32$. Player A wins, so $G_{A/B} = 1$. The expected score is calculated using equation (32) and $D_{A/B} = 1704 - 1623 = 81$. We obtain $S_{A/B} \approx 0.61$. Player A's new rating is thus:

$$\begin{aligned} E'_A &= E_A + K(G_{A/B} - S_{A/B}) && (37) \\ &= 1704 + 32(1 - 0.61) && (38) \\ &\approx 1716.34, && (39) \end{aligned}$$

which results in a gain of 12.34 Elo. Symmetrically, Player B loses the same number of Elo points and their new rating is $E'_B \approx 1610.66$.

## 2.2 Tournament calculation

Ratings are often kept unchanged during a tournament and their evolution is only calculated once the tournament is complete, as if all the games had been played simultaneously. This method is convenient as it does not necessitate the recomputation of players ratings at the end of each game. Although it yields slightly different results to the systematic use of the one-game calculation explained above, these results are qualitatively similar and both methods are sound.

To calculate rating changes after a tournament, formula (35) is simply modified into:

$$E'_A = E_A + K(G_A - S_A), \tag{40}$$

[5] it is not necessary for the rating difference to be infinite due to the fact that the result will be rounded before finalized

where $E_A$ and $E'_A$ are the Elo ratings of Player A before and after the tournament, $G_A$ is the number of points scored in the tournament by Player A and $S_A$ is the expected number of points scored in the tournament by Player A. The quantity $S_A$ is simply calculated as the sum of the expected scores of Player A against each of their opponents.

Example: Rating change after a tournament

We assume that Player A is rated $1704$ Elo, Player B is rated $1623$ Elo, Player C is rated $1851$ Elo, Player D is rated $1471$ Elo and $K = 32$. Player A won against Player B, drew against Player C and won against Player D. The expected scores for Player A in each of these games are: $S_{A/B} \approx 0.61$, $S_{A/C} \approx 0.30$ and $S_{A/D} \approx 0.79$. The sum of these expected scores is $S_A \approx 1.71$ and Player A scored $G_A = 2.5$ points during the tournament. Player A's new rating is thus:

$$\begin{aligned} E'_A &= E_A + K(G_A - S_A) && (41)\\ &= 1704 + 32(2.5 - 1.71) && (42)\\ &\approx 1729.36, && (43) \end{aligned}$$

which results in a gain of $25.36$ Elo.

## 3 What is a performance rating?

If you ever looked at chess tournament statistics, you probably noticed that results are often complemented with a performance rating. The performance rating of a player in a tournament reflects on how well they played: it is the rating that the player should have had at the start of the tournament so that their tournament results would not generate any rating change. It is useful to compare the rating of a player before a tournament with their performance rating at the end of it: if the latter if larger than the former, the player has performed beyond the expectations associated with his initial rating. At the end of a tournament, a player's rating evolves toward their performance rating.

In theory, the performance rating of a player is searched by finding their initial rating such that:

$$S_A = G_A. \tag{44}$$

This equation just stems from assuming that equation (40) does not produce any rating change. By using equation (28) and replacing the rating difference by the difference between the performance rating of Player A, $P_A$ ("$P$" for performance), and the rating of Player A's opponent, for example $D_{A/B} = P_A - E_B$, we get an equation like:

$$\frac{1}{1+10^{-\left(\frac{P_A - E_B}{400}\right)}} + \frac{1}{1+10^{-\left(\frac{P_A - E_C}{400}\right)}} + \frac{1}{1+10^{-\left(\frac{P_A - E_D}{400}\right)}} + \cdots = G_A, \tag{45}$$

where $E_B$, $E_C$, $E_D$, $\ldots$ are the known ratings of the opponents of Player A, and where $G_A$ is the tournament score of Player A. Here, we need to solve for $P_A$, which is a trivial task using a simple root searching algorithm.

For some, probably historical, reasons, performance rating computations are nearly always carried out in a simplified, non-equivalent way. However, most of the methods used provide qualitatively similar results. One's performance rating should be interpreted in comparison to one's actual rating to answer the following questions: (i) Did I perform above or below expectation? and (ii) Was my performance slightly above/below or clearly above/below expectation?

Perhaps the most thematic performance rating computation method consists in considering that Player A only faced opponents whose rating is the average rating of the opponents actually faced. Under this assumption:

$$S_A = NS_{A/avg}, \tag{46}$$

where $N$ represents the number of games played by Player A and $S_{A/avg}$ is the expected score of Player A against an opponent whose rating is the average rating of the opponent Player A actually faced. Combining equations (40) and (28), we get:

$$\frac{N}{1 + 10^{-\left(\frac{P_A - E_{avg}}{400}\right)}} = G_A \tag{47}$$

$$\Longrightarrow P_A = E_{avg} - 400 \log_{10}\left(\frac{N}{G_A} - 1\right), \tag{48}$$

where $E_{avg}$ is the average rating of Player A's opponents.

Another method to compute the performance rating of a player relies on the formula:

$$P_A = E_{avg} + \alpha\frac{(W_A - L_A)}{N}, \tag{49}$$

where $W_A$ (resp. $L_A$) is the number of wins (resp. losses) scored by Player A during the tournament. With $\alpha = 400$, this method is known as the *algorithm of* 400. It provides a very quick way to compute a performance rating on the spot.

Example: Performance rating calculation

We assume that Player A is rated 1704 Elo, Player B is rated 1623 Elo, Player C is rated 1851 Elo and Player D is rated 1471 Elo. Player A won against Player B, drew against Player C and won against Player D.
Using the "ideal" method of equation (44), the performance rating of Player A is: $P_A \approx 1973.76$.
Using the average method of equation (48) with $G_A = 2.5$, $N = 3$, $E_{avg} \approx 1648.33$, we get: $P_A \approx 1927.92$.
Using the algorithm of 400 of equation (49) using $E_{avg} \approx 1648.33$, $\alpha = 400$, $W_A = 2$, $L_A = 0$ and $N = 3$, we get: $P_A \approx 1915.00$.
All three methods provide a performance rating substantially larger than the rating of Player A before the tournament, so Player A should be proud to have performed substantially beyond the expectations associated with their initial rating. The quantitative disagreement between the three methods can be attributed to (i) the rather disparate ratings of the players involved, (ii) the extreme score of Player A (performance ratings become more accurate when the score is closer to even) and (iii) the inherent approximations made by each method.