ELO Rating System and MMR in Modern Games Literature Review

William Barthell and Kyllan Wunder

Introduction

The ELO system is a widely recognized and popular rating system used to rank players in various competitive games. Initially developed for chess, the ELO system has since been adapted and used in various games, including video games. This write-up will comprehensively analyze the ELO rating system, its history, and how it is currently used in modern video games.

One of the primary purposes of the ELO rating system is to provide a fair and accurate way to match players with similar skill levels. Players are assigned a numerical rating based on their past performance, which determines their relative skill level compared to other players. The ELO system is used in popular video games such as League of Legends, Overwatch 2, and Dota 2. It matches players with similarly skilled opponents and provides a more enjoyable gameplay experience.

While the ELO system has been praised for its effectiveness in ranking players and providing suitable matches, there are also concerns that it may need to be more optimal for providing the best possible player experience. In this write-up, we will argue that the ELO system should be changed lightly to improve the player experience. We will explore different modifications that could be made to the system to address some of the raised concerns. By the end of this write-up, readers will have a better understanding of the ELO rating system, its current use in video games, and potential optimizations.

History

Arpad E. Elo was a skillful chess player in the 1930-40s and was bothered by the chess ratings given by the United States Chess Federation (USCF) as they used a numerical rating system. Elo and other members of the USCF were unhappy with the rating system as it sometimes gave players too big of a rating gain. On behalf of the USCF, Elo devised a new rating system with a solid statistical foundation, which uses a statistical approach to ranking players based on their past performance. Elo's background in physics and mathematics helped him design a system based on solid statistical principles and accurately predict the outcomes of chess matches based on a player's skill level (Elo 242).

One of the core principles of Elo's new rating system was based on this idea; "From general experience in sports, we know that the stronger player does not invariably outperform the weaker. A player has good days and bad, good tournaments and bad. By and large, at any point in his career, a player will perform around some average level. Deviations from this level occur, large deviations less frequently than small ones. These facts suggest the basic assumption of the Elo system" (Elo 7).

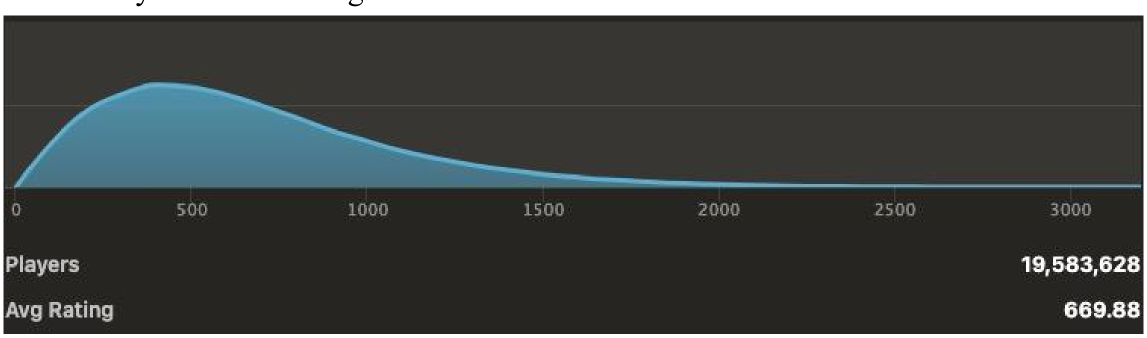
The ELO rating system quickly gained popularity within the chess community and was adopted by the USCF in 1960. The system was initially used to rank players in national tournaments, but it soon spread to other countries and became the standard rating system for international chess competitions.

The success of the ELO rating system in chess led to its adoption in other competitive games and sports. The system has been modified and adapted to suit different games' needs and addresses concerns about accuracy and fairness. For example, in some video games, the ELO system is used to match players with similar skill levels, while in sports, the system is used to predict the outcomes of games and to rank teams.

Despite its widespread use, the ELO rating system has also been criticized and debated. Some have argued that the system does not accurately reflect a player's skill level, especially in games where luck or chance plays a significant role. Others have suggested that the system may need to be more flexible and allow for more flexibility in ranking players.

Despite these concerns, the ELO rating system remains widely recognized and respected for ranking players in various competitive games and sports. Its solid statistical foundation and widespread adoption make it a valuable tool for players, coaches, and fans.

Current day chess Elo rating distribution:



How ELO works

At its core, the ELO system is based on the idea that the relative skill level of two players can be determined by the outcome of their match. If a player with a high ELO rating defeats a player with a lower ELO rating, then both players' ELO ratings will be adjusted to reflect the match's outcome. The higher-skill player only gains a few points and the lower-skill player only loses a few points as the difference in their Elo is taken into account. If two equally skilled players duel, they would lose and gain the same number of points respectively. The ELO rating system assigns a numerical rating to each player, representing their skill level. The initial rating is typically set to 1000 for new players, ranging from a minimum of 0 to a maximum of 3000 (Joe 85). The rating is calculated based on the player's performance in previous games. The amount that their rating changes after a game is determined by the difference between their rating and their opponent's rating (Marcus 195).

The basic formula for calculating the change in a player's rating after a match is as follows (Fu-Hsing Tsai 214):

$$R_{new} = R_{old} + K * (S - E)$$

Where:
R_{new} = The new rating of the player
R_{old} = The old rating of the player
K = A constant value that determines the weight of the rating change (usually set between 10 and 40)
S = The score of the player in the match (1 for a win, 0.5 for a draw, and 0 for a loss)
E = The expected score of the player in the match, based on their rating relative to their opponent's rating.

The expected score E is calculated using the following formula (Fu-Hsing Tsai 214):

$$E = 1 / (1 + 10^{((R_{Opponent} - R_{Player})/400)})$$

Where:
$R_{Opponent}$ = The rating of the opponent
R_{Player} = The rating of the player

The ELO system is designed to be self-correcting over time, meaning that a player's rating will tend to converge towards their true skill level as they play more matches. This is because the more matches a player plays, the more accurate their rating becomes, as it is based on a larger sample of their performance (Langville et al. 56). Additionally, the system accounts for fluctuations in performance over time by adjusting the weight of the rating change based on the player's recent performance.

Overall, the ELO rating system is a widely used and well-regarded method for determining relative skill levels in competitive games. It provides a fair and consistent way to rank players based on their performance and allows for meaningful comparisons between players of different skill levels.

Different types of distributions formed in ELO systems

Depending on the assumptions made about player performance, different distributions can be formed from players within an ELO ranking system.

1. Normal distribution: The most common assumption is that player performance follows a normal distribution, with the majority of players being of average skill level and fewer players being either significantly better or significantly worse. This assumption implies that players' performances are symmetrically distributed around the mean ELO rating.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f(x) = probability\ density\ function$

$\sigma = standard\ deviation$

$\mu = mean$

2. Skewed distribution: In some cases, player performance may not be normally distributed. For example, the distribution may be skewed to the left or right, indicating that the majority of players are either stronger or weaker than average. This could happen in games where few players dominate the competition.

$$\mu_3 = \frac{\sum_{i}^{N}(X_i - X)^3}{(N-1)*\sigma^3}$$

$\mu_3 = skewness$

$\sigma = standard\ deviation$

$N = number\ of\ variables\ in\ the\ distribution$

$X_i = random\ variable$

$X = mean\ of\ the\ distribution$

3. Multimodal distribution: In other cases, the distribution may be multimodal, indicating multiple distinct groups of players with different skill levels. This could happen in games where players are separated into different leagues or tiers based on their skill level.

   $f(x) = pg_1(x) + (1 - p)g_2(x)$
   $g_i = \text{Probability distribution}$
   $p = \text{mixing parameter}$
   (Bimodal distribution shown, i can increase without loss of generality)

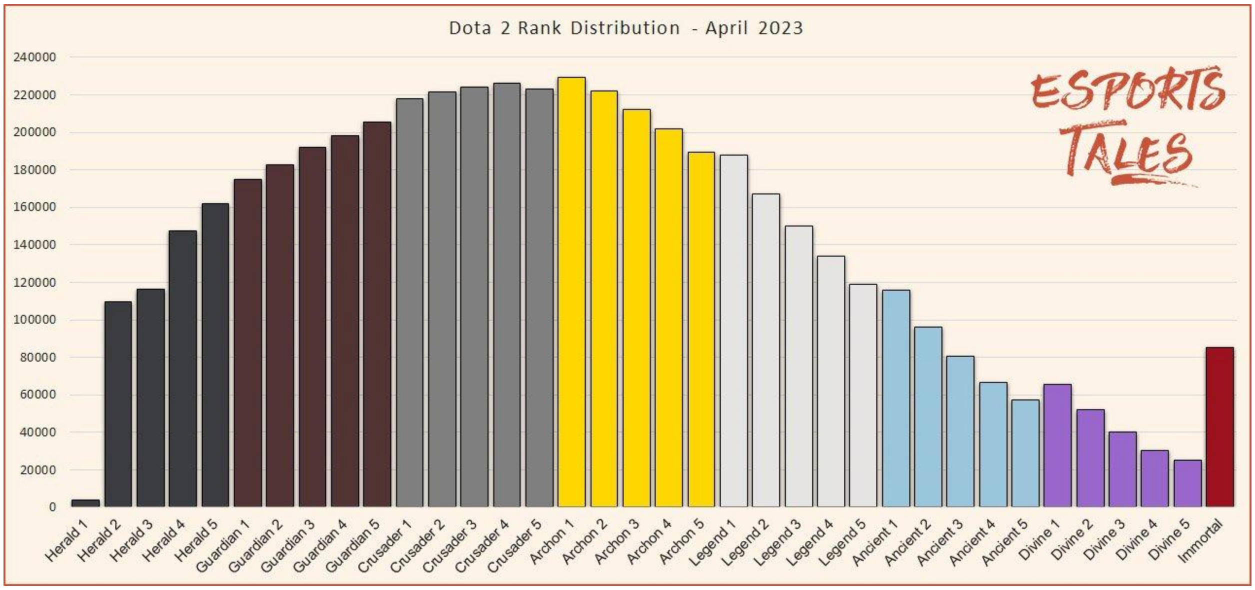
Assumptions about player performance are critical in producing accurate rankings and can affect these distributions. For example, if we assume that player performance is normally distributed, but this assumption is incorrect, or if we assume that player performance is skewed, but the skewness is not accurately represented in the distribution, it could lead to inaccurate ratings and rankings.

Strengths and weaknesses of different distributions

Choosing the best distribution to use in an ELO ranking system requires consideration of several factors, including accuracy, fairness, and ease of implementation. As a result, we will analyze the performance of the previously outlined distributions.

Normal Distribution

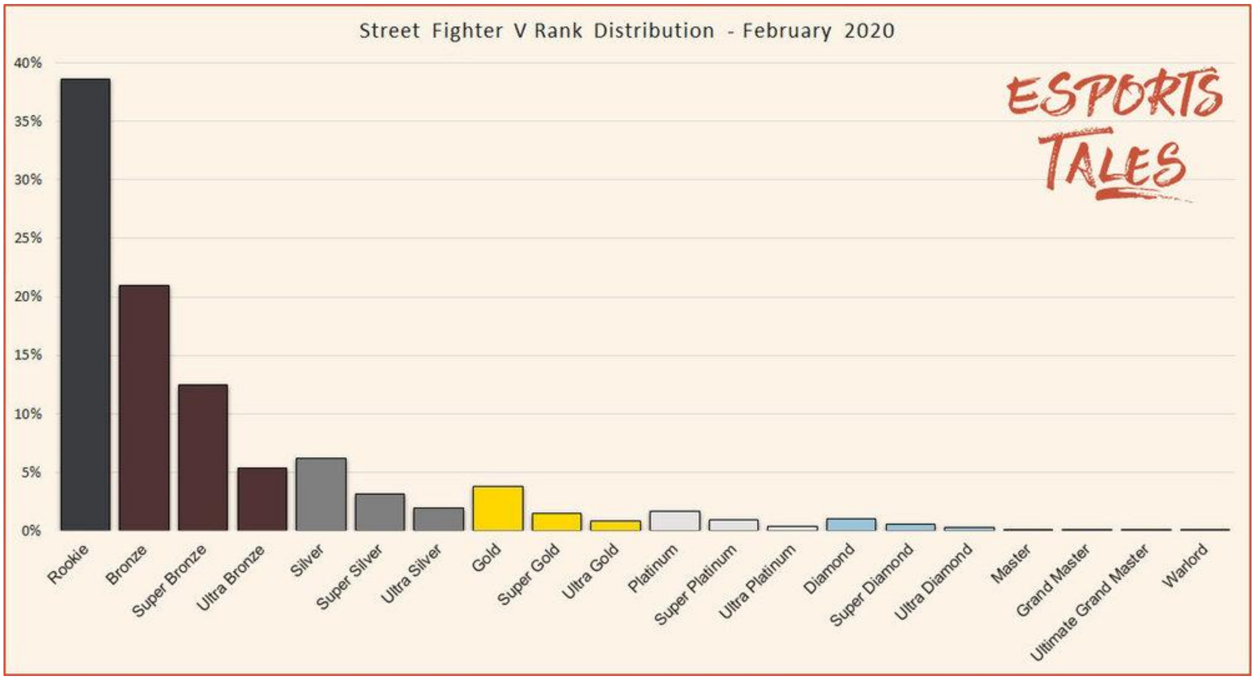
The normal distribution is a well-known and widely used statistical distribution, making it a straightforward choice for many ranking systems. It assumes that player performance follows a bell curve, where most players perform at an average level, and fewer players perform at the extremes. This assumption can be a good fit for certain games where players' skill level is relatively evenly distributed. Additionally, the normal distribution can be easy to implement, as it requires minimal additional data or complex calculations beyond the ELO rating system itself. However, several issues arise when using a normal distribution. It assumes that the performance of all players is independent and identically distributed, which may be untrue in some instances and lead to significant issues in ranking accuracy. Consequently, the normal distribution is unsuitable for games with an extensive skill gap between the best players and the rest.



The distribution of Dota 2, an online 5v5 multiplayer video game, currently uses a normal distribution to model players' skill levels. However, as we can see, there is a large volume of players in the highest rank. A skewed distribution could fix this issue by spreading out the highest-ranked players over a larger area.

Skewed Distribution

Skewed distribution can be a suitable choice for an ELO ranking system in games where few players dominate the competition. In such cases, the distribution of player performance may be skewed to the right, indicating that a small number of players perform at a significantly higher level than the rest of the players. In this case, using a skewed distribution can help to capture the differences in player performance better, as it can more accurately reflect the distribution of performance levels in the game. However, using a skewed distribution in an ELO ranking system has several potential weaknesses. One potential issue is that it may be challenging to determine when the distribution is actually skewed and how much it is skewed. Determining whether a distribution is skewed requires careful data analysis, which can be time-consuming and challenging to interpret. Another potential issue is that using a skewed distribution can lead to unfair rankings, where players with above-average performance receive less recognition than they deserve. In a skewed distribution, the majority of players are likely to have below-average performance, which can lead to a situation where players with above-average performance are not rewarded as much as they should be. Additionally, using a skewed distribution may make comparing player performance across different games or competitions more difficult because the distribution of player performance can vary widely, a simple example would be if you have a competition where only 50% of players are above or below the mean compete at allowing for larger player skill discrepancy.

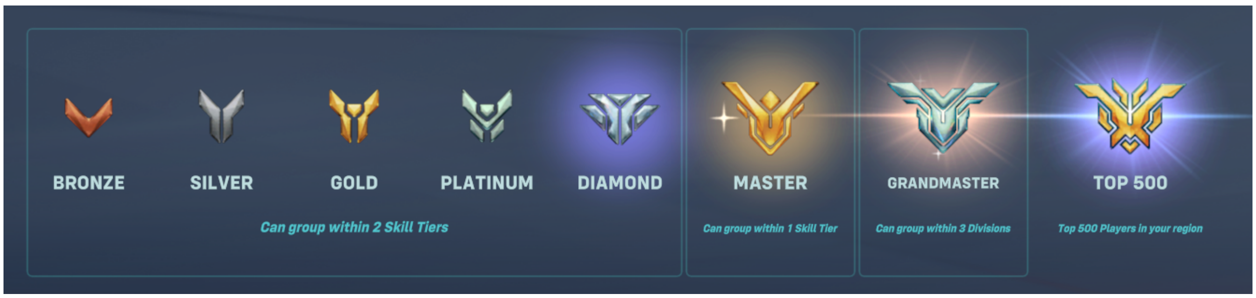


Here, we can see that Street Fighter, a 1v1 2D fighting game, uses a skewed distribution as the player base is small but with a high average skill level. A new player would face stronger opponents as many players are grouped at the bottom, but higher-skilled matchups are fairer.

Multimodal Distribution

The multimodal distribution represents player skill levels by separating players into different tiers. This can help prevent highly skilled players from dominating lower leagues and lower-skilled players from being matched against much higher-skilled players, leading to a more competitive and fair environment. By having clear tiers for players to move up to, a multimodal distribution can motivate players to improve their skills and move up the ranks. This can lead to a more competitive environment and overall improvement in player performance (Fu-Hsing Tai 214). Also, a multimodal distribution can be flexible regarding the number of tiers. However, multimodal distribution has several issues similar to the other two discussed distributions. One of the biggest challenges in using a multimodal distribution is determining the skill level cutoffs for each tier. This requires careful analysis of player performance data and may involve trial and error. Additionally, the skill level cutoffs may need to be adjusted over time as player skill levels change. By separating players into different tiers, another challenge arises. This may result in slower matchmaking times as players can only be matched with others in their own tier. This can be especially true for higher-tier players who may have fewer opponents to match against. Lastly, a multimodal distribution can be more complex for players to understand compared to a normal distribution. This may require more education and communication with players to help them understand how the ranking system works and how they can move up the ranks.

Most modern games provide a combination of a ranked distribution, such as the normal, with a multimodal on top of it. Games will typically have a true hidden rank called Match Making Rank (MMR) or Skill Ranking (SR) that can range from 0 to roughly 4000. This uses the Elo system based on a normal distribution. However, this is hidden with a faux multimodal distribution to incentivize players. These distributions usually include 5-10 skill tiers with divisions within each tier, all with a unique badge to incentivize players to get to a high tier. However, this faux multimodal distribution is used to match players, we assume, because it is easier and gives faster queue times. For example, Overwatch 2, an online 5v5 multiplayer video game, has eight skill tiers, each with five divisions. Shown below is how this multimodal is used to group players into matches.



Each tier has the following SR requirements(which is derived from Elo rating):
Bronze - 1 to 1499 SR
Silver - 1500 to 1999 SR
Gold - 2000 to 2499 SR
Platinum - 2500 to 2999 SR
Diamond - 3000 to 3499 SR
Master - 3500 to 3999 SR
Grandmaster - 4000+ SR
Top 500 - Based on the top 500 players in your region regardless of tier or SR

Our Opinion

After analyzing the strengths and weaknesses of each distribution, we have concluded that a multimodal distribution is the best choice for an ELO ranking system due to several key points. The multimodal distribution provides a more accurate representation of player performance by separating players into skill-based tiers. This consequently results in a fairer ranking system, as the system ensures that players are matched against equally-skilled opponents. In a normal distribution, players slightly better than average may be ranked much lower than those slightly worse due to its bell curve distribution. A skewed distribution may favor specific types of players over others depending on the direction of the skew. Separating players into different tiers based on their skill level can help prevent top players from dominating lower leagues, and it can also help motivate lower-tier players to improve their skills. Also, by outlining clear tiers and explaining to players what they must do to rise in the rankings, a clear motive and path of milestones are set out for players, which may incentivize them to play and improve their rating. This feature is unavailable in normal and skewed distribution ranking systems. Also, a multimodal distribution is much more flexible than a normal or skewed distribution. A multimodal distribution can alter the number of tiers established, adjust tiers depending on the player pool size and the skill gap between players, and scale for larger player pools if the game becomes more popular. Finally, most games already have skill tiers defined by which players are familiar with and which are used to place players in matches, so why continue to use a normal distribution underneath when it has to be converted to a multimodal distribution? We propose a multimodal distribution where normal distributions represent most players and skewed distribution models the highest-skilled players.

Multiplayer Elo

The Elo rating system was originally designed for one player versus another player. However, in modern video games, it is used in a team format. So now, even if one player does extremely well and outperforms their average play skill, they could still be punished if the rest of their team does not do as well and they lose. For this reason, it is yet to be known how the ELO system can be optimized for multiplayer online games. Nate Silver devised a margin of victory multiplier (MoVM) that allows for some correction. If the game has a point system, it can be used by checking the point differential and seeing how close the victory was so as not to take as many points away if it was a close game.

$$MoVM = ln(WinnerPointDiff + 1) * 2.2/(WinnerEloDiff * 0.001 + 2.2)$$

This is combined with the Elo rating system as follows:

$$R_{new} = R_{old} + K * (S - E) * MoVM$$

Conclusion

When choosing which distribution to use for a game's ELO ranking system, it is imperative to analyze game data and the nature of the competition. Careful consideration of the strengths and weaknesses of different distributions is necessary to ensure the validity and fairness of the ranking system. Although the optimal choice of distribution in an ELO ranking system is highly dependent on the specific game being played and the nature of the competition, we have concluded that a multimodal distribution is the most suitable choice due to its ability to represent player skill levels better, provide a fairer ranking system, offer clear milestones for players to achieve, and provide flexibility in the number of tiers established. While a normal or skewed distribution may work for other ranking systems, a multimodal distribution that combines a normal and skewed distribution is likely the best choice for an ELO ranking system. Ultimately, choosing the appropriate distribution for an ELO ranking system is critical to ensuring the success of the ranking system, and a multimodal distribution provides a strong foundation for this.

Works Cited

Elo, Arpad E. (August 1967). "The Proposed USCF Rating System, Its Development, Theory, and Applications" Chess Life. XXII (8): 242–247. https://uscf1-nyc1.aodhosting.com/CL-AND-CR-ALL/CL-ALL/1967/1967_08.pdf#page=26

Elo, Arpad E. *The Rating of Chessplayers Past and Present*. Second ed., ARCO Publishing Inc., 1986.

Fu-Hsing Tsai. "The Effectiveness Evaluation among Different Player-Matching Mechanisms in a Multi-Player Quiz Game." *Journal of Educational Technology & Society*, vol. 19, no. 4, 2016, pp. 213–24. *JSTOR*, http://www.jstor.org/stable/jeductechsoci.19.4.213.

Jabin, Pierre-Emmanuel, and Stephane Junca. "A Continuous Model for Ratings." SIAM Journal on Applied Mathematics, vol. 75, no. 2, 2015, pp. 420–42. JSTOR, http://www.jstor.org/stable/24511455.

Joe, Harry. "Extended Use of Paired Comparison Models, with Application to Chess Rankings." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 39, no. 1, 1990, pp. 85–93. *JSTOR*, https://doi.org/10.2307/2347814.

Langville, Amy N., and Carl D. Meyer. "ELO's System." *Who's #1?: The Science of Rating and Ranking*, Princeton University Press, 2012, pp. 53–66. *JSTOR*, http://www.jstor.org/stable/j.ctt7rwdt.8.

Marcus, David J. "New Table-Tennis Rating System." *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 50, no. 2, 2001, pp. 191–208. *JSTOR*, http://www.jstor.org/stable/2681094.

Milella, Vincenzo "SKULZ." "Dota 2 and Street Fighter Distribution Graphs." *Esports Tales*, https://www.esportstales.com/.

Silver, Nate. "How Our NFL Predictions Work." *FiveThirtyEight*, 5 Sept. 2018, https://www.fivethirtyeight.com/methodology/how-our-nfl-predictions-work

"Current day chess Elo rating distribution." *Chess.com*, https://www.chess.com/leaderboard/live.