Statistical Theory
Chess Dataset Analysis

Dor Boker, Itamar Nakar
I.D: 209271279 , 325829000
Email: dorboker@gmail.com, itamar.nakar@gmail.com

I. INTRODUCTION

Chess is one of humanity's oldest board games [1], played by two players on an 8×8 grid. Each player, controlling 16 pieces of their color, aims to checkmate the opponent's King, making it impossible for the King to escape capture. Unlike many games, chess does not involve luck or hidden information; the outcome is determined solely by the players' knowledge, strategy, and analytical skills.

The length of a chess match is influenced by several factors, including the players' skill levels, strategic choices, and in-game dynamics. In this project, we investigate the relationship between player ratings and the duration of chess matches, aiming to determine whether higher-rated players tend to play shorter or longer games. In addition to player ratings, we examine other variables that may impact match length, such as opening strategies, game outcomes (win, loss, draw), and time controls.

Our dataset [2] comprises 19,113 games from Lichess, an online chess platform, each described by 16 features. The most relevant features to our analysis are:

- **Turns**: A positive integer representing the length of the game, where one turn includes a move by both White and Black.
- **ELO Rating**: A system that quantifies player skill, widely used by FIDE and online chess platforms [3], which we use to assess its correlation with match length.
- **Time Controls**: Competitive games are categorized into four main time control classes: Bullet, Blitz, Rapid, and Classical, based on initial clock times and increments [4].
- **Winner and Victory Status**: These columns capture whether White or Black won and the method of victory (checkmate, resignation, draw).

Through statistical analysis, we aim to understand how these factors influence the duration of chess matches and to provide insights into the relationship between skill level and game dynamics.

II. RESULTS

A. Data Transformation

To optimize the quality of our analysis, we undertook a cleaning process to remove entries that could distort or obscure meaningful insights from the dataset. Below are the key transformations:

- **Uncompetitive Games**: Games marked as casual or non-competitive were excluded. Such games may not reflect players' full effort or strategic depth, and are thus not representative of a standard chess match.
- **Elo Adjustments**: New players are initially assigned an Elo rating of 1500, which can skew the distribution of player strength. To mitigate this, each player's rating was updated to reflect their most recent Elo score within the dataset.
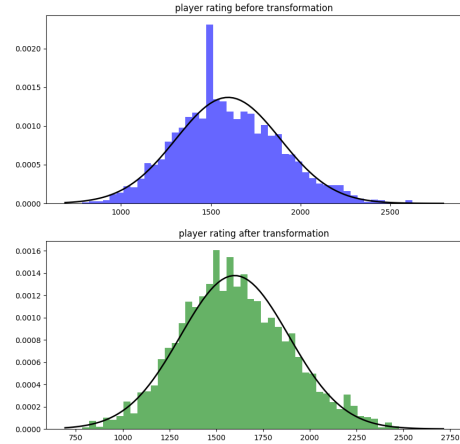

Fig. 1: Transformation's effected on Rating and Turns

After these trasformations, we are left with 5,827 games which we will now begin to analyze.
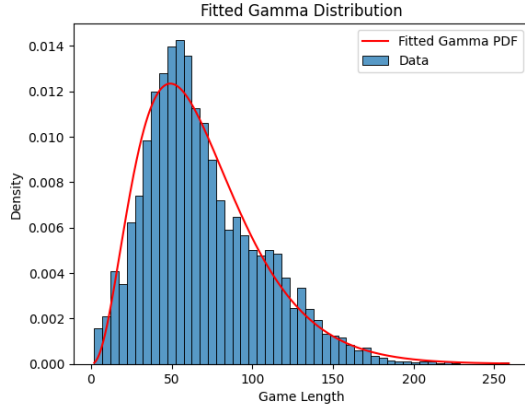
Fig. 2: Gamma distribution fitted to the 'Turns' data

B. Distribution of 'Turns'

To better understand the distribution of the 'Turns' variable, we attempted to fit it to several probability distribution families. The Gamma distribution was selected first due to its ability to model skewed data. However, as shown in Figure 2, the Gamma distribution is too broad for this dataset, suggesting it does not fully capture the underlying structure of 'Turns.'

To refine the fit, we combined the Gamma distribution with a Poisson distribution, which has a narrower shape. This combination results in a more accurate fit to the empirical data, as shown in Figure 3. The Gamma-Poisson mixture balances the excess spread of the Gamma distribution with the narrowness of the Poisson, providing a better representation of the 'Turns' distribution.
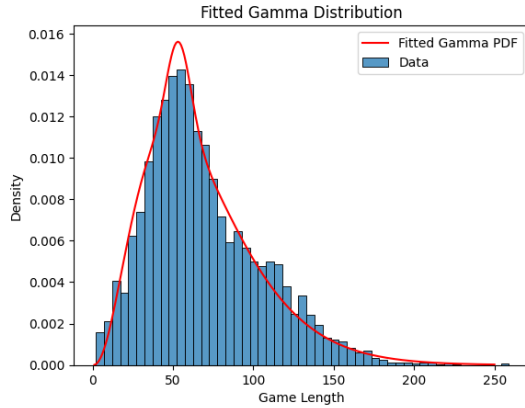

Fig. 3: Gamma-Poisson distribution fitted to the 'Turns' data

C. Rating Distribution

The distribution of player ratings in the dataset initially resembles a normal distribution with a peak around the mean. However, this peak diminishes when games involving unrated players are excluded, bringing the distribution closer to normality. This is illustrated in Figure 1. Additionally, the QQ plot (Figure 4) supports the similarity between the player rating distribution and a normal distribution, albeit with deviations at the extremes.
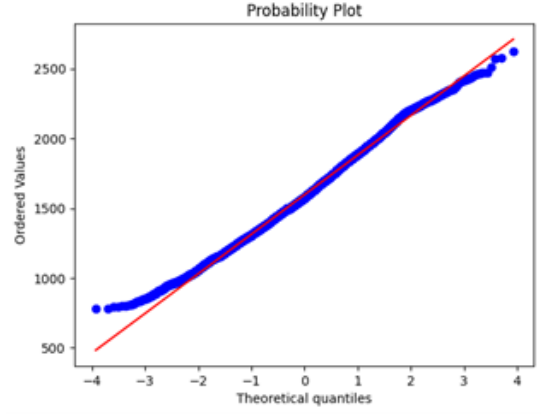

Fig. 4: Gamma-Poisson distribution fitted to the 'Turns' data

1) Kolmogorov-Smirnov (KS) Test for Normality: We applied the Kolmogorov-Smirnov (KS) test to evaluate whether the player rating distribution follows a normal distribution. The test returned a p-value of 8.9649×10^{-13}, which is significantly lower than the threshold of 0.03 for a confidence level of 97%, thus rejecting the null hypothesis that the rating distribution is normally distributed. However, when the test was performed on a random 10% sample of player ratings, the p-value increased to 0.49, meaning we could not reject the null hypothesis at that sample size. This result indicates that while the entire population does not fit a normal distribution, smaller samples may approximate normality [5].

D. Relations Between Features

Analyzing the dataset's convolution matrix reveals linear relationships between certain features. For example, there is a strong correlation between the ratings of white and black players in a game, which aligns with the expectation that players of similar ratings tend to compete against each other. However, no significant linear relationship was observed between the winner and the ratings of either player.
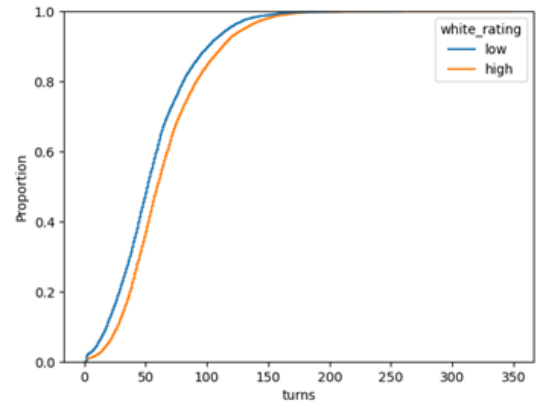

Fig. 5: Gamma-Poisson distribution fitted to the 'Turns' data

E. Connection Between Player Rating and Game Length

To investigate whether player ratings are associated with game length, we divided players into two groups based on their

ratings: those with ratings higher than 1500 (high-rated) and those with ratings lower than 1500 (low-rated). The threshold of 1500 was chosen because it represents the starting rating on the Lichess platform. The empirical cumulative distribution function (ECDF) of game lengths for both groups (Figure 5) shows that high-rated players generally play longer games.

1) Mann-Whitney U Test for Game Length: Since we could not confirm the normality of the rating distribution, we applied the non-parametric Mann-Whitney U test to compare the game lengths between high-rated and low-rated players. Our null hypothesis was that high-rated players do not play longer games than low-rated players on average. The test returned a p-value of 2.7×10^{-73}, which allows us to reject the null hypothesis. Therefore, we conclude that player rating significantly influences game length, with higher-rated players tending to play longer games.

F. Summary of Results

The analysis demonstrates a clear statistical relationship between player rating and game length. The Mann-Whitney U test confirmed that higher-rated players play longer games. This finding supports the hypothesis that player rating is a factor in game duration, with the distribution of ratings further informing our understanding of game dynamics.

REFERENCES

[1] Wikipedia, "Chess," https://en.wikipedia.org/wiki/Chess, 2024.
[2] J. Mitchell, "Chess game dataset (lichess)," https://www.kaggle.com/datasets/datasnaek/chess, 2017.
[3] Unknown, "Elo rating system," https://www.chess.com/terms/elo-rating-chess, 2020.
[4] R. Bell, "Time controls - everything you wanted to know," https://www.chess.com/blog/RussBell/time-controls-everything-you-wanted-to-know, 2024.
[5] L. G. Larson, "Investigating statistical vs. practical significance of the kolmogorov-smirnov two-sample test using power simulations and re-sampling procedures," *ndsu*, 2018.