

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 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, 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.
- **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.
- **Resignations:** Games concluded by resignation were removed. A resignation introduces a psychological factor into game length and outcome, which can skew the statistical distribution. By removing these games, we aim to focus on games decided by normal play.
- **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.

After these transformations, we are left with 5,827 games which we will now begin to analyze.

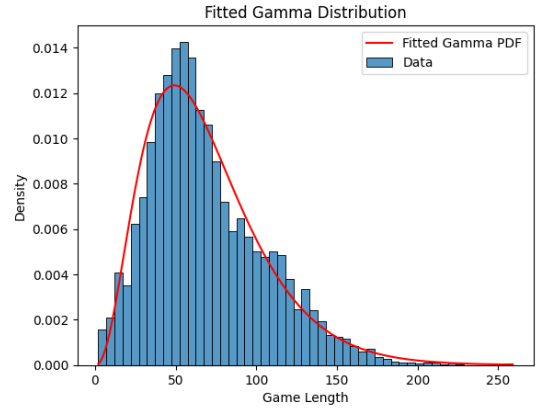


Fig. 1: Gamma distribution fitted to the 'Turns' data

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 1, 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 2. 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.

REFERENCES

[1] Wikipedia, "Chess," <https://en.wikipedia.org/wiki/Chess>, 2024.

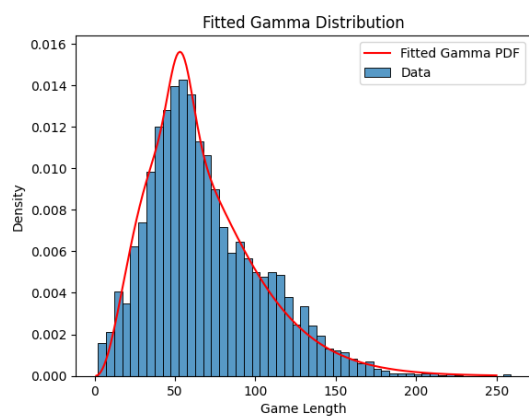


Fig. 2: Gamma-Poisson distribution fitted to the 'Turns' data