Analysis of Trends in Chess Games Between Selected Grandmasters

Xiaoqing Li

June 3, 2015

Abstract

Chess is a deep and well-studied game with a wealth of content and data. Because of the relative simplistic nature of the game states, games themselves are relatively easy to record, and thus there are large databases of games, dating back to Francesco di Castellvi vs Narciso Vinyoles, 1475.[1] The vast majority of published games are between masters and grandmasters, usually recorded by tournaments; recorded games of individual players are also often collected throughout their lifetime and later compiled. In this paper, we aim to perform basic analysis of several trends over time: opening moves and their replies, the win rates of knight-bishop imbalances, and the win rates of uncastled sides. and create simple models of behavior based on the era of play. While this is by no means a complete analysis, our goal is to demonstrate several techniques of statistical and data analysis applied to chess and quantitatively show some trends which were previously assumed to exist by the community.

1 Introduction

Chess is a popular strategy board game played between two players. The game is played on a discrete 8-by-8 board made up of alternating light and dark squares on which identical sets of black and white pieces are placed. Each player is assigned either the white or black pieces (it is common to refer to the players in a particular game as either "white" or "black"). Starting with white, each player takes alternating turns moving one of their pieces according to the rules of the game.

The 1851 International Tournament, which took place at the Crystal Palace of the London Exhibition, is often regarded as the birthplace of competitive chess.[2] Since then, tournament chess has grown to be a worldwide competition.

The games played at these tournaments are almost always recorded and stored in databases. For the purposes of this study, we used two online databases, www.pgnmentor.com and www.endgame.nl, which contain various collections of chess games played by grandmasters in .pgn format. Because this file format does not always list the date of the match, we divided the data into "eras" through which we can track time. The selection methods of the data will be discussed in the following section.

Our goal is to perform statistical data analysis on this data, and then use the results of the analysis to create three predictors:

1. http://www.chessgames.com/perl/chessgame?gid=1259987, *Francesco di Castellvi vs Narciso Vinyoles, 1475.*
2. Robert Byrne, "Chess," *The New York Times* (1997).

1. Given an opening move played by white and the era of the game, predict black's move.
2. Given an era and an opening, predict if a bishop-for-knight exchange will occur in that game.
3. Given an era, predict the number of "forcing moves" in the game relative to the number of moves total.

2 Data Selection

Our goal for this paper is to analyze changes in chess throughout history. However, because the date of a particular game is not always trivial to find, and because players tend to carry tendencies with them from previous eras, we instead used as our data individual games played by grandmasters, divided into five eras of twenty-five years each until 2015 based on each individual player's peak rating. An exception was made for the earliest era, as the number of recorded games during this era were relatively small – players from between 1850 and 1890 were admitted to the dataset after all players from 1890 to 1915 were exhausted (though competitive chess began in the mid-19th century, very little data is available until the beginning of the 20th century). Players were selected from the list of all time top players by Elo rating[3] first and then by Chessmetrics ranking[4] (Elo rating is heavily skewed towards players in the last two decades) and were then placed into eras based on the year the player attained his or her peak Elo or peak Chessmetrics three year average rating; if the latter range of a particular player was on the border of an era split, they were placed in the earlier of the two possible eras. The collection of a particular player's games, if it exceeded 1500, was truncated to 1500 by randomly removing members of his or her collection; then players were added to each era until the total number of games exceeded 10000 for a given era, and the remainder games of the lowest rated player were truncated at random from his or her collection.

We then converted the data contained in the pgn files into a much simpler form. We parsed the game data using the python-chess library, and for each game, stored the following data:

- The era of the game
- The opening white played
- Black's reply to white's opening
- Whether at some point in the game, an "equal" bishop-for-knight exchange was made. For this to be true, the following must have occured: a bishop must have captured a knight, or a knight a bishop; the capture must have occurred when the material difference between the two players was no more than one pawn; and the game must revert to a state where the material difference between the two players was no more than one pawn within four moves. These stipulations serve to isolate the target of our study – "equal" bishop-for-knight exchanges.
- The number of "forcing moves" in the game. A forcing move is defined as either a check or the capture of a non-pawn piece.
- The total number of moves in the game.

3. http://chess-db.com/public/top100alltime.jsp, *All time Top 100 Ranklist by Highest ELO Rating.*
4. http://www.chessmetrics.com/cm/CM2/PeakList.asp?Params=, *Peak Average Ratings: 3 year peak range.*

Era	Player Name (by rating)	Number of Games
	Emanuel Lasker	900
	Harry Pillsbury	388
	Géza Maróczy	756
	Siegbert Tarrasch	704
	Wilhelm Steinitz	590
	Akiba Rubinstein	797
1850-	Johannes Zukertort	265
1915	Mikhail Chigorin	688
	David Janowski	769
	Carl Schlecher	739
	Joseph Blackburne	738
	Richard Teichmann	536
	Isidor Gunsberg	319
	Jaan Ehlvest	1500
	Rudolf Spielmann	1050
	José Capablanca	597
	Alexander Alexhine	1500
	Aron Nimzowitsch	512
	Frank Marshall	1027
	Reuben Fine	305
1915-	Efim Bogoljubow	973
1940	Max Euwe	1122
	Salo Flohr	986
	Grigory Levenfish	354
	Richard Réti	646
	Andor Lilienthal	649
	Savielly Tartakower	1029
	Mikhail Botvinnik	891
	Vasily Smyslov	1500
	Tigran Petrosian	1500
1940-	Mikhail Tal	1500
1965	Samuel Reshevsky	1267
	Miguel Najdorf	1500
	Paul Keres	1500
	David Bronstein	342
	Garry Kasparov	1500
	Bobby Fischer	827
	Anatoly Karpov	1500
1965-	Viktor Korchnoi	1500
1990	Boris Spassky	1500
	Lev Polugaevsky	1500
	Alexander Beliavsky	1500
	Jay Timman	173
10	Magnus Carlsen	1500
	Fabiano Caruana	1305
1990-	Levon Aronian	1500
2015	Viswanathan Anand	1500
	Veselin Topalov	1500
	Vladimir Kramnik	1500
	Alexander Grischuk	1195

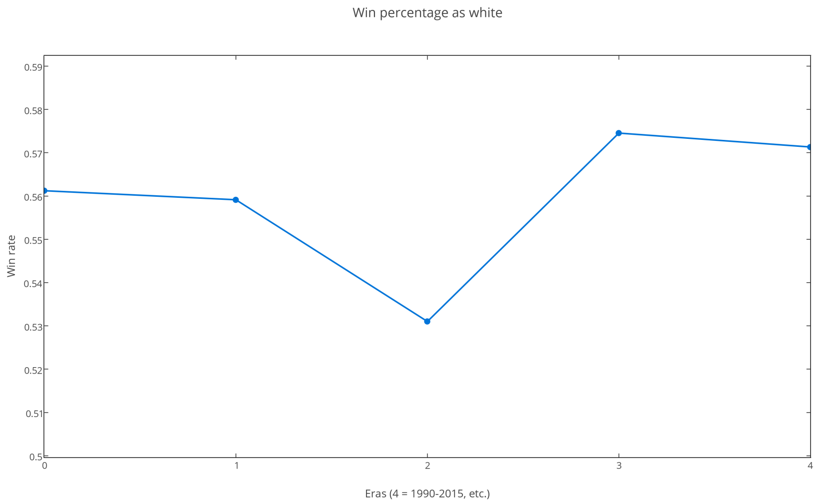

Figure 1: White's advantage in our dataset

3 Exploratory Analysis

We performed several cursory analyses of the data before making our predictive tasks.

3.1 Win Rate for White

First, since it is a well-studied topic, we created our own win rate for white over time. We calculated the average win rate of white in our entire dataset by summing the points earned per game (1 for a win, 0.5 for a draw, and 0 for a loss) by white and dividing it by the size of the dataset in each era. Figure 1 shows this data.

The data is fairly fitting with general consensus of white having a small advantage over black, as well as Olson's research (discussed in a later section).

3.2 Percentage of Games Involving Castling

An early task that we explored was castling, and whether there was a shift in the prevalence of the practice throughout the history of chess. Our results are plotted in Figure 2:

It seems that very little has changed throughout the years and castling is almost ubiquitous at the grandmaster level.

3.3 Modern Openings

As part of our modeling, we examined openings. For instance, Figure 3 shows the most popular openings and replies to those openings made by black. The popularity of black's most played reply shows slight correlation with the popularity of white's opening, but in general was much higher; however, it is not necessarily possible to infer that white has more good options than black, as the fifth most common opening appeared in just 1% of games.

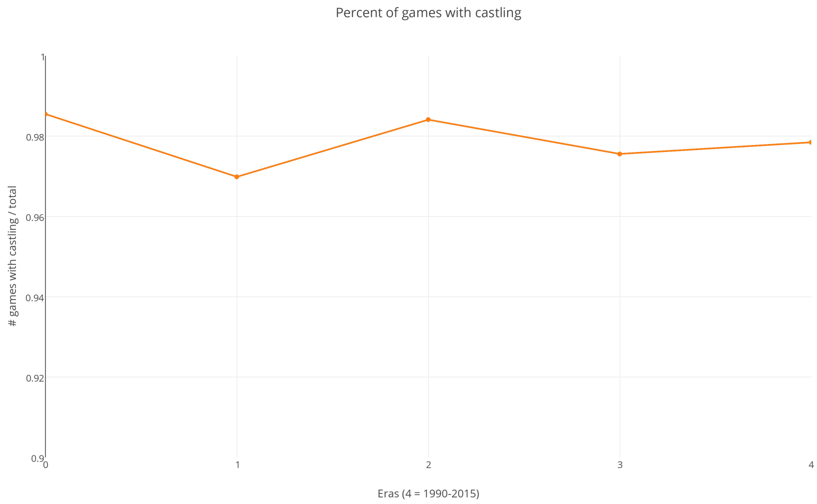

Figure 2: Percentage of games involving castling by at least one player

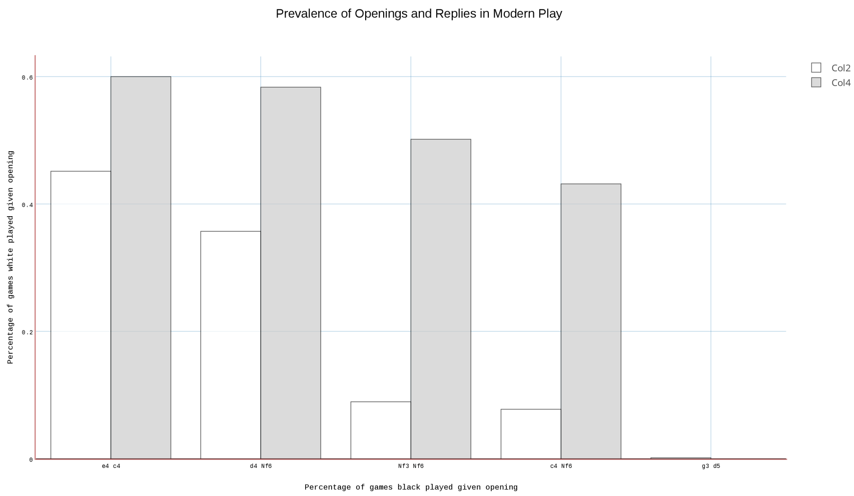

Figure 3: Percentage of most common openings by white and most common reply by black given that opening

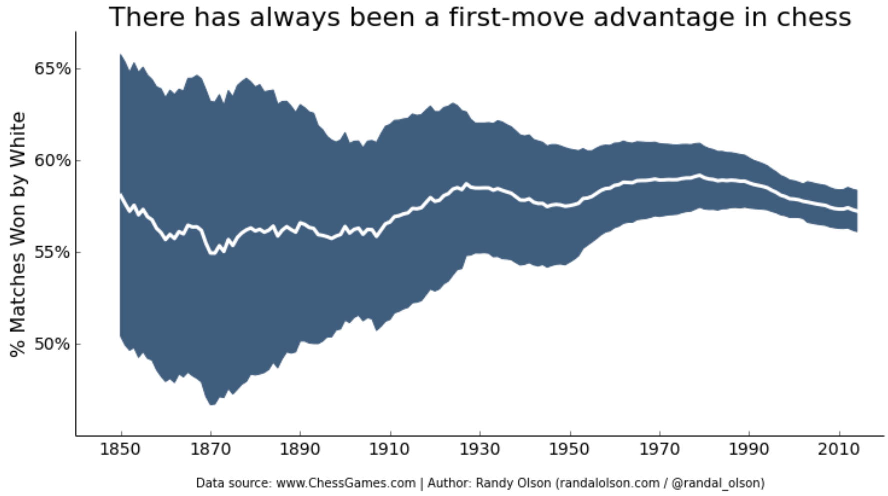

Figure 4: Olson

4 Literature and Research

Chess is an extremely well-studied game, and there is a wealth of existing literature about the game. However, at a cursory glance, the vast majority of these are strategic guides and case studies created primarily by high-level players. Another large number of literature deals with chess-playing AI. Neither of these fields are particularly relevant to our task, which deals with statistical analysis of Chess, so we will not make mention of it.

Even so, there are a wealth of studies which perform statistical analysis of the game. It would be impossible to list them all here, but we will mention significant ones which affected this paper in some way.

4.1 Randal Olson

Dr. Randal Olson is a data visualizer at the University of Pennsylvania who performed a similar task to our in his blog post,[5] which analyzed trends among recorded chess games throughout history. He took data from games over a similar time scale and plotted the data in order to show trends such as game length (in moves) or white's advantage (Figure 1). Though we did not draw explicitly from his data, it is worth noting the trends he highlights in his article.

4.2 Dangauthier et al.

Dangauthier et al. extended the TrueSkill chess ranking system,[6] a statistical algorithm which ranks players based on their historical performances and estimates that player's current skill rating. This is very similar to the tasks we perform, but deal more with a player's overall "skill" rather than the state of a particular game. Though we did not use TrueSkill rankings in

5. Randal Olson, *A data-driven exploration of the evolution of chess: Game lengths and outcomes*, http://www.randalolson.com/2014/05/24/a-data-driven-exploration-of-the-evolution-of-chess-match-lengths-and-outcomes/, 2014.

6. Pierre Dangauthier et al., "TrueSkill Through Time: Revisiting the History of Chess," in *Advances in Neural Information Processing Systems 20* (MIT Press, 2008), 931–938, `http://research.microsoft.com/apps/pubs/default.aspx?id=74417`.

our study, it is nevertheless a worthwhile study to note, as it performs a predictive task using historical knowledge about a player.

4.3 Regan et al., Guid and Bratko

Regan et al. created a model to rank players based on quality of moves rather than the outcome of the game by calculating the probabilities of particular moves made by a stochastic "perfect" agent playing at particular assumed skill levels.[7] While their work is not directly related to ours, it is likewise important to note in its methodology, which makes predictions on players of particular skill levels in given situations.

Likewise, Guid and Bratko used a modified chess engine to measure the performance of selected world champions[8] in order to attempt to quantify who the "best" players are. Of particular note in their study is their examination of the popularity of various board states amongst players.

5 Predictive Task

We first split our data into training (80%) and test (20%) sets. We built our models using our training set, and the validity of our models and errors were checked by running our predictors against the test set. We randomly divided our data into these sets.

Using our data, we selected three predictive tasks. In these tasks, our goal was to maximize our true positives, as it is rather difficult with only the information given to target another goal. Thus, we ended up with rather simple models, but in such a dataset with relatively few features that is to be expected. To make up for it, we performed three predictive tasks instead of one.

5.1 Opening Reply

Our first task is to predict black's reply to a given white opening in some era. There are a discrete number of moves available at each moment, and in these first two moves of the game, there are only 20 options for either player: 16 pawn moves and 4 knight moves. Thus, there are a total of 400 possible states of the game after the first two moves, though only a small handful of these are seen in grandmaster play.

This problem is a discrete classification problem with categorical labels (i.e. given 1. e4, there are 20 possible replies for black). However, with relatively few features with which to tackle this problem (we only know the eras of the games), we could only use simplistic models. We tried three approaches: randomly selecting the move (as a control), always selecting the most popular move in our entire training data, always selecting the most popular move of the era, and by using a basic one-vs-rest classifier of logistical regression for multiclass data. We used the sklearn library for this calculation. Our results can be found in the table below.

Surpringly, one-vs-rest classification performed worse than simply guessing the most popular choice. Because the feature vector was one-dimensional, the predictor suffered greatly from overfitting and generally predicted the most popular choice anyway.

7. Kenneth Wingate Regan and Guy McCrossan Haworth, "Intrinsic Chess Ratings.," in *AAAI* (2011).
8. Matej Guid and Ivan Bratko, "Computer Analysis of World Chess Champions.," *ICGA journal* 29, no. 2 (2006): 65–73.

Method	% correct predictions
Random selection	3.7%
Most popular overall	58.1%
Most popular era	64.6%
One-vs-rest	61.2%

5.2 Knight-for-Bishop Exchanges

Our second task is to predict whether a knight-for-bishop exchange will occur given an era and an opening two moves. The percentage of games in which these exchanges occured in our training data is the following:

Era	% games with equal k-b exchanges
1850-1915	27.8%
1915-1940	30.3%
1940-1965	24.8%
1965-1980	33.1%
1980-2015	32.9%

For this data, we chose to try both logistic regression and a support vector machine, both again implemented using the sklearn library. We also used a random binary classifier as a control. The results of each of these approaches is shown in the following table:

Method	% correct predictions
Logistic regression	67.3%
SVM	63.3%
Random	28.5%

The random variable hit fairly close to the baseline, and the results of logistic regression and SVM were very similar. This can be explain by two factors – first, the opening two moves don't necessarily provide much information, and second, the data was highly overfitted because there weren't many features.

5.3 Forcing Moves

Our final task is to predict the number of forcing moves in a game given the era of the game. The data from our training set is show below:

Era	Average number of forcing moves per game
1850-1915	7.3
1915-1940	6.3
1940-1965	7.1
1965-1980	6.7
1980-2015	6.5%

We applied simple linear regression on this data. Figure 6 shows our fitted line. On the test data, our fit had a mean-squared error of 0.4425, which is very similar to the MSE of our training set. From this data, we can conclude that as time passes, the number of forcing moves generally decreases.

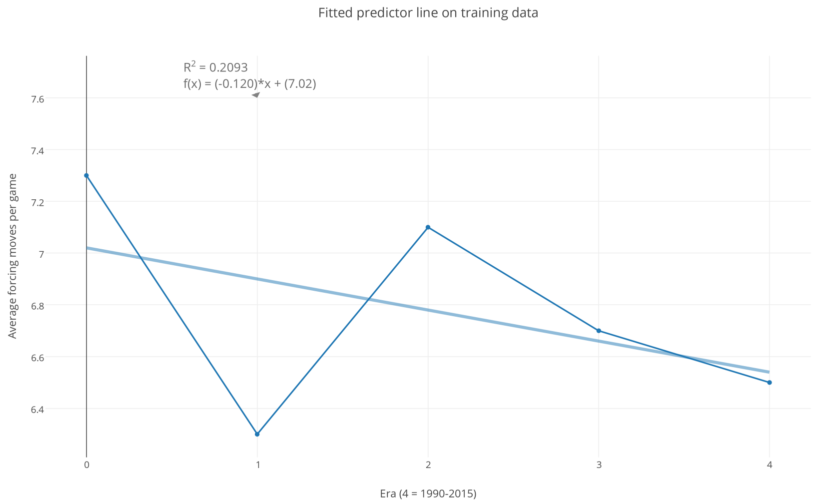

Figure 5: Result of linear regression on forcing move data

6 Conclusions

Our predictive tasks suffered greatly from overfitting of data. Without going in-depth into the moves of the games themselves, it was very difficult for us to make predictions based on just the era alone. A different approach might have been to try using only player data rather than year data, or making predictions based on more accurate time knowledge rather than 25-year blocks. Still, our goal was to isolate some trends in chess in the past 125, and we believe we have taken a small step in the right direction.

References

Byrne, Robert. "Chess." *The New York Times* (1997).

Dangauthier, Pierre, Ralf Herbrich, Tom Minka, and Thore Graepel. "TrueSkill Through Time: Revisiting the History of Chess." In *Advances in Neural Information Processing Systems 20*, 931–938. MIT Press, 2008. `http://research.microsoft.com/apps/pubs/default.aspx?id=74417`.

Guid, Matej, and Ivan Bratko. "Computer Analysis of World Chess Champions." *ICGA journal* 29, no. 2 (2006): 65–73.

http://chess-db.com/public/top100alltime.jsp. *All time Top 100 Ranklist by Highest ELO Rating.*

http://www.chessgames.com/perl/chessgame?gid=1259987. *Francesco di Castellvi vs Narciso Vinyoles, 1475.*

http://www.chessmetrics.com/cm/CM2/PeakList.asp?Params=. *Peak Average Ratings: 3 year peak range.*

Olson, Randal. *A data-driven exploration of the evolution of chess: Game lengths and outcomes.* Http://www.randalolson.com/2014/05/24/a-data-driven-exploration-of-the-evolution-of-chess-match-lengths-and-outcomes/, 2014.

Regan, Kenneth Wingate, and Guy McCrossan Haworth. "Intrinsic Chess Ratings." In *AAAI.* 2011.