# ♟ Chess Data Analysis

Jae-Hyun Lee

# A Bayesian Approach on Opening Tier

## Objective

The main objective of this study is to compare the strength of chess openings, using a Bayesian multinomial model. Opening tier under current and modified win rate policies may give a helpful guide for chess players to prepare the opening. This article is expected to be read by both chess players, who are uninformed of statistics, and statisticians, who are uninformed of chess.

## Introduction to chess openings and win rates

An opening is a series of initial moves in chess. *The Oxford Companion to Chess* lists more than a thousand openings in chess and the well known *Encyclopedia of Chess Openings*, or simply '*ECO*', consists of five thick volumes. [1] Players have evaluated the early moves since the advent of chess, however it is still impossible to specify the move winning by force in the beginning of the game. In fact, even the latest chess engines can compute the forced result of the game only when seven pieces or less are left on board. Exponential amount of calculation will be required every time a piece is added. Therefore, statistical approach on the strength of each opening is reasonable, since it cannot be fully calculated mathematically.

One way of evaluating the strength of an opening is to obtain the results of the games that started out with the opening. If black has won more games than white, obviously the opening favors for the black player. However such opening does not exist. Statistics observe that white is more advantageous even if black counters with the best move. Although it is not proven whether the advantage of moving a piece first leads to a forced win, there is no doubt that white is not disadvantageous at least. Therefore, the threshold of the strength of an opening is not a specific value of win rate, and should be evaluated relatively to other openings.

Another approach would be evaluating the positional and tactical consequences of an opening. For example, novices in chess are taught not to play a4 on move 1, because it helps the opponent to take control of the center of the chessboard first. Controlling the center enables the player to move the major and minor pieces more freely than the opponent. In other words, the player prevails in the number of possible moves on each turn. This is an example of an approach called positional evaluation.

Then how can we compare the strengths of the openings in statistics? First, collect the game samples and classify the opening played. The number of wins, draws and losses can easily be obtained. We can also sort the openings by win-rates.

$$win\ rate = 1 \cdot Pr(win) + \frac{1}{2} \cdot Pr(draw) + 0 \cdot Pr(loss) = p_w + \frac{1}{2}p_d$$

We would also like to apply the idea of positional and tactical evaluation. However, it is not an easy task to evaluate a position objectively in numerics. Stockfish, the strongest engine of today evaluates each position with difference in the number of pawns. For example, +1.0 means that white's position is as advantageous as a position with an extra pawn. However, this doesn't necessarily mean that the number of pawns differ by one. The number is an overall evaluation of the coordination of all pieces, not a simple weighted sum of piece values. Even though this Stockfish statistic is intuitive and useful for understanding the position as a human being, the problem is that it is merely an arbitrary number. The number itself hardly gives information of the distribution of wins, draws and losses. Exceptions are situations like forced checkmate or stalemate, but they don't belong to the opening phase.



*White plays 1. a4 and immediately the Stockfish evaluation drops from +0.2 to -0.3.*

As an alternative, we can simulate numerous  sample games with the Stockfish engine. The sampled data would give answers to how the computer evaluates each opening and the corresponding distribution of wins, draws and losses. This could be used as an objective comparison or a supplement for the user data.

# Data collection

## Data1: lichess.org user data [2]

This is a dataset that summarizes 200,000 games that are played from May 2019 on lichess.org. Lichess is the second biggest online platform for chess. `200k_blitz_rapid_classical_bullet.csv` contains the following information.

- Black and white's Elo ratings and the difference after the game
- Date and time of the game
- Event (classical, rapid, blitz and bullet) and time control
- Opening name
- All moves played and the result of the game
- Stockfish evaluation per each move
- Time used per each move

| Black | BlackElo | BlackRatingDiff | Date | ECO | Event | Opening | Result | Round | Site | Termination | TimeControl |
|-------|----------|-----------------|------|-----|-------|---------|--------|-------|------|-------------|-------------|
| albertoPlasta | 906 | 13 | 2019.04.30 | B15 | Rated Blitz game | Caro-Kann Defense | 0-1 | - | https://lichess.org/g9rQgEgj | Normal | 300+3 |
| Luckystriker | 1296 | 28 | 2019.04.30 | C50 | Rated Blitz game | Italian Game | 0-1 | - | https://lichess.org/SHT7fNep | Normal | 300+0 |
| Verascardoso | 1761 | -13 | 2019.04.30 | C41 | Rated Rapid game | Philidor Defense #2 | 1-0 | - | https://lichess.org/x3FcpWTb | Normal | 600+0 |
| Stockfish94 | 2404 | 8 | 2019.04.30 | B06 | Rated Blitz game | Modern Defense | 0-1 | - | https://lichess.org/TNkviuh3 | Normal | 60+0 |
| deneme12376 | 1595 | -10 | 2019.04.30 | B32 | Rated Blitz game | Sicilian Defense: Loewenthal Variation | 1-0 | - | https://lichess.org/rmTPOL2Q | Normal | 180+0 |
| abusado | 1715 | -9 | 2019.04.30 | A40 | Rated Rapid game | English Defense #2 | 1-0 | - | https://lichess.org/EPK6N2CB | Normal | 600+0 |

### Preprocessing and visualization

Only the relevant variables we are interested in are excerpted. In this study we are interested in the classical high-level games. This is because often objectively inferior openings tend to have a merit in shorter games, since they pressure the opponent to take time to discover the move that could punish the bad move. Since the number of classical games are too small, we tolerate the rapid games to be included in our dataset. Only the games between players over 2,000 Elo rating and under 200 Elo difference are included.
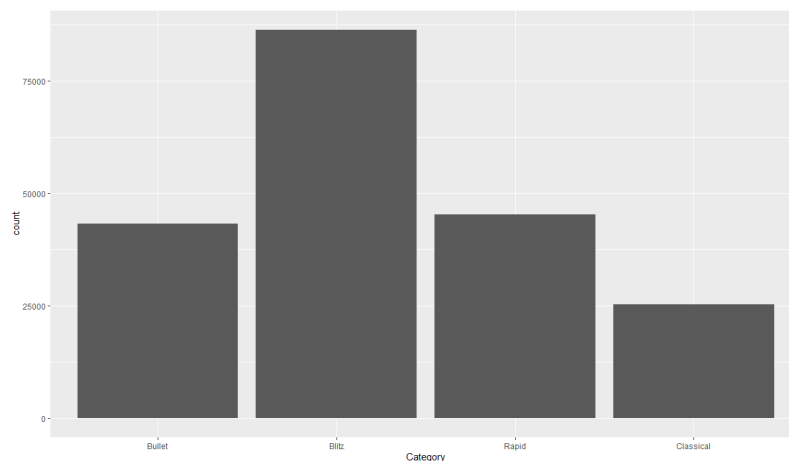


As explained in the introduction, there are many openings in chess, but they are not played in equal rates. We would only focus on the openings that have been played 10 times or more.
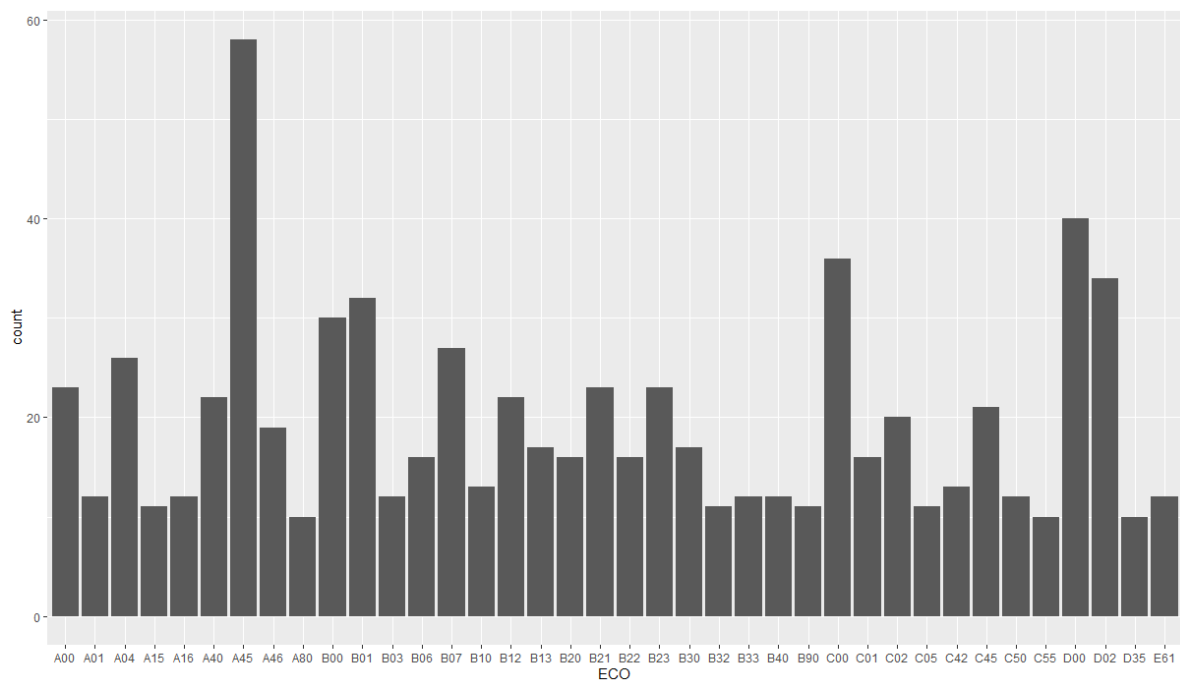
Only 38 openings are played 10 times or more, and the opening that is played the most often is A45, which is known as the "Indian game": 1. d4 Nf6 .



We can also find out which opening has the best win rate. If the number is closer to 1, the opening favors for white.
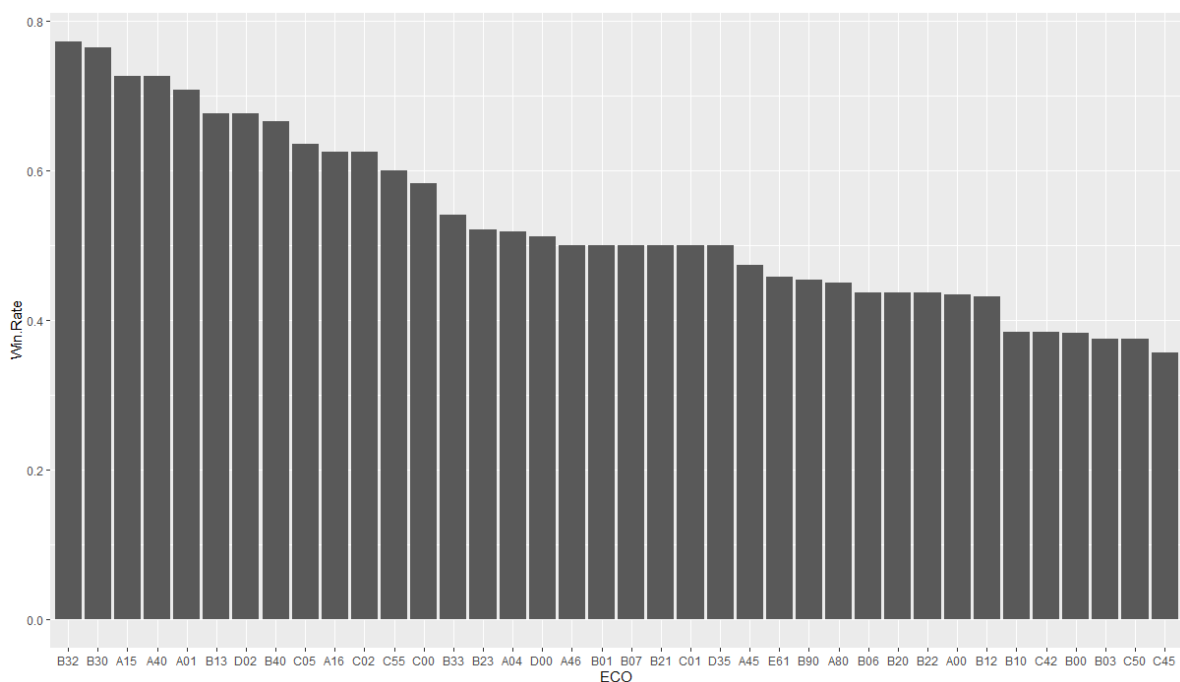


- B32 (Open Sicilian) and B30 (Old Sicilian) are observed to have the best win rate as white.

- B32: 1. e4 c5 2. Nf3 Nc6 3. d4 cxd4 4. Nxd4 e5
- B30: 1. e4 c5 2. Nf3 Nc6



Then shouldn't white just play the Open or Old Sicilian every game according to the data? Unfortunately, this is impossible because chess is a game of interaction. To enter the lines above, black must play c5 and Nc6 on the first and second move. The move 2... Nc6 is considered inferior by the latest engines. Therefore, it is important to find which side has the initiative to get into a specific opening. In other words, white has the advantage to punish 2... Nc6 only if black plays the specific move.

- C45 (Scotch Game) is observed to have the worst win rate as white.
    - C45: 1. e4 e5 2. Nf3 Nc6 3. d4 exd4 4. Nxd4
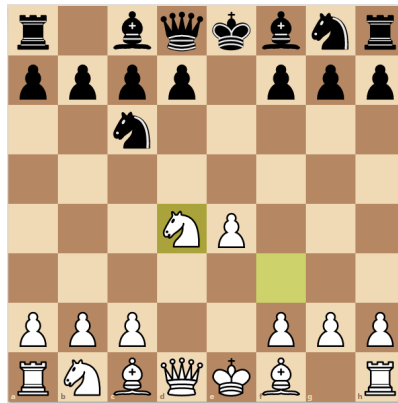


Likewise, it is impossible for black to play C45 every game. However, if white plays an inaccurate move 3. d4, ... exd4 could be recommended for black by statistics.

It seems that we can answer which openings to play and not play just by taking a look at the bar plot above. However, a reasonable doubt arises when we investigate the scatter plot below.
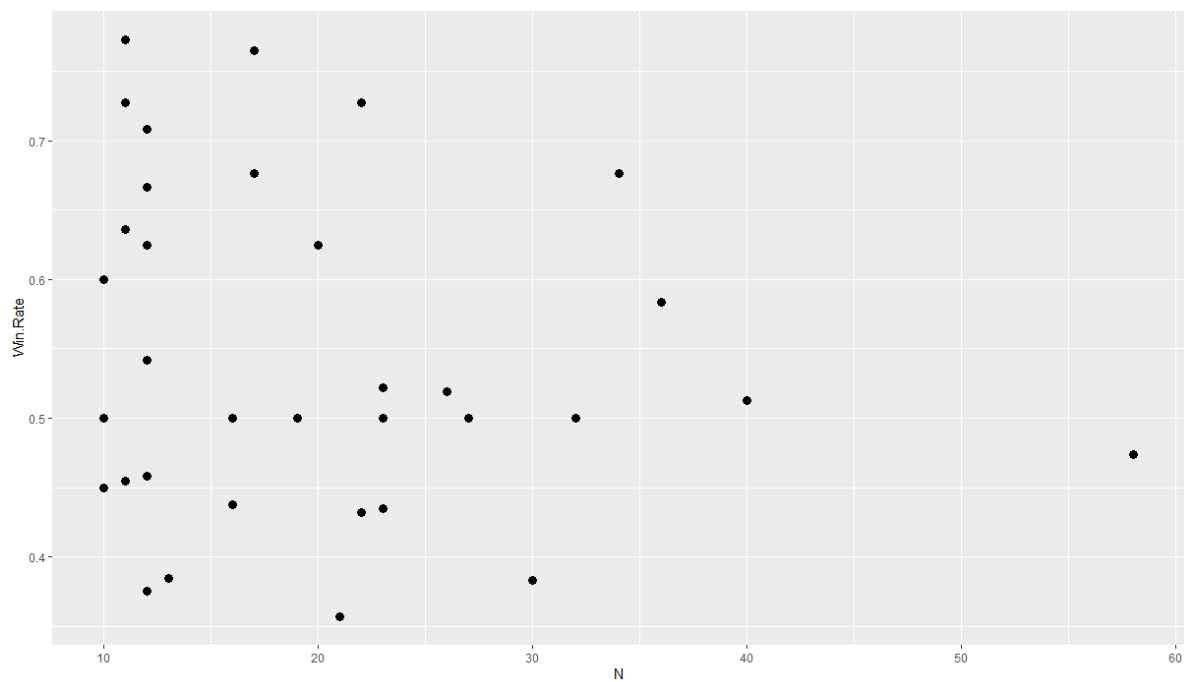
We can easily notice that the smaller the sample size, the more dispersion of win rates. In other words, the reason behind extreme win rates may simply be small sample size, not the intrinsic superiority or inferiority of the opening.

## Data2: Stockfish simulation [3]

As an alternative to using a single set of samples as above, a Bayesian approach can be applied. To set up a prior belief, we can use the information of how the best player in the world evaluates each opening. The best player is, of course, Stockfish. Using python and the Stockfish package, we can generate a Stockfish vs Stockfish game. For each opening that is played ten times or more in the user data, Stockfish will be playing itself ten times starting from the specific opening position.

The reason we simulate ten games per opening is as following.

- We want to give same weight to the prior belief and the obtained user data, when the sample size is at its minimum. (N=10)
- The speed of simulation is slower than expected. One of the reasons is that Stockfish is reluctant to take draws, thus plays meaningless random moves in obviously drawn positions like King+Bishop vs King+Bishop.
- The simulated games are not deterministic in theory but empirically they shared great similarity. If Stockfish were deterministic, the ten simulated games would be identical. To ensure that the simulated games are indeterministic, handicaps and multiple cores should be validated just as in the default settings.

Here is an example of a simulated game between two identical Stockfish engines. They had to play from a given position, in this case ECO B90 Sicilian Najdorf, one of the most popular openings among top grandmasters.

- B90: 1. e4 c5 2. Nf3 d6 3. d4 cxd4 4. Nxd4 Nf6 5. Nc3 a6

The game ended up in a checkmate, white to win. Full codes and game logs are attached on `Stockfish_Simulation.ipynb`. We are only interested in the results, therefore we only saved the results of the 380 games after the simulation.
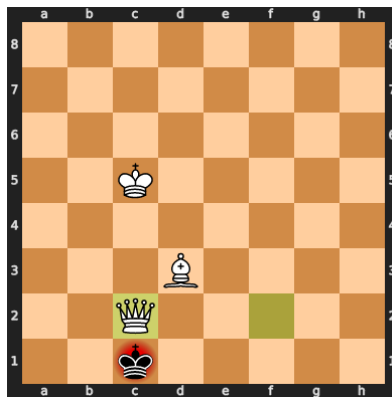


```python
for i in range(len(eco)):
    for j in range(10):
        board = chess.Board()
        board.set_fen(eco['FEN'][i])
        engine = chess.engine.SimpleEngine.popen_uci(r'E:\anaconda3\Lib\site-packages\stockfish\stockfish_13_win_x64.exe')
        while not board.is_game_over():
            result = engine.play(board, chess.engine.Limit(time=0.1))
            board.push(result.move)
        engine.quit()
        eco.iloc[i,j+2] = board.result()
```

| | ECO | FEN | G1 | G2 | G3 | G4 | G5 | G6 | G7 | G8 | G9 | G10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | A00 | rnbqkbnr/pppppppp/8/8/1P6/8/P1PPPPPP/RNBQKBNR ... | 1/2-1/2 | 1/2-1/2 | | 0-1 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 |
| 1 | A01 | rnbqkbnr/pppppppp/8/8/1P6/P1PPPPPP/RNBQKBNR ... | 0-1 | 1/2-1/2 | | 0-1 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | | 0-1 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 |
| 2 | A04 | rnbqkbnr/pppppppp/8/8/5P2/8/PPPPP1PP/RNBQKBNR ... | 0-1 | | 0-1 | 1/2-1/2 | 1/2-1/2 | | 1-0 | 1/2-1/2 | 1/2-1/2 | | 0-1 | 1/2-1/2 | 1/2-1/2 |
| 3 | A15 | rnbqkb1r/pppppppp/5n2/8/2P5/8/PP1PPPPP/RNBQKBN... | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | | 1-0 | 1/2-1/2 | | 1-0 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 |
| 4 | A16 | rnbqkb1r/pppppppp/5n2/8/2P5/2N5/PP1PPPPP/R1BQK... | 1/2-1/2 | | 0-1 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | | 1-0 |
| 5 | A40 | rnbqkbnr/pppp1ppp/8/4p3/3P4/8/PPP1PPPP/RNBQKBN... | 1-0 | 1-0 | 1-0 | 1-0 | 1-0 | 1-0 | 1-0 | 0-1 | 1-0 | 1-0 |
| 6 | A45 | rnbqkb1r/pppppppp/5n2/8/3P4/8/PPP1PPPP/RNBQKBN... | 1/2-1/2 | 1/2-1/2 | | 1-0 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | | 1-0 | 1-0 | 1/2-1/2 | 1/2-1/2 |
| 7 | A46 | rnbqkb1r/pppppppp/5n2/8/3P4/5N2/PPP1PPPP/RNBQK... | 1/2-1/2 | | 0-1 | 1/2-1/2 | | 1-0 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | | 1-0 | 1/2-1/2 |
| 8 | A80 | rnbqkbnr/ppppp1pp/8/5p2/3P4/8/PPP1PPPP/RNBQKBN... | 1-0 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | | 0-1 | 1/2-1/2 | 1/2-1/2 | | 0-1 | 1/2-1/2 | | 1-0 |
| 9 | B00 | rnbqkbnr/ppppp1pp/5p2/8/4P3/8/PPPP1PPP/RNBQKBN... | 1-0 | 1-0 | | 0-1 | 1/2-1/2 | | 1-0 | 1/2-1/2 | | 1-0 | 1/2-1/2 | 1-0 | 1-0 |
| 10 | B01 | rnbqkbnr/ppp1pppp/8/3p4/4P3/8/PPPP1PPP/RNBQKBN... | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | 1/2-1/2 | | 1-0 | 1/2-1/2 | | 1-0 | 1/2-1/2 | 1-0 | 1-0 |

10 games per each of the 38 openings are simulated and the results are saved as `Stockfish_Simulation.csv`. Before moving on to a statistical analysis, we spotted some interesting features within the dataframe.

- There were overall 254 draws, 84 wins and 42 losses for white. Since two identical chess engines collide, it was expected to see many draws if the opening is legitimate.
- Games under ECO A40, the Englund Gambit, indicated greatest win rate for white. White won 9 games and lost one with no draws.
- Meanwhile, 10 games under ECO B10, the Caro-Kann Defense were all drawn. In chess, we may call openings as the Englund Gambit sharp, since players tend to win or lose but not draw. On the other hand, we may call openings as Caro-Kann 'drawish', since often times there are few imbalances within the game.

## Bayesian multinomial model [4]

There are three possible results in each game of chess for white, a win, draw and loss. Therefore the result after n games follows a trinomial distribution.

$$Y|\theta \sim Trinomial(\theta)$$

$$p(y|\theta) = p(w, d|p_w, p_d) = \frac{n!}{w!d!(n-w-d)!} p_w^w p_d^d p_l^l$$

$$l = n - w - d$$

$$p_l = 1 - p_w - p_d$$

Via simulation we can collect reliable game data between two highest-level players. We construct the prior distribution with Stockfish vs Stockfish dataset, since the simulated results can provide information about the evaluation on each opening before collecting the user data.

For simpler calculation we use a conjugate model. The prior distribution follows a Dirichlet distribution, with parameters equivalent to the number of wins, draws and losses in the simulated data.

$$\theta \sim Dirichlet(\alpha, \beta, \gamma)$$

$$p(\theta) = p(p_w, p_d) = \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} p_w^{\alpha-1} p_d^{\beta-1} (1 - p_w - p_d)^{\gamma-1}$$

$$\alpha, \beta, \gamma = number\ of\ wins, draws, losses\ in\ simulation\ (constants)$$

Using the Bayes theorem, we can obtain the posterior distribution of closed form.

$$p(\theta|y) = p(p_w, p_d|w, d) = \frac{p(p_w, p_d)p(w, d|p_w, p_d)}{\int_\Theta p(p_w, p_d)p(w, d|p_w, p_d)d\Theta}$$

$$
\begin{aligned}
p(\theta|y) = p(p_w, p_d|w, d) &\propto p(p_w, p_d) \times p(w, d|p_w, p_d) \\
&\propto p_w^w p_d^d (1 - p_w - p_d)^{n-w-d} \times p_w^{\alpha-1} p_d^{\beta-1} (1 - p_w - p_d)^{\gamma-1} \\
&\propto p_w^{\alpha+w-1} p_d^{\beta+d-1} (1 - p_w - p_d)^{\gamma+n-w-d-1} \\
&= p_w^{\alpha+w-1} p_d^{\beta+d-1} (p_l)^{\gamma+l-1}
\end{aligned}
$$

Note that the kernel of the posterior distribution indicates another Dirichlet distribution. Therefore conjugacy holds, since the prior and posterior distribution belong to the same family of distribution.

$$\theta|y \sim Dirichlet(\alpha + w, \beta + d, \gamma + l)$$

$$p(\theta|y) = \frac{\Gamma(\alpha + \beta + \gamma + n)}{\Gamma(\alpha + w)\Gamma(\beta + d)\Gamma(\gamma + l)} p_w^{\alpha+w-1} p_d^{\beta+d-1} (p_l)^{\gamma+l-1}$$

Using the posterior distribution obtained, we can answer the following list of questions in this study. The expected win rate can be obtained using Monte Carlo approximation instead of numerical method.

$$E\left(p_w + \frac{1}{2}p_d \,\middle|\, w, d\right) = \int_\Theta \left(p_w + \frac{1}{2}p_d\right) p(p_w, p_d | w, d)\, d\Theta$$

**Monte Carlo approximation**

i. From the posterior distribution, sample S sets of p_w and p_d.

ii. Calculate p_w + 0.5 p_d for each sample.

iii. By the law of large numbers, the mean of the obtained values is asymptotically equal to the expectation.

$$\theta^{(1)} = \{p_w^{(1)}, p_d^{(1)}\}$$
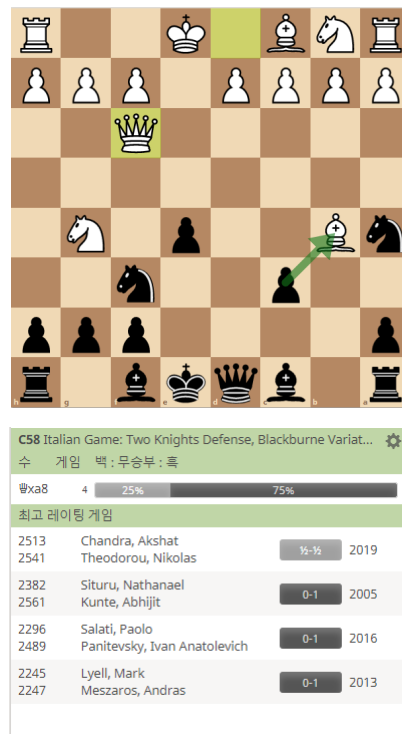$$\theta^{(2)} = \{p_w^{(2)}, p_d^{(2)}\}$$
$$\cdots$$
$$\theta^{(S)} = \{p_w^{(S)}, p_d^{(S)}\}$$

$$\frac{1}{S}\sum_{s=1}^{S} g(\theta^{(s)}) \to E(g(\theta)|y) \quad as\ S \to \infty$$

$$\frac{1}{S}\sum_{s=1}^{S} \left(p_w^{(s)} + \frac{1}{2}p_d^{(s)}\right) \to E\left(p_w + \frac{1}{2}p_d \,\middle|\, w, d\right) \quad as\ S \to \infty$$

**Interpretation and Justification**

How can we interpret the Bayesian trinomial model and what's the point of it? The result of the model is very simple. We merge the Stockfish simulation dataset (prior) and lichess user data (likelihood), then add the wins, draws and losses respectively. The consequent three numbers indicate the parameters of the posterior Dirichlet distribution. The idea beneath is that players and Stockfish find a reasonable consensus on strengths of each opening.



This is an example of when the Bayesian approach could be useful. It is a line named 'C58: Italian Game, Two Knights Defense, Blackburne Variation.' Strange as it seems, it is actually one of the mainlines after white goes for the well-known Fried Liver Attack. Black's rare response after Qf3 is cxb5, which looks almost impossible at a glance, because it simply hangs a rook on a8. However, Stockfish's evaluation on this line is 0.0. Though black may be down on exchange (rook - bishop),

Stockfish thinks there are enough compensation for Black to keep the game going. Again, the number 0.0 doesn't mean anything metric. It may infer that there is a forced draw line for black after Qxa8, or it may be simply telling us the winning chances are about equal. Deeper analysis is required to find out, but the distribution of the game results will still be unknown.

In a frequentist's point of view, the parameters of the distribution are estimated from the games that are actually played. There are only four games in lichess masters database: 1 draw and 3 wins for black. User data not only indicates that this line is playable, but also suggests it is almost winning for black! However, there are two main problems in the conclusion. Firstly, the sample size is only four, which is too small for a frequentist approach. Secondly, black's Elo rating is higher than white's in every game. To guess the true distribution of the game result, we add Stockfish simulation results as prior belief to the model, so that the extreme likelihoods are diluted when sample sizes are too small. This is the reason why we take a Bayesian approach on evaluating the strengths of chess openings.

## Posterior intervals of draw rates

Draw is a big issue in chess. Unlike any other sports, chess is notorious for its high draw rates, especially among games between top grandmasters. For example, in 2019 World Chess Championship Finals, GM Magnus Carlsen and GM Fabiano Caruana tied in all 8 classical games and had to play more rapid games to decide the world champion. There is a growing opinion in chess community, where some people argue that the 'drawish' tendency among top-level games of chess makes professional, classical chess lame. Some chess fans even consider reducing the compensation of a draw, which is currently half the point of a win.

It is a common strategy as a chess player to fight for a win as white, and pursue draw as black. There are openings like the Berlin Defense which are called the 'drawing weapon' for black, which satisfies the idea. As the first research in the study, we will find out the posterior draw rates to distinguish which openings are 'drawish.'

To obtain the posterior intervals we could follow the Bayesian trinomial model, but the binomial model is used instead.

$$draw\ rate = Pr(draw) = p_d$$
$$1 - p_d = p_w + p_l = p_{wl}$$

Basically, we are going to build up a binomial model, where success is a draw and failure is a win or a loss. The likelihood function of a binomial distribution is given as following.

$$Y|\theta \sim Binomial(\theta)$$
$$p(y|\theta) = p(d|p_d) = \frac{n!}{d!(n-d)!} p_d^d p_{wl}^{wl}$$
$$wl = n - d$$
$$p_{wl} = 1 - p_d$$

The prior distribution is the beta distribution with given constant parameters a and b.

$$\theta \sim Beta(a, b)$$
$$p(\theta) = p(p_d) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p_d^{a-1}(1 - p_d)^{b-1}$$
$$a, b = number\ of\ draws\ and\ elses\ in\ simulation\ (constants)$$

Using the Bayes theorem the posterior distribution of beta can be obtained.

$$p(\theta|y) = p(p_d|d) = \frac{p(p_d)p(d|p_d)}{\int_{\Theta} p(p_d)p(d|p_d)d\Theta}$$
$$\propto p(p_d) \times p(d|p_d)$$
$$\propto p_d^d(1-p_d)^{n-d} \times p_d^{a-1}(1-p_d)^{b-1}$$
$$\propto p_d^{a+d-1}(1-p_d)^{b+n-d-1}$$

The kernel of the distribution is that of a beta distribution.

$$\theta|y \sim Beta(a+d, b+n-d)$$
$$p(\theta|y) = \frac{\Gamma(a+b+n)}{\Gamma(a+d)\Gamma(b+n-d)}p_d^{a+d-1}(1-p_d)^{b+n-d-1}$$

$$E(p_d|d) = \int_{\Theta} p_d \, p(p_d|d)d\Theta$$
$$= \int_0^1 p_d \frac{\Gamma(a+b+n)}{\Gamma(a+d)\Gamma(b+n-d)}p_d^{a+d-1}(1-p_d)^{b+n-d-1}dp_d$$
$$= \frac{a+d}{a+b+n}$$

Few changes are made compared to the trinomial model. The only difference is that we use the specific form of the multinomial and Dirichlet distribution, which are the beta and binomial distribution. The reason we build up a simpler model from scratch is to use the closed form of the posterior distribution of draw rates. Note that to obtain a win rate Monte Carlo simulation is used, but to obtain a draw rate, taking the exact values from the beta distribution is available.

```
posterior %>%
  mutate(WL_post = Win_post + Loss_post,
         post_Mean = Draw_post / (Draw_post+WL_post),
         post_lb = qbeta(0.025, Draw_post, WL_post),
         post_ub = qbeta(0.975, Draw_post, WL_post)) %>%
  arrange(Draw_post) -> DRAWISH_post
```
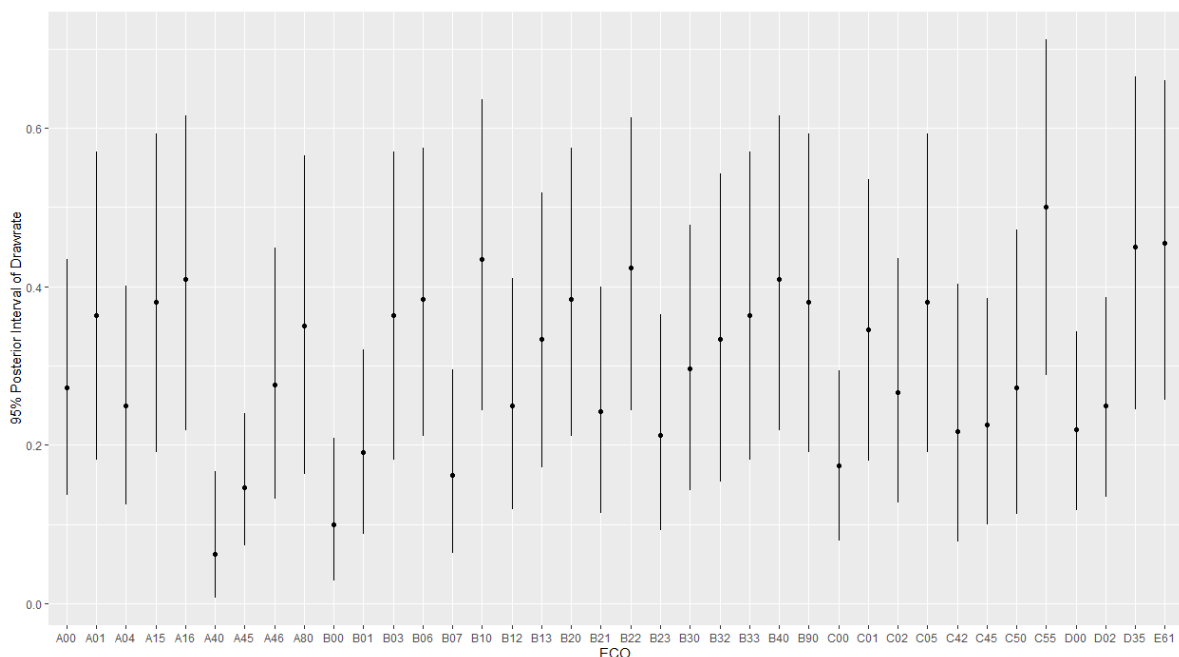


- Recall that under Stockfish simulations, ECO A40: the Englund Gambit had 10 decisive games out of all 10 games. Likewise, the posterior distributions indicate that is is the sharpest opening among the 38 openings.
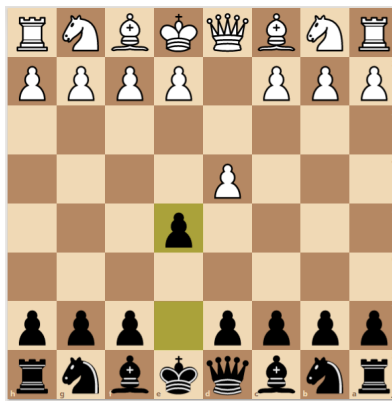
- ECO C55: the Two Knights Defense had the greatest 'drawish' tendency, with a draw rate of 50% on average.



Likewise, expected win, draw and loss rates can be obtained respectively. Sum of the three expected values equals 1 and the results can be shown graphically. Note that the wins and losses are from white's perspective.
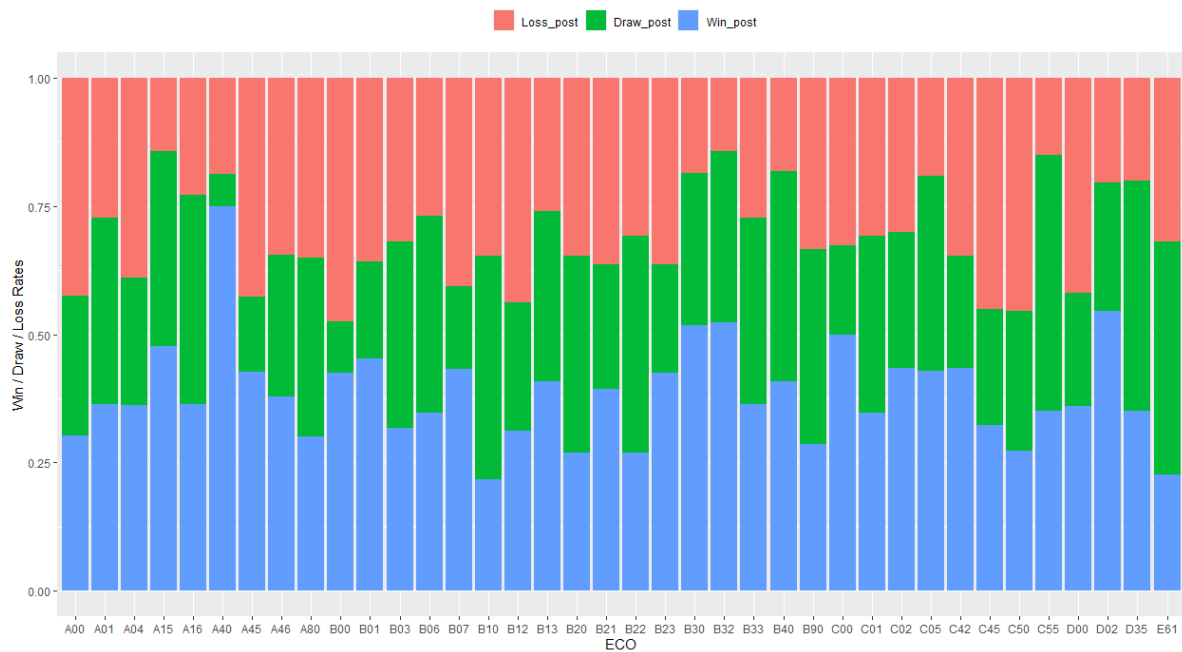
# Opening tier in conventional win rates

$$win\ rate = 1 \cdot Pr(win) + \frac{1}{2} \cdot Pr(draw) + 0 \cdot Pr(loss) = p_w + \frac{1}{2}p_d$$

This is a formula for the conventional win rate in chess. 1 point for a win, and 0.5 point for a draw is given. As mentioned before, to compare the posterior win rates of the openings, Monte Carlo approximation is applied.

```
for(i in 1:nrow(posterior)){
  w = posterior$Win_post[i]
  d = posterior$Draw_post[i]
  l = posterior$Loss_post[i]
  MC_Sampling = rdirichlet(50000, c(w,d,l))
  MC_Winrate = as.vector(MC_Sampling %*% c(1, 0.5, 0))
  posterior$MC_lb[i] = quantile(MC_Winrate, 0.025)
  posterior$MC_ub[i] = quantile(MC_Winrate, 1-0.025)
  posterior$MC_Mean[i] = mean(MC_Winrate)
}
```
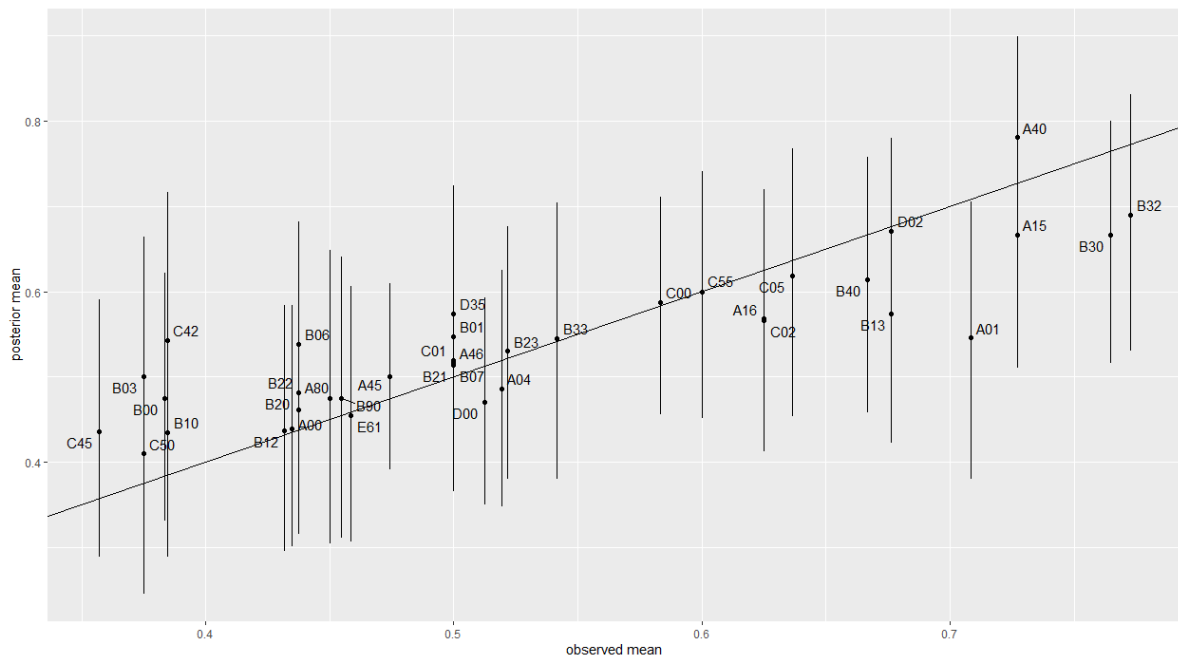


- This is a plot where the x-axis is the observed mean from the user data and the y-axis is the posterior mean.
- Vertical segments and center points are posterior intervals and the means respectively.
- The diagonal line is a 45 degree line from the origin.
- ECO A40: the Englund Gambit has the greatest posterior mean. It ranked third when the user data was inspected, but because the prior belief is highly biased to white, the posterior mean of the win rate is greater than that of any other openings.

- ECO C50: Giuoco Piano has the least posterior mean. It was originally the second worst opening for white according to the user data. On the other hand, in black's perspective, Giuoco Piano is proven to be the best counterattack.



- We can observe the **shrinkage effect**. The expected value of win rate is pulled a bit from the observed mean towards the prior mean by the amount depending on the sample size. As mentioned before, it was doubted that the reason behind the extreme values of win rates can simply be small sample sizes. The extreme values on the left hand side of the plot have mostly shrunken upward and the extreme values on the right hand side have mostly shrunken downward, where the 45 degree diagonal line is the threshold.

## Opening tier in modified win rates

Chess players earn certain points after the game: 1 point for a win, 0.5 point for a draw and 0 point for a loss. This is the reason why the win rate in chess can be constructed as the weighted sum of win ratio and draw ratio. Now consider other sports such as football. In many football leagues, loss is considered as 0 point as well. The difference is that in football, a win is thrice as valuable as a draw: 3 points for a win, 1 point for a draw. Perhaps the 'drawish' tendency in professional chess is due to the amount of compensation to a draw. This lead people to consider manipulating the points in chess.

Now we will define modified win rates. The corresponding point for a draw is set to δ, $0 \le \delta \le 1$. We will track the changes in the opening tier as the value of δ shifts.

$$modified\ win\ rate = p_w + \delta \cdot p_d$$

- Note that when δ is set to zero, it means that a draw is equal to a loss to a player. In player's perspective, this is often the case at a tournament. If the player loses or draws, he/she disqualifies. Therefore, it is important for the player to choose an opening that can maximize the probability of victory. The posterior distributions of each opening with δ equal to zero would be a useful guide.

- Likewise, when δ is equal to one, it means that a draw is equal to a win to a player. The information obtained would be useful when a draw is enough for the qualification at a tournament. For both extreme cases marginalizing the posterior Dirichlet distribution to a beta distribution is possible as in 1.
- When δ does not equal 0.5, the sum of win rates for white and black does not equal to 1. Thus research on both sides of the board is required.

The difference in the codes compared to the codes for the conventional win rates is that the Monte Carlo sample from the Dirichlet distribution is now multiplied to [1, δ, 0]. Again, the modified win rate also follows the form of weighted sum of ratios.

```
for(i in 1:nrow(posterior_delta)){
  w = posterior_delta$Win_post[i]
  d = posterior_delta$Draw_post[i]
  l = posterior_delta$Loss_post[i]
  delta = posterior_delta$delta[i]
  MC_Sampling = rdirichlet(50000, c(w,d,l))
  Winrate_delta = as.vector(MC_Sampling %*% c(1, delta, 0))
  posterior_delta$Mean_delta[i] = mean(Winrate_delta)
}
```

The followings are the ranks of the opening strengths. As δ shifts from zero to one, some iterations in the ranks are made.



Opening Tier in Rankings as delta Changes [White]
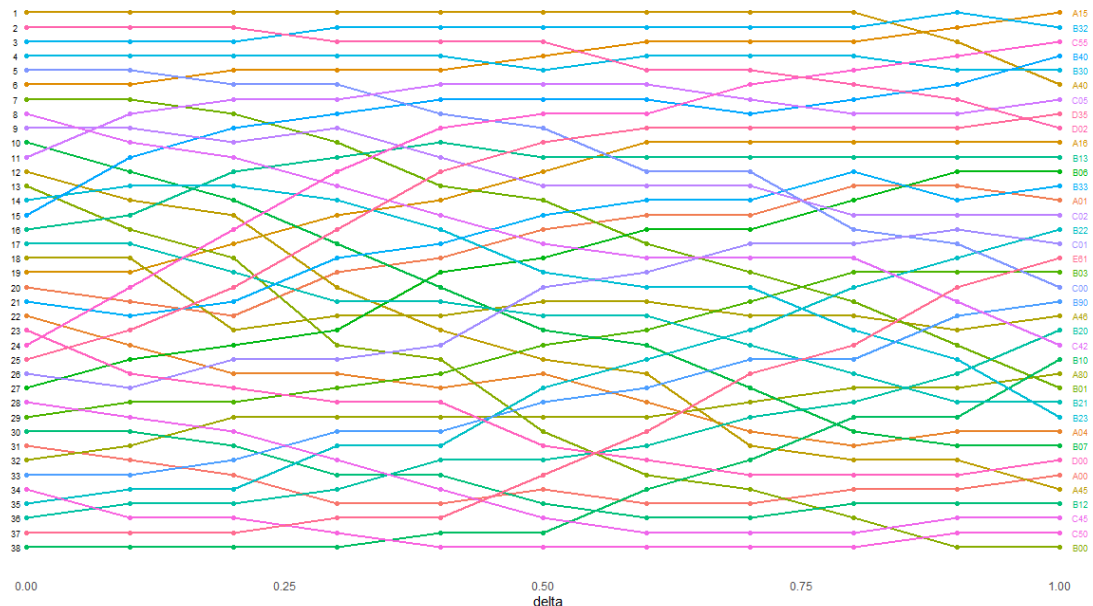
- δ = 0
    - ECO A40: the Englund Gambit is the strongest opening as white.
    - ECO B10: Caro-Kann is the worst opening to face as white.
    - In this must-win situation, the sharpest opening, the Englund Gambit ranked first on the list. On the other hand solid but drawish Caro-Kann ranked the last.
- δ = 0.33
    - The interval of possible δ is set to 0.1, thus the case δ = 0.33 is disregarded. Checking the results of δ = 0.3 can be a close guess.

- - Nevertheless, 0.33 does have a certain meaning. It is the point where the win is 3 times more valuable than a draw, just like in football leagues. Since draw is not as compensating as it used to be, it is expected to see sharper openings in chess tournaments.
- δ = 0.5
  - δ set to 0.5 is the win rate computed conventionally, thus the results are also the same. The best opening is ECO A40: the Englund Gambit and the worst is ECO C50: Giuoco Piano.
- δ = 1
  - ECO A50: English opening is the strongest.
  - ECO B00: Uncommon King's Pawn opening is the worst option for white.
  - Since draw means a win in this case, the positional and solid English took the first place. On the other hand, King's Pawn opening with uncommon first moves has proven to be too risky.



Opening Tier in Rankings as delta Changes [Black]
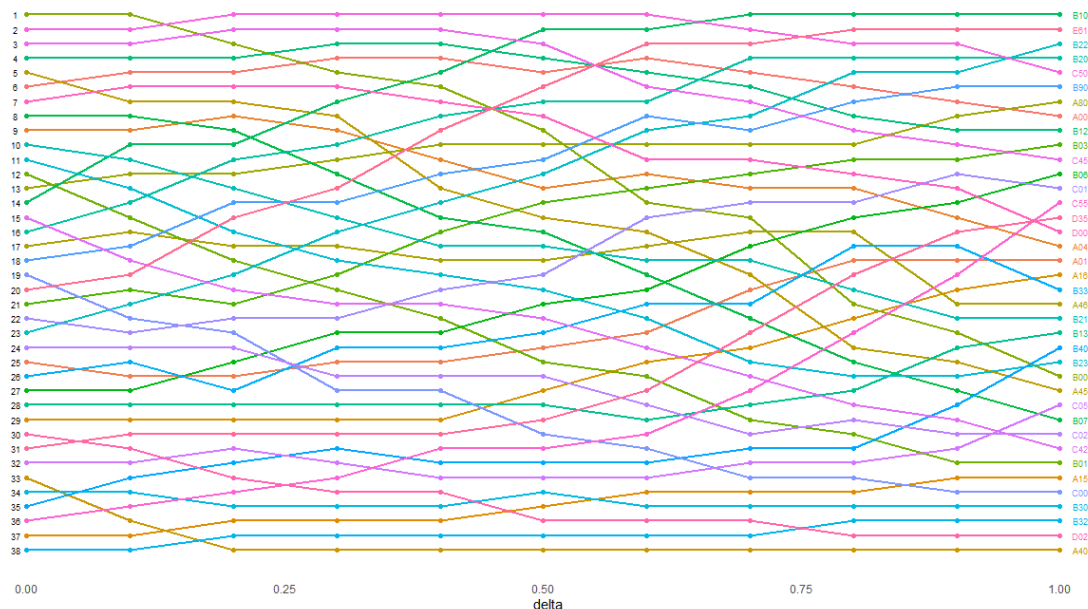
- δ = 0
  - ECO B00: Uncommon King's Pawn opening is the strongest opening as black.
  - ECO B32: the Open Sicilian is the least successful for black.
- δ = 0.5
  - In this case the ranks for black and white are direct opposites. Now the worst opening is ECO A40: the Englund Gambit and the best is ECO C50: Giuoco Piano.
- δ = 1
  - ECO B10: Caro-Kann is the strongest opening as black.
  - ECO A40: the Englund Gambit is the worst option to play as black.
- The results are almost symmetrical for black and white. As δ increases, rather drawish openings that are solid and positional are more successful.

By moderating δ, we were able to find out the opening tier on specific situations. In most of the chess tournaments the point system is the same thus the win rate formula remains conventional. However, due to increasing demands for decisive games among chess fans, the point system itself may be reconsidered. Also, even without any changes in the rulebook, the approach can still be

useful because as an individual δ doesn't necessarily equal 0.5. GM Teimour Radjabov, top tenth grandmaster criticized for his mundane 'drawish' tendency within the games, explained the reason behind his draws on his twitter. When he loses a game, it is not a simple loss with zero point. His FIDE ratings may fall and he might not be invited in the consequent tournaments because he is on the verge of being invited. As a metaphor, Tiemour's δ is close to 1, only because the aftermath of a loss is too critical. These situations are not limited to top grandmasters; all chess players may face different, yet similar situations like a must-win case. Setting up δ and searching for adequate opening options based on the rank plots may be a good choice.

## Limitations

**Limitations within the ECO system**

An opening is distinguished by its ECO code during the study. Openings aren't distinguished enough under ECO. For example, ECO A40 includes the Englund Gambit but random openings such as 1. d4 a5 and 1. d4 a6 are also designated to A40. Also, most ECO codes are determined by first 3 moves or less. However professional chess players prepare openings in deeper depths. To target grandmasters of chess, a verified list of deep mainlines can be studied.

The good news is that the openings don't have to be refrained to ECO codes. Once the line that a person wants to study is prepared, the whole process of Bayesian approach can be easily repeated. Recall the opening C58: Italian Game, Two Knights Defense, Blackburne Variation. Black loses a rook for a bishop but Stockfish says the position its completely fine for black. This is a specific line in the C58 Italian Game and the whole line could be put into inspection. To build up the prior belief, the exact position can be put into Stockfish simulation to obtain the result of the games. 4 games in lichess masters database play a role as the likelihood, thus posterior distribution of the result of the specific line can be obtained.

**Limitations of the user data**

lichess user data is used and games between players with over 2000 Elo ratings are excerpted. However there is a huge gap between 2000 Elo rating online (especially in lichess) and 2000 FIDE rating. The latter would hardly lose a single game. The top grandmasters are over 2800 in FIDE ratings, and chess engines are expected to place over 3600. Because of their relative weaknesses, it is doubtful if the evaluations based on user data over 2000 Elo is even credible. Limiting to higher Elo ratings is one option to solve the issue, but now we will face critical sample size problems.

The second option is using other improved datasets. lichess provides monthly user data for free, and the data that is used in this study is only a fraction of the data that is over 566 GB. If the budget allows, accessing master database like Chessbase would also be a good idea. If as a professional player an uncontaminated analysis is needed, such better datasets may be used instead.

1. https://en.wikipedia.org/wiki/Chess_opening ↪
2. https://web.chessdigits.com/data ↪
3. https://stockfishchess.org/ ↪
4. Gelman, A. (2004). *Bayesian data analysis*. Boca Raton, Fla: Chapman & Hall/CRC. ↪