Analiza statystyczna czasów na wykonywanie ruchów w szachach

Piotr Rogula

Politechnika Wrocławska

29 stycznia 2022

promotor: prof. dr hab. inż. Marcin Magdziarz

Spis treści

Motywacja

1. Szachy jako hobby
2. popularny temat
3. niedosyt literatury opisującej dane zagadnienie

Kluczowe wyniki innych autorów

- System Elo [2] (Arpad Elo)

System Elo

- przyznawanie punktów bazujące na różnicy rankingu graczy
- pierwszy system mający podłoże probabilistyczne

Kluczowe wyniki innych autorów

- System Elo [2] (Arpad Elo)
- System Glicko-2 [3] (Mark Glickman)

System Glicko-2

- ulepszenie systemu Elo.
  - wzięcie pod uwagę przedziału ufności rankingu każdego z graczy.
- używany w dużej liczbie gier MMO.

Kluczowe wyniki innych autorów

- System Elo [2] (Arpad Elo)
- System Glicko-2 [3] (Mark Glickman)
- Silnik Stockfish [1]

Silnik Stockfish

- funkcja oceny

wynik liniowej funkcji ważonej sumy cech, na którą składają się między innymi:
f_b, f_c – wartość figur odpowiednio białych i czarnych
k_b, k_c – bezpieczeństwo króla odpowiednio białych i czarnych
m_b, m_c – mobilność figur odpowiednio białych i czarnych
z_b, z_c – potencjalne zagrożenia wykonane odpowiednio białych i czarnych

$$f(f_b, f_c, k_b, k_c, m_b, m_c, \dots) = c_1(f_b - f_c) + c_2(k_b - k_c) + c_3(m_b - m_c) + \dots$$

gdzie: c_i są stałymi określającymi wagę danej pary zmiennych.

Silnik Stockfish

- rodzaje błędów szachowych
  - ?? – błąd (ang. *blunder*)
  - ? – pomyłka (ang. *mistake*), posunięcie błędne w mniejszym stopniu niż „błąd"
  - ?! – niedokładność (ang. *innacuracy*), posunięcie, które można zastąpić zdecydowanie lepszym.

Potrzebne oznaczenia

- **ruch** składa się z dwóch posunięć - 1 białych i 1 czarnych
  - wyjątkiem może być ostatni ruch, gdy po posunięciu białych nastąpił koniec partii.
- oznaczenia posunięcia jako „błąd” i „duży błąd” są tożsame
- oznaczenie „pomyłka” jest rozróżnialne od oznaczenia „błąd” – drugie jest wg silnika gorszym posunięciem

sformułowanie problemu

- zbadanie zależności pomiędzy czasem poświęconym na wykonanie ruchu, a jego dokładnością
- zbadanie zależności między numerem ruchu, a czasem na jego wykonanie oraz jego dokładnością
- ~~próba wyznaczenia optymalnego czasu na wykonanie ruchu minimalizacja ryzyka wystąpienia błędu~~

dane

- baza danych **Lichess.com** - 1 plik 72Gb
- zbadanie 2 najczęściej granych formatów (600+0, 300+0)
- stworzenie bazy ok. 7% gier ocenionych przez silnik
- stworzenie bazy wszystkich ruchów ze wszystkich gier
  - 17,52 mln posunięć z 275,94 tyś gier

wstępna analiza

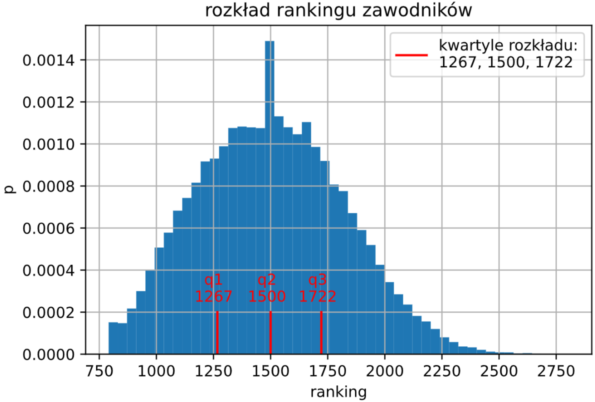
	game_ID	score	delta_time	WhiteElo	BlackElo	TimeControl	color	move	Result
510	9	0	0	1192	1204	300+0	w	23	0-1
511	9	0	9	1192	1204	300+0	b	23	0-1
512	9	0	2	1192	1204	300+0	w	24	0-1
513	9	0	18	1192	1204	300+0	b	24	0-1
514	9	0	7	1192	1204	300+0	w	25	0-1
515	9	0	10	1192	1204	300+0	b	25	0-1
516	9	blunder	8	1192	1204	300+0	w	26	0-1
517	9	blunder	8	1192	1204	300+0	b	26	0-1
518	9	mistake	4	1192	1204	300+0	w	27	0-1
519	9	0	2	1192	1204	300+0	b	27	0-1
520	9	0	29	1192	1204	300+0	w	28	0-1

Rysunek: Fragment bazy zawierającej ruchy z gier o formatach czasowych „300+0” i „600+0”.

dane

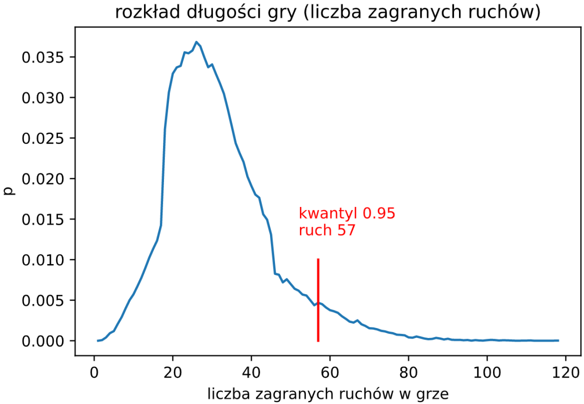
- czego nie ma w danych?
  - digitalizacja czasu – czas na posunięcie zaokrąglony do pełnych sekund
  - brak informacji i odpowiedniej miary dotyczącej skomplikowania sytuacji na szachownicy

Analiza problemu - wstępny przegląd danych



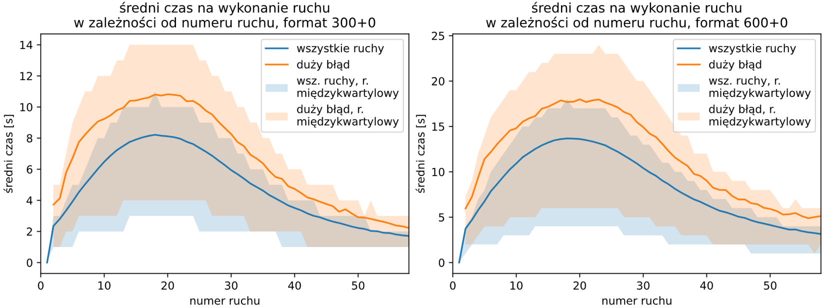
Rysunek: rozkład rankingu zawodników wraz z zaznaczonymi kwartylami

Analiza problemu - wstępny przegląd danych



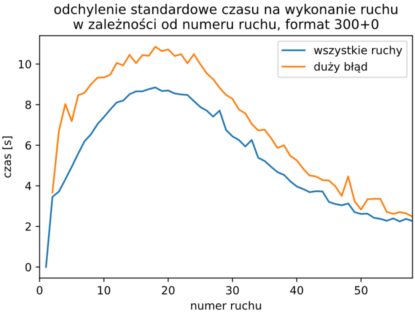
Rysunek: rozkład długości gier wraz z zaznaczonym kwantylem rzędu 0.95, dla gier z formatu 300+0 oraz 600+0.

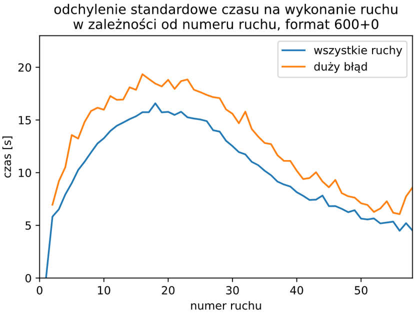
Zależność między indeksem ruchu, a poświęconym czasem



Rysunek: Średni czas na wykonanie ruchu dla analizowanych formatów czasowych wraz z zaznaczonym rozstępem międzykwartylowym. Osobno wszystkie ruchy i ruchy oznaczone przez silnik jako błąd.

Zależność między indeksem ruchu, a poświęconym czasem





Rysunek: Odchylenie standardowe czasu na wykonanie ruchu dla analizowanych formatów czasowych. Osobno wszystkie ruchy i ruchy oznaczone przez silnik jako błąd.

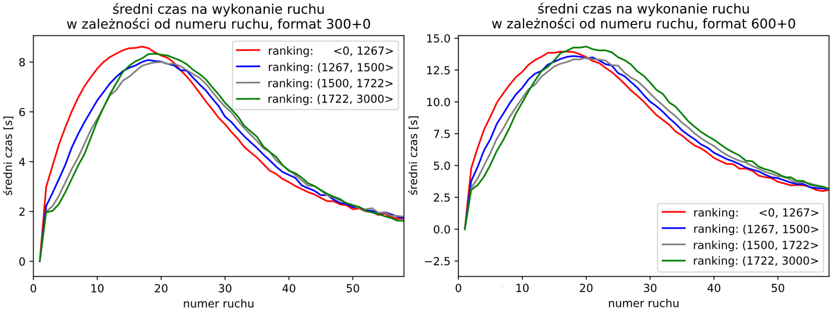
Zależność między indeksem ruchu, a poświęconym czasem

Po oznaczeniu:
D jako zmienną określającą skomplikowanie pozycji,
T - zmienną określającą czas na wykonanie ruchu w danej pozycji,
B - zdarzenie polegające na tym, że posunięcie jest błędne
i przy założeniu, że T jest silnie dodatnio skorelowane z D, według prawdopodobieństwa otrzymujemy:

$$P(B|D > d_0) > P(B|D \leqslant d_0) \rightarrow P(B|T > t_0) > P(B|T \leqslant t_0)$$

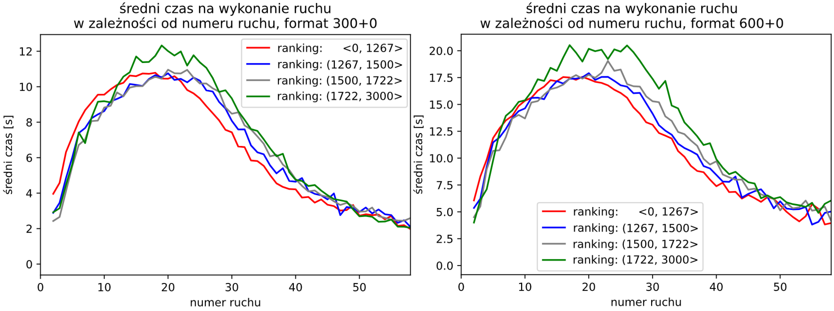
gdzie d_0 i t_0 określają punkty, od których według wybranej miary można określić, że pozycja jest skomplikowana $(D > d_0)$ i czas na wykonanie posunięcia jest długi $(T > t_0)$.
Może to tłumaczyć wyższą średnią czasu na wykonanie błędnego posunięcia w porównaniu do zbioru wszystkich posunięć.

Zależność między indeksem ruchu, a poświęconym czasem



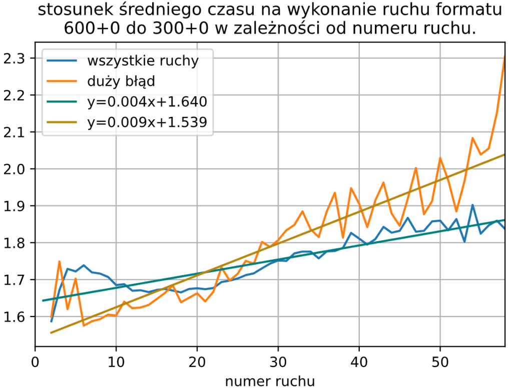
Rysunek: Średni czas na wykonanie ruchu dla dla graczy z różnych przedziałów rankingowych. Wszystkie posunięcia.

Zależność między indeksem ruchu, a poświęconym czasem



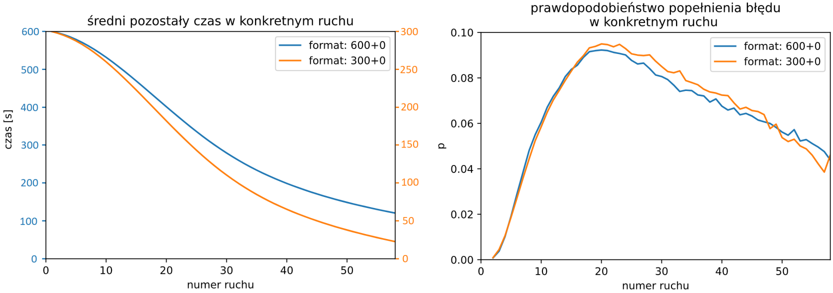
Rysunek: Średni czas na wykonanie ruchu dla dla graczy z różnych przedziałów rankingowych. Posunięcia oznaczone przez silnik jako błąd.

Zależność między indeksem ruchu, a poświęconym czasem



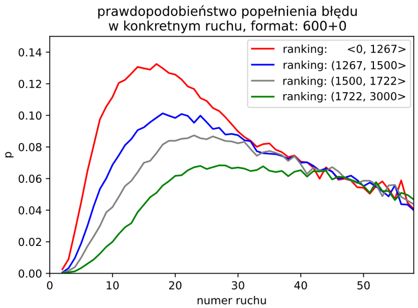
Rysunek: Stosunek średniego czasu na wykonanie ruchu w formacie 600+0 do czasu w formacie 300+0. Zaznaczone proste regresji obrazujące trend.

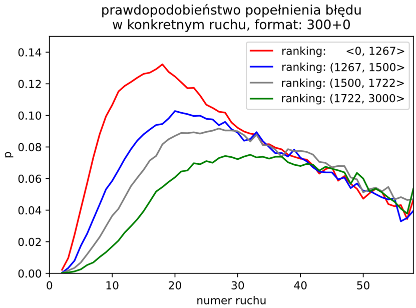
Zależność między indeksem ruchu, a poświęconym czasem



Rysunek: Pierwszy wykres – zestawienie średniego pozostałego czasu w formatach 600+0 oraz 300+0. W celu lepszego porównania zastosowane osobne osie dla każdego formatu. Drugi wykres – empiryczne prawdopodobieństwo popełnienia błędu w konkretnym ruchu.

Zależność między indeksem ruchu, a poświęconym czasem





Rysunek: Prawdopodobieństwo popełnienia błędu w konkretnym ruchu dla różnych przedziałów rankingowych.

Zależność między indeksem ruchu, a poświęconym czasem

$$X_i = \begin{cases} 1, & \text{gdy w } i\text{-tym ruchu został popełniony błąd} \\ 0, & \text{gdy w } i\text{-tym ruchu nie został popełniony błąd} \end{cases}$$

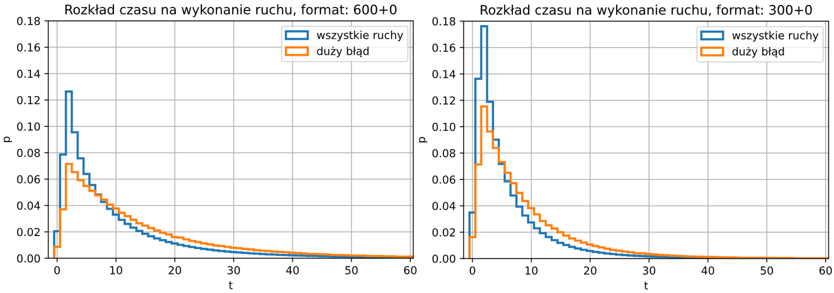
$$P(X_1 = 1 \vee X_2 = 1 \vee \cdots \vee X_n = 1) =$$
$$= 1 - P(X_1 = 0, X_2 = 0, \ldots, X_n = 0) =$$
$$= P(X_1 = 0)P(X_2 = 0)\ldots P(X_n = 0) =$$
$$= \prod_{i=1}^{n} P(X_i = 0)$$

Zależność między indeksem ruchu, a poświęconym czasem

Tabela: Szansa na popełnienie błędu w pierwszych *n* ruchach dla gracza z określonego przedziału rankingowego z rozdzieleniem na formaty „600+0” i „300+0”

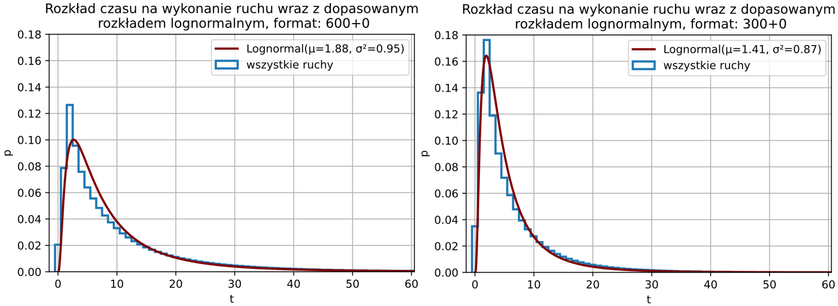
***n* pierwszych ruchów**	**n = 10**		**n = 20**		**n = 30**	
centyl \ format	**600+0**	**300+0**	**600+0**	**300+0**	**600+0**	**300+0**
0-25%	43,36%	40,72%	85,50%	84,33%	95,32%	94,96%
25%-50%	25,52%	22,09%	71,56%	68,63%	89,38%	88,66%
50%-75%	15,58%	12,42%	58,86%	56,27%	83,08%	82,93%
75%-100%	6,76%	5,63%	40,78%	38,66%	70,27%	70,40%

Zależność między czasem poświęconym na ruch, a jego oceną



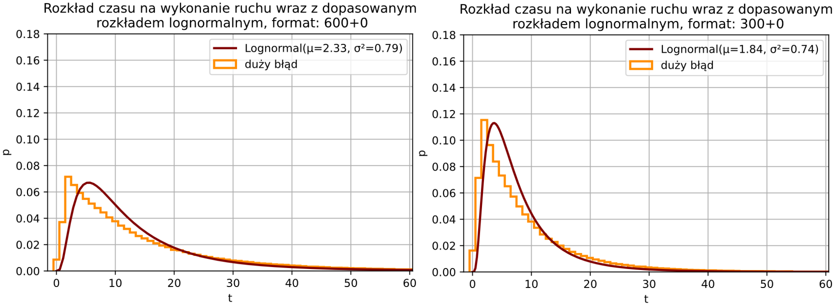
Rysunek: Rozkład czasu poświęconego na wykonanie posunięcia dla formatów „600+0" i „300+0". Osobno wszystkie posunięcia i posunięcia błędne

Zależność między czasem poświęconym na ruch, a jego oceną



Rysunek: Rozkład czasu poświęconego na wykonanie posunięcia dla formatów „600+0” i „300+0” z dopasowanym rozkładem log-normalnym.

Zależność między czasem poświęconym na ruch, a jego oceną



Rysunek: Rozkład czasu poświęconego na wykonanie posunięcia dla formatów „600+0" i „300+0" z dopasowanym rozkładem log-normalnym. Zestaw ruchów błędnych.

Zależność między czasem poświęconym na ruch, a jego oceną

Jiahang Lyu, Saraless Nadarajah:
„Discrete lognormal distribution with application to insurence data” [4]

$$p_X(i) = \Phi\left(\frac{\log i - \mu}{\sigma}\right) - \Phi\left(\frac{\log(i-1) - \mu}{\sigma}\right) \tag{1}$$

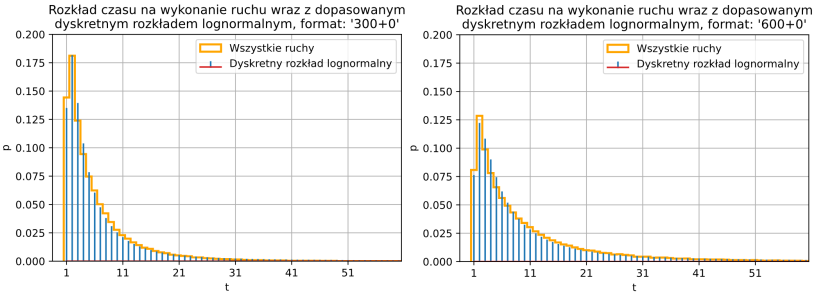
dla $i \in \mathbb{N}_+$, gdzie $\Phi(x)$ jest wartością dystrybuanty standardowego rozkładu normalnego w punkcie x.

Zależność między czasem poświęconym na ruch, a jego oceną

Funkcja największej wiarygodności

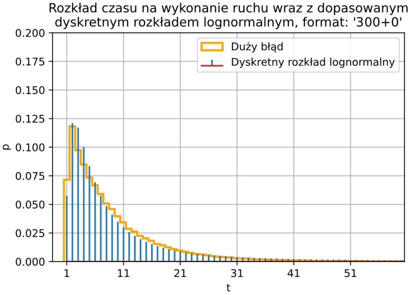
$$L(\mu,\sigma) = \sum_{i=1}^{n} \log\left[\Phi\left(\frac{\log x_i - \mu}{\sigma}\right) - \Phi\left(\frac{\log(x_i - 1) - \mu}{\sigma}\right)\right] \quad (2)$$

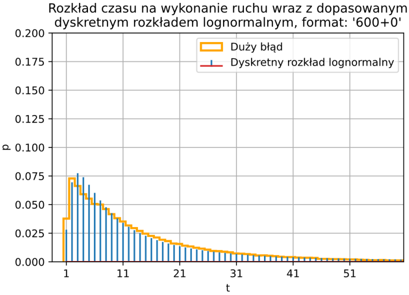
Zależność między czasem poświęconym na ruch, a jego oceną



Rysunek: Rozkład czasu poświęconego na wykonanie posunięcia dla formatów „600+0” i „300+0” z dopasowanym dyskretnym rozkładem log-normalnym.

Zależność między czasem poświęconym na ruch, a jego oceną





Rysunek: Rozkład czasu poświęconego na wykonanie posunięcia dla formatów „600+0" i „300+0" z dopasowanym dyskretnym rozkładem log-normalnym. Zestaw ruchów błędnych.

Zależność między czasem poświęconym na ruch, a jego oceną

Tabela: Średnie wyniki testu zgodności Chi-kwadrat, średnia ze 100 prób Monte-Carlo dla losowych próbek o długości 1000. Poziom istotności $\alpha = 0,05$.

dane	**parametry**	**p-wartość**	**wynik**
„300+0”,	$\mu = 1,22,\ \sigma = 1,09$	22,28%	przyjęcie
„600+0”,	$\mu = 1,70,\ \sigma = 1,19$	20,40%	przyjęcie
„300+0”, błędne	$\mu = 1,67,\ \sigma = 1,06$	7,14%	przyjęcie
„600+0”, błędne	$\mu = 2,16,\ \sigma = 1,12$	6,49%	przyjęcie

Wnioski

- 95% gier kończy się w 57 ruchach
- błędne posunięcie zajmuje średnio dłużej niż standardowe
  - dla każdego rankingu i numeru ruchu
  - błędne posunięcie ma też zawsze większą wariancje
- największa szansa na błąd umiejscowiona jest w okolicach 20 ruchu

Wnioski

- gracze z wyższym rankingiem szybciej wykonują ruchy początkowe, dłużej te z największą szansą na błąd
  - w odniesieniu do graczy z niższym rankingiem
  - więcej czasu poświęconego na posunięcie w okolicy 20 ruchu $\rightarrow$ mniejsza szansa na błąd
  - więcej poświęconego czasu przy skomplikowanej pozycji
  - mniej przy znanej pozycji początkowej

Wnioski

- największa szansa na błąd
  - słabsi gracze – dużo więcej błędów na początku (ruchy 1 – 20)
  - słabsi gracze – więcej błędów w środkowej części gry (ruchy 20 – 30)
  - słabsi gracze – podobna liczba błędów w fazie końcowej (ruchy 30+)
- różnice między formatami
  - dla „600+0” ruchy zajmują nieliniowo więcej czasu, niż dla „300+0”
  - dla „300+0” gracze są zmuszeni szybko wykonywać ruchy późniejsze.

Wnioski

- 10 pierwszych ruchów (otwarcie),
  - istotna różnica w popełnieniu przynajmniej jednego błędu między rankingami
  - przykładowo: format „300+0”
    5,63% szansy na błąd dla graczy powyżej centylu 75,
    40,72% szansy na błąd dla graczy poniżej centylu 75,

Dalsza praca

- stworzenie odpowiedniej miary i wzięcie pod uwagę poziomu skomplikowania pozycji
  - potrzebna dużo większa moc obliczeniowa
- stworzenie odpowiedniej miary do określania błędów (innej niż Stockfish)
  - sprawdzenie zgodności wyników
  - potrzebna dużo większa moc obliczeniowa

Bibliografia

Repozytorium zawierające kod źródłowy silnika Stockfish.
https://github.com/official-stockfish/Stockfishr.
Dostęp: 11/10/2021.

Arpad E. Elo.
The rating of chessplayers, Past & Present (second edition).
FIDE, New York, United States, 1986.

Mark Glickman.
The glicko system.
Boston University, 1999.

Jiahang Lyu and Saralees Nadarajah.
Discrete lognormal distributions with application to insurance data.
Springer, 7 2021.