

Misurare le prestazioni agonistiche: il caso degli Scacchi

Paolo Ciancarini
Università di Bologna

Evento metrico GUFPI-ISMA
Università di Bologna
12 settembre 2013



Chi sono io?

- Insegno Ingegneria e Architettura del Software presso il Dipartimento di Informatica dell'Università di Bologna
 - Mi occupo inoltre di Informatica Ludica e Agonistica (“Entertainment Computing”)
 - Ho scritto un libro sulle macchine che giocano a scacchi: *Giocatori Artificiali*, pubblicato da Mursia
-

Alcune domande

- Dati due giocatori che si sfidano a Scacchi, come si calcola la probabilità del risultato (come misuriamo il **rating** di un giocatore)?
- Dato un gruppo di giocatori, come facciamo a “metterli in fila” (come stabiliamo il loro **ranking**) ?

Rating: misurare la qualità (es. di un entità)

Ranking: definire la posizione in una classifica (ordinare un insieme di entità)

Rating vs ranking

UNDERSTANDING ONLINE STAR RATINGS:



Rank	Rating	Title	Votes
1.	9.2	Le ali della libertà (1994)	1,025,319
2.	9.2	Il Padrino (1972)	722,303
3.	9.0	Il Padrino - Parte II (1974)	467,711
4.	8.9	Pulp Fiction (1994)	794,486
5.	8.9	Il buono, il brutto, il cattivo. (1966)	308,533
6.	8.9	Il cavaliere oscuro (2008)	1,000,196
7.	8.9	La parola ai giurati (1957)	251,597
8.	8.9	Schindler's List (1993)	521,466
9.	8.8	Il signore degli anelli - Il ritorno del re (2003)	728,018
10.	8.8	Fight Club (1999)	778,735

imdb.com

Rating: misurare la qualità (es. di un entità)

Ranking: definire la posizione in una classifica (ordinare un insieme di entità)

Sommario

- Gli Scacchi ed il sistema Elo
 - La storia del sistema Elo
 - La teoria
 - Aspetti pratici
 - Applicazioni oltre gli Scacchi
-

Chi gioca?

- Giocano a Scacchi milioni di persone
- Alcuni paesi hanno una grande tradizione
- Esiste una letteratura immensa sul gioco, dell'ordine di oltre 100.000 libri, tesi e articoli scientifici
- Il gioco è di solito associato all'intelligenza e come tale si ritrova in molti libri e film



Persone famose che gioca(va)no a Scacchi



Goethe



Lenin con Gorkj



Fischer con FidelCastro



Nicholas Cage



Humphrey Bogart



Bono (U2)



Tolstoy



Madonna

NAPOLÉON A L'ILE S^T HÉLÈNE.



Enlève. L'art.



Lehr. de P. 1818.

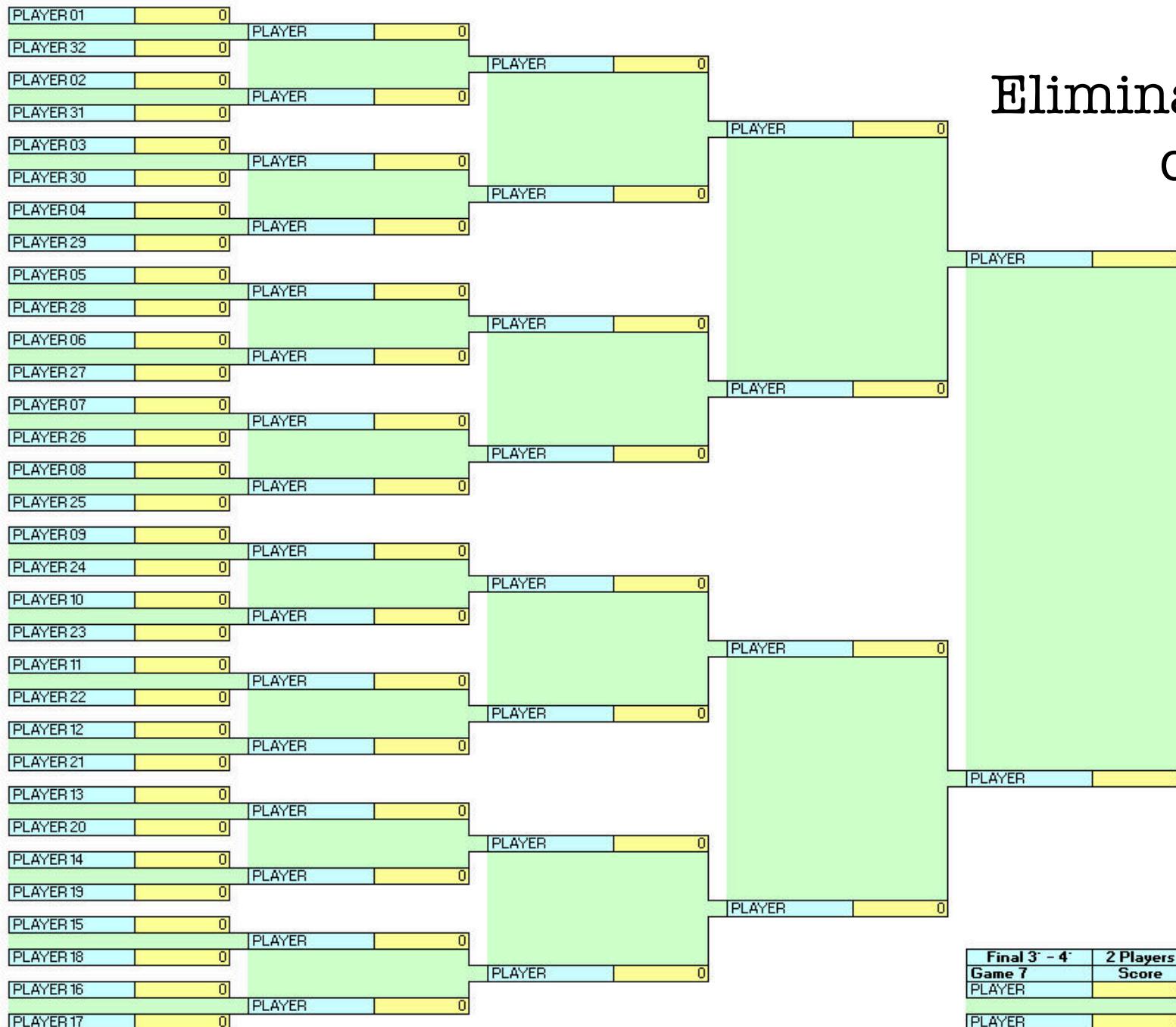
Gli Scacchi come sport

- Si gioca a Scacchi da circa 1500 anni: il gioco venne inventato in India, non si sa da chi
 - Il gioco come lo conosciamo oggi è stato codificato verso la fine del XV secolo in Italia
 - Le regole internazionali sono state promulgate nel 1929 dalla FIDE (*Fédération Internationale des Echecs*), che gestisce i tornei “ufficiali” e il Campionato del Mondo
 - Anche se gli Scacchi non sono uno sport olimpico, la FSI – Federazione Scacchistica Italiana – è associata al CONI, e ne segue le regole generali
-

Necessità dei metodi di ranking

- In molti sport si pone il problema di “mettere in fila” i giocatori
 - Esempio: definire il Campione del Mondo
 - Nel Calcio, occorre selezionare le squadre che arrivano alla finale del Campionato del Mondo
 - Nel Pugilato e negli Scacchi, occorre selezionare chi può sfidare il campione in carica
 - Nel Tennis, occorre stabilire chi può partecipare ad un torneo prestigioso senza passare dalle eliminatorie o senza “wild card”
-

Round 2	32 Players	Round 3	16 Players	Quarter Final	8 Players	Half Final	4 Players	Final	2 Players
Game 3	Score	Game 4	Score	Game 5	Score	Game 6	Score	Game 7	Score



PLAYER **Winner**

PLAYER **3rd**

ATP Ranking

Tournament category	W	F	SF (3rd/4th)	QF	R16	R32	R64	R128	Additional qualifying points	
Grand Slam	2000	1200	720	360	180	90	45	10	25	
ATP World Tour Finals	1500 [^] 1100 ^m	1000 [^] 600 ^m	600 [^] 200 ^m	(200 for each round robin match win, +400 for a semifinal win, +500 for the final win)						
Masters 1000	1000	600	360	180	90	45	10 (25)	(10)	25 (16)	
Olympics	750	450	340 (bronze) 270 (4th)	135	70	35	5			
500 Series	500	300	180	90	45	(20)			20 (10)	
250 Series	250	150	90	45	20	(5)			12 (5)	
ATP Challenger Tour Finals	125 [^] 95 ^m	75 [^] 45 ^m	45 [^] 15 ^m	(15 for each round robin match win, +30 for a semifinal win, +50 for the final win)						
Challenger 125,000 +H	125	75	45	25	10				5	
Challenger 125,000	110	65	40	20	9				5	
Challenger 100,000	100	60	35	18	8				5	
Challenger 75,000	90	55	33	17	8				5	
Challenger 50,000	80	48	29	15	7				5	
Challenger 35,000 +H	80	48	29	15	6				5	
Futures 15,000 +H	35	20	10	4	1					
Futures 15,000	27	15	8	3	1					
Futures 10,000	18	10	6	2	1					

22:00

Torneo all'italiana

SCHEMA GARA GIRONE ALL'ITALIANA															
COPPIE	0	1	2	3	4	5	6	7	8	9	10	11	12	PUNTI	CLAS.
Michele ILARIO	1	X	0	3	0	0	0	0	0					3	
MASSIMO ANTONIO	2	3	X	0	0	0	0	0	0						
FEDERICO SILVANO G	3	0	3	X	2	3	3	3	3					17	1
MAURO EMILIO	4	3	3	1	X	3	3	3	1					17	2
GIUSEPPE CHIARA	5	3	3	0	0	X	3	3	2					14	3
GIULIANO ANDREA	6	3	0	0	0	1	X	1	3					7	
FEDERICO FLAVIO	7	3	3	0	0	0	2	X	0					8	
SILVANO GIULIA	8	3	3	0	2	1	0	3	X					12	4
	9								X						
	10									X					
	11										X				
	12											X			

Problemi

- Eliminatoria diretta: un singolo torneo trova solo il più forte, non forma un ranking completo perché non ordina in sequenza i restanti giocatori
 - Round-robin: troppo lungo con tanti giocatori; inoltre occorre comunque gestire gli *ex.aequo*
-

Dal rating al ranking

- Per formare un *ranking* di un universo di giocatori si può usare una funzione di *rating del singolo giocatore* basata su una distribuzione di probabilità
- Ipotesi di Arpad Elo: la prestazione attesa (*rating*) di un giocatore è una variabile casuale che assume una forma normale (detta *gaussiana*, o *curva a campana*)



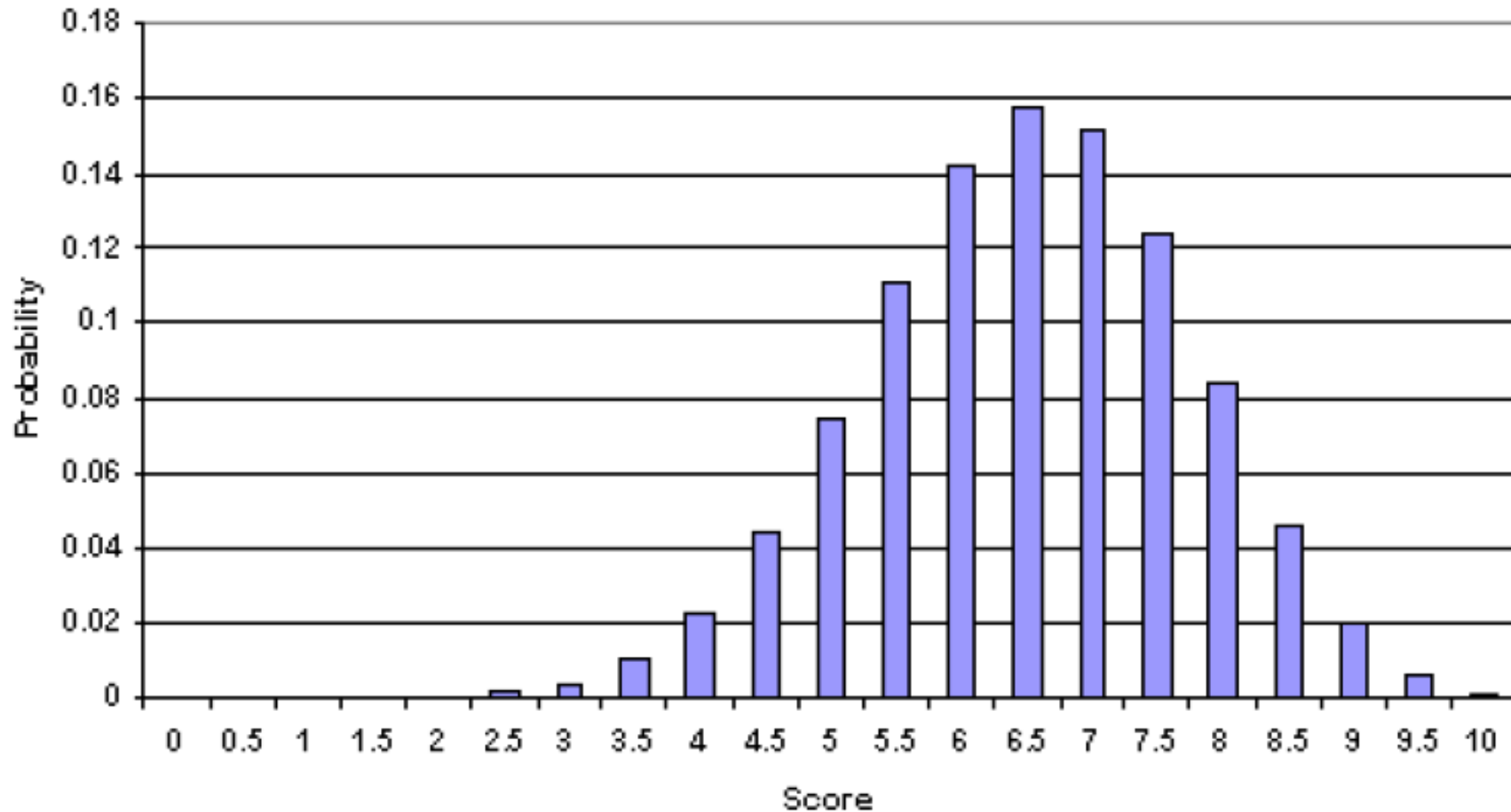
Arpad Elo 1903-1992

Prima assunzione: distribuzione normale

- Negli scacchi ogni vittoria (W) dà un punto, una sconfitta (L) zero punti, mentre la patta (D) divide il punto metà per uno
- per una serie di N partite, se conosciamo le probabilità di W, D, L possiamo usare la formula delle permutazioni con elementi ripetuti

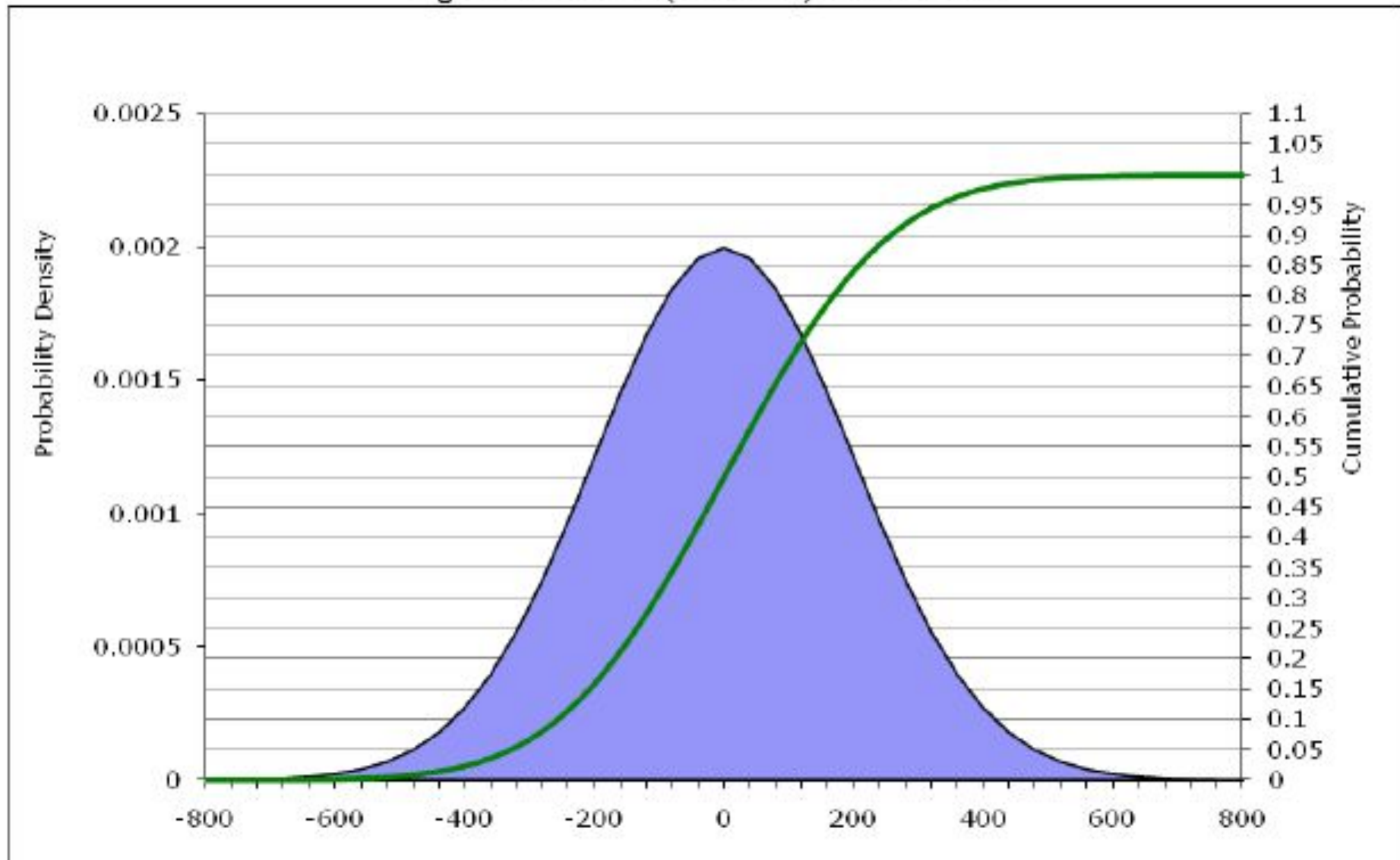
$$P(W, L, D) = \frac{N!}{W! \cdot L! \cdot D!} \cdot P(\text{win})^W \cdot P(\text{loss})^L \cdot P(\text{draw})^D$$

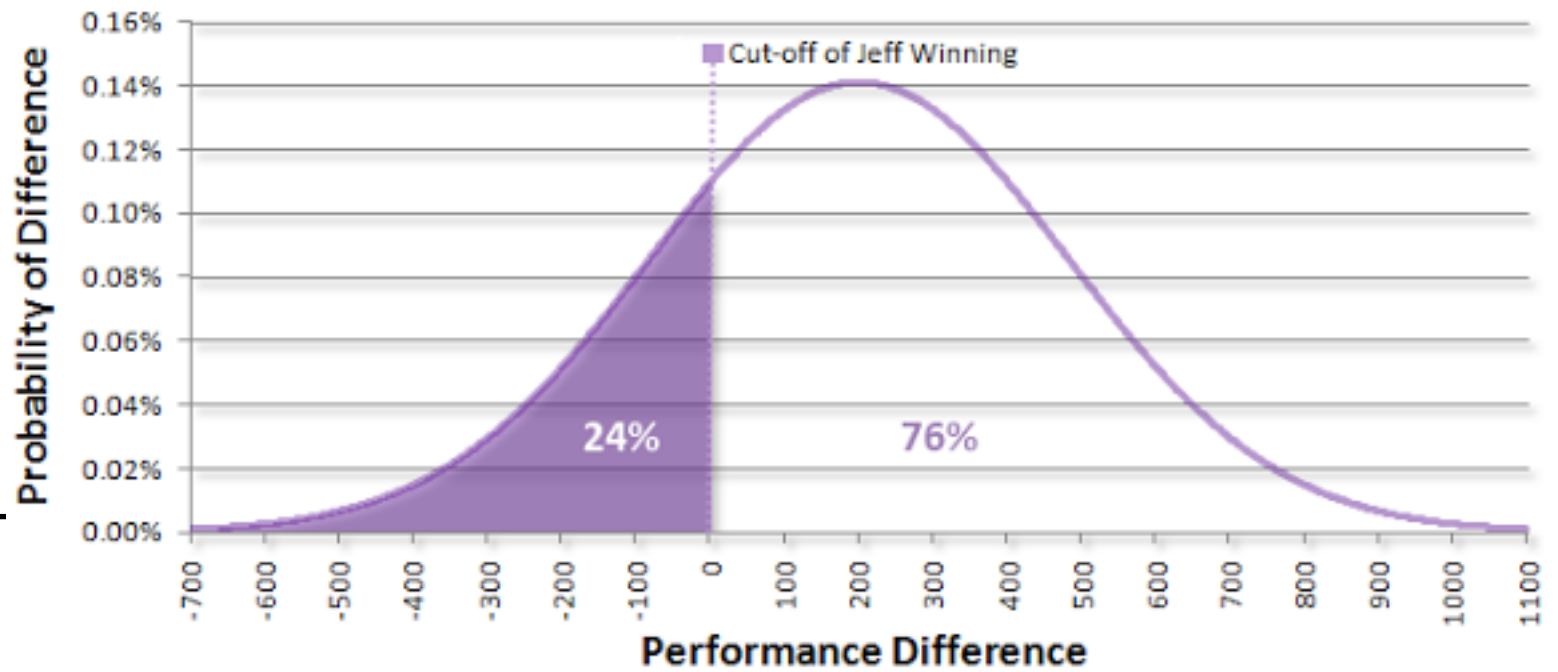
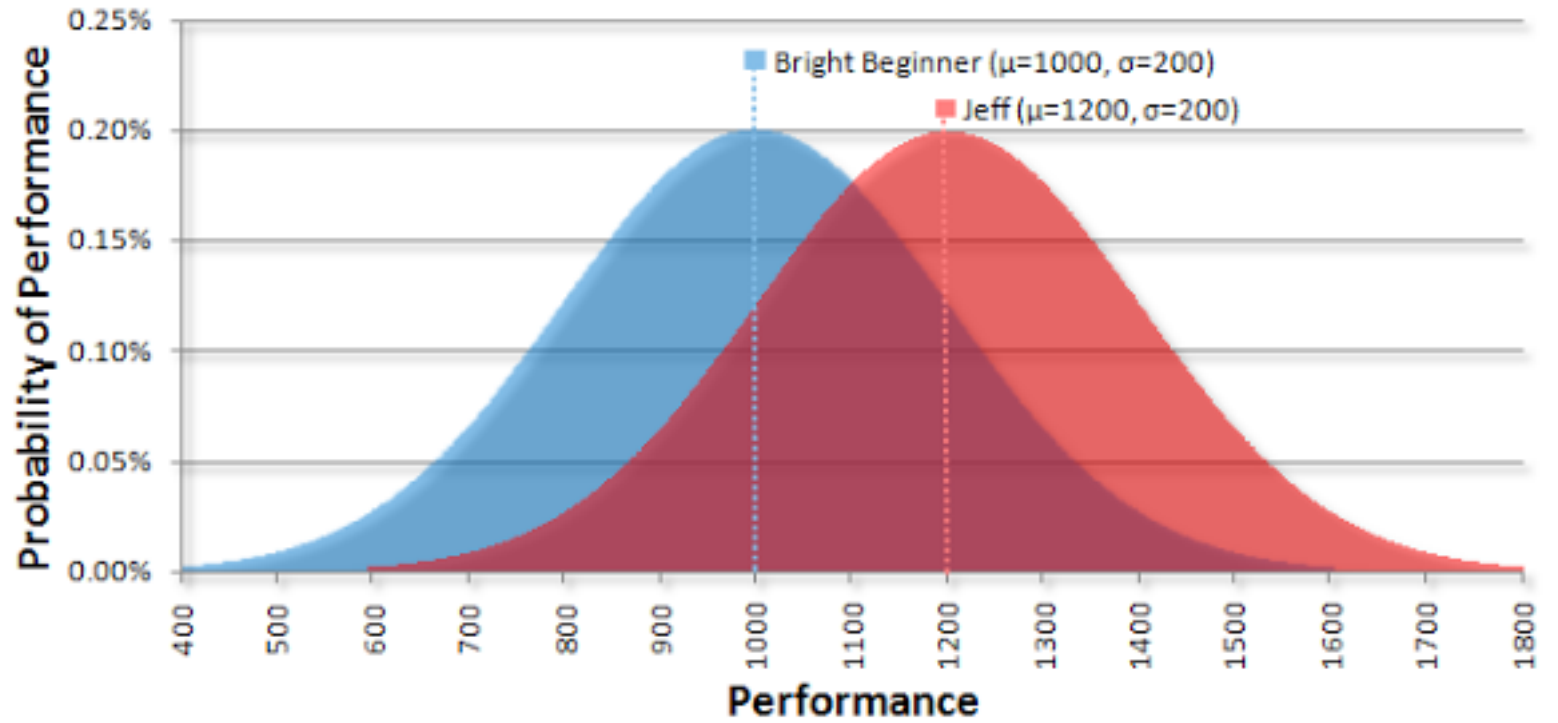
Distribuzione dei possibili risultati di un giocatore
con $N=10$, $P(\text{win})=0.5$, $P(\text{draw})=0.3$, $P(\text{loss})=0.2$



Seconda assunzione: la stima del risultato di una partita è approssimabile con la gaussiana

Figure 1. Normal (Gaussian) Distribution





Convenzioni nel sistema Elo per gli Scacchi

- Un principiante ottiene un punteggio iniziale di 1000
 - Questo numero è “adimensionale”, cioè ha senso solo in confronto con altri giocatori: la distribuzione normale è centrata a 1500
 - Una differenza di 200 punti tra due giocatori A e B (es. A ha 2000, B ha 1800) prevede una probabilità del 76% che A vinca su B
-

Table 1. Expected Winning Probabilities by Rating Difference (Elo, 1978)

Diff.	Higher	Lower	Diff.	Higher	Lower	Diff.	Higher	Lower
0-3	50	50	122-129	67	33	279-290	84	16
4-10	51	49	130-137	68	32	291-302	85	15
11-17	52	48	138-145	69	31	303-315	86	14
18-25	53	47	146-153	70	30	316-328	87	13
26-32	54	46	154-162	71	29	329-344	88	12
33-39	55	45	163-170	72	28	345-357	89	11
40-46	56	44	171-179	73	27	358-374	90	10
47-53	57	43	180-188	74	26	375-391	91	9
54-61	58	42	189-197	75	25	392-411	92	8
62-68	59	41	198-206	76	24	412-432	93	7
69-76	60	40	207-215	77	23	433-456	94	6
77-83	61	39	216-225	78	22	457-484	95	5
84-91	62	38	226-235	79	21	485-517	96	4
92-98	63	37	236-245	80	20	518-559	97	3
99-106	64	36	246-256	81	19	560-619	98	2
107-113	65	35	257-267	82	18	620-735	99	1
114-121	66	34	268-278	83	17	735+	100	0

Esempi

Fischer-Spassky 1972

Giocatore	Risultato finale	Percentuale	Differenza D_p
Fischer	12.5	62.5%	?
Spassky	7.5	37.5%	?

Karpov-Korchnoi 1974

Giocatore	Rating iniziale	Differenza	Percentuale P_D
Karpov	2715	70	?
Korchnoi	2645	-70	?

Esempi

Fischer-Spassky 1972

Giocatore	Risultato finale	Percentuale	Differenza D_p
Fischer	12.5	62.5%	90
Spassky	7.5	37.5%	-90

Karpov-Korchnoi 1974

Giocatore	Rating iniziale	Differenza	Percentuale P_D
Karpov	2715	70	60%
Korchnoi	2645	-70	40%

Risultato finale 12.5-11.5,
pari a 52% vs 48%

L'equazione del rating

$$R_n = R_i + K \times (W - W_e)$$

dove

- R_n è il nuovo rating,
 - R_i è il rating iniziale,
 - K è un parametro di variazione (alto per chi inizia, es. 45 - basso per gli esperti - es. 10)
 - W è il punteggio finale,
 - W_e è il punteggio atteso secondo la tabella 1
-

Esempio 1: singolo giocatore

Un giocatore con Elo 1830 partecipa ad un torneo a otto turni

Turno	Elo avversario	Differenza	%Attesa	Risultato
1	1824	+6	51	1
2	1898	-68	41	1
3	1830	0	50	0
4	1850	-20	47	0,5
5	1720	+110	65	0
6	1800	+30	54	1
7	1860	-30	46	1
8	1790	+40	56	0,5
			A=4,1	R=5

Nuovo rating: $1830 + K(R-A)$; se $K = 30$ si ha $1830 + (30 \cdot 0.9) = 1857$

Esempio 2: torneo e variazioni

Giocatore	R_i	R_m	P_e	W_o	P_o	ΔR	R_f
Anand	2792	2746	0.565	9	0.643	+7	2799
Kramnik	2769	2749.3	0.528	8	0.571	+6	2775
Morozevich	2758	2750.9	0.510	6	0.429	-11	2747
Leko	2751	2751.9	0.499	7	0.500	0	2751
Aronian	2750	2752	0.497	6	0.429	-10	2740
Svidler	2735	2754.1	0.473	6.5	0.464	-1	2734
Gelfand	2733	2754.4	0.470	8	0.571	+14	2747
Grischuk	2726	2755.4	0.459	5.5	0.393	-9	2717

Campionato del Mondo, Città del Messico, 2007

Applicazioni Elo di “demografia agonistica”

- È possibile profilare la carriera di un singolo giocatore plottando l'andamento temporale del suo Elo (che cambia solo se gioca partite ufficiali)
 - È possibile confrontare le carriere dei campioni anche di epoche diverse
 - È possibile definire le curve medie di progresso delle prestazioni dei vari tipi di giocatori
-

Carlsen, Magnus (NOR)



[Profile](#)

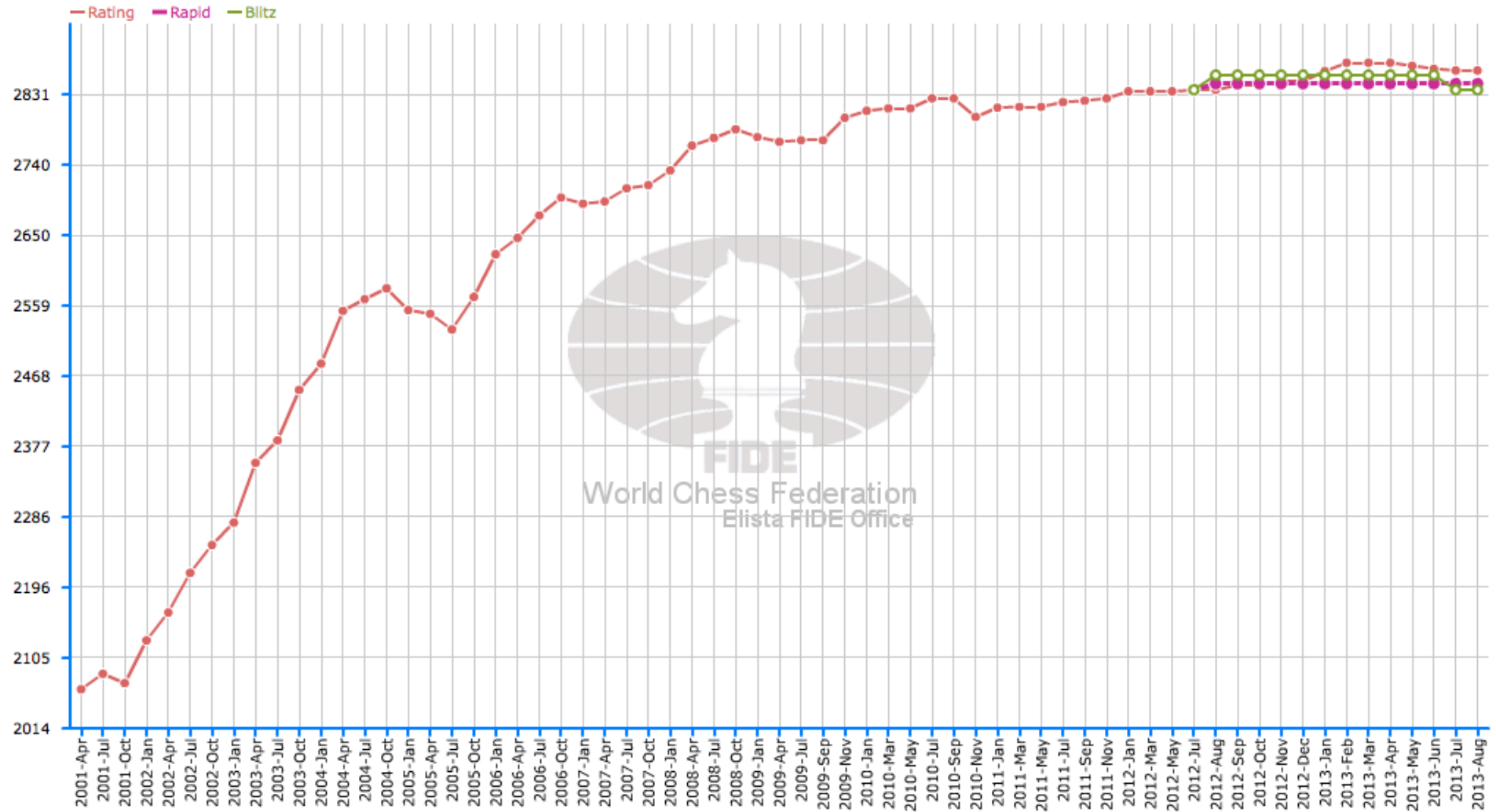


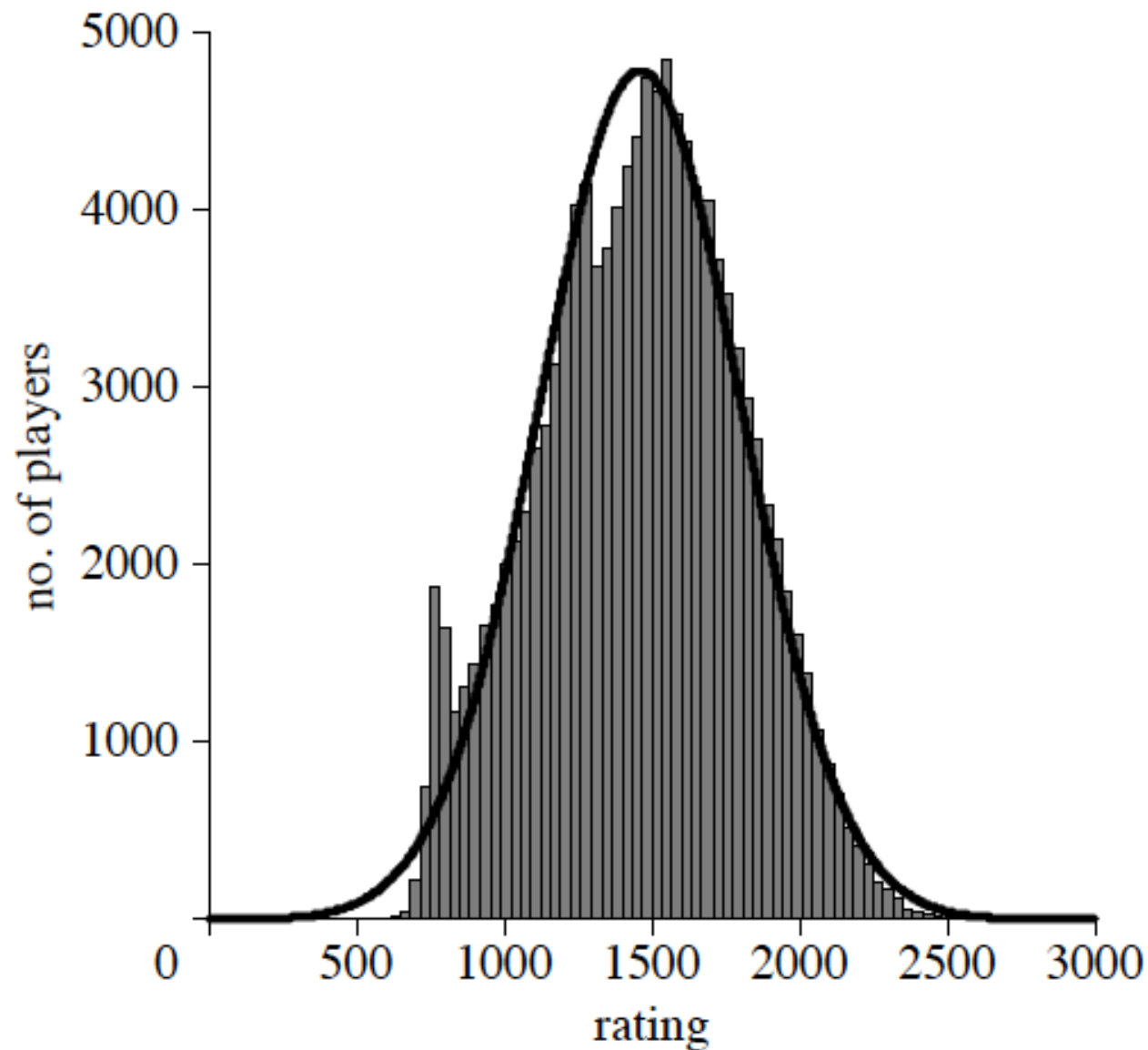
[Rating Progress](#)



[Game Statistics](#)

Rating Progress Chart

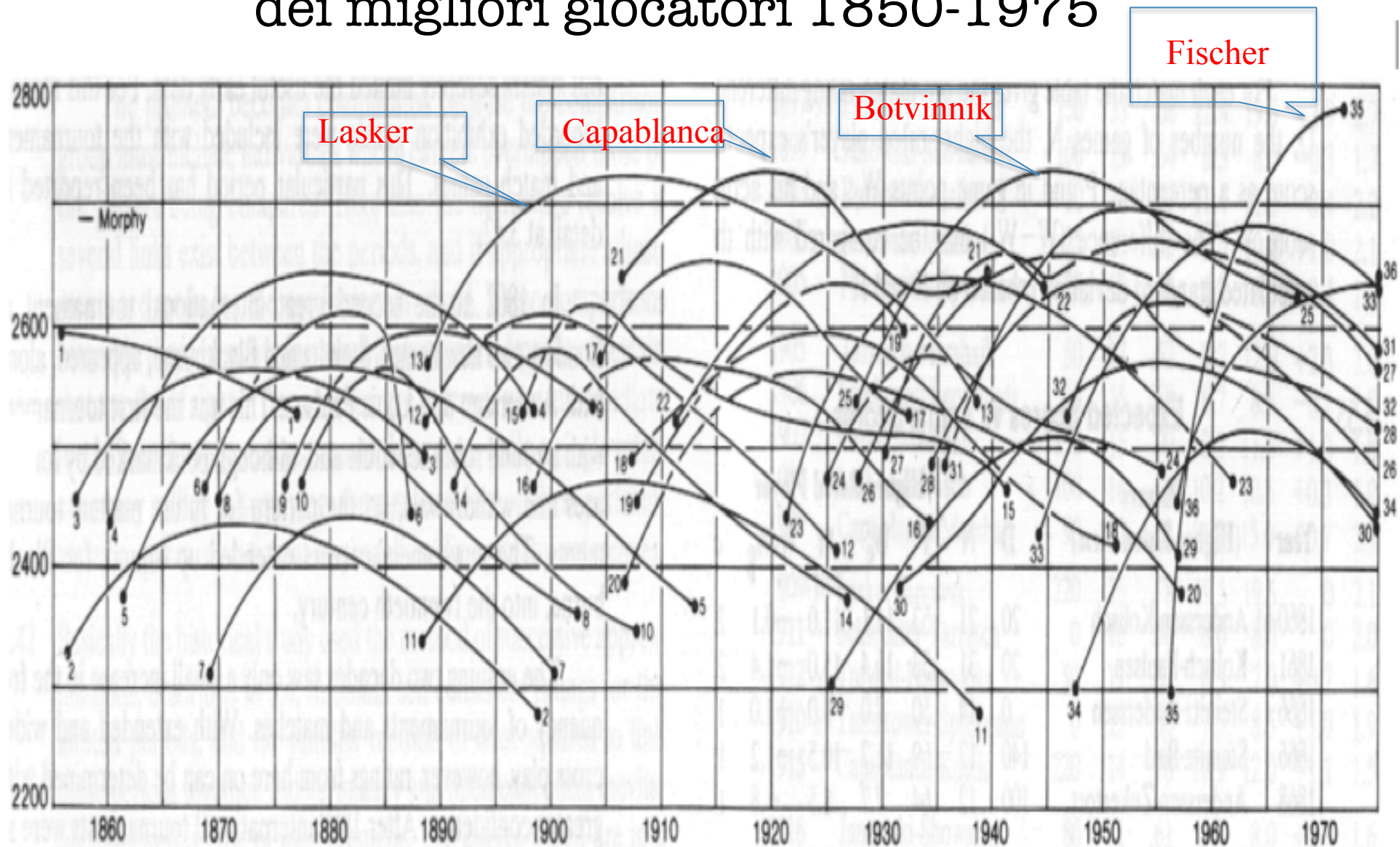




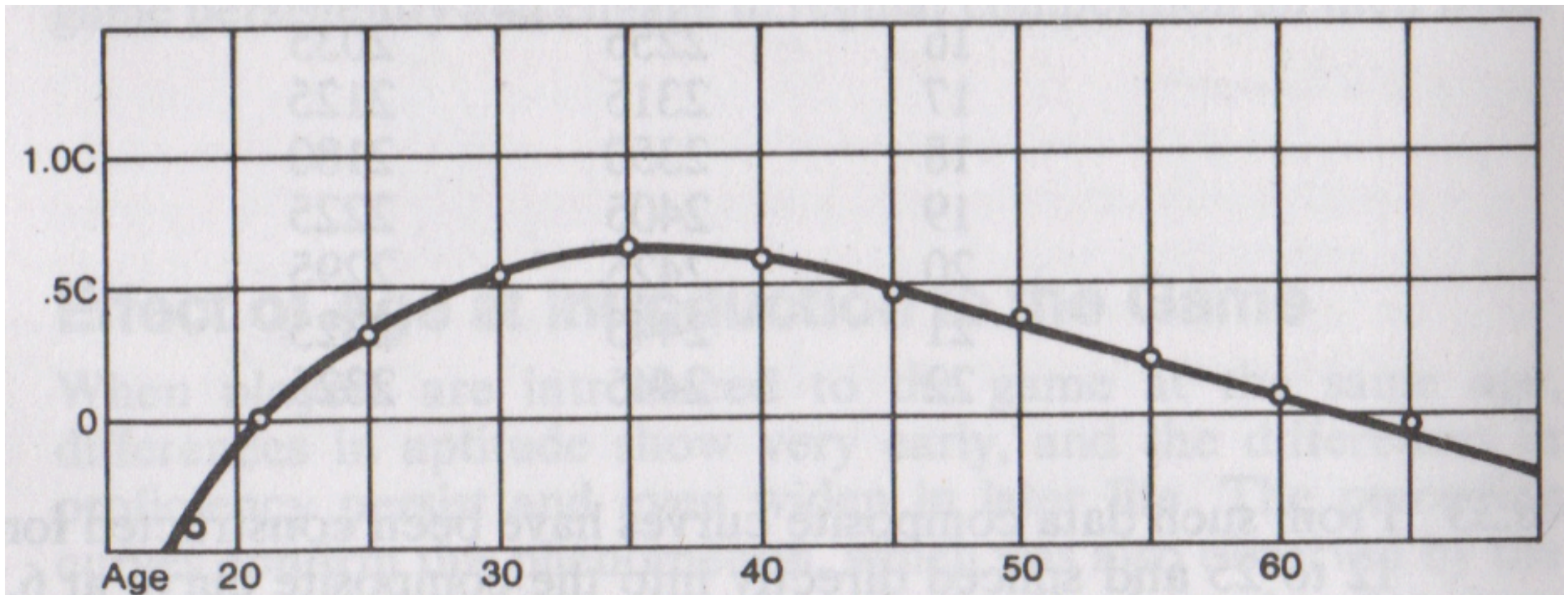
Fonte: Bilalic 2009, Why are (the best) women so good at chess? Participation rates and gender differences in intellectual domains

Figure 1. The distribution of the German chess rating with the best-fit normal curve superimposed. $n=120\,399$, $\mu=1461$, $\sigma=342$, 16 : 1 men to women ratio.

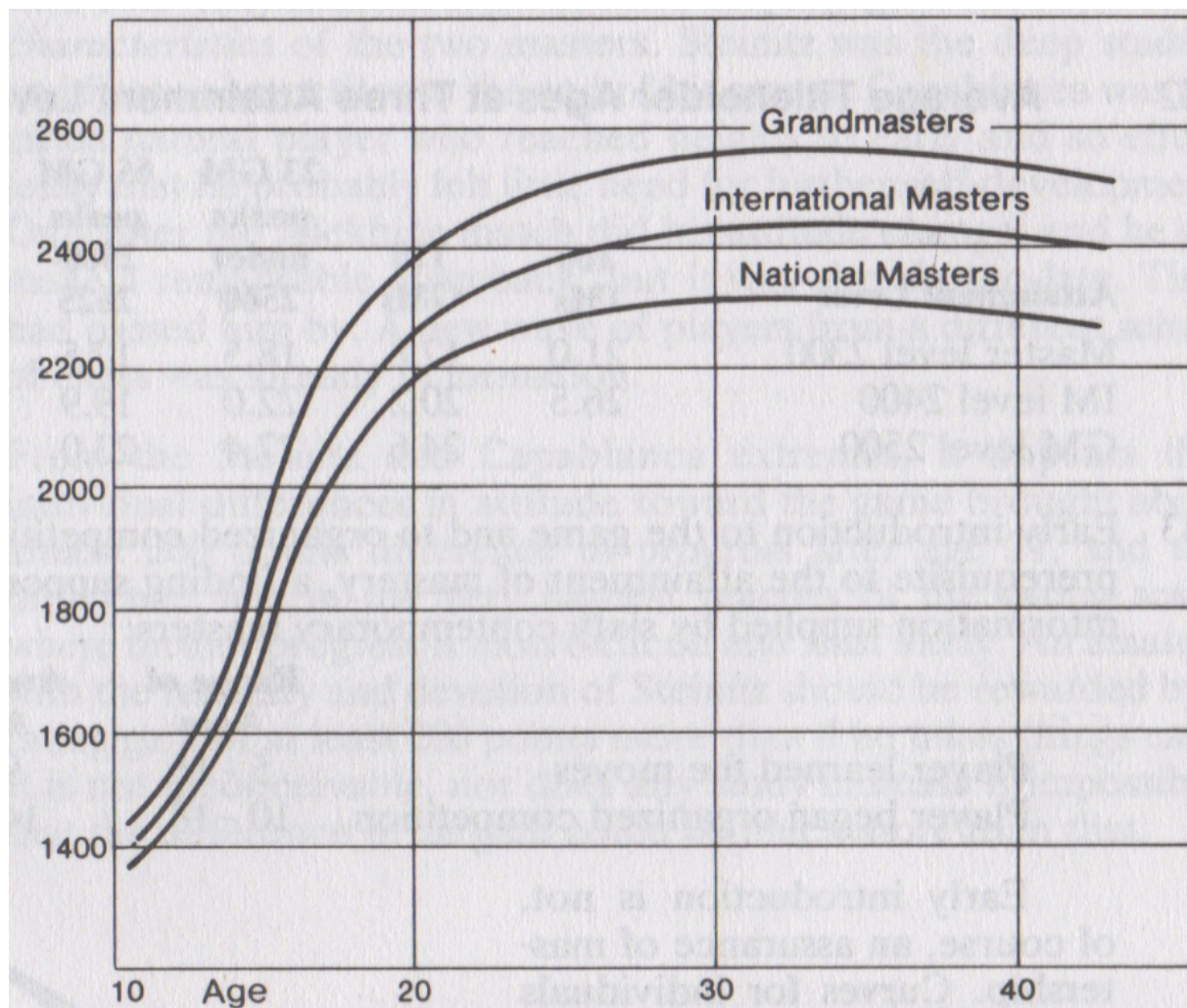
Le variazioni Elo della carriera dei migliori giocatori 1850-1975



Curva media dell'Elo durante la vita (composita, sui campioni)

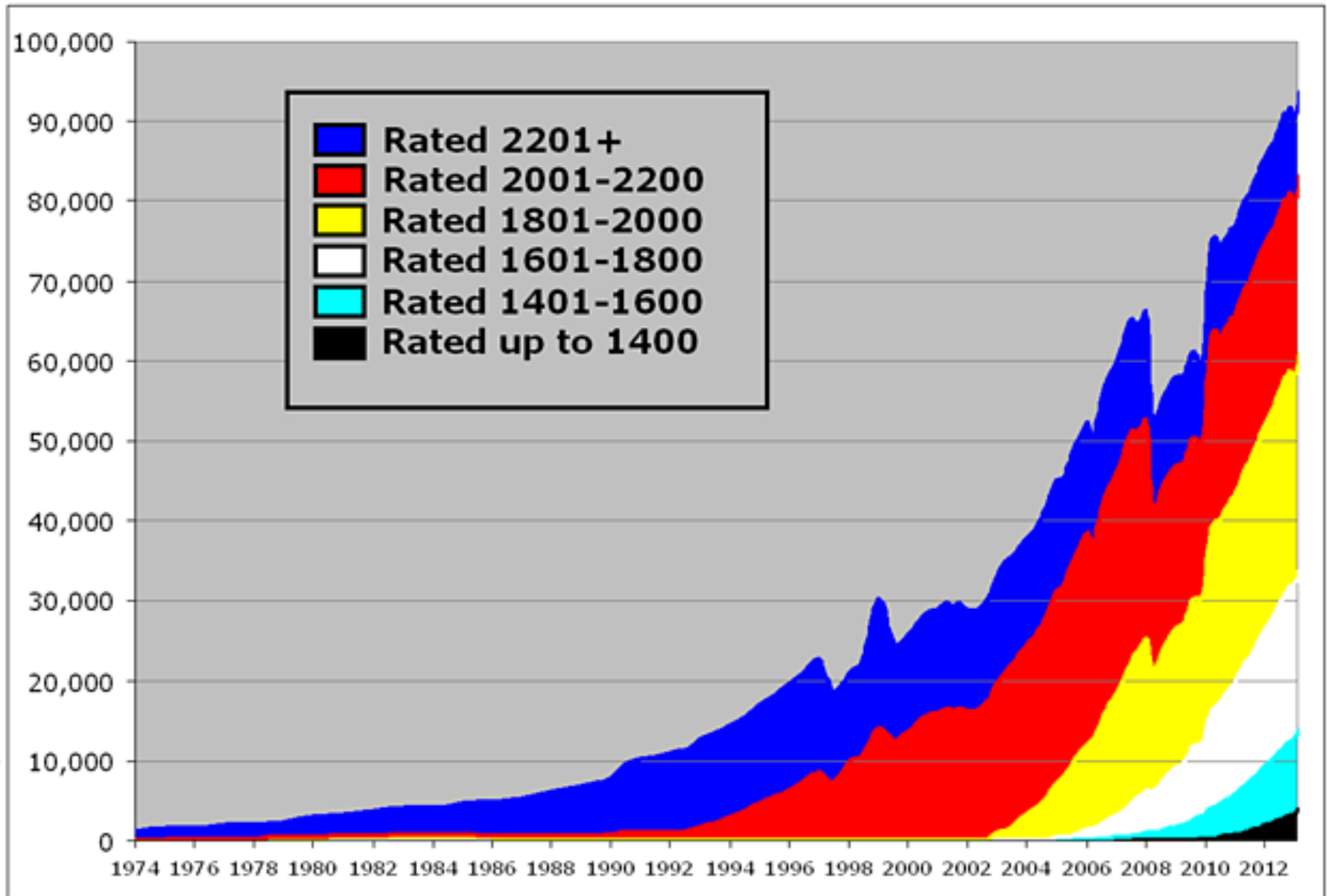


Curve di riferimento delle varie classi (solo giocatori che raggiungono 2300 o più)



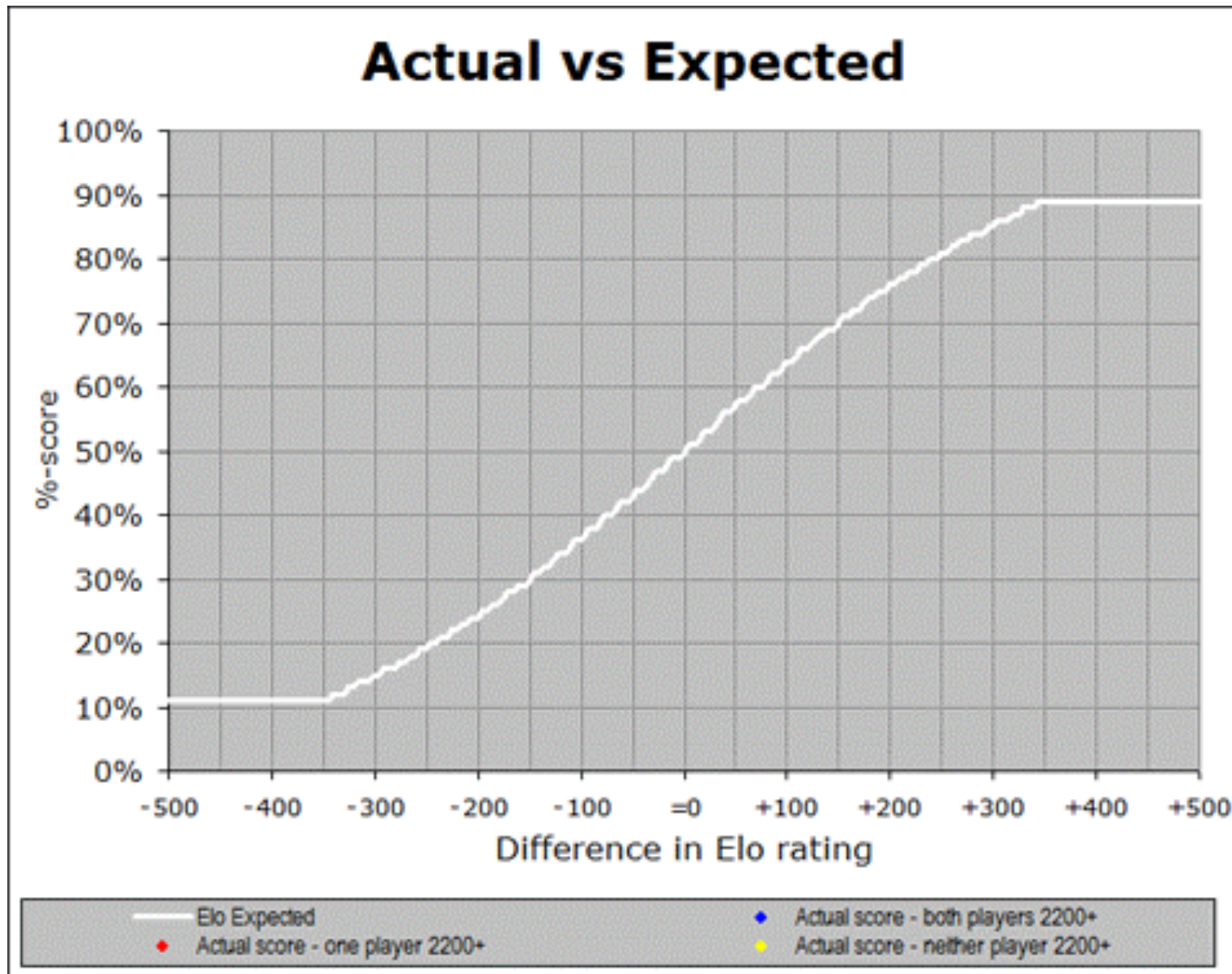
I giocatori sono partizionati in “classi”: es. i maestri hanno rating tra 2100 e 2300, i grandi maestri da 2500 in su

Distribuzione Elo FIDE negli anni



Critiche al sistema Elo

- Il sistema è tuttora usato dalla FIDE nella sua forma originale, applicando diversi valori di K per diverse classi di giocatori; il valore “giusto” di K è spesso soggetto a discussioni
 - La distribuzione normale è stata criticata a favore della distribuzione logistica
 - La simmetria di guadagni e perdite in una singola partita in molte situazioni risulta inadeguata
 - Il sistema è inflazionabile, e necessita di periodiche revisioni
 - Non è chiaro quanto il sistema sia estendibile alla valutazione di squadre di giocatori
-



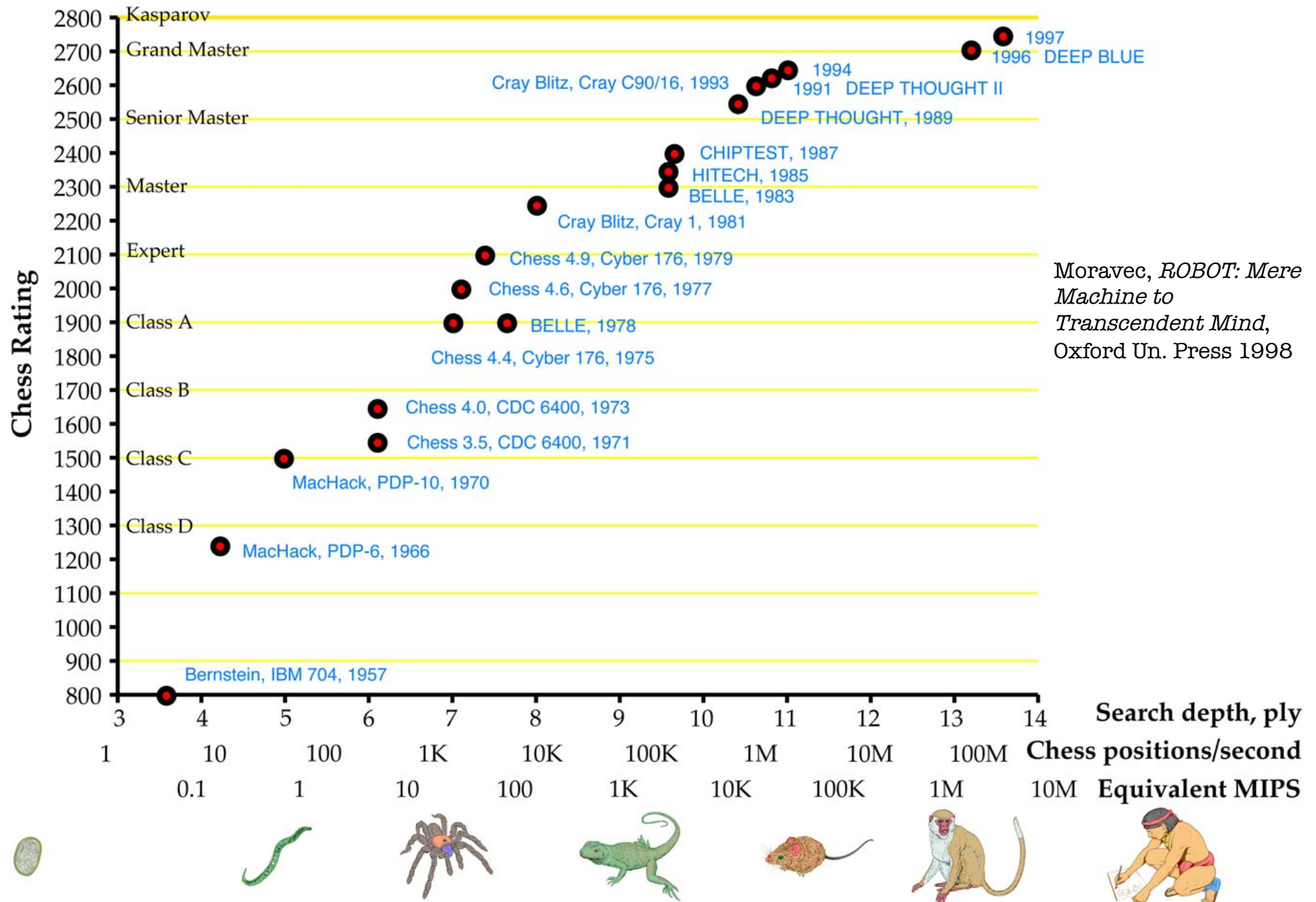
Autore: Jeff Sonas

en.chessbase.com/Home/TabId/211/PostId/4010815/sonas-overall-review-of-the-fide-rating-system-220813.aspx

Applicazioni: Intelligenza Artificiale

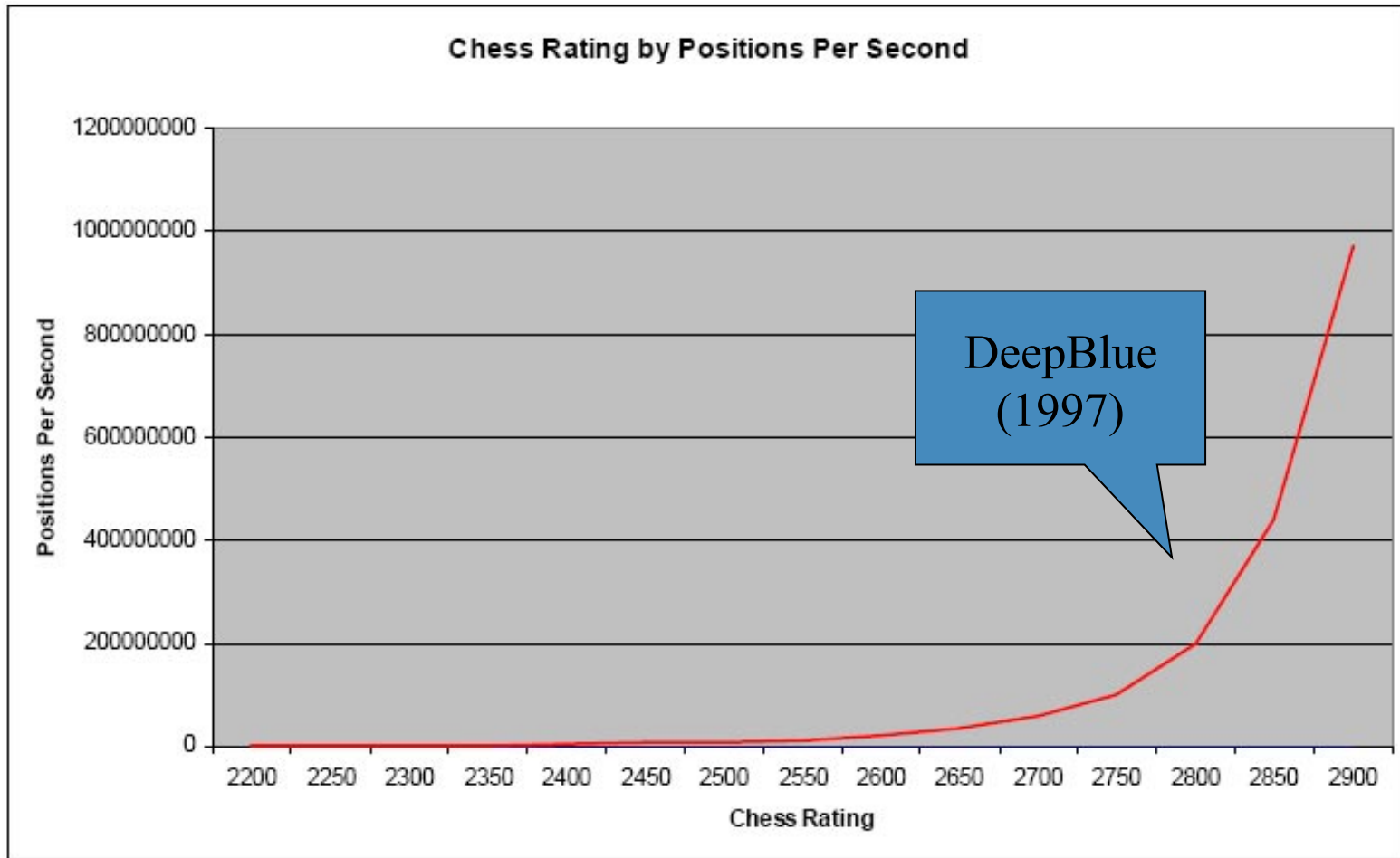
- Negli anni '70 il rating Elo venne applicato alla valutazione dei programmi per giocare a Scacchi
 - E' stato ed è un formidabile misuratore del progresso della ricerca in Intelligenza Artificiale e della tecnologia hardware
-

Chess Machine Performance versus Processing Power



Moravec, *ROBOT: Mere Machine to Transcendent Mind*, Oxford Un. Press 1998

Il progresso della tecnologia



Un rating Elo dei programmi attuali

Color legend: **Commercial**, **Free**, **Open source**, **Private**.

Bold font - tested with 200 games or more. Normal font - less than 200 games.

Rank	Name	Rating			Score	Average Opponent	Draws	Games
		Elo	+	-				
1	Houdini 3 64-bit 4CPU	3248	+16	-16	69.8%	-128.1	44.0%	1359
2	Komodo 5.1 MP 64-bit 4CPU	3207	+25	-25	63.2%	-79.5	54.5%	473
3	Stockfish 3 64-bit 4CPU	3176	+23	-23	56.5%	-38.1	58.3%	516
4	Critter 1.6a 64-bit 4CPU	3173	+14	-14	58.8%	-53.9	58.0%	1594
5	Rybka 4 64-bit 4CPU	3164	+14	-14	61.0%	-71.3	48.7%	1609
6	Bouquet 1.5 64-bit 4CPU	3144	+27	-27	47.9%	+12.3	63.0%	378
7	Strelka 5.5 64-bit	3115	+12	-12	56.8%	-41.6	56.4%	2174
8	Naum 4.2 64-bit 4CPU	3079	+12	-12	48.8%	+5.8	49.9%	2309
9	Hiarcs 14 4CPU	3075	+18	-18	43.1%	+41.3	51.3%	876
10	Chiron 1.1a 64-bit 4CPU	3071	+24	-24	43.0%	+43.3	49.8%	508
11	Protector 1.5.0 64-bit 4CPU	3070	+32	-32	45.9%	+23.8	46.0%	302
12	Gull R375 64-bit 4CPU	3053	+23	-23	41.8%	+45.8	56.7%	543
13-14	Deep Fritz 13 4CPU	3048	+16	-16	40.8%	+56.0	49.9%	1161
13-14	Hannibal 1.3 64-bit 4CPU	3048	+23	-23	41.7%	+51.4	48.6%	587
15	Deep Junior 13 64-bit 4CPU	3041	+17	-17	42.3%	+48.2	49.7%	1082

Un altro rating

		Rating	+	-	Games Won	Av.opp
1	Deep Rybka 4 x64 2GB Q6600 2,4 GHz	3212	27	-25	930	75% 3022
2	Deep Hiarcs 14 2GB Q6600 2,4 GHz	3205	29	-27	716	71% 3049
3	Deep Rybka 3 x64 2GB Q6600 2,4 GHz	3202	24	-22	1251	77% 2991
4	Naum 4.2 MP x64 2GB Q6600 2,4 GHz	3152	24	-23	919	64% 3050
5	Naum 4 x64 2GB Q6600 2,4 GHz	3137	25	-23	1030	74% 2960
6	Deep Junior 13.3 2GB x64 Q6600 2,4 GHz	3123	25	-25	760	57% 3074
7	Deep Shredder 12 x64 2GB Q6600 2,4 GHz	3111	20	-19	1381	65% 3003
8	Spike 1.4 MP 2GB Q6600 2,4 GHz	3106	20	-19	1320	64% 3009
9	Deep Hiarcs 13.2 2GB Q6600 2,4 GHz	3106	28	-27	632	60% 3034
10	Hiarcs 13.1 2GB Q6600 2,4 GHz	3104	25	-24	788	60% 3037
11	Deep Fritz 13 2GB Q6600 2,4 GHz	3095	28	-27	624	58% 3037
12	Deep Fritz 12 2GB Q6600 2,4 GHz	3094	21	-21	1040	56% 3050
13	Deep Rybka 3 256MB Athlon 1200 MHz	3078	39	-37	332	58% 3022
14	Deep Junior 12 x64 2GB Q6600 2,4 GHz	3074	24	-23	938	63% 2979
15	Deep Fritz 11 2GB Q6600 2,4 GHz	3066	19	-18	1424	63% 2977

<http://ssdf.bosjo.net/list.htm>, 23.3.2013

I programmi Campioni del Mondo

• Kaissa		1974	Stoccolma
• Chess		1977	Toronto
• Belle		1980	Linz
• Cray Blitz		1983	New York
• Cray Blitz		1986	Colonia
• Deep Thought		1989	Edmonton
• Rebel		1992	Madrid
• Fritz		1995	Hong Kong
• Shredder		1999	Paderborn
• Junior		2002	Maastricht
• Shredder		2003	Graz
• Junior		2004	TelAviv
• Zappa		2005	Reykjavik
• Junior		2006	Torino
• Rybka		2007-2010	squalificato (in quanto clone di Fruit)
• Junior		2011	Tilburg
• Junior		2013	Yokohama



Il Trofeo Shannon, che va all'autore del programma Campione del Mondo

Applicazioni: altri sport

- College football
- Baseball
- Calcio <http://www.eloratings.net>
- Rugby <http://www.rugbyleagueratings.com>
- Competizioni di programmazione

In genere i relativi rating si usano per calcolare le quote di scommesse

Applicazioni: TrueSkill[®]

- TrueSkill è un algoritmo di ranking brevettato da Microsoft per matchmaking su XBOX in giochi multiutente
- Si basa sul sistema di rating Gliko, un'evoluzione del sistema Elo
- Il sistema definisce un giudizio su ciascun giocatore (rappresentato da una distribuzione normale), e lo modifica dopo ciascun match

Altre applicazioni

- In crowdsourcing, per determinare le fotografie di miglior qualità
- In algoritmi biometrici, per determinare l'identità
- In ricerche etologiche, per determinare le gerarchie di maschi dominanti
- Il film “The social network” mostra Zuckerberg che usa una variante del sistema Elo per Facemash, predecessore di Facebook

Conclusioni

- Il sistema Elo è stato inventato per dare un rating ai giocatori e definire un sistema “oggettivo” per assegnare titoli agonistici
 - È stato usato per valutare il progresso della tecnologia delle macchine che giocano
 - È oggi usato in moltissimi sport, per esempio per valutare le quote di scommesse su match
 - È molto studiato e criticato, ed esistono molte proposte migliorative; tuttavia la sua base resiste alla prova del tempo e di diversi domini
-

Riferimenti

- Elo, *The Rating of Chessplayers*, ARCO, 1978 e 1986
- Ciancarini, *Giocatori Artificiali*, Mursia, 1992



Siti utili

- www.fide.com
 - www.chessmetrics.com
 - www.statsilk.com/maps/elo-football-ranking-interactive-world-map
 - www.eloratings.net
 - www.moserware.com/2010/03/computing-your-skill.html
 - www.wizards.com/magic/magazine/Article.aspx?x=mtg/daily/twtw/159
-



Grazie per l'attenzione! Domande?

Paolo Ciancarini
Università di Bologna

Evento metrico GUFPI-ISMA
Università di Bologna
12 settembre 2013

