Il sistema Elo

R. Carosi
Pisa

Introduzione

Il sistema Elo è un sistema di classificazione elaborato dal Prof. Arpad Elo, fisico americano di origine ungherese. La scala dei valori va da 0 punti a circa 3000. Il più forte giocatore del mondo, Garry Kasparov, ha 2851 punti; un maestro nazionale è intorno ai 2200; una prima nazionale è circa 1700; e così via.

Il funzionamento, almeno in linea di principio, non è complicato. Contrariamente al sistema di punteggio usato per le classifiche dei tornei, in cui la vittoria vale 1 punto, la patta 1/2 punto e la sconfitta 0 punti, ogni partita non ha lo stesso valore. Il numero dei punti che si perdono o si guadagnano dipende dalla forza dell' avversario, o meglio dalla differenza di forza tra il giocatore e il suo avversario Una vittoria contro un avversario più forte vale più punti che contro un avversario più debole, e naturalmente perdendo contro un avversario più debole si perdono più punti che perdendo contro un avversario più forte.

Il numero di punti che si guadagnano o si perdono in una partita si basa sul confronto tra il risulato ottenuto e il "risultato atteso", quello che cioè ci si aspetta in base alla differenza tra i punti del giocatore e quelli dell' avversario. Se il risulato è migliore di quello atteso, allora il giocatore guadagna punti, mentre se il risultato è peggiore perde punti. Il risultato atteso viene espresso in termini di punti percentuali (infatti si parla di percentuale attesa).

La ricetta

Per illustrare il funzionamento del sistema Elo facciamo subito un esempio pratico.

Un giocatore ha Elo pari a 1600 punti. In un torneo a 5 turni vince al primo turno contro un giocatore, A, che ha 1300 punti; al secondo turno vince ancora contro B che ha 1650 punti; al terzo turno perde contro C che ha 1600 punti; al quarto turno vince contro D che ha 1500 punti; al quinto ed ultimo turno patta contro E che ha 1700 punti. Il nostro giocatore ha realizzato 3.5/5. Vediamo ora come si calcola la sua variazione Elo. Secondo la ricetta dobbiamo calcolare, partita per partita, le differenze tra l' Elo del giocatore in questione e gli aversari:

- Prima partita: 1600 - 1300 = 300
- Seconda partita: 1600 - 1650 = -50
- Terza partita: 1600 - 1600 = 0
- Quarta partita: 1600 - 1500 = 100
- Quinta partita: 1600 - 1700 = -100

Quando la differenza è positiva vuol dire che l' avversario è (almeno sulla carta) più debole, quando è negativa vuo dire che l' avversario è più forte e quando è nulla vuol dire che la forza è pari.

A questo punto dobbiamo usare una apposita tabella, la tabella delle percentuali attese in funzione della differenza punti. Prendiamo la prima partita: una differenza di 300 punti corrisponde ad una percentuale attesa dell' 85% a favore del più forte (15% nel caso in cui la differenza sia negativa). Il significato della percentuale attesa è intuitivo: alla lunga, contro avversari che hanno 300 punti Elo meno di noi, ci si aspetta di realizzare l' 85% dei punti partita. In una singola partita la percentuale attesa e il numero di punti attesi coincidono, quindi scriviamo 0.85, anche se sappiamo che in pratica possiamo realizzare 1 (vittoria), 0.5 (patta) o 0 (sconfitta).

Ecco quindi le percentuali attese per ogni partita:

- Prima partita: 0.85
- Seconda partita: 0.43
- Terza partita: 0.50
- Quarta partita: 0.64
- Quinta partita: 0.36

Il passo successivo è quello di passare dalle percentuali attese ai punti attesi. Questo si ottiene semplicemente facendo la somma: i punti attesi sono 0.85 + 0.43 + 0.50 + 0.64 + 0.36 = 2.78, che viene arrotondato a 2.8. I punti realizzati in pratica, 3.5, sono superiori ai punti attesi, e ciò di dice che il risultato del torneo è stato buono e che quindi il giocatore in questione dovrebbe guadagnare punti.

A questo punto si fa l' aggiornamento vero e proprio del punteggio Elo. Si prende la differenza tra punti fatti e punti attesi: 3.5 - 2.8 = 0.7. Questa differenza, che in questo caso è positiva, viene moltiplicata per una quantità, il "fattore K", che per le categorie al di sotto di quelle magistrali, è uguale a 30: $0.7 \times 30 = 21$. Questa è la variazione Elo: il nuovo Elo sarà quindi 1600 + 21 = 1621.

Poiché l' aggiornamento dell' Elo viene fatto in Italia ogni sei mesi, questo calcolo viene ripetuto per le partite di ogni torneo del semestre, e la variazione totale sarà la somma delle variazioni dovute ad ogni singolo torneo.

Questo è, in breve il funzionamento pratico del sistema Elo. Nel seguito di questo articolo cercheremo di chiarire da dove vengono le tabelle delle percentuali attese, il fattore K, e tutto il resto.

MINI-INTRODUZIONE AL SISTEMA ELO

Il sistema Elo si propone di valutare numericamente la ”forza” dei giocatori. Qui vediamo semplicemente di dare delle semplici giustificazioni teoriche alle tabelle e alle formule che si usano di solito, senza troppo perderci nei dettagli della teoria (per chi vuole approfondire c'è comunque una bibliografia piuttosto vasta). I ragionamenti che seguiremo non sono necessariamente quelli seguiti a suo tempo da Arpad Elo, ma vengono proposti semplicemente a scopo didattico.

Un modo diretto per quantificare la forza è quello di usare la probabilità di vittoria (eventualmente stimata mediante i risultati), che in pratica è la percentuale attesa che abbiamo introdotto precedentemente. Dire, per esempio, che il giocatore A è più forte del giocatore B, significa che la probabilità di vittoria di A contro B è maggiore del 50%.

Inoltre, se per esempio due giocatori giocano tra di loro un match di 10 partite col risultato di 8-2 in favore di uno dei due, possiamo dare una valutazione approssimativa della probabilità di vittoria del più forte, che risulterebbe dell' 80%. Analogamente quella del più debole sarebbe del 20%.

Nella discussione che segue tralasceremo la possibilità della partita patta per semplicità. L' introduzione della patta complica molto le cose, in maniera tale che dal punto di vista teorico la questione è ancora aperta. Da un punto di vista pratico ció è poco importante percheé quello che interessa è il risultato atteso, non la probabilità di ottenere un dato risultato. La partita patta può essere inserita nella discussione considerandola come ”metà vittoria e metà sconfitta”. Due patte, cioè, equivalgono a tutti gli effetti, ad una vittoria e una sconfitta. In una singola partita che finisce patta si può ”interpolare” tra vittoria e sconfitta. Vedremo per esempio nell' analisi del sistema Glicko che la probabilità di patta, data la probabilità di vittoria v, può essere considerata come $\sqrt{v(1-v)}$.

Intuitivamente possiamo ritenere valida la proprietà transitiva delle probabilità di vittoria: se A è più forte di B, e B è più forte di C, ci asettiamo che anche A sia più forte di C. Inoltre, se la probabilità di vittoria del giocatore A contro il giocatore C è 75%, mentre quella del giocatore B contro il giocatore C è del 60%, allora ci aspettiamo che A sia più forte di B.

Se però vogliamo conoscere esattamente la probabilità di vittoria di A contro B, non abbiamo nessuna regola che ci dice come vanno combinate le varie probabilità.

Un altro problema è che la probabilità di vittoria ha una scala di valori molto scomoda da usare, in quanto presenta ”appiattimenti” al 100% e a 0. Per esempio, un Grande Maestro e un giocatore di 1^a categoria nazionale hanno tutti e due probabilità di vittoria praticamente del 100% rispetto ad un principiante, ma la loro forza relativa è tutt' altro che simile.

C'è da dire che esiste un sistema, il sistema Ingo, in cui le differenze tra i punteggi danno direttamente le probabilità di vittoria (a meno del 50%). Se ad esempio due giocatori hanno rispettivamente 100 e 120 punti Ingo, vuol dire che la probabilità di vittoria è 50+(120-100) = 70% a favore del primo (nel sistema Ingo più il punteggio è basso, più un giocatore è forte).

Quindi occorrono due cose: un cambiamento di scala e un meccanismo per combinare le probabilità. Cambiare scala significa mettere in relazione la probabilità di vittoria con un rating che da' una misura la forza. Inoltre facciamo l' ipotesi che la probabilità di vittoria di un giocatore contro un altro dipenda esclusivamente dalla differenza dei loro ratings. Nel caso di due giocatori A e B:

- probabilità di vittoria per A: $p_A = f(a-b)$
- probabilità di vittoria per B: $p_B = g(a-b) = 1 - f(a-b)$

dove a e b sono i ratings, e f e g due funzioni.

Le funzioni f e g devono avere certe caratteristiche, per esempio, $f+g=1$; e $g(a-b) = f(b-a)$. Inoltre, per $a = b, p_A = p_B = 0.5$, mentre per $a-b$ tendente a $\pm\infty$ f deve tendere a 1 o a 0 (analogamente g deve tendere a 0 o a 1). Di funzioni con queste proprietà ne esistono una infinità.

Per arrivare alla relazione tra probabilità di vittoria e rating, vediamo in dettaglio il meccanismo di una singola partita. Il risultato di una singola partita è di carattere probabilistico. Se un giocatore A è più forte di un giocatore B, alle lunghe vincerà una frazione di partite contro B data dalla probabilità di vittoria. In ogni singola partita, però, anche B può vincere. Questo significa che le abilità dei due giocatori, a e b, non sono costanti, ma variano da partita a partita. Se in una partita si ha che $a > b$, allora vince A. Se invece $a < b$, allora vince B. Se A è più forte di B, allora il valore medio di a sarà maggiore dle valore medio di b e ciò si rifletterà sul risultato. Quindi, in una partita, non è tanto l' abilità media di un giocatore che conta, ma l' abilità ”nascosta”, quella che viene effettivamente usata nella partita e che fluttua in modo casuale attorno alla media. È ovvio che le fluttuazioni dell' abilità lontane dal valor medio (cioè il rating del giocatore) sono poco probabili, mentre quelle vicine sono più probabili. Quindi, se un giocatore è molto più forte di un altro, la probabilità che perda è comunque molto piccola.

Arpad Elo ipotizzò che l' abilità di un giocatore seguisse un andamento gaussiano (modello di Thurstone-Mosteller), con media uguale al suo rating e una deviazione standard uguale a 200 circa e costante per tutti. Quindi, se un giocatore è classificato a 1700 punti Elo, significa che in media giocherà con un'abilità corrispondente a 1700, ma partita per partita questo valore cambierà. Una volta giocherà come un 1650, un' altra volta come un 1710, e così via. La distribuzione di questi valori nelle varie partite è gaussiana, con centro a 1700.

Il fatto che la deviazione standard sia stata ipotizzata uguale per tutti è una cruda approssimazione. In realtà ogni giocatore dovrebbe avere la sua deviazione standard: certi giocatori hanno un rendimento più uniforme e giocano quasi sempre al loro livello; altri invece hanno rendimenti più incostanti. Per un giocatore da 1800 Elo perdere contro uno da 2200 e poi vincere contro uno da 1400 non è la stessa cosa che vincere contro uno da 2200 e poi perdere contro uno da 1400. In media

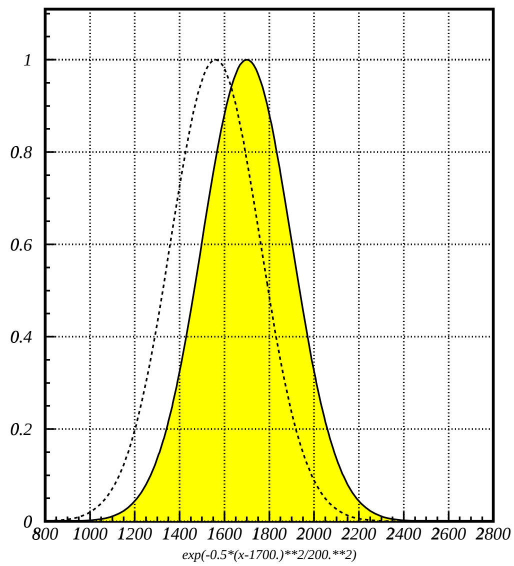

Figure 1: Due gaussiane centrate a 1700 e a 1560

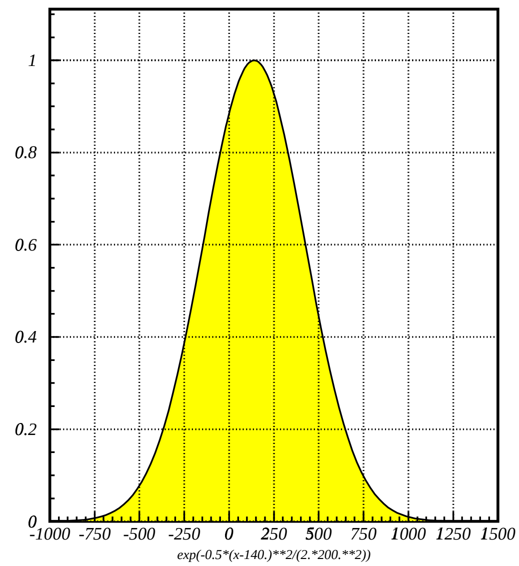

Figure 2: Gaussiana centrata a 140 (=1700-1560)

è la stessa cosa, ma per la deviazione standard è molto diverso. La giustificazione a questa ipotesi, della deviazione standard uguale per tutti, è esclusivamente numerica, e non di principio: anche ipotizzando deviazioni standard diverse tra di loro, anche di un fattore due, si può vedere che il risultato finale (la probabilità di vittoria) cambia di poco, al massimo di qualche %, raggiunge il 6% in caso di fattore 3. Inoltre il risultato dipende dalla somma delle due varianze, per cui è lo stesso che prendere due distribuzioni con la stessa varianza, uguale alla media delle due varianze.

Una volta stabilita la distribuzione dell' abilità "nascosta", è possibile calcolare la probabilità di vittoria dati i ratings dei due giocatori. Esempio: due giocatori, uno con 1560 punti, l' altro con 1700. Il gioco possibile del primo sarà rappresentato da una gaussiana centrata a 1560, mentre per l' altro la gaussiana sarà centrata a 1700 (fig. 1). Per calcolare la probabilità di vittoria del giocatore con 1700 punti si prende la distribuzione della differenza delle due abilità (che, per inciso è ancora una gaussiana, con media uguale alla differenza delle due abilità e varianza uguale alla somma delle varianze delle singole distribuzioni, fig 2).

La frazione dei casi in cui tale differenza è positiva è l' integrale della gaussiana risultante su tutti i valori positivi. Quindi:

$$p_A = \frac{1}{\sqrt{2\pi}\sigma}\int_0^{+\infty} e^{-\frac{1}{2}\frac{(x-d)^2}{\sigma^2}}dx =$$

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{d} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}dx =$$

$$\frac{1}{2}\left[1+erf\left(\frac{d}{\sigma\sqrt{2}}\right)\right]$$

dove $d = a - b$, $\sigma \simeq 200\sqrt{2}$.

Questa è la relazione tra la probabilità di vittoria e la differenza tra i punteggi. Poiché si parla sempre di differenza tra i punteggi, per definire la scala dobbiamo prendere un punto di riferimento, per esempio il punteggio di un giocatore medio e riferirsi sempre a lui. La costante σ invece stabilisce i rapporti di forza tra due giocatori data la loro differenza di punteggio. È ovvio che anche in questo caso si tratta di un parametro che viene scelto arbitrariamente. Se scegliessimo un altro valore, per esempio uguale al doppio, la scala dei punteggi Elo risulterebbe "espansa" di un fattore due: la differenza tra i punteggi di due giocatori dovrebbe essere doppia per corrispondere alle stesse probabilità di vittoria. Il valore 200 che abbiamo utilizzato è la differenza di punti tra i limiti delle varie categorie negli Stati Uniti. Le categorie sono infatti cosı̀ definite: Master 2200+, Expert 2000-2199, cat. A 1800-1999, cat. B 1600-1799, e cosı̀ via fino alla E. Il Master corrisponde grossomodo ai nostri Maestri nazionali, l' Expert ai nostri Candidati Maestri, la categoria A alla nostra 1^a categoria nazionale ... ma il paragone non è cosı̀ evidente, anche perché in Italia le categorie non sono uniformemente spaziate.

La formula per le percentuali attese che viene usata nella pratica non è quella riportata sopra. La FIDE e la FSI usano delle tabelle precalcolate; altre federazioni usano la seguente funzione (funzione logistica):

$$p_A = \frac{1}{10^{(-\frac{d}{400})}+1}$$

che è molto simile numericamente all' integrale di una gaussiana con media 0 e $\sigma = (400/\log 10)\cdot(\pi/\sqrt{3}) \simeq 200\sqrt{2}$ (infatti l' integrale tra $-\infty$ e σ per tale gaussiana vale circa 0.84, che è circa uguale al valore di p_A quando $d = 200\sqrt{2}$).

Questa formula ci dice per esempio che se la differenza punti è 400, ci si aspetta un risultato di 10-1 a favore del più forte (probabilità =10/11).

Questa formula risulta dall' uso, anziché del modello gaussiano usato da Elo, di un altro modello, detto di Bradley-Terry. Secondo questo modello, dati i giocatori A e B, le abilità dei quali sono date rispettivamente w_A e w_B, le probabilità di vittoria sono rispettivamente:

$$p_A = \frac{w_A}{w_A + w_B}$$

$$p_B = \frac{w_B}{w_A + w_B}$$

cioè le probabilità di vittoria sono proporzionali alle abilità:

$$\frac{p_A}{p_B} = \frac{w_A}{w_B}$$

Ciò è semplice e intuitivo, ma pone un problema: la probabilità è funzione non della differenza tra i ratings ma del loro rapporto. Infatti tutte e due dipendono da $r = w_A/w_B$:

$$p_A = \frac{r}{1+r}$$

$$p_B = \frac{1}{1+r}$$

Quindi ci vuole un altro cambiamento di ”scala”, qualcosa che trasformi un rapporto in una differenza. Non è difficile: $w_A = 10^{a/s}$, $w_B = 10^{b/s}$, dove a e b sono i ratings nella nuova scala e s è una costante che ci permette di aggiustare i rapporti di forza tra i due giocatori. Sostituendo, si ottiene la funzione logistica, con $s = 400$.

Le ipotesi di partenza del modello di Bradley-Terry sono quindi molto semplici. Il modello di Bradley-Terry fornisce le regole per comporre le probabilità di vittoria dei giocatori. Se prendiamo tre giocatori, A, B, e C in ordine decrescente di abilità, che chiamiamo rispettivamente a, b, c, date la probabilità di vittoria di B contro C, p_B, e la probabilità di vittoria di A contro C, p_A, si pu‘øcalcolare la probabilità di vittoria di A contro B. Infatti, poiché $p_A = p_A(a - c)$ e $p_B = p_B(b - c)$, invertendo tali relazioni e prendendo la differenza si può calcolare $a - b$ in funzione di p_A e di p_B, e da $a - b$ si può calcolare la probabilità di vittoria di A contro B. Essa risulta

$$\frac{p_A(1 - p_B)}{p_A(1 - p_B) + p_B(1 - p_A)}$$

Il modello di Bradley-Terry può essere ricavato anche secondo un procedimento simile a quello usato per il modello gaussiano, partendo cioè dalle distribuzioni delle abilità ”nascoste” dei singoli giocatori. In questo caso però la distribuzione dell' abilità non è esattamente gaussiana, ma è tale che la funzione di distribuzione cumulativa è la seguente:

$$F_A(x) = exp(-e^{-(x - \log(w_A))})$$

La funzione di distribuzione, che ne è la derivata, è addirittura leggermente asimmetrica verso destra (favorisce cioè valori più alti).

La funzione di distribuzione cumulativa della differenza della abilità di A e B, $x = x_A - x_B$, è data dalla funzione logistica:

$$F_{AB}(x) = \frac{1}{1 + e^{-(x - (\log(w_A) - \log(w_B)))}}$$

Il risultato per la probabilità di vittoria è

$$Pr(x_A - x_B > 0) \ = \ 1 - F_{AB}(0) \ =$$

$$1 - \frac{1}{1 + e^{(\log(w_A) - \log(w_B))}} \ =$$

$$\frac{w_A}{w_A + w_B}$$

Da un punto di vista numerico, comunque, la differenza tra i due modelli è trascurabile.

Per completezza, diamo un' altra versione della formula per la percentuale attesa, una formula approssimata e linearizzata a tratti, quindi molto maneggevole. Indichiamo con E_{avv} l' Elo dell' avversario, e poniamo

$$d = E - E_{avv}$$

se $-350 < E - E_{avv} < 350$,

$$d = 350$$

se $E - E_{avv} > 350$,

$$d = -350$$

se $E - E_{avv} < -350$. Si ha allora per la percentuale attesa:

$$p_A = 0.5 + d/800$$

Questa funzione rappresenta lo sviluppo in serie della funzione logistica al primo ordine attorno a $d = 0$, con l' approssimazione $\log(10)/2 \simeq 1$ invece di 1.15 (la derivata della funzione logistica in $d = 0$ è $\log(10)/1600$). È facile vedere che la percentuale attesa è ”tagliata” inferiormente al 6.25% e superiormente al 93.75%.

La figura 3 rappresenta gli andamenti delle 3 funzioni che abbiamo finora visto che danno la probabilità di vittoria in funzione della differenza tra i punteggi. La differenza tra la gaussiana e la logistica è raffigurata nella figura 4, e la differenza tra la funzione linearizzata e la logistica in figura 5.

Si vede quindi che la differenza tra la gaussiana e la logistica è sempre piccola e raggiunge l' 1.1% attorno ai 350 punti di differenza. Le cose vanno peggio per la funzione linearizzata, per la quale la differenza raggiunge anche il 4% quando la differenza è circa 350 punti. Oltre i 400 punti di differenza la discrepanza dipende dal fatto che la differnza d è ”tagliata” a ± 350.

PERFORMANCE RATING

Giunti a questo punto, siamo in grado di calcolare la probabilità di vittoria di un giocatore contro un altro conoscendo la differenza tra i loro punteggi. Utilizzando tale

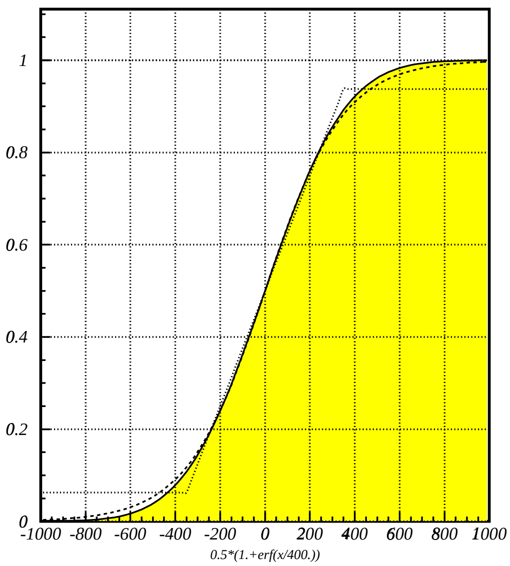

Figure 3: Andamento delle funzioni cumulative gaussiana (linea continua), logistica (linea tratteggiata) e linearizzata (line punteggiata).

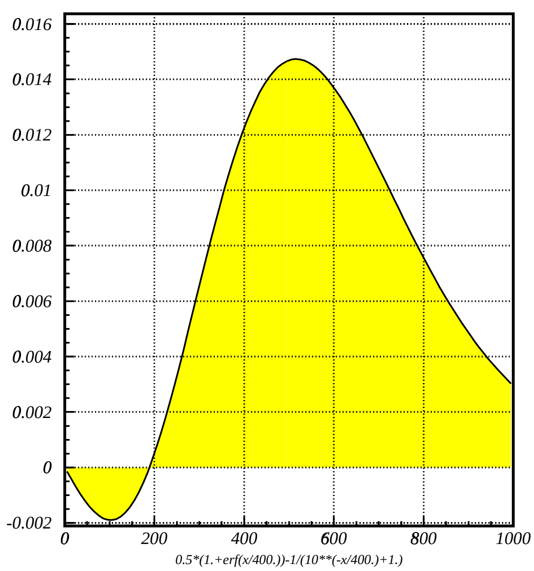

Figure 4: Differenza tra la funzione cumulativa gaussiana e logistica.

relazione al contrario, possiamo dare una valutazione numerica dell' abilità di un giocatore in base ai suoi risultati e ai punteggi dei suoi avversari, utilizzando come stima della percentuale attesa la percentuale realizzata.

Facciamo ancora un esempio pratico. Il giocatore A, con rating sconosciuto, perde un match per 3.5-6.5 contro il giocatore B, che ha 1700 punti. La forza relativa tra A e B può essere valutata invertendo la formula della percentuale attesa:

$$a - b = 400 \cdot Log(\frac{p_A}{1 - p_A})$$

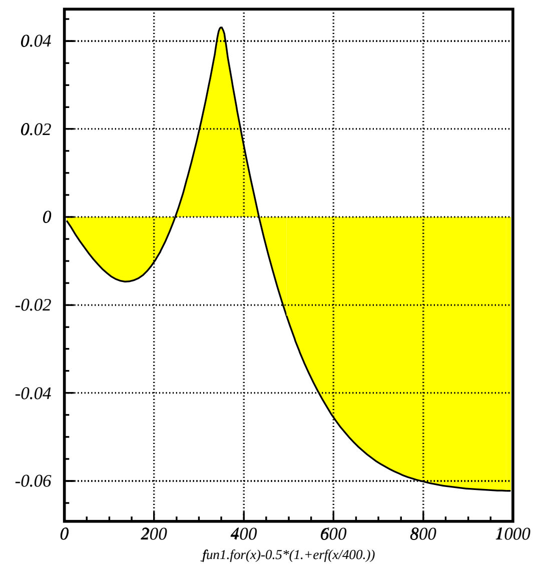

Figure 5: Differenza tra la funzione linearizzata e la funzione cumulativa logistica.

dove p_A, normalmente la percentuale attesa di A, viene posta uguale a 3.5/10=0.35. Quindi il risultato corrisponde ad una differenza pari a -108 punti. Questo significa che A si è comportato come se avesse una forza corrispondente ad un rating di 1700 - 108 = 1592.

Introduciamo così un altro concetto, quello di "performance rating" (PR): Il PR è il punteggio Elo corrispondente ad una percentuale realizzata uguale alla percentuale attesa.

Quando di un giocatore, in questo caso A, non conosciamo nulla, il PR è l' unico modo di trarre informazioni dai risultati. In pratica si tratta di stimare i "punti attesi" ponendoli uguali ai "punti realizzati".

Il caso si complica leggermente se neanche di B conosciamo l'Elo. A questo punto tutto cio che possiamo dire è che la stima della differenza tra i punteggi di A e di B è 108. Per saperne di più occorre far giocare A e B contro un terzo giocatore, C con punteggio noto.

Una caratteristica preoccupante della definizione di PR risulta evidente quando un giocatore vince tutte le partite contro l' avversario (o gli avversari). In tal caso il suo PR risulterebbe infinito (analogamente, se perde tutte le partite, il suo PR risulta $-\infty$). Nella pratica, quindi, si fa un taglio, che varia nelle differenti organizzazioni, ma che in genere è al 90% o al 95%. Questo significa che se la percentuale realizzata supera il 90% (o il 95%) viene posta a questo valore.

Un' altra osservazione sul PR riguarda il fatto che è molto facile da calcolare nel caso di un solo avversario o di molti avversari con lo stesso punteggio Elo. Se invece abbiamo a che fare con molti avversari con punteggi diversi la cosa diventa più complicata, in quanto ad ogni avversario corrisponde una diversa percentuale attesa.

Normalmente si considera, come approssimazione

$$PR = \langle E_{avv} \rangle + 400 \cdot Log(\frac{p_A}{1-p_A})$$

con $\langle E_{avv} \rangle$ uguale alla media dei punteggi Elo degli avversari.

Se si approssima la relazione tra percentuale attesa e differenza tra i ratings con la funzione lineare a tratti, si ha:

$$PR = E_{avv} + 400 \cdot (W - L)/N$$

dove N è il numero di partite giocate, W il numero di vittorie, L il numero di sconfitte e R_{avv} l' Elo medio degli avversari. Questa formula deriva direttamente da quella linearizzata per la percentuale attesa, $p_A = 0.5 + (E - E_{avv})/800$ (considerando che $2p_A - 1 = (W - L)/N$) e corrisponde a tagli superiore e inferiore sulle percentuali attese di 91% e di 9% (± 400 punti).

Queste approssimazioni funzionano abbastanza bene quando i punteggi degli avversari non sono troppo diversi tra loro. Quando i punteggi variano molto, invece, l' approssimazione diventa troppo grossolana. Facciamo un esempio. Per avere il titolo di Maestro, in Italia è necessario in un torneo, realizzare un PR di 2100. Prendiamo il caso di un giocatore (con un punteggio per esempio di 2000), in un torneo a 5 turni, che affronta 4 avversari con 1950 punti ciascuno, contro i quali realizza 3 punti su 4 (3 vittorie e 1 sconfitta). Con il quinto avversario, che ha 1400 punti, vince (Ovviamente un "quinto" avversario del genere può capitare anche al primo turno di un torneo a sistema svizzero). Se applichiamo la formula di cui sopra, vediamo che il nostro giocatore ha realizzato, con 4 punti su 5, l' 80% dei punti contro degli avversari che hanno 1840 punti di media. Il suo PR sarebbe perciò 2080, quindi non otterrebbe il titolo di Maestro.

Se però andiamo a vedere i suoi risultati, è evidente che con i primi 4 avversari ha realizzato un PR di 2141. Col quinto, più che vincere non poteva: ciò nonostante, a causa del punteggio basso, la sua performance è stata rovinata.

Dove sta l' incongruenza? In questo caso l' incongruenza sta proprio nel fare la media aritmetica dei punteggi degli avversari, che equivale a dare all' ultimo avversario lo stesso peso dei quattro precedenti. È chiaro che l' ultima partita, data l' enorme differenza di punteggio tra i due giocatori, non da' nessuna informazione sulla prestazione.

Ci sono naturalmente dei modi abbastanza semplici per correggere questo effetto. Certe federazioni considerano solo partite tra giocatori la cui differenza di punteggio non eccede i 350 punti. Una alternativa potrebbe essere quella di considerare una media "corretta", in cui i punteggi degli avversari che differiscono più di 350 punti (per esempio) dal giocatore di cui si vuole valutare la performance sono corretti in modo da portare questa differenza a 350 (nell' esempio sopra, nel calcolo della performance, il punteggio dell' ultimo avversario andrebbe considerato come 1650 invece che 1400).

Più esattamente, il PR dovrebbe essere calcolato risolvendo la sequente equazione

$$\sum_{i=0}^{N} \left[r_i - \frac{1}{10^{(-d_i/400)} + 1} \right] = 0$$

(r_i è il risultato della i-esima partita, d_i la differenza tra il PR e l' Elo dell' i-esimo avversario, $E_{perf} - E_i$), cosa che può essere fatta numericamente con metodi iterativi. Il valore approssimato che si ottiene dalla formula linearizzata a tratti, $E_0 = E_{avv} + 400 \cdot (W - L)/N$, può essere usato come valore di partenza. Il PR "vero" può essere espresso come $E_{perf} = E_0 + \delta$, con δ piccolo. Si può prendere quindi lo sviluppo in serie di potenze di δ dell' equazione al primo ordine. Se

$$f(\delta) = \sum_{i=0}^{N} \frac{1}{10^{(-(\delta + E_0 + E_i)/400)} + 1}$$

si ha

$$f(0) = \sum_{i=0}^{N} \frac{1}{10^{(-(E_0 - E_i)/400)} + 1} = \sum_{i=0}^{N} p_i = R_a$$

$$f'(0) = \sum_{i=0}^{N} \frac{(\log(10)/400) 10^{(-(E_0 - E_i)/400)}}{(10^{(-(E_0 - E_i)/400)} + 1)^2} =$$

$$(\log(10)/400) \sum_{i=0}^{N} p_i (1 - p_i)$$

dove p_i è la probabilità di vittoria contro l' $i-$esimo avversario calcolata con la performance approssimata E_0, e R_a la somma di queste probabilità, cioè il risultato atteso in base a E_0. L' equazione diventa, ponendo $\sum_{i=0}^{N} r_i = R$ (risultato ottenuto)

$$f(0) + \delta f'(0) - R = 0$$

$$\delta = \frac{400}{\log 10} \frac{R - R_a}{\sum_{i=0}^{N} p_i (1 - p_i)}$$

Naturalmente il valore ottentuto $E_{perf} = E_0 + \delta$ può essere usato come valore di partenza per un' altra iterazione. Con questo metodo, il PR per l' esempio fatto sopra, risulta, dopo la prima iterazione, 2139.

Mark Glickman, matematico dell' Università di Boston, fornisce delle formule ancora più precise. Si parte da una valutazione approssimata del PR:

$$R_g = \Sigma \frac{Ri}{N} + 400 \cdot (2 \cdot W - N)/N$$

(è la formula linearizzata) dove R_i è il rating dell'avversario nella i-esima partita, N è il numero totale di partite giocate, W sono i punti fatti, e la somma è fatta su tutte le partite. Poi, per ogni partita, si calcolano le seguenti quantità:

$$H_i = 10^{(R_i/400)}$$

$$P_i = H_g/(H_g + H_i)$$

dove $H_g = 10^{(R_i/400)}$. Poi si calcolano le seguenti 4 quantità:

$$a = \Sigma P_i$$

$$b = \Sigma[P_i \cdot (1 - P_i)]$$

$$c = \Sigma[P_i \cdot (1 - P_i) \cdot (H_i - H_g)/(H_i + H_g)]$$
$$D = \sqrt{b^2 + 2 \cdot c \cdot (W - a)}$$

se $b^2 > 2 \cdot c \cdot (W - a)$, $D = 0$ altrimenti. Il PR è quindi

$$PR = R_g + K\frac{D - b}{c}$$

dove $K = 400/\log(10)$, circa 173.4

Il PR del giocatore dell' esempio di cui sopra, calcolato con questo metodo, sarebbe 2143.

Il performance rating ha un significato statistico ben preciso. Per capirlo, scriviamo l' espressione della probabilità di realizzare un certo risultato. Consideriamo un giocatore con Elo E che gioca N partite, ogni partita contro un avversario che ha Elo pari a E_i, ottenendo il risultato r_i ($i = 1, 2, ..., N$), con $r_i = 0$ in caso di sconfitta, $r_i = 1$ in caso di vittoria e $r_i = 0.5$ in caso di patta. Il totale dei punti realizzati sarà $r = \Sigma_{i=1}^{N} r_i$. Supponiamo per il momento che non ci siano patte, e utilizziamo la funzione logistica del modello di Bradley-Terry. Poniamo $x = 10^{E/400}$, $x_i = 10^{E_i/400}$. In ogni singola partita la probabilità di vittoria è:

$$p_{1i} = \frac{x}{x + x_i}$$

mentre la probabilità di sconfitta è:

$$p_{0i} = \frac{x_i}{x + x_i}$$

Si può quindi scrivere più genericamente:

$$p_{r_i} = \frac{x^{r_i} \cdot x_i^{1-r_i}}{x + x_i}$$

Il prodotto di tutte queste probabilità ci da' la probabilità totale per i risultati r_i in funzione di E, cioè la likelihood. Essa è, a meno di un fattore costante:

$$L = \prod_{i=1}^{N} \frac{x^{r_i}}{x + x_i}$$

dove r è il risultato totale. Questa espressione può essere estesa anche al caso in cui vi siano patte, considerando la patta come mezza vittoria e mezza sconfitta:

$$p_{0.5i} = \frac{x^{0.5} \cdot x_i^{0.5}}{x + x_i}$$

In questo caso l' espressione per L è ancora valida, con le patte che contribuiscono al risultato totale r.

Calcoliamo ora il valore di E che rende massima la likelihood. Per fare questo prendiamone il logaritmo:

$$\log(L) = r \cdot \log(x) - \sum_{i=1}^{N} \log(x + x_i)$$

e calcoliamone la derivata rispetto a x:

$$\frac{d\log(L)}{dx} = \frac{r}{x} - \sum_{i=1}^{N} \frac{1}{x + x_i}$$

È facile verificare a questo punto che l' equazione

$$\frac{d\log(L)}{dx} = 0$$

è identica all' equazione per il calcolo del PR. Il PR è quindi il punteggio che massimizza la likelihood di un certo risultato. Questo è vero rigorosamente nel caso del modello di Bradley-Terry. Nel caso del modello di Thurstone-Mosteller ciò è vero solo approssimativamente, anche se in pratica le differenze numeriche tra i due modelli sono molto piccole.

Questo importante risultato ci permente anche di valutare la precisione con la quale si può calcolare il PR. Si calcola la derivata seconda:

$$\frac{d^2\log(L)}{dx^2} = \frac{-r}{x^2} + \sum_{i=1}^{N} \frac{1}{(x + x_i)^2}$$

Utilizzando il fatto che

$$\frac{d^2\log(L)}{dE^2} = \frac{d^2\log(L)}{dx^2} \cdot \left(\frac{dx}{dE}\right)^2 + \frac{d\log(L)}{dx}\frac{d^2x}{dE^2}$$

e che nel punto di massimo

$$\frac{d\log(L)}{dx} = 0$$

si ottiene

$$\frac{d^2\log(L)}{dE^2} = -\left(\frac{400}{\log(10)}\right)^2 \cdot \sum_{i=1}^{N} p_i(1 - p_i)$$

dove abbiamo indicato con p_i la probabilità di vittoria nella $i-$esima partita (precedentemente indicata con p_{1i}).

Se facciamo l' ipotesi che $\log(L)$ può essere approssimata ad una parabola nella zona del massimo (cioè L ha andamento gaussiano), e di calcolare l' errore su E come la distanza dal massimo per la quale $\log(L) = \log(L_{Max}) - 1/2$, si ha che tale errore è la radice quadrata dell' inverso della derivata seconda cambiata di segno. Quindi, indicando l' errore con σ:

$$\sigma = \frac{400}{\log(10)} \frac{1}{\sqrt{\sum_{i=1}^{N} p_i(1 - p_i)}}$$

Facendo una stima piuttosto rozza, si può approssimare il termine sotto radice quadrata facendo lo sviluppo in serie dei termini $p_i(1 - p_i)$ attorno al punto $p_i = 0.5$ e trascurare i termini dal secondo ordine in poi, per cui ogni termine risulta uguale a 1/4. Si trova che l' errore è grossomodo $\sigma \simeq c/\sqrt{N}$, con $c \simeq 800/\log(10) \simeq 350$. Perché le nostre approssimazioni siano valide occorre che il risultato r non sia troppo vicino a 0 o a N, e che N non sia troppo vicino a 1. Se ciò non accade, L non può essere approssimata ad una gaussiana e quanto detto finora non risulta valido. Nei casi limite, $r = 0$, $r = N$, e anche $N = 1$, non è neanche possibile calcolare la performance.

In generale, comunque, L può essere approssimata ad una gaussiana nelle vicinanze del massimo. Questo si può vedere

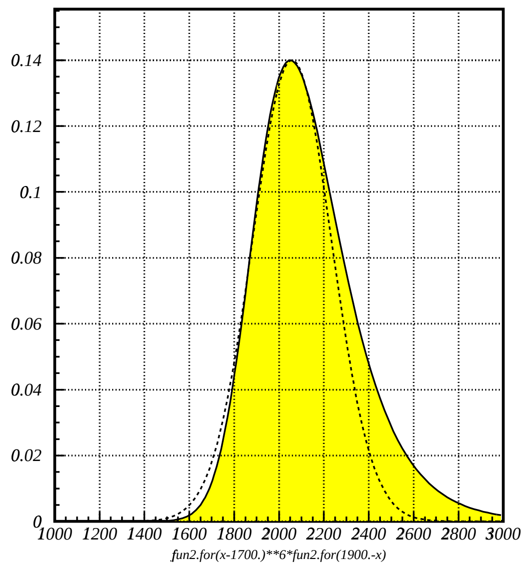

Figure 6: Likelihood, primo caso (ved. testo).

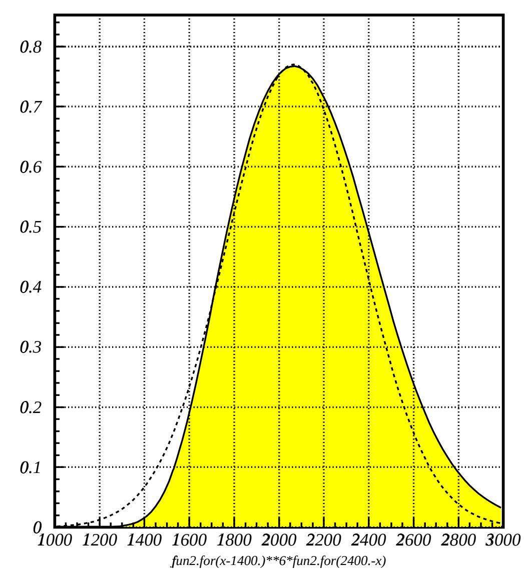

Figure 7: Likelihood, secondo caso (ved. testo).

con qualche esempio. Le figure 6, 7 e 8 rappresentano la likelihood (linea continua) e la gaussiana che la approssima (linea tratteggiata) in tre casi diversi.

Nel primo caso si hanno 6 vittorie contro 6 avversari, tutti e 6 a 1700 punti, e una sconfitta contro un avversario a 1900 punti. Nel secondo caso si hanno 6 vittorie contro 6 avversari, tutti e 6 a 1400 punti, e una sconfitta contro un avversario a 2400 punti. Nel terzo caso si hanno 6 sconfitte contro 6 avversari, tutti e 6 a 1400 punti, e una vittoria contro un avversario a 2400 punti. Quest' ultimo caso è ovviamente poco probabile, come si vede dall' altezza del massimo.

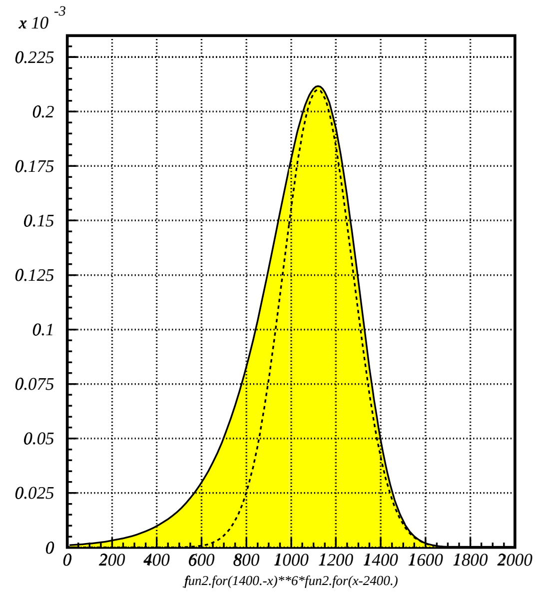

Figure 8: Likelihood, terzo caso (ved. testo).

Abbiamo visto che nel modello di Bradley-Terry il performance rating è anche il punteggio che massimizza la likelihood di un certo risultato su N partite. Vediamo ora se questo è vero anche per altri tipi di relazione tra rating e probabilità di vittoria.

Se $p_i = p_i(E - E_i)$ è la probabilità di vittoria nella i-esima partita contro un avversario il cui rating è E_i (E è il rating del giocatore), la lakelihhod sarà data dal prodotto di termini del tipo p_i per le partite vinte e del tipo $1 - p_i$ per le partite perse. In generale, per un risultato r_i (uguale a 1, 0.5, o 0) possiamo scrivere la probabilità di quel risultato come

$$P(r_i) = p_i^{r_i}(1 - p_i)^{1-r_i}$$

relazione che è valida anche nei casi di patta.

La likelihood sarà quindi

$$L = \prod p_i^{r_i}(1 - p_i)^{1-r_i}$$

Come al solito prendiamo il logaritmo:

$$\log(L) = \sum r_i \log(p_i) + \sum (1 - r_i) \log(1 - p_i)$$

Per la condizione di massimo abbiamo

$$\frac{\partial \log(L)}{\partial E} = \left(\sum \frac{r_i}{p_i} - \sum \frac{1 - r_i}{1 - p_i} \right) \frac{dp_i}{dE}$$

$$= \sum \frac{p_i'}{p_i(1 - p_i)}(r_i - p_i) = 0$$

Perché questa condizione sia equivalente all' equazione che definisce il performance rating occorre che il termine

$$\frac{p_i'}{p_i(1 - p_i)} = q$$

sia costante. L' equazione differenziale che ne risulta è facilmente integrabile per separazione delle variabili, e il risultato è proprio la funzione logistica del modello di Bradley-Terry. Non esitono quindi altre funzioni che godono di questa proprietà.

AGGIORNAMENTO DEL PUNTEGGIO

Per aggiornamento del punteggio si intende il ricalcolo periodico del punteggio Elo di un giocatore in modo da tenere conto dei risultati ottenuti.

Il procedimento standard per l' aggiornamento del punteggio prevede l' uso della seguente formula:

$$E_{new} = E_{old} + K \cdot (R - R_a)$$

dove E_{new} è il punteggio aggiornato, E_{old} il punteggio vecchio, R i punti fatti e R_a i punti attesi, e K un coefficiente. Osserviamo innanzitutto che se si usa la formula linearizzata per la percentuale attesa si può fare l' aggiornamento dell' Elo senza passare attraverso il calcolo di R_a. Si può infatti sostituire l' espressione di sopra per tale quantità direttamente nella formula dell' aggiornamento:

$$E_{new} = E + (K/2) \cdot [(W - L) + \Sigma(E_{avv} - E)/400]$$

dove W è il numero di vittorie, L è il numero di sconfitte e $\Sigma(E - E_{avv})$ è la somma delle differenze tra l' Elo vecchio del giocatore e l'Elo di ciascun avversario, e tale differenza viene "troncata" a ± 350 (cioè posta = 350 se è maggiore di tale valore, =-350 se è minore).

Vediamo qual è l' origine della formula per l' aggiornamento.

Il problema dell' aggiornamento dell' Elo può essere posto in questi termini: un giocatore con punteggio E_0 gioca un certo numero di partite N in un torneo o in più tornei durante un periodo di valutazione (per esempio sei mesi). I suoi risultati nelle N partite possono essere sintetizzati in un numero, il PR, che poniamo uguale a E_N. Qual è la nuova stima dell' abilità del giocatore, dovo avere messo insieme queste informazioni?

La nuova stima dovrebbe essere una media tra la nostra conoscenza prima del semestre, E_0, e quella relativa alle ultime N partite solamente, E_N, in modo da constituire una via di mezzo tra due possibilità estreme: (a) considerare solo le ultime N partite, quindi attribuire come nuovo punteggio E_N, ignorando le informazioni che il punteggio precedente ci fornisce; (b) calcolare il nuovo punteggio includendo le ultime N partite nell' insieme di tutte le partite giocate fino a quel momento, dando cosi̇̀ la stessa importanza alle ultime partite e a quelle giocate a inizio carriera.

Inoltre E_N dovrebbe avre un peso maggiore rispetto a E_0 quanto più grande è N.

Il modo più naturale di procedere è quindi quello di introdurre dei pesi, che dipendono dal numero di partite giocate. Si avrebbe quindi

$$E_{new} = \frac{f_1 E_0 + f_2 E_N}{f_1 + f_2}$$

con f_1 e f_2 pesi positivi. Ridefinendo $f = f_2/(f_1 + f_2)$ (con $0 < f < 1$) si ha

$$E_{new} = E_0 + f \cdot (E_N - E_0)$$

In questo modo ogni volta che il punteggio viene aggiornato il peso del punteggio iniziale diminuisce con un andamento grossomodo esponenziale, per cui i nuovi risultati contano sempre di più rispetto a quelli vecchi.

Vediamo ora come scegliere i pesi per la media. I pesi dovrebbero dipendere dal numero di partite, quindi $f \propto N$ oppure $f_2 \propto N$. Si può scrivere $f = N/N_0$, con N_0 costante e maggiore di N.

Il nuovo paramentro N_0 può essere interpretato come il numero di partite necessario per avere una valutazione della forza del giocatore il più attendibile possibile. Dalla formula sopra, infatti, si vede che se $N \geq N_0$, allora $E_{new} = E_N$, dato che il peso non può essere maggiore di 1. In tal caso, quindi, il nuovo PR ha un' attendibilità tale che possiamo completamente trascurare E_0.

Se un giocatore gioca N_0 partite, l' errore stastistico sul E_N è proporzionale a $1/\sqrt{N_0}$, saremmo tentati di dire che N_0 deve essere il più grande possibile. Se però N_0 è troppo grande, occorre troppo tempo per giocare N_0 partite, nel frattempo l' abilità di un giocatore può essere cambiata. Nella formula dell' aggiornamento questo si riflette in un peso eccessivo per il punteggio iniziale anche dopo molto tempo.

In linea di massima N_0 dovrebbe essere dell' ordine del numero massimo di partite che vengono giocate in un periodo in cui siamo ragionevolmente sicuri che il rating reale del giocatore rimanga stabile. Negli Stati Uniti questo numero è posto uguale a 25. Un neofita, nelle sue prime 25 partite ha un rating provvisorio basato sul PR, o meglio sulla media dei PR nel corso delle 25 partite. Dopo 25 partite il rating diviene definitivo e viene aggiornato con la formula standard, quella col K. In Italia N_0 vale circa 26.7, vedremo dopo perché.

La formula dell' aggiornamento è quindi:

$$E_{new} = \frac{(N_0 - N) \cdot E_0 + N \cdot E_N}{N_0}$$

$$= E_0 + \frac{N}{N_0} \cdot (E_N - E_0)$$

perlomeno quando N è minore di N_0, altrimenti $E_{new} = E_N$.

Per ottenere la formula con il fattore K che abbiamo dato all' inizio del capitolo dobbiamo fare alcuni sviluppi in serie al primo ordine. Sia E_N che E_0 possono essere espressi in termini della percentuale realizzata o attesa e dell'Elo medio degli avversari E_{avv}.

Il secondo termine al secondo membro può essere riscritto utilizzando la formula approssimata linerizzata pe rla performance che abbiamo dato nel capitolo precedente:

$$E_N = E_{avv} + 800 \cdot (p - 0.5)$$

$$E_0 = E_{avv} + 800 \cdot (p_a - 0.5)$$

dove p è la percentuale realizzata ($p = R/N$) e p_a la percentuale attesa ($p_a = R_a/N$). Abbiamo utilizzato la relazione $(W-L)/N = 2\cdot p - 1$, con W uguale al numero di vittorie e L al numero di sconfitte.

Quindi la formula finale è:

$$E_{new} = E_0 + \frac{N}{N_0}\cdot 800\cdot(p-p_a)$$
$$= \frac{800}{N_0}(R-R_a)$$

Confrontando con la formula standard per l' aggiornamento di vede che $K = 800/N_0$.

A questa formula si può arrivare anche esprimendo i PR come

$$E_N = E_{avv} + 400\cdot Log\frac{p}{1-p}$$
$$E_0 = E_{avv} + 400\cdot Log\frac{p_a}{1-p_a}$$

Sempre facendo l' approssimazione di considerare la media dei punteggi degli avversari anziché i punteggi uno per uno, e prendendo lo sviluppo in serie di potenze al primo ordine della funzione $\log[p/(1-p)]$ attorno al punto $p = 0.5$ (e analogamente per la parte che contiene p_a), facendo l' approssimazione $\log(10)/2 \simeq 1$.

Vediamo anche un terzo modo per la derivazione della formula dell' aggiornamento, in cui si fa uso della definizione esatta del PR come soluzione dell' equazione

$$\sum_{i=0}^{N}\left[r_i - \frac{1}{10^{(-(E_N-E_i)/400)}+1}\right] = 0$$

Poniamo $E_N = E_0 + \delta$ e sviluppiamo in serie di potenze il primo membro dell' equazione. Analogamente ai calcoli che abbiamo fatto per la performance nel capitolo precedente e con le stesse notazioni, si trova

$$\delta = \frac{400}{\log 10}\frac{R-R_a}{\sum_{i=0}^{N} p_i(1-p_i)}$$

dove p_i è la probabilità di vittoria nella $i-$esima partita e R_a il risultato atteso (Σp_i), calcolati usando E_0, il punteggio iniziale.

Il punteggio aggiornato diventa quindi

$$E_{new} = E_0 + \frac{N}{N_0}\cdot\delta = E_0 + K\cdot(R-R_a)$$

dove

$$K = \frac{400}{\log 10}\frac{N}{N_0}\frac{1}{\sum_{i=0}^{N} p_a^i(1-p_a^i)} = \left(\frac{\log 10}{400}\right)\frac{N}{N_0}\sigma^2$$

dove σ è l' errore sul PR calcolato nel capitolo precedente. Ancora una volta approssimiamo $p_i(1-p_i) \simeq 1/4$, per cui $N\sigma^2 \simeq 4(400/\log 10)^2$ e

$$K = \frac{400}{\log 10}\frac{4}{N_0} \simeq \frac{800}{N_0}$$

Il termine $800/N_0$ è il fattore K che già conosciamo. Tra parentesi, questo spiega perché in Italia $N_0 = 26.7$, semplicemente perché è stato scelto un valore di K tondo (30), mentre negli Stati Uniti è stato scelto un N_0 tondo ($N_0 = 25$ che implica K=32).

Questa formula è valida quando:

- la differenza tra percentuale attesa e percentuale realizzata non è troppo grande;
- il numero di partite non supera N_0, come già detto.

Riepiloghiamo i vari casi. Se il numero di partite giocate è 0, si riprende pari pari l' Elo vecchio. Se il numero di partite giocate è piccolo, il PR avrà poco peso rispetto all' Elo vecchio, che quindi cambierà poco. Se il numero di partite è uguale a N_0 o supera questo valore, allora abbiamo un numero sufficiente di partite per un valore attendibile, e possiamo assegnare direttamente come nuovo rating il PR, dimenticandoci del rating vecchio.

A parità di numero di partite giocate, se la differenza tra l' Elo vecchio e il PR è piccola vuol dire che il vecchio Elo dava comunque una buona stima della forza del giocatore, e questo si rifletterà in una piccola differenza tra tra percentuale attesa e percentuale realizzata. Se invece la differenza tra l'Elo vecchio e il PR è grande, vuol dire che l' Elo vecchio non descriveva così bene la forza del giocatore e l' Elo subirà una variazione maggiore, la percentuale attesa e la percentuale realizzata saranno molto diverse.

Quindi la formula sembra dare, almeno qualitativamente, risultati ragionevoli. Tale formula tuttavia viene applicata da tutte le federazioni o quasi indipendentemente dalle condizioni di validità. Se la formula viene applicata impropriamente, l' Elo aggiornato di un giocatore può essere sovrastimato o sottostimato. Ciò comunque non è poi così grave, in quanto l' Elo tenderà ad aggiustarsi nel torneo o semestre successivo, sempre che l' errore non sia di grossa entità.

Un' ultima nota, per quanto riguarda l' aggiornamento dell' Elo. Spesso, nel calcolare le differenze tra il punteggio di un giocatore e i punteggi degli avversari, si usa "troncare" a 350, cioè la differenza viene posta uguale a 350 se supera questo valore, e a -350 se è inferiore. Questo ha come conseguenza che la percentuale attesa non è mai superiore a 88% e mai inferiore a 12%. Il vantaggio di ciò è che se un giocatore vince contro un avversario molto più debole di lui (cioè con Elo molto inferiore, anche 500-600 punti), guadagna comunque qualche punto.

IL FATTORE K

Abbiamo visto che $K = 800/N_0$, dove N_0 è il numero di partite necessario per dare una valutazione attendibile della forza del giocatore. Cerchiamo di vedere meglio il significato di tale definizione. Partiamo da una osservazione banale: se il giocatore di cui si vuole aggiornare il punteggio gioca $N_0/2$ partite durante il periodo di valutazione, allora il punteggio aggiornato sarà numericamente uguale alla media aritmetica tra il PR su quel periodo e il punteggio vecchio. Essi hanno cioè lo stesso peso. Questo vuol dire che in qualche modo al PR su $N_0/2$ partite e al punteggio vecchio diamo lo stesso grado di fiducia. Per dirla in un altro modo, $N_0/2$, e quindi N_0 e anche K quantificano in un certo senso il grado di fiducia che abbiamo nel punteggio vecchio, o, se vogliamo, con quanta precisione rappresenta la forza del giocatore.

Prendiamo il caso di un esordiente, che entra nella graduatoria Elo con un punteggio di partenza fisso, per esempio 1500. È chiaro che il suo punteggio di partenza, come valutazione della sua forza, ha una grossa indeterminazione. Quando questo esordiente partecipa al suo primo torneo, giocando per esempio 5 partite, è chiaro che il PR basato su di esse fornisce una valutazione della sua forza ben più attendibile del punteggio di partenza. I valori limite per K potrebbero essere tali che il corrispondente valore di N_0 (cioè $800/K$) sia $N_0 > 5$ (in modo che il punteggio iniziale abbia comunque un peso), e $N_0/2 < 5$ (in modo che il suo peso non superi quello del PR nel torneo). Questi valori sono $K = 130$ e $K = 90$, volendo dare una valutazione della forza del giocatore dopo quel torneo. In generale, se il numero di partite giocate durante il periodo di valutazione (sia esso la durata di un solo torneo o un semestre) è N, deve essere $800/N > K > 800/2N$.

I giocatori confermati, soprattutto ai livelli più alti, hanno il loro punteggio basato su molte partite e molti anni di esperienza. È quindi lecito supporre che il punteggio abbia una indeterminazione piuttosto piccola. Il grado di fiducia dovrebbe essere equivalente ad un N_0 più grande. In effetti, sopra i 2400 punti si ha $K = 10$ (nella FSI e nella FIDE).

Supponiamo ora che un giocatore confermato (cioè il cui Elo è noto con la precisione dovuta) incontri un esordiente. Supponiamo che la precisione con la quale è noto l' Elo dell' esordiente sia tale che per lui sia $K = 90$, mentre per il giocatore confermato $K = 30$. Per l' esordiente si può applicare la formula dell' aggiornamento dell' Elo con $K = 90$. Che facciamo però per il giocatore confermato?

Facciamo il seguente ragionamento: il K dell' esordiente è il triplo di quello del giocatore confermato. Questo significa che per l' esordiente occorrono il triplo di partite per avere una valutazione nuova attendibile rispetto al giocatore confermato. Ai fini del peso per la valutazione, ogni partita del giocatore confermato vale come 3 dell' esordiente. Possiamo quindi supporre che quando si trovano uno di fronte all' altro, per il giocatore confermato valga la seguente regola: 3 partite contro l' esordiente pesano come una contro un giocatore confermato, e quindi il K deve essere diviso per 3. In generale quindi, il K di un giocatore confermato quando gioca contro un giocatore non confermato (con K maggiore), deve essere moltiplicato per il rapporto N_{avv}/N_0, dove N_{avv} è l' N_0 dell' avversario.

Vediamo ora il caso in cui il numero di partite giocate N durante il periodo di valutazione è molto diverso da N_0. Se $N \ll N_0$ la formula standard dell' aggiornamento è applicabile, e in linea di principio non ci sono particolari problemi. L' unico problema consiste nel fatto che ogni volta che viene fatto l' aggiornamento il peso dell' Elo di inizio periodo è grande, e questo vuol dire che se N rimane troppo basso daremo troppo peso a partite giocate troppo lontane nel tempo. In pratica, cioè, per giocare N_0 partite occorre troppo tempo, e una valutazione basata su prestazioni troppo lontane nel passato è soggetta a grande indeterminazione. In tal caso sarebbe preferibile usare un N_0 più piccolo (meglio una valutazione basata su poche partite recenti che su molte partite lontane nel passato).

Se invece $N \gg N_0$ il problema è diverso. Innanzitutto la formula dell' aggiornamento non è applicabile. Il procedimento corretto sarebbe di prendere come punteggio aggiornato il PR sulle N partite, come discusso nel capitolo precedente.

Se succede sistematicamente che $N \gg N_0$, è chiaro che N_0, e quindi K, deve essere assegnato diversamente. Una situazione possibile di questo tipo si può avere per esempio considerando un ipotetico Elo per partite lampo. In un semestre il numero di partite giocate può essere molto grande, per cui K deve essere piccolo.

Chiudiamo la discussione sul fattore K riepilogando il significato di questo parametro.

- sono necessarie almeno $N_0 = 800/K$ partite per avere una valutazione attendibile del rating;
- in punteggio calcolato è basato su circa $N_0 = 800/K$ partite equivalenti;
- K è legato all' incertezza con la quale è noto il punteggio iniziale E_0;
- K è legato al peso relativo delle prestazioni recenti rispetto al punteggio iniziale E_0.

Frequenza di aggiornamento

L' aggiornamento della graduatoria Elo viene fatta al termine di ogni periodo di valutazione. In linea di principio il periodo di valutazione può variare da un minimo pari alla durata di una partita o di un torneo, fino ad un massimo variabile purché consistente con N_0 e con K. Per esempio, con $K = 30$ avrebbe poco senso usare un periodo di valutazione di due anni (il numero di partite giocate in tale periodo potrebbe essere sistematicamente maggiore di N_0).

Il periodo di valutazione può essere espresso in termini di partite (aggiornamento dell' Elo ogni N partite), di tornei (per esempio dopo ogni torneo), o di tempo (per esempio ogni sei mesi, come fanno molte federazioni, tra cui la FSI e la FIDE).

Vediamo brevemente le caratteristiche dei vari tipi di periodo.

L' aggiornamento ogni N partite implica un aggiornamento individuale dell' Elo, anziché dell' intera graduatoria. Inoltre N dovrebbe essere minore di $N0$, e sicuramente non superiore. Se $N = N_0$, come discusso precedentemente, si tratta di cancellare l' Elo precedente e sostituirlo con il PR sulle N ultime partite. $N = N_0/2$ potrebbe essere una buona scelta per i giocatori di livello medio ($K = 30$), e in questo caso avrebbe valori come 13 o 14, mentre per i giocatori forti ($K = 10$) dovrebbe essere ridotto, perché $N = 40$ renderebbe l' aggiornamento troppo poco frequente. C'è la spiacevole possibilità (piuttosto probabile però) che l' N-esima partita cada nel mezzo di un torneo, causando problemi di carattere pratico.

L' aggiornamento dopo un numero fisso di tornei ha pressappoco le stesse caratteristiche del caso precedente (in particolare l' aggiornamento individuale). A complicare le cose c'è il fatto che stavolta il numero di partite non è fisso. Comunque è ragionevole supporre che il numero di partite giocate può arrivare a N_0 dopo 3 o 4 tornei. Con 2 tornei è molto improbabile, ma può avvicinarcisi. In pratica, quindi, in questo caso, si tratterebbe di fare l' aggiornamento dell' Elo dopo ogni torneo.

L' aggiornamento dopo un periodo di tempo fisso è quello più diffuso, anzi, bisogna dire che non si hanno notizie di federazioni che usino metodi diverse, eccetto alcune organizzazioni per il gioco per corrispondenza. Il periodo di tempo più usato è quello di sei mesi, ma vengono usati anche periodi minori, fino ad un mese. È sicuramente il sistema più pratico, consente un aggiornamento globale della graduatoria a tempi fissati. La caratteristica di questo sistema è che il numero di partite giocate durante il periodo di valutazione è variabile.

L'uso di un periodo di valutazione di durata fissa, può comportare dei problemi quando il numero di partite giocate diventa troppo elevato (cioè superiore a N_0).

Prendiamo il caso di un giocatore con Elo pari a 2100. Questo giocatore, improvvisamente comincia a perdere punti, e alla fine di un periodo di valutazione di sei mesi arriva a 1700. Nel semestre successivo gioca diversi tornei, i suoi avversari hanno una forza media di 2100, come il vecchio punteggio del nostro giocatore. Supponiamo anche che il nostro giocatore sia tornato a giocare "bene" per il suo vecchio livello di gioco, ma non troppo, per cui realizza un 45% (invece del 50%) contro i suoi avversari. Ora, il nostro giocatore parte da 1700, per cui la sua percentuale attesa contro i suoi avversari è del 9%. È chiaro quindi che guadagnerà punti. Vediamo quanti. Considerando N partite giocate (K=30):

$$E_{new} = 1700 + 30 \cdot N \cdot (0.45 - 0.09) = 1700 + 10.8 \cdot N$$

Giocando 40 partite il nostro giocatore si riprenderebbe tutti i suoi punti persi. Giocandone di più andrebbe ben oltre, e con $N = 100$ guadagnerebbe oltre 1000 punti.

Questo è quanto è accaduto realmente in Italia nel 1998. Il protagonista della "scalata" è stato un Candidato Maestro torinese, Roberto Ricca, che è arrivato a 2780 punti.

Un salto di 1000 punti è in effetti molto grosso, e sebbene sia stato realizzato senza violare neanche una regola, si può avere il dubbio che qualcosa non abbia funzionato a dovere. In base alla nostra discussione dovrebbe essere evidente che il problema sta nel fatto che in questo caso N è superiore a N_0, e quindi la formula per l' aggiornamento non è valida. Nei regolamenti delle varie federazioni, però non si parla mai di limiti di validità della fromula.

Si può modificare il regolamento per impedire questo tipo di scalate? Le possibilità sono: (a) applicare la definizione di N_0 e usare il PR quando $N > N_0$, (b) mettere un limite massimo alla viariazione dell' Elo (per esempio 100 punti), e (c) ridurre il periodo di valutazione in modo da rendere improbabile o impossibile avere grandi valori di N.

Il metodo (a) è quello corretto matematicamente, quindi applicabile senza nessun problema. Il metodo (b) non ha nessuna base matematica: perché proprio 100 punti, e non 200, o addirittura 300? In pratica si introduce un altro parametro difficile da giustificare teoricamente. Il metodo (c) può essere applicato in diversi modi: si può ridurre il periodo di valutazione a 4 mesi, 3 mesi o addirittura a un mese, o aggiornare l' Elo dopo ogni torneo invece che dopo un periodo di tempo fisso.

La soluzione scelta dalla FSI, come è noto, è la celebre "regola dei 4 tornei", che ricade nella categoria (c): l' Elo di un giocatore viene aggiornato, all'interno del semestre, ogni 4 tornei (se ne ha fatti 4 o più). Ai fini del calcolo delle variazioni Elo degli avversari, però, rimane valido il punteggio dell' inizio del semestre.

Il numero di partite giocate in 4 tornei può variare, tuttavia raramente dovrebbe superare 27.

Se immaginiamo di applicare la regola dei 4 tornei durante la scalata di Ricca, vediamo che riusciamo a bloccare completamente l'accumulo dei circa 1000 punti, anche se non riusciamo a evitare la perdita degli stessi punti da parte dei suoi circa 100 avversari. Complessivamente, comunque, la regola dei 4 tornei riesce ad evitare le scalate, anche se le basi statistiche sono scarse.

Una proposta che a suo tempo è stata fatta per impedire scalate "alla Ricca" è l' aggiornamento del punteggio Elo dopo ogni torneo. Questa proposta, che rientra ancora nella categoria (c), rappresenta il massimo frazionamento possibile del periodo di valutazione.

Per capire la differenza tra l' aggiornamento torneo per torneo e l' aggiornamento dopo un periodo fisso di sei mesi, vediamo un esempio pratico. Prendiamo un giocatore con Elo E_0 all' inizio di un semestre che partecipa ad un torneo di 5 turni, realizzando una performance E_5. Aggiornare l'Elo dopo il torneo vuol dire fare una media pesata tra E_0 e E_5, dove E_0 pesa all' 81.5%, mentre E_5 pesa al 18.5%.

Questo giocatore partecipa ad un ulteriore torneo, poniamo ancora di 5 turni, con una performance di E_{10}. Se aggiorniamo il suo Elo, E_{10} peserà al 18.5%, mentre l' Elo precedente all' 81.5%. L'Elo precedente era però una media pesata in cui l' Elo originario E_0 compariva già all' 81.5%. Quindi nel nuovo Elo il peso risultante di E_0 sarà 66.4%, mentre il peso risultante di E_5 sarà 15.1%.

Supponiamo che questi due tornei rimangano i soli ai quali partecipa il nostro giocatore nel corso del semestre. Proviamo ora a fare l' aggiornamento classico di fine semestre. Ciò equivale a prendere il PR del giocatore nei due tornei a 5 turni ciascuno ai quali ha partecipato, e fare la media pesata con E_0. Stavolta, poiché ai due tornei diamo lo stesso peso di un solo torneo a 10 turni, E_0 avrà un peso pari a 63% (minore quindi rispetto al caso dell'aggiornamento torneo per torneo), mentre E_5 e E_{10} peseranno 18.5% ciascuno.

Confrontiamo ora i pesi ottenuti con i due metodi: per E_0 abbiamo 66.4% per l' aggiornamento dopo ogni torneo contro 63% per l' aggiornamento dopo sei mesi, per E_5 15.1% contro 18.5%, per E_{10} 18.5% in tutti e due i casi.

Che cosa abbiamo imparato da questo esempio? L' aggiornamento dell' Elo dopo un semestre significa considerare contemporanei, quindi con lo stesso peso, i tornei che si sono svolti in quel semestre. L' aggiornamento dopo ogni torneo invece da' un peso maggiore all' ultimo torneo, e pesi sempre più piccoli ai tornei man mano precedenti.

Nel nostro esempio l' Elo iniziale ha un peso maggiore con l' aggiornamento torneo per torneo, rispetto all' aggiornamento dopo un semestre. Come si spiega questo? Il punteggio iniziale è in realtà dato dalle performances in un certo numero di tornei del semestre precedente, e naturalmnte un altro punteggio iniziale di due semestri prima. Se facciamo l' aggiornamento ad ogni semestre, ognuno dei tornei del semestre precedente avrà peso minimo, proprio con lo stesso meccanismo secondo il quale tutti i tornei dell' ultimo semestre hanno il peso massimo. Con l' aggiornamento torneo per torneo, invece, anche i tornei del semestre precedente hanno pesi diversi, con l' ultimo torneo che avrà peso maggiore e poi pesi via via minori, fino al valore minimo. È chiaro quindi che il peso complessivo dei tornei del semestre precedente sarà più grande che nel caso in cui siano considerati contemporanei.

Le differenze tra i due metodi di aggiornamento possono anche essere più considerevoli, almeno per quanto riguarda la distribuzione dei pesi. Prendiamo per esempio il caso di un giocatore che gioca 4 tornei in un semestre, e supponiamo per semplicità che il numero di partite per torneo sia sempre lo stesso e tale da comportare un peso del 20% per ciascun torneo. Con l' aggiornamento alla fine del semestre avremo che il peso del punteggio iniziale è del 20%, e il peso dei 4 tornei complessivi dell' 80%.

Con l' aggiornamento torneo per torneo si trova, facendo un po' di calcoli, che alla fine del semestre il peso del punteggio iniziale è del 41%, il peso del primo torneo del semestre del 10%, quello del secondo torneo del 13%, quello del terzo torneo del 16% e quello del quarto del 20%. Questo naturalmente non vuol dire che l' Elo aggiornato sarà molto differente nei due casi, però la differenza nei pesi, soprattutto in quello del punteggio iniziale, è drammatica.

Ovviamente è del tutto inutile chiedersi qual è il modo giusto di fare l' aggiornamento. Ognuno dei due metodi può essere preferibile all' altro per certi aspetti. La differenza sostanziale consiste comunque nel dare pesi diversi a tornei diversi: nel caso dell' aggiornamento ogni sei mesi il peso dipende solo dal semestre in cui viene giocato il torneo ed è costante entro il semestre stesso; nel caso dell' aggiornamento dopo ogni torneo il peso è funzione esclusivamente del numero d' ordine del torneo (andando in ordine di tempo nel passato), indipendentemente dal periodo.

Nel caso dell' aggiornamento dopo ogni torneo possono crearsi piccoli paradossi, dovuti al fatto che tornei vicini temporalmente hanno peso diverso. Questo può essere visto col seguente esempio.

Un giocatore con Elo pari a 1700 partecipa, in un semestre, a due tornei, tutti e due a 5 turni, e in tutti e due la media Elo degli avversari è 1700. Nel primo torneo realizza 5 punti su 5, e nel secondo realizza 0 punti. Supponiamo inoltre che sia $K = 32$.

Se facciamo l' aggiornamento alla fine del semestre, si trova che il punteggio aggiornato è ancora 1700 (infatti ha realizzato il 50% dei punti contro avversari della stessa forza).

Facciamo invece l' aggiornamento dopo ogni torneo. Dopo il primo torneo il giocatore ha guadagnato 80 punti. All' inizio del secondo torneo ha quindi 1780 punti, per cui la sua percentuale attesa non è più 50% come all' inzio del primo torneo, ma 61%. Quindi perde 96 punti anziché 80, e il suo punteggio finale sarà 1684. Quello che è successo è che l' ultimo torneo, che è andato male, ha avuto un peso maggiore rispetto a quello meno recente.

Abbiamo visto che un' altra caratteristica dell' aggiornamento dopo ogni torneo è che il peso dell' Elo iniziale diminuisce più lentamente col procedere dei vari aggiornamenti rispetto all' aggiornamento dopo ogni semestre. Questo non è un problema grave nel caso dei giocatori confermati, per i quali l' Elo iniziale costituisce una buona stima della loro abilità. Nel caso di giocatori appena inseriti nella graduatoria è però possibile (anzi probabile) che l' Elo che è stato loro assegnato non rispecchi la loro forza. Il problema non è gravissimo, poiché l' Elo tende a stabilizzarsi al giusto valore col tempo. Però nel caso dell' aggiornamento dopo ogni torneo questo richiederà più tempo rispetto al caso dell' aggiornamento ogni sei mesi. È opportuno quindi assegnare dei valori di K più grandi per gli esordienti.

Un' altra considerazione riguarda la seguente domanda: ha senso fare un aggiornamento del punteggio Elo dopo un numero troppo piccolo di partite? Non sarebbe meglio aspettare che il giocatore abbia accumulato un numero statisticamente significativo di partite? Questa in effetti è stata una obiezione da parte della FSI alla proposta dell' aggiornamento dell' Elo torneo per torneo.

Per rispondere, ripartiamo dalla definizione di N_0: questa quantità rappresenta il numero di partite necessario (minimo) per poter dare una valutazione attendibile della forza di un giocatore. Consideriamo per semplicità solo giocatori confermati, cioè che hanno già un punteggio Elo e che hanno già giocato le loro N_0 partite. Se uno di questi giocatori gioca N_0 partite in un semestre, siamo in grado di dare una valutazione attendibile della forza di quel giocatore sulla base dei suoi risultati nelle N_0 partite. Questa valutazione sostituisce completamente il suo punteggio precedente. Questo è il significato di N_0.

Se invece il giocatore gioca un numero di partite minore di N_0, allora la valutazione della sua forza basata solo sui risultati di quelle partite dovrà essere combinata con il punteggio iniziale del giocatore, con dei pesi che dipendono dal numero di partite giocate.

In ogni caso l' Elo finale, risultato dell' operazione di aggiornamento, sarà comunque basato un un numero statisticamente sufficiente di partite, solo che esse possono non essere giocate tutte nell' ultimo semestre (o torneo). Una parte di esse può essere "nascosta" nell' Elo di partenza.

Quindi non è sbagliato fare l' aggiornamento dell' Elo dopo un numero piccolo di partite, in quanto le informazioni che abbiamo a disposizione vengono mischiate nella maniera giusta. Non è neanche sbagliato attendere che venga giocato un numero più grande di partite, se non è necessaria una valutazione della forza del giocatore nel frattempo. Quello che è sbagliato invece è considerare N_0 come il numero "ideale" di partite per poter fare l' aggiornamento.

Esiste tuttavia una critica più sottile all' aggiornamento dell' Elo dopo un numero piccolo di partite. La critica riguarda l' uso di certe approssimazioni che sono state fatte. Se il numero di partite giocate è uguale a 1, o se i punti realizzati sono esattamente 0 oppure il 100%, la stima del PR diventa troppo grossolana. Nei casi limite non si può fare di meglio che usare la formula linearizzata approssimata. In questi casi, sebbene una tale approssimazione sia veramente brutale, l' aggiornamento è giustificato dal fatto che l' la formula linearizzata è contenuta implicitamente nella formula col fattore K (è stata ricavata proprio facendo quest' approssimazione).

Qualunque sia il numero di partite giocate durante il periodo di valutazione, la formula dell' aggiornamento, essendo lineare, può essere scomposta in un certo numero di termini, uno per partita, e valutare la variazione partita per partita, che esattamente ciò che viene fatto in pratica quando sifa l' aggiornamento.

Facciamo un esempio numerico molto semplice, in cui un giocatore con Elo pari a 2000 incontra 5 avversari tutti con Elo 2000, ottenendo 4 vittorie e una sconfitta. Il PR del giocatore è quindi 2240. Aggiornare l' Elo significa fare la media pesata tra 2000 (punteggio vecchio) e 2240. Se $K = 32$ i pesi relativi sono 4/5 e 1/5, e il risultato finale sarà 2048. Questo è esattamente il risultato che si ottiene applicando la formula standard (infatti siamo in condizioni buone di applicabilità di questa formula). La variazione Elo trovata, 48, può essere interpretata come la somma di 5 diverse variazioni, una per ogni partita, in particolare +16 per ciascuna vittoria e -16 per la sconfitta. Quindi si può attribuire una variazione a ciascuna partita, e pensare la formula dell' aggiornamento come applicata volta per volta.

Se immaginiamo il che giocatore dell' esempio riporti 5 vittorie su 5 partite, si può quindi applicare la variazione di +16 moltiplicata per 5. Questo giustifica l' applicazione della formula anche nei casi in cui vengono realizzati il 100% dei punti o 0, e anche al caso in cui viene giocata una sola partita (che non finisce patta). In tal caso l' aggiornamento ha senso se combinato con altri risultati (che possono anche essere contenuti nel punteggio iniziale), altrimenti il risultato sarà affetto da una grossa indeterminazione.

Vediamo ora una critica all' aggiornamento del punteggio dopo un numero N di partite troppo grande, anche se non supera N_0. Nel procedimento dell' aggiornamento dell' Elo si usano i punteggi iniziali degli avversari, anziché i loro punteggi attuali aggiornati. Questa è un' approssimazione universalmente accettata anche se non viene quasi mai messa in evidenza. Per capire le conseguenze di questa approssimazione e la sua relazione con N, facciamo un esempio semplice, in cui abbiamo solo due giocatori, A e B, con punteggi iniziali a_0 e b_0 che giocano N partite tra di loro durante il periodo di valutazione. Supponiamo che A realizzi P punti, quindi che B ne realizza $N - P$. La procedura standard dell' aggiornamento del punteggio equivale a calcolare le performances di A e di B, che poniamo a_N e b_N e a calcolare così i punteggi finali:

$$a = a_0 + \frac{N}{N_0}(a_N - a_0)$$

$$b = b_0 + \frac{N}{N_0}(b_N - b_0)$$

Le performances dovrebbero essere calcolate come

$$a_N = b + 400 \cdot Log\frac{P}{N-P}$$

$$b_N = a + 400 \cdot Log\frac{N-P}{P}$$

Normalmente, però, calcoliamo le performances approssimate in questo modo:

$$a_P = b_0 + 400 \cdot Log\frac{P}{N-P}$$

$$b_P = a_0 + 400 \cdot Log\frac{N-P}{P}$$

Facendo le opportune sostituzioni si ottiene:

$$a_N = a_P + \frac{N}{N_0}(b_N - b_0)$$

$$b_N = b_P + \frac{N}{N_0}(a_N - a_0)$$

Per conoscere le performances bisogna risolvere questo sistema. Nel caso di due soli giocatori ciò è semplice, purché N non sia uguale a N_0, nel qual caso il sistema è indeterminato. Quando si ha un torneo con più giocatori conviene usare metodi iterativi. Anche senza risolvere il sistema si può avere un' idea del grado di approssimazione. Facciamo un' ulteriore sostituzione:

$$a = a_0 + \frac{N}{N_0}(a_P - a_0) + \left(\frac{N}{N_0}\right)^2 (b_N - b_0)$$

$$b = b_0 + \frac{N}{N_0}(b_P - b_0) + \left(\frac{N}{N_0}\right)^2 (a_N - a_0)$$

Si vede quindi che per l' aggiornamento corretto dell' Elo bisogna tenere conto di un termine aggiuntivo, dato dalla variazione dell' Elo dell' avversario, pesata secondo un fattore pari a $\left(\frac{N}{N_0}\right)^2$. Se N è un numero piccolo rispetto a N_0, allora il termine con $\left(\frac{N}{N_0}\right)^2$ può essere trascurato rispetto al termine standard con $\frac{N}{N_0}$, e l' aggiornamento normale costituisce una buona approssimazione. Se invece $N \simeq N_0$, allora il termine aggiuntivo assume la stessa importanza del termine standard. Nel caso di più avversari in un torneo, il termine aggiuntivo può essere interpretato come la variazione della media del punteggio degli avversari. In tal caso possono esserci variazioni individuali in un verso o nell' altro, per cui la variazione della media può anche essere piccola, ma ciò non è assolutamente detto.

In conclusione, quindi, fare l' aggiornamento dopo un numero di partite N può portare a delle imprecisioni quando N si avvicina troppo a N_0 pur non superandolo.

È possibile fare l' aggiornamento dell' Elo evitando di calcolare la variazione partita per partita, considerando gli N avversari come un solo avversario con punteggio pari alla media dei punteggi e contro il quale vengono giocate N partite. Questo procedimento approssimato è usato dalla FIDE.

Per capire le possibili differenze tra i due metodi, quello del calcolo della variazione partita per partita e quello del calcolo sulla media degli avversari, facciamo un esempio.

Supponiamo che un giocatore con 1700 punti partecipi ad un torneo a 5 turni e incontri avversari rispettivamente con 2100, 2100, 2100, 2100, e 1300 punti. Qual è la sua percentuale attesa? Considerando la media degli avversari, in questo caso 1940, la percentuale attesa risulta 20% (1 punto su 5).

Calcolando invece la variazione in ciascuna delle partite si ha che contro i 2100 la percentuale attesa è 9%, mentre in quella contro il 1300 è 91%; complessivamente quindi dal giocatore ci si aspettano 0.09+0.09+0.09+0.09+0.91 = 1.27 punti, cioè il 25.4%. I risultati ottenuti con i due metodi sono quindi diversi.

I due procedimenti danno risultati quasi identici quando i punteggi dei giocatori variano poco l' uno dall' altro. Quando però i punteggi degli avversari sono molto diversi tra di loro la discrepanza tra i due metodi risulta evidente.

ARROTONDAMENTI

La formula che da' la percentuale attesa in funzione della differenza tra i rating non viene quasi mai utilizzata in maniera diretta, ma in forma tabulata. Tali tabelle sono anche molto compatte, poiché danno solo percentuali arrotondate all' unità. Spesso, come nel caso della FSI, l' arrotondamento finale è al 5%, per esempio 20%, 25%, e così via.

Come si ripercuote tale approssimazione sul calcolo del punteggio? Vediamo di fare qualche calcolo. Prendiamo un giocatore di cui vogliamo aggiornare l' Elo, che ha giocato N partite, ed ha il coefficiente K noto. Supponiamo di fare il calcolo per ogni partita. Chiamiamo Δ l' errore di arrotondamento della percentuale attesa in ogni singola partita. La variazione dell' Elo sarà data da una somma di N termini del tipo $K \cdot (P - P_a)$, dove $P = 1$, 0.5 o 0 a seconda del risultato, P_a la percentuale attesa. L' errore su ciascuno di questi termini è $K \cdot \Delta$. Poichè ciascun termine è indipendente dagli altri, l' errore di arrotondamento sull' Elo aggiornato sarà $K \cdot \Delta \cdot \sqrt{N}$.

Facciamo un esmpio pratico: $K = 30$, $N = 25$, $\Delta = 0.025$ (arrotondamento al 5%). L' errore totale è 3.75 . Se invece prendiamo $\Delta = 0.005$ (arrotondamento all' 1%) allora l' errore diventa 0.75.

Nel sistema Elo FSI l' arrotondamento è all' 1% durante il calcolo delle singole partite, e poi al 5% sul totale. Esempio: tre partite, supponiamo tutte e tre vinte, e percentuali attese di 45%, 76%, 51% in ciascuna delle tre. La formula ci da', per la variazione dell' Elo, $K \times (1 + 1 + 1 - 0.45 - 0.76 - 0.51) = K \times (3 - 1.72)$. È a questo punto che si fa l'arrotondamento della somma delle percentuali attese. Nel caso in cui si approssima al 10% il contributo all' errore totale è 3, mentre per il 5% è 1.5.

Molte organizzazioni per corrispondenza aggiornano l' Elo dopo ogni partita. Poiché $N = 1$ la precisione con cui devono conoscere la percentuale attesa non deve essere grande per avere un errore sull' Elo aggiornato dell' ordine di 1: $\Delta = 1/K = 0.03$ nella maggior parte dei casi.

È applicabile addirittura la formula approssimata linearizzata invece della funzione logistica per la percentuale attesa senza perdere molto in precisione. Quando la differenza di rating è 300, la differenza tra le percentuali attese date dalla formula linearizzata e dalla funzione logistica è 2.6%, mentre a 350 la differenza sale a 5.5% che si riflette in un errore circa 2 sull'Elo aggiornato. L' uso intelligente della formula approssimata da' risultati più precisi della formula esatta se viene usata male. Aggiornare l' Elo dopo un numero N di partite relativamente grande richiede la conoscenza più precisa della percentuale attesa rispetto a valori piccili di N.

Vediamo ora il caso in cui per fare l' aggiornamento si usa la media dei puntegi Elo degli avversari. La formula da'

$$E_{new} - E_{old} = K \cdot N \cdot (P - P_a)$$

L' errore di arrotondamento è quindi $K \cdot N \cdot \Delta$. Con i numeri di prima, $K = 30, N = 25, \Delta = 0.025$, si ha 18.75, che è un'enormità. Naturalmente questo è un caso pessimistico, perché questo modo di procedere, cioè di fare la media degli Elo degli avversari, si fa generalmente per un singolo evento e non su tutte le partite del semestre o qualunque sia il periodo di valutazione. In questo caso quindi per N andrebbero presi valori come $N = 5$, $N = 7$ o $N = 10$. Per tali valori si ha rispettivamente 3.75, 5.25, 7.5. Comunque sia, l' uso della media dei ratings degli avversari richiede una conoscenza più precisa numericamente della percentuale attesa. La FSI usa il calcolo partita per partita.

INFLAZIONE E DEFLAZIONE

Il valore del K, come è noto, non è lo stesso per tutte le categorie. Nella FSI, K=30 sotto i 2100 punti, K=20 fino a 2400, k=10 oltre. Variazioni simili sihanno anche nella FIDE.

Osserviamo che il fatto che esistano coefficienti K diversi implica che il valore medio dell' Elo non si conserva. Quando si incontrano giocatori con uguale K, qualunque sia il risultato, quello che un giocatore guadagna (in termini di punti Elo), l' altro giocatore lo perde, per cui la media dell' Elo si conserva. Se il K non è lo stesso per i due giocatori, uno dei due guadagna un certo numero di punti, l' altro ne perde un numero diverso, e la media non è più la stessa. Con buona probabilità vincerà il giocatore con K più basso, e quindi di fascia Elo più alta. Quindi, chi guadagna punti ne guadagna meno di quanti nel perda il suo avversario. I fattori K diversi hanno quindi un leggero effetto deflattivo sull' intero sistema. Tale effetto è comunque molto piccolo

La deflazione del sistema, cioè la diminuzione del punteggio medio col passare del tempo, ha dimensioni più vaste dovute ad altre cause. Arpad Elo ha fornito una spiegazione, sicuramente la più importante.

L'argomento è il seguente: in una graduatoria, i giocatori che escono sono in genere quelli a "fine carriera", quindi con punteggi relativamente alti, e comunque più alti di quando sono entrati. I giocatori che invece entrano in graduatoria in genere vi entrano con punteggi bassi. Questo continuo ricambio ha come effetto quello di dimimuire la media generale.

La media Elo non ha niente a che vedere con il livello medio di gioco, ma è solo un punto di riferimento (sono significative solo differenze di punteggio). Quindi, per esempio, un giocatore che nel 1980 aveva 1800 punti non è confrontabile con giocatore che nel 2000 ha lo stesso punteggio. A causa dell' effetto deflattivo, quest' ultimo dovrebbe essere più forte.

Se vogliamo stabilire una corrispondenza assoluta tra punteggio Elo e abilità, allora è necessario correggere la deflazione. Il modo ovvio di correggere sarebbe aggiustare periodicamente i punteggi di tutti i giocatori in modo da riportare la media al valore desiderato. Questo avrebbe come conseguenza la variazione dell' Elo anche di giocatori che non hanno giocato partite. Per questo motivo questo tipo di correzione non viene usato.

In genere l' effetto deflattivo è in parte controbilanciato da altri fattori inflattivi. Esempi sono: (1) possibili "bonus" (cioè "premi" in punti per chi realizza prestazioni superiori), (2) la sovrastima del punteggio di ingresso in graduatoria , (3) il non considerare variazioni negative di giocatori che vincono tornei, (4) l' eventuale impossibilità di retrocedere di categoria ("rating floors").

PUNTEGGIO DI INGRESSO

Affrontiamo ora il problema del punteggio di ingresso nella graduatoria Elo. Il problema si pone, per esempio, nel caso di un esordiente che partecipa al primo torneo.

Questo problema viene affrontato in vari modi dalle varie federazioni.

Il modo più semplice per attribuire un punteggio di ingresso è quello di attribuire un punteggio fisso per tutti, per esempio 1400 e usarlo come punteggio di partenza per le variazioni Elo sia del giocatore in questione sia degli avversari.

Com'è ovvio, ci sono argomenti sia a favore, sia contro questo sistema. Sicuramente a suo favore c'è la semplicità di applicazione. Inoltre c'è il fatto che il punteggio, dopo vari aggiornamenti, tenderà a stabilizzarsi ad un valore che rispecchia di più la vera forza del giocatore in questione. Facciamo qualche rapido calcolo su un esempio concreto: supponiamo che un giocatore venga inserito con 1400 punti nella graduatoria, ma che la vera forza del giocatore corrisponda a 1600 punti. Supponiamo quindi che questo giocatore giochi realizzando un PR pari appunto a 1600. Supponiamo che i suoi avversari abbiano punteggio medio 1400. Questo significa che nonostante il punteggio iniziale di 1400, il nostro giocatore realizzerà il 75% dei punti contro i suoi avversari. La sua percentuale attesa è 50%. Se gioca N partite il suo punteggio aggiornato sarà (con $K = 30$), $1400 + K \cdot N \cdot (075 - 0.50) = 1400 + 7.5 \cdot N$. Per arrivare a 1600 dovrebbe quindi giocare circa 27 partite (per essere esatti 26.7, numero che ci è già noto. In effetti a questo risultato ci si poteva arrivare dalla definizione di N_0). Tale numero è ragionevole, ed è alla portata di tutti i giocatori.

Bisogna considerare però che se il punteggio di ingresso è molto diverso da quello corrispondente alla forza reale del giocatore, questo porterà ad una valutazione sbagliata delle prestazioni degli avversari. L' aggiustamento del punteggio al suo "vero" valore comporterà una ridistribuzione dei punti in eccesso o in difetto.

La cosa da fare, nel caso di un punteggio di ingresso fisso, è comunque una assegnazione del fattore K differente, in particolare molto più grande di 30, dato che il punteggio di ingresso fornisce la forza del giocatore con una grossa indeterminazione. Per esempio, $K = 50$ corrisponde ad una attendibilità del punteggio di ingresso equivalente alla performance su 8 partite. Il fattore K degli avversari dovrebbe essere diminuito dello stesso fattore.

Vediamo come viene stabilito il punteggio di ingresso dalla USCF (United States Chess Federation). Durante le prime 20 partite si attribuisce un punteggio provvisorio ad un giocatore. Tale punteggio provvisorio è dato dalla media dei PR sulle partite giocate fino a quel momento. Alla 20a partita tale punteggio diventa definitivo e da quel momento l'aggiornamento sarà fatto con le formule note. Questo metodo fornisce sicuramente dei valori più accurati che non il punteggio fisso, ma comporta lo stesso qualche problema. Prima di tutto durante il periodo del punteggio provvisorio le variazioni dei punteggi degli avversari devono essere trattate in maniera diversa dal solito (perché è chiaro che i punteggi provvisori sono molto instabili). Inoltre il fatto che i PR nelle prime partite hanno lo stesso peso e senza nessun limite di tempo può costituire un problema se per esempio un principiante ottiene punteggi molto bassi nelle prime partite, si scoraggia, lascia il gioco per un certo tempo e torna a giocare solo dopo essere migliorato sostanzialmente.

Se guardiamo bene, fare la media dei PR dopo le prime N partite vuol dire usare dei fattori K grandi che diminuiscono progressivamente con il numero delle partite. Utilizziamo per semplicità le formule semplificate linearizzate. Supponiamo che nelle prime N partite un giocatore abbia ottenuto un PR uguale a E. Nella sua $(N+1)$-esima partita gioca contro un avversario con Elo pari a E_a e ottiene un risultato R (=1,0,-1 nei casi di vittoria, patta, sconfitta rispettivamente). Il PR relativo all' ultima partita solamente è $E_a + R \cdot 400$.

Facciamo la media con le N partite precedenti:

$$\frac{E \cdot N + E_a + R \cdot 400}{N+1} = E + \frac{E_a - E + R \cdot 400}{N+1}$$

Confrontiamo questa espressione con la formula lineare per l'Elo aggiornato:

$$E + (K/2) \cdot R + (K/800) \cdot (E_a - E)$$

Le due formule sono identiche per $K = 800/(N+1)$. Quindi alla seconda partita si ha $K = 400$, alla decima $K = 80$, alla ventesima K=40. Questo dimostra anche che non ha senso usare un numero elevato di partite in questa fase di punteggio provvisorio, altrimenti il K scende al di sotto del valore normale, 32.

Secondo Mark Glickman, una soluzione di gran lunga superiore al problema del punteggio di ingresso, consiste nell'assegnare un punteggio preliminare ai giocatori. Tale punteggio non deve essere fisso, ma essere basato su qualche stima anche grossolana. Glickman suggerisce un punteggio preliminare a seconda dell' età. Il grado di fiducia che avremo in tale punteggio sarà molto basso. Per il primo torneo del giocatore sarà utilizzato come punteggio Elo questo punteggio preliminare, ma con un fattore K molto grande, per esempio 150. Questo corrisponde ad un grado di fiducia nel punteggio preliminare pari ad un PR basato su 2.7 partite.

Ci si può chiedere perché una soluzione simile debba essere superiore in senso statistico a quella basata sui PR nelle prime partite. Il PR basato sulle prime partite non contiene nessuna informazione a priori sul giocatore. Supponiamo che un certo numero di giocatori giochino un torneo, e che questi giocatori siano tutti allo stesso livello (cioè con punteggio Elo uguali). Ci aspettiamo che in media ciascuno di loro realizzi il 50% dei punti. Questo però è vero solo in media, per cui è possibilissimo che qualcuno di loro vinca il torneo anche con una percentuale alta. Se gli stessi giocatori giocano un secondo torneo è anche possibile che la classifica del primo venga completamente ribaltata. Quindi sarebbe sbagliato assegnare punteggi definitivi sulla base del primo torneo solamente, in quanto potrebbe trattarsi di fluttuazioni statistiche. L' idea è quindi di utilizzare il punteggio preliminare per riportare queste fluttuazioni verso la media. È chiaro che il peso del punteggio preliminare diminuirà man mano che le informazioni attendibili (cioè le partite giocate) si accumulano.

Nel caso della FSI, l' ingresso nella graduatoria Elo si ha solo per i giocatori che realizzano un certo risultato positivo in certi tornei di promozione. Si è ammessi nella graduatoria solo con la promozione alla 3a categoria nazionale, con 1500 punti, o alla 2a categoria nazionale con 1600 punti. La promozione si può ottenere o realizzando una certa percentuale (che corrisponde grossomodo ad un PR di 1450 per la 3a nazionale e 1550 per la seconda) in tornei misti, oppure, per la 3a nazionale, realizzando il 70% dei punti nei tornei di promozione, riservati ai non classificati. Quelli che non hanno punteggio Elo sono considerati, ai fini della variazione dei punteggi degli avversari, con Elo pari a 1400 o 1300, a seconda della categoria sociale.

Questo suggerisce che il punteggio di ingresso nella graduatoria Elo FSI, sovrastimi in media la forza reale del giocatore. Nel caso di tornei open misti ciò è evidente dal piccolo salto di circa 50 punti tra il PR realizzato e il punteggio che si acquisisce. Nel caso dei tornei di promozione la sovrastima è dovuta alla eventualità di giocatori la cui prestazione del 70% sia occasionale, magari dopo una serie di prestazioni inferiori. È comunque ovvio che l' Elo assegnato, anche se sovrastimato, si stabilizzerà, dopo un certo numero di partite ad un valore molto più vicino a quello corrispondente alla vera forza del giocatore, ma questo processo sarà più rapido se la discrepanza iniziale è minima. Inoltre i punti in eccesso vengono ridistribuiti agli avversari del giocatore, con effetto inflattivo.

Quando a suo tempo è stato proposto alla FSI l' aggiornamento della graduatoria Elo dopo ogni torneo, è stata proposta anche l' estensione del sistema Elo a tutti i giocatori. Quali sono i possibili vantaggi dell' estensione del sistema Elo? In primo luogo c'è il vantaggio, così evidente che sembra banale dirlo, che il sistema Elo verrebbe usato per lo scopo per il quale è stato inventato, cioè di dare una valutazione numerica della forza dei giocatori. Questo sarebbe sicuramente di grande beneficio per i non classificati essi stessi, ma anche per i loro avversari nei tornei. Infatti, ai fini del calcolo delle variazioni Elo degli avversari, viene attribuito ai non classificati un punteggio fittizio convenzionale indipendentemente dalla loro forza, rendendo impreciso il calcolo.

Il problema dell' estensione dell' Elo a tutti i giocatori è connesso a quello dell' organizzazione dei tornei. In Italia, infatti, semplificando al massimo, vengono organizzati essenzialmente due tipi di tornei: gli open integrali, aperti a tutte le categorie, e i festival, in cui ci sono tornei separati per le diverse categorie. Naturalmente ci sono anche vie di mezzo, come torneo aperti a più categorie, ma ai nostri fini questo non è rilevante.

In un open integrale un giocatore incontra avversari appartenenti a varie categorie (compresi i non classificati), e la classifica finale è unica (anche se poi ci premi distinti per le varie categorie). Nei festival un giocatore incontra esclusivamente avversari della propria categoria. In un festival gli incontri tra non classificati e giocatori con punteggio Elo sono rari o talvolta impossibili, mentre sono abbastanza frequenti negli open.

Nell' ipotesi di una estensione del sistema Elo a tutti i giocatori, per poter assegnare un punteggio Elo ai non classificati in modo non banale, occorre che questi giochino un numero sufficiente di partite contro giocatori in possesso di punteggio Elo. Un sistema del genere quindi sarebbe legato logicamente più a tornei open che ai festival.

Il sistema Elo FSI attuale, invece, è più legato logicamente ai festival, tenendo separati i non classificati. Gli open addirittua comportano problemi nel calcolo dell' Elo, dovendo ricorrere a punteggi convenzionali.

L' estensione dell' Elo a tutti i giocatori potrebbe comportare qualche svantaggio? Un possibile svantaggio è che, alle lunghe, l' Elo dei giocatori sotto i 1500 punti, potrebbe essere assegnato solo su basi locali. Facendo una semplificazione forse un po' estrema, si può dire che giocatori più forti tendono a partecipare a competizioni più sparse sul territorio italiano, mentre giocatori più deboli tendono a partecipare a competizioni locali o addirittura di circolo. Un esempio di come questo può succedere sono le varie fasi del campionato italiano, in cui solo i giocatori più forti accedono alle fasi successive (regionali e nazionali).

Il rischio è che i giocatori più deboli possano formare dei gruppi di giocatori indipendenti. Giocatori appartenenti a gruppi lontani geograficamente non giocherebbero praticamente mai tra di loro, e il confronto tra questi gruppi sarebbe solo indiretto, attraverso avversari comuni di livello nazionale. La conseguenza di questo fatto è che potrebbe risultare difficile confrontare i punteggi dei giocatori di differenti località.

L' unico modo per evitare che questo succeda è di far competere tra di loro scacchisti di varie località, o almeno che ci sia una connessione attraverso i loro avversari. Più questa connessione è diretta, meglio è.

In Italia, comunque, il problema non dovrebbe essere particolarmente grave, in quanto i giocatori delle categorie superiori possono costituire una buona connessione tra le varie regioni. Il problema sarebbe invece molto più serio nel caso della FIDE, se, come si dice, si vuole abbassare la soglia per l' Elo FIDE a 1500 punti.

Il problema del punteggio di ingresso si pone non solo per gli esordienti, ma anche per i giocatori confermati provenienti da altre federazioni. In questi casi la cosa migliore sarebbe di assegnare come punteggio di ingresso il vecchio punteggio assegnato dalla federazione di appartenenza, eventualmente corretto. La correzione andrebbe calcolata su basi empiriche.

SISTEMA ELO FRANCESE

Fino a qui abbiamo visto come il sistema Elo è stato implementato da varie federazioni come la FIDE, la FSI, la USCF. Vediamo ora brevemente e a grandi linee come è stato implementato dalla Federazione Francese (FFE - Fédération Française des Echecs).

La particolarità è che per l' Elo viene usato solo il performance rating, e non si parla mai di percentuale attesa. L' aggiornamento dell' Elo viene fatto senza la formula standard col fattore K, facendo direttamente la media pesata tra il PR ottenuto nel periodo di valutazione (che è di 4 mesi) e l' Elo precedente. I pesi sono dati secondo il numero di partite giocate N, come abbiamo discusso precedentemente:

$$E_{new} = \frac{(N_0 - N) \cdot E_{old} + N \cdot PR}{N_0}$$

con $N_0 = 16$. Questo equivale a utilizzare un K=50. Ovviamente se $N \geq 16$, $E_{new} = PR$. Altri due piccoli dettagli: solo le partite dette "compatibili" sono valide per la variazione dell' Elo. Una partita si dice compatibile se la differenza tra i punteggi dei due giocatori non supera 350. Inoltre quando il PR è minore di E_{old}, la variazione viene "frenata" pesando E_{old} con un fattore 16:

$$E_{new} = \frac{16 \cdot E_{old} + N \cdot PR}{16 + N}$$

La prima valutazione dell' Elo si ottiene a 8 partite (compatibili) giocate. Fino a quel momento i dati vengono inseriti in un database. Per il calcolo dell'Elo dei non classificati (cioè quelli che non hanno un Elo precedente) si parte dal database. Ai giocatori viene assegnato un punteggio arbitrario, e sulla base di esso vengono calcolati i PR. Questi PR vengono assegnati ai giocatori, e sulla base di essi viene eseguito un nuovo calcolo dei PR. Se uno dei nuovi PR differisce da quello calcolato precedentemente di meno di 20 punti, viene considerato come finale e non viene più toccato. Per gli altri viene fatta un' altra iterazione. Il numero massimo di iterazioni è 25. Alla fine di ogni iterazione viene sempre controllata la differenza tra i PR nuovi e quelli vecchi. Quando tale differenza è meno di 20 punti, il PR viene congelato. Il calcolo è quindi molto laborioso e impossibile da fare senza un computer. A parte ciò, ci sono dei vantaggi in questo modo di implementare l'Elo. Prima di tutto, è a prova di scalate "alla Ricca". Inoltre la graduatoria è estesa a tutti i giocatori.

SISTEMA GLICKO

Il sistema Glicko è stato sviluppato da Mark Glickman del Dipartimento di Matematica dell' Università di Boston tra il 1993 e il 1995, e proposto come sistema per l' USCF. È il primo grosso cambiamento al sistema Elo che sia stato proposto. È stato costruito su basi statistiche molto solide.

Il Glicko usa la stessa scala dell' Elo, quindi i punteggi individuali hanno lo stesso significato. La differenza principale consiste nel fatto che nel Glicko viene fornita anche una "rating deviation" (RD), che rappresenta l' incertezza con la quale è noto il rating del giocatore. Qui occorre precisare che quando si parla di rating non si deve confondere il "parametro" rating con la "stima" del rating. Il parametro rappresenta la vera abilità del giocatore, ed è sconosciuto. Le formule che abbiamo discusso in questo articolo si applicano per i parametri. Una RD=100 per un giocatore valutato 1800 significa che il suo punteggio vero è compreso tra 1700 e 1900 con una probabilità del 68% (1 sigma), oppure che è compreso tra 1600 e 2000 con una probabilità del 95% (2 sigma).

L'idea che sta dietro all' introduzione della RD è la seguente: due giocatori con uguale punteggio giocano una partita. Uno di loro è stato inattivo per un certo tempo, diciamo un anno. L' altro invece ha giocato 50 partite nell' ultimo anno. È chiaro che il punteggio del secondo giocatore è più affidabile, in quanto è stato calcolato recentemente e sulla base di un certo numero di partite. Quindi il punteggio del primo giocatore è conosciuto con poca precisione, mentre quello del secondo è conosciuto con una precisione maggiore. Il grado di affidabilità si riflette nel fattore K. Dopo la partita, il punteggio Elo del secondo giocatore dovrebbe essere variato di poco. Inoltre egli ha giocato una partita contro un giocatore il cui punteggio è conosciuto con una grande indeterminazione, per cui il peso che dobbiamo dare al risultato della partita dovrebbe essere ancora più piccolo.

Per il primo giocatore è vero il contrario. Già il fatto che il suo punteggio è noto con una notevole indeterminazione implica che il suo peso nell' aggiornamento dell' Elo deve essere basso, e quindi dobbiamo dare molto peso al risultato della partita. Inoltre partita è stata giocata contro un giocatore il cui punteggio è noto con precisione, per cui il risultato fornisce una stima relativamente attendibile della forza del primo giocatore, e questo è a favore di un peso maggiore.

Ne concludiamo che in genere il fattore K per un giocatore dovrebbe aumentare con l'aumentare della RD propria e con il diminuire della RD dell'avversario.

Il sistema Glicko, quindi, corregge un difetto del sistema Elo originale, l' arbitrarietà del fattore K. Nel sistema Glicko, infatti, K è una funzione ben precisa delle RD, e l' unica assunzione che si deve fare è quella sulla RD del punteggio iniziale di un giocatore. Gli Internet Chess Servers (ICS) usano il sistema Glicko, e per i giocatori che giocano per la prima volta viene assunto un rating iniziale pari a 1720 (grossomodo il punteggio medio del totale dei giocatori) e una RD pari a 350, sufficiente quindi a coprire un largo intervallo di punteggi. Man mano che il giocatore gioca le sue partite, il rating viene aggiornato e anche la RD, che diminuisce progressivamente con l' aumentare del numero delle partite.

La RD aumenta col passare del tempo, per tenere conto del fatto che il rating vero di un giocatore può cambiare. Dato che non sappiamo dire niente sulla variazione del punteggio, includiamo questa nostra ignoranza addizionale nella RD. La RD aumenta in particolare durante i periodi di inattività.

Nel sistema Glicko c'è quindi bisogno di un' altra ipotesi, cioè la legge con la quale la RD aumenta nel tempo. Glickman consiglia di assumere funzioni che danno l' aumento del quadrato della RD di tipo lineare o logaritmico, con costanti che vanno determinate empiricamente. Se un giocatore gioca ad un ritmo tale per cui la RD diminuisce con l' aggiornamento più di quanto aumenti col tempo, la sua RD è destinata a diventare troppo piccola, e quindi anche il K, cosí che le partite che gioca saranno praticamente ininfluenti sul suo punteggio. Quindi anche qui è consigliabile introdurre un nuovo parametro, la RD minima di un giocatore. La RD diminuisce meno quando la differenza tra i rating del giocatore e dell' avversario è maggiore.

Una conseguenza dell' introduzione delle RD, è che in ogni partita, anche tra giocatori della stessa forza, non è più vero che quello che viene guadagnato da uno (in termini di punti), viene perso dall' altro, poiché i valori di K possono essere differenti per i due giocatori. Intuitivamente, infatti, è il punteggio noto con minore precisione che è soggetto ad un maggiore cambiamento, avendo un maggiore intervallo di valori possibili a disposizione.

Le formule matematiche nel sistema Glicko sono basate su una analisi Bayesiana approssimata per il problema del calcolo dell' abilità del giocatore a posteriori dopo un certo numero di partite. Seguendo la notazione di Glickman, indichiamo con θ l' abilità di un giocatore, che assumiamo distribuita in modo gaussiano con media μ e varianza σ^2:

$$\theta \sim N(\mu, \sigma^2)$$

Dopo un certo tempo t, cosniderando possibili variazioni dell' abilità, si assume che l' incertezza sull' abilità sia aumentata, e questo si riflette in un aumento della varianza, per esempio in modo lineare col tempo:

$$\theta^{(t)} \sim N(\mu, \sigma^2 + \nu^2 t)$$

Questa è quella che assumiamo come distribuzione a priori dell' abilità del giocatore nell' analisi bayesiana. La distribuzione a posteriori, dopo aver giocato N partite, è data da

$$f(\theta|s_i, \theta_i) \propto \phi(\theta|\mu, \sigma^2) \cdot L(\theta|s_i, \theta_i)$$

dove s_i sono i risultati 0, 1/2, 1) contro gli avversari di abilità θ_i durante il periodo di valutazione (gli indici i si riferiscono all' insieme di tutti gli avversari), ϕ la funzione di distribuzione gaussiana con le date medie e varianze per il giocatore e per gli avversari, e L è la likelihood:

$$L(\theta|s_i, \theta_i) = \prod_i w_i \cdot \phi(\theta_i|\mu_i, \sigma_i^2)$$

dove abbiamo posto

$$w_i = \frac{(10^{(\theta-\theta_i)/400})^{s_i}}{1 + 10^{(\theta-\theta_i)/400}}$$

Le partite contro lo stesso giocatore vengono considerate come partite diverse contro diversi avversari con la stessa abilità. Il passo successivo è l' integrazione della likelihood rispetto ai punteggi degli avversari per ottenere la likelihood marginale. Per fare questo approssimiamo i vari fattori w_i a funzioni cumulative gaussiane, il che, come abbiamo visto precedentemente, è lecito, e la media e la varianza corrispondenti sono rispettivamente θ_i e $\tau^2 = (400/\log 10)^2(\pi^2/3)$. La likelihood diventa quindi il prodotto di tanti integrali indipendenti, ciascuno sulla variabile θ_i, ognuno costituito dal prodotto di una gaussiana ϕ per una funzione cumulativa gaussiana F.

A questo punto si sfrutta la seguente relazione:

$$\int F(\theta|\theta_i, \tau^2)\phi(\theta_i|\mu_i, \sigma_i^2)d\theta_i = F(\theta|\mu_i, \tau^2 + \sigma^2)$$

Quindi i risultati delle integrazioni saranno tante funzioni cumulative gaussiane (che approssimiamo di nuovo a funzioni logistiche) con medie μ_i e varianze

$$\left(\frac{400}{\log 10}\right)^2 \frac{\pi^2}{3} + \sigma_i^2$$

Ogni integrale è quindi uguale a:

$$\psi_i(\theta, \mu_i) = \frac{(10^{g(\sigma^2)(\theta-\mu_i)/400})^{s_i}}{1 + 10^{g(\sigma^2)(\theta-\mu_i)/400}}$$

dove si è posto

$$q = \frac{400}{\log 10}$$

$$g(\sigma^2) = \frac{1}{\sqrt{1 + 3q^2\sigma^2/\pi^2}}$$

Gli integrali con i valori $s_i = 0.5$ devono essere trattati diversamente, ma il loro trattamento simile a quelli con $s_i = 0$ o $s_i = 1$ è giustificabile per interpolazione, secondo Glickman. In ogni caso, i risultati che si ottengono possono essere estesi a valori seminteri. La likelihood marginale è quindi uguale approssimativamente a

$$L(\theta, s_i) = \prod_i \psi_i$$

Il passo successivo in questa analisi è l' approssimazione di L ad una gaussiana e la determinazione della media e della varianza. La discussione sulla legittimità di questa approssimazione è simile a quella fatta a proposito del PR calcolato con la likelihood. Ricordiamo quindi che è necessario che il numero di partite non sia troppo piccolo e il risultato non troppo vicino a 0 o al 100%. Cerchiamo prima il massimo di L ponendo uguale a 0 la derivata del logaritmo:

$$\frac{\partial}{\partial\theta}\log(L) = q\sum_i g(\sigma_i^2)(s_i - \epsilon_i) = 0$$

dove abbiamo posto

$$\epsilon_i = \frac{1}{1 + 10^{-g(\sigma^2)(\theta-\mu_i)/400}}$$

Quest' ultima quantità rappresenta la probabilità di vittoria contro l' avversario i tenendo conto dell' incertezza sul punteggio di questi. L' effetto è quello di ridurre in valore assoluto la differenza tra il rating del giocatore e quello dell' avversario. Rispetto alla probabilità di Bradley-Terry pura, questa probabilità è quindi più vicina a 0.5. Se l' incertezza sul rating dell' avversario è molto piccola, allora il risultatto numerico sarà molto vicino a quello della probabilità di Bradley-Terry. Se invece l' errore è molto grande (cioè l' avversario ha punteggio praticamente sconosciuto), allora la probabilità è 50%.

Per la derivata seconda si ha:

$$\frac{\partial^2}{\partial\theta^2}\log(L) = q^2\sum_i (g(\sigma_i^2))^2\epsilon_i(1-\epsilon_i) = \frac{1}{\delta^2}$$

dove δ è la varianza della distribuzione gaussiana che approssima la likelihood vicino al massimo.

Assumiamo ora che il punto di massimo di $\log(L)$, $\hat{\theta}$, non sia molto distante da μ, per cui δ può essere calcolato per $\theta = \mu$ invece che $\hat{\theta}$. Questa approssimazione è ragionevole perché i prodotti $\epsilon(1-\epsilon)$ cambiano poco al variare di ϵ.

Il risultato che abbiamo ottenuto è quindi che la distribuzione marginale a posteriori è data dal prodotto di due gaussiane, con medie μ e $\hat{\theta}$, e varianze σ^2 e δ^2.

$$f(\theta|s_i) = \phi(\theta|\mu,\sigma^2)\cdot\phi(\theta|\hat{\theta},\delta^2)$$

La media e la varianza a posteriori sono date da

$$\sigma'^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\delta^2}\right)^{-1}$$

$$\mu' = \sigma'^2\left(\frac{\mu}{\sigma^2} + \frac{\hat{\theta}}{\delta^2}\right)$$

$$= \mu + \frac{1/\delta^2}{1/\delta^2 + 1/\sigma^2}(\hat{\theta} - \mu)$$

Cerchiamo ora una approssimazione per $\hat{\theta}$. Per fare ciò, torniamo all' equazione $d\log(L)/dx = 0$. Poniamo

$$h(\theta) = \sum_i g(\sigma_i^2)\epsilon_i$$

così che l' equazione diventa

$$h(\hat{\theta}) = \sum_i g(\sigma_i^2)s_i$$

Sviluppiano h in serie di potenze attorno al punto μ e prendiamo i termini fino al primo ordine:

$$h(\hat{\theta}) \approx h(\mu) + (\hat{\theta} - \mu)h'(\mu)$$

$$h'(\mu) = q\sum_i (g(\sigma_i^2))^2\epsilon_i(1-\epsilon_i) = q/\delta^2$$

Quindi si ha per la media aggiornata

$$\mu' = \frac{q}{1/\delta^2 + 1/\sigma^2}\left(h(\hat{\theta}) - h(\mu)\right)$$

Questa è la formula finale per l' aggiornamento del punteggio Glicko. Il termine tra parentesi può essere considerato come una differenza tra il risultato realizzato s_i e il risultato atteso. Tali differenze, moltiplicate per i fattori g, vengono sommate e il risultato moltiplicato per l 'analogo del fattore K del sistema Elo. Tale fattore aumenta quando le incertezze sui punteggi diminuiscono. Le incertezze sui punteggi degli avversari sono contenute anche nei fattori g. Se tali incertezze sono molto grandi, i fattori g sono piccoli, per cui il punteggio aggiornato differirà poco dal punteggio vecchio. Se invece tali incertezze sono molto piccole, per cui si può porre $g \simeq 1$, allora si ritrova la formula di aggiornamento per il sistema Elo, con

$$K = \frac{q}{1/\delta^2 + 1/\sigma^2}$$

Molto approssimativamente, si trova $K \simeq 800/(N + N_0)$, dove N_0 è il numero di partite sul quale il punteggio iniziale è basato. Ciò equivale grossomodo ad una media pesata tra punteggio vecchio e performance nel periodo di valutazione con pesi N_0 e N. Nel sistema Elo standard il peso del punteggio iniziale viene ridotto artificialmente a $N_0 - N$, in modo da evitare che risultati troppo vecchi influiscano troppo. Nel sistema Glicko ciò è ottenuto aumentando la varianza della distribuzione a priori.

Il sistema usato inizialmente dalla PCA e poi dalla WCC, che ora si chiama Professional World Chess Ranking, per la classifica mondiale dei top 150 GM, è una variante del Glicko, sviluppata da un gruppo di persone tra cui Mark Glickman e Ken Thompson. Una delle principali differenze è che viene considerato il fatto che col Bianco le probabilità di vittoria aumentano. Su basi statistiche (per i top GM) il vantaggio del Bianco è stato quantificato come 32 punti a favore. Inoltre è diverso il peso delle varie partite giocate durante il periodo di valutazione, che è di un mese. Il peso è proporzionale a $1/N$, con $N = 1$ per l' ultima partita, $N = 2$ per la penultima, $N = 3$ per la terzultima, e così via.

Analisi Bayesiana della singola partita

In questo capitolo affrontiamo il problema dell' aggiornamento del punteggio dopo una sola partita. Seguiremo lo stesso metodo usato nella discussione sul sistema Glicko, cioè un' analisi bayesiana approssimata. In tale discussione la likelihood marginale (integrata sulle abilità degli avversari) veniva approssimata ad una gaussiana. Questa approssimazione non ha senso quando un giocatore realizza 0 punti oppure il 100% dei punti. Questo succede nel caso di una singola partita, se non finisce patta (la patta comunque non viene contemplata dai modelli di coppie comparabili e richiede un trattamento a parte).

Riferiamoci al modello di Bradley-Terry. Prendiamo il caso di un giocatore A, con punteggio a che gioca una partita contro un giocatore B con punteggio b.

Introduciamo, come nel sistema Glicko, le rating deviations, σ_a e σ_b per i due giocatori A e B rispettivamente. Questo significa che le abilità dei due giocatori, che chiameremo x_a e x_b, sono da considerarsi variabili casuali che seguono distribuzioni gaussiane con medie a e b e varianze σ_a^2 e σ_b^2

$$f_a(x_a|a,\sigma_a^2) = \phi(x_a|a,\sigma_a^2)$$

$$f_b(x_b|b,\sigma_b^2) = \phi(x_b|b,\sigma_b^2)$$

dove

$$\phi(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

La probabilità di vittoria per A è data da

$$V = \frac{1}{1+10^{-\left(\frac{(x_a-x_b)}{400}\right)}}$$

$$= \frac{1}{1+\exp[-q(x_a-x_b)]}$$

dove $q = \log 10/400$. Con buona approssimazione si ha

$$V \simeq F(x_a - x_b, \frac{\pi^2}{3q^2}) =$$

$$\int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{1}{\pi/q\sqrt{3}} \exp\left[-\frac{1}{2}\frac{(z-(x_a-x_b))^2}{\pi^2/3q^2}\right] dz$$

$$= F(x_a|x_b, \pi^2/3q^2)$$

Consideriamo per ora la vittoria di A su B. In tal caso la distribuzione a posteriori dell' abilità di A sarà data da

$$f_v(x_a|a,b) = L_v(x_a|x_b) \cdot f_a(x_a|a,\sigma_a^2)$$

$$L_v(x_a|x_b) = V \cdot f_b(x_b|b,\sigma_b^2)$$

Integrando $L_v(x_a|x_b)$ su x_b si ottiene la probabilità marginale. Facciamo uso ancora una volta della relazione

$$\int F(x|y,\tau^2)\phi(y|\mu,\sigma^2)dy = F(x|\mu,\tau^2+\sigma^2)$$

Nel nostro caso x_a rappresenta x, x_b rappresenta y, μ rappresenta b, $q^2\pi^2/3$ rappresenta τ^2 e σ^2 rappresenta σ_b^2, e se approssimiamo la probabilità di vittoria con la funzione F si ha

$$L_v(x_a) = F(x_a|b, \pi^2/3q^2 + \sigma_b^2)$$

Approssimando di nuovo alla funzione logistica si ha

$$L_v(x_a) = \frac{1}{1+\exp(-qg(x_a-b))} =$$

$$\frac{1}{1+10^{-\frac{g}{400}(x_a-b)}} = W$$

dove abbiamo posto

$$g = \left(1 + \frac{3\sigma_b^2 q^2}{\pi^2}\right)^{-1/2}$$

W rappresenta la probabilità di vittoria corretta per il fatto che l' abilità dell' avversario non è nota con precisione. Se fosse nota con precisione tale che $\sigma_b \simeq 0$, allora si riduce al valore che si ottiene con la formula di Bradley-Terry. Se invece σ_b è grandissimo, allora $g \simeq 0$ e $W \simeq 0.5$.

La distribuzione a posteriori dell' abilità di A diventa quindi

$$f_v(x_a) = W \cdot \phi(x_a|a,\sigma_a^2)$$

Questa funzione, prodotto di una gaussiana per una funzione cumulativa gaussiana (con buona approssimazione), avrà ancora un andamento a campana, con un massimo per un certo valore di x_a, e tenderà a 0 per $x_a \to \pm\infty$. Essa approssima molto bene una gaussiana se x_a è molto lontano da b, mentre è piuttosto asimmetrica se x_a è vicino a b.

Per avere il valore aggiornato del punteggio di A basterà prendere il valore di aspettazione di x_a, cioè la media. I calcoli sono molto più semplici se invece della media prendiamo la moda, cioè il valore di x_a per il quale $f_v(x_a)$ è massima. Per semplificare ulteriormente consideriamo il logaritmo di $f_v(x_a)$

$$l(x_a) = \log(f_v(x_a)) = \log W - \frac{1}{2}\frac{(x_a-a)^2}{\sigma_a^2} + cost.$$

Facendo uso della seguente relazione

$$\frac{\partial W}{\partial x_a} = qgW(1-W)$$

si ottiene
$$\frac{\partial l(x_a)}{\partial x_a} = qg(1-W) - \frac{x_a - a}{\sigma_a^2}$$
Per la derivata seconda si ha
$$\frac{\partial^2 l(x_a)}{\partial x_a^2} = -q^2 g^2 W(1-W) - \frac{1}{\sigma_a^2}$$
La nuova stima per l' abilità a' sarà data dalla soluzione dell' equazione in x_a
$$x_a - a = \sigma_a^2 qg(1-W)$$
mentre per la nuova varianza $\sigma_a'^2$ si ha
$$\frac{1}{\sigma_a'^2} = \frac{1}{\sigma_a^2} + q^2 g^2 W(1-W)$$
Consideriamo ora il caso in cui A viene sconfitto da B. La distribuzione a posteriori è data da
$$f_s(x_a|a,b) = L_s(x_a|x_b) \cdot f_a(x_a|a,\sigma_a^2)$$
$$L_s(x_a|x_b) = (1-V) \cdot f_b(x_b|b,\sigma_b^2)$$
L' integrazione su x_b ci da'
$$L_s(x_a) = 1 - W$$
Prendendo il logaritmo e procedendo come nel caso della vittoria si ottiene l' equazione in x_a:
$$x_a - a = \sigma_a^2 qg(-W)$$
mentre per la varianza il risultato è esattamente lo stesso. Le due equazioni per la vittoria e per la sconfitta possono essere unificate chiamando R il risultato della partita, cioè 0 in caso di sconfitta e 1 in caso di vittoria:
$$x_a - a = \sigma_a^2 qg(R-W)$$
e possiamo addirittura fare di più, cioè generalizzare l' equazione al caso di patta, $R = 0.5$. In questo caso la variazione del punteggio starebbe a metà strada tra vittoria e sconfitta, come si intuisce.

Occupiamoci ora della soluzione dell' equazione. L' incognita x_a è contenuta sia al primo membro che al secondo membro, in $W(x_a)$.

Sviluppiamo $W(x_a)$ in serie di potenze attorno ad a e fermiamodi al primo ordine:
$$W \simeq W(a) + (x_a - a)W'(a) =$$
$$W(a) + qg(x_a - a)W(a)(1 - W(a))$$
Risolvendo in $x_a - a$ si ottiene
$$x_a - a = \frac{\sigma_a^2 qg(R-W)}{1 + \sigma_a^2 q^2 g^2 W(1-W)}$$
Si potrebbe essere tentati di approssimare $W(x_a) \simeq W(a)$, cioè di prendere W all' ordine 0 in x_a. Per capire se ciò è lecito, vediamo numericamente quanto può valere il secondo termine del denominatore. Se consideriamo che $qg \simeq 1/200$ e $W(1-W) \simeq 1/4$ (anzi è il suo massimo valore), $\sigma_a \simeq 80$ (punteggio iniziale ottenuto sulla base di 25 partite, con $K = 32$, e considerando l' errore come $\simeq 400/\sqrt{N}$), tale termine risulta dell' ordine di 0.04, quindi abbastanza piccolo nel caso di una singola partita.

Se confrontiamo questa formula con quella del sistema Elo, vediamo che approssimativamente
$$K = \sigma_a^2 qg$$
Poniamo $\sigma_a^2 \simeq 400/N_0$, con N_0 uguale al numero di partite sul quale è basato il punteggio iniziale del giocatore. Il fattore g dipende dalla rating deviation dell' avversario. Quando questa è uguale a 80 (rating basato su 25 partite, $K = 32$), $g = 0.97$. Se l' avversario ha giocato un numero piccolo di partite, g diventa più piccolo. Se $\sigma_b \simeq 300$, allora $g = 0.7$. In generale si può approssimare $g \simeq (1 + 1.2/N_b)^{-1/2} \simeq 1 - 0.6/N_b$ (N_b è il numero di partite sulle quali è basato il punteggio di B). Se l' avversario ha punteggio non provvisorio, $g \simeq 1$, quindi
$$K \simeq \frac{800}{N_0}$$
Se invece l' avversario ha punteggio provvisorio, il fattore K deve essere ridotto di un fattore g. Analogamente le probabilità di vittoria devono essere corrette, riducendo la differenza tra i punteggi di un fattore g.

Se confrontiamo le formule ottenute con quelle del sistema Glicko, vediamo che sono identiche. In effetti i calcoli sopra possono essere rifatti in modo generale al caso di N partite contro altrettanti avversari, e si ottengono esattamente le stesse formule del Glicko. Nel nostro caso, però non è stata fatta l' approssimazione della likelihood marginale a gaussiana. Abbiamo quindi dimostrato la validità del sistema Glicko anche quando un giocatore realizza 0 punti o il 100% dei punti.

Il sistema Elo e il sistema Glicko danno sostanzialmente gli stessi risultati. Le differenze si hanno quando i giocatori hanno delle rating deviations non trascurabili. Questo si verifica quando per esempio non sono state accumulate abbastanza partite per dare una valutazione precisa della forza e quando i giocatori restano inattivi per un certo periodo.

Naturalmente l' Elo può essere corretto in modo tale da incorporare i casi con rating meno preciso, per esempio usando diversi valori di K, ma la teoria non prescrive come. Già il fatto che K sia diverso nelle tre fasce di punteggi limitate da 2400 e 2100, appare un po' arbitrario. Il sistema Glicko, invece, ci dice esattamente come trattare questi casi.

Per esempio, nel sistema Elo non c'è nessuna indicazione su come trattare le partite di un giocatore contro avversari con punteggio provvisorio (o che hanno giocato poche partite). Nel sistema Glicko, come abbiamo visto, l' incertezza sul punteggio dell' avversario è contenuta nel fattore g. Facendo una cruda approssimazione, si può stimare l' effetto dell' incertezza sul punteggio dell' avversario come una riduzione di K di un fattore $N_b/(N_b + 1)$. È chiaro che tale riduzione è significativa solo se il numero di partite giocate dall' avversario, N_b, è molto piccolo, diciamo minore di 5.

In linea di principio, quindi, quando un avversario è da considerarsi esordiente, si possono considerare le partite giocate contro di lui con un K ridotto di 1/2. Dopo il primo torneo o il primo periodo di valutazione, il numero di partite giocate dovrebbe essere sufficiente per considerare il K normale senza nessuna riduzione.

L' effetto dell' incertezza del rating del giocatore di cui si vuole calcolare la variazione, invece è già stato discusso.

IL CASO ELO ITALIA

In questo capitolo prendiamo in considerazione alcune statistiche relative all' Elo dei giocatori italiani, basate sulla graduatoria Elo Italia del gennaio 2000.

Il numero totale dei giocatori, 7481, è ripartito nelle varie categorie come mostrato dalla tabella 1. Nella cetegoria Maestri (M) sono inclusi anche quelli in possesso dei titoli FIDE (Maestro FIDE, Maestro Internazionale e Grande Maestro). La categoria più popolata è la 2^a nazionale.

Table 1
Giocatori nelle varie categorie nel gennaio 2000

Categoria	Giocatori
3N	1425
2N	2947
1N	1651
CM	1123
M	335
Totale	7481

La distribuzione dei giocatori nelle varie fasce Elo (ogni fascia è di 100 punti) è invece riportata dalla figura 9.

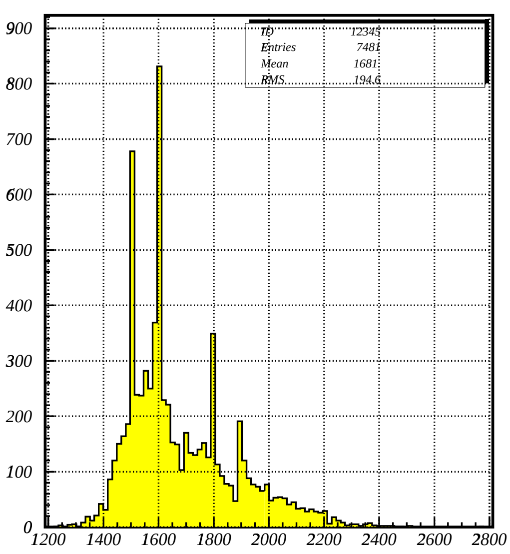

Figure 9: Distribuzione dell' Elo dei giocatori

La massima popolazione si ha nella fascia 1500-1600. La correlazione con la tabella 1 suggerisce quindi che i limiti Elo della fascia relativa alla 2^a categoria nazionale sono diversi da quelli "teorici" 1600-1700.

Questo può essere visto più in dettaglio dalle distribuzioni dell' Elo per categoria. Le figure dalla 10 alla 14 riportano tali distribuzioni per la 3^a nazionale, la 2^a, la 1^a, i Candidati Maestri e i Maestri. Se vogliamo stabilire dei limiti Elo "pratici" entro i quali si trovano la maggior parte dei giocatori di ciascuna categoria, si ha 1400-1550 per la 3N, 1500-1700 per la 2N. Per la 1N e i CM si ha una soddisfacente corrispondenza con i limiti teorici, cioè 1700-1900 e 1900-2100. Per i Maestri, invece si ha 2000-2500.

In ciascuna categoria il picco si trova sul punteggio teorico, a parte la 1N che ha un doppio picco di difficile spiegazione.

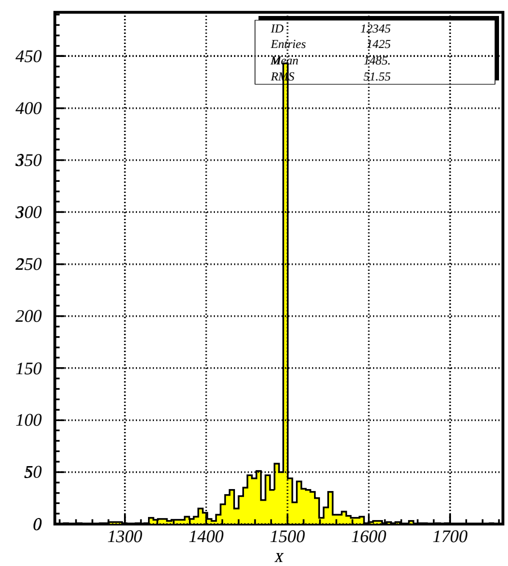

Figure 10: Distribuzione dell' Elo nella terza cat. nazionale

Table 2
Elo medio delle categorie

Categoria	Media	Deviaz. standard
M	2162	118
CM	1944	92
1N	1753	72
2N	1580	57
3N	1485	52
Tot.	1681	195

La tabella 2 riassume i valori medi dei punteggi Elo per le varie categorie e per il totale dei giocatori. La deviazione standard dipende dalle ampiezze delle fasce Elo delle categorie.

È interessante vedere il numero di giocatori "retrocessi", cioè con Elo inferiore a quello teorico della categoria. Per la 3N il giocatori con meno di 1500 punti sono 45.5%, mentre per la 2N quelli con meno di 1600 sono il 50.7%. Si può fare il confronto con gli stessi dati per la 1N e CM, per le quali si ha rispettivamente 16.1% e 19.5%. Questi numeri, però non

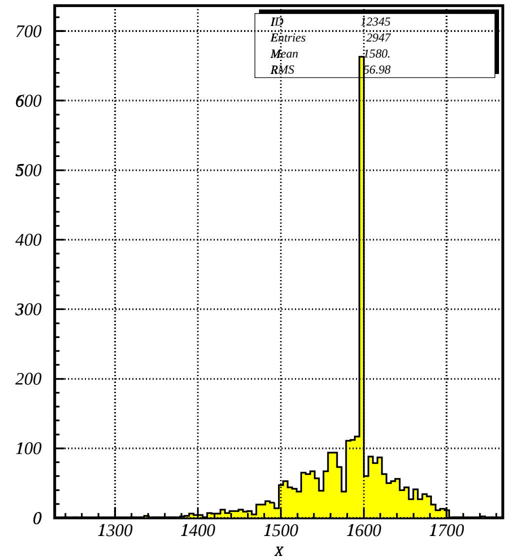

Figure 11: Distribuzione dell' Elo nella seconda cat. nazionale

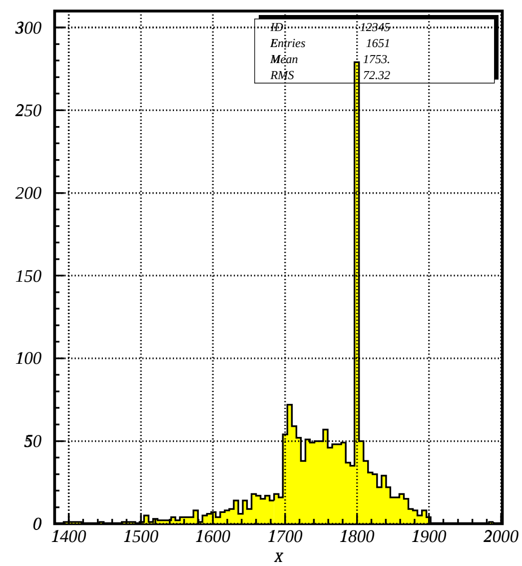

Figure 12: Distribuzione dell' Elo nella prima cat. nazionale

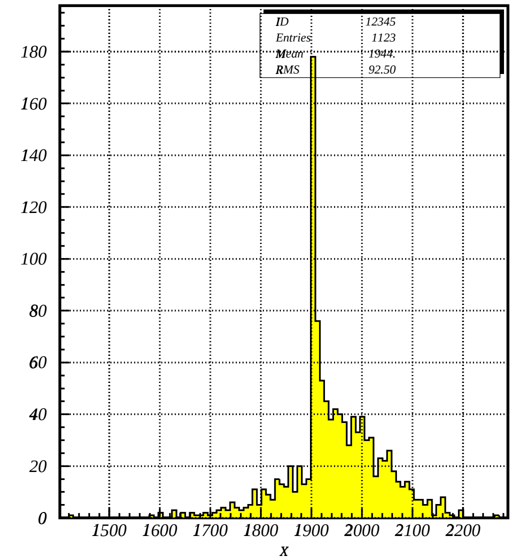

Figure 13: Distribuzione dell' Elo per i Candidati Maestri

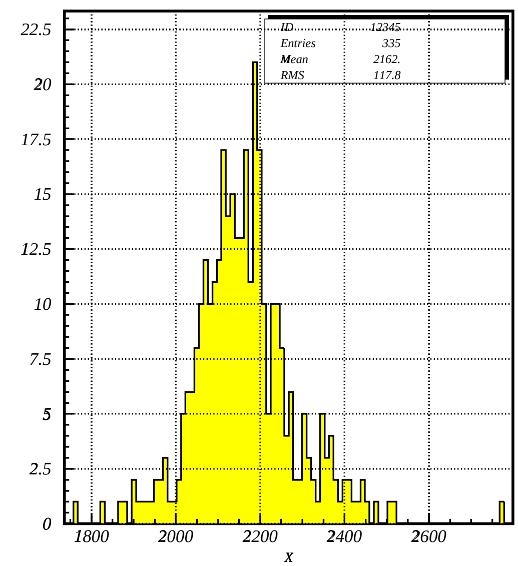

Figure 14: Distribuzione dell' Elo per i Maestri

sono direttamente confrontabili con quelli per la 3N e la 2N in quanto corrispondono a fasce Elo più larghe (200 punti). Per fare il confronto prendiamo quindi, per 1N e i CM, solo i giocatori che hanno rispettivamente meno di 1800 e 2000 punti. Le percentuali diventano quindi 24.8% per la 1N e 26.3% per i CM, quindi ancora molto al di sotto dei valori della 3N e la 2N.

Una spiegazione di questo fatto può essere il meccanismo con il quale si è promossi 3N e 2N. Abbiamo osservato, nella discussione sul punteggio di ingresso nella graduatoria Elo Italia, che questi è molto probabilmente sovrastimato, in quanto viene assegnato trascurando eventuali precedenti prestazioni negative. Successivamente, quindi, il punteggio si stabilizza ad un valore più giusto, anche se il titolo di categoria rimane. Il perché la percentuale dei ”retrocessi” sia maggiore nella 2N che nella 3N non è molto chiaro. In linea di principio ci si potrebbe aspettare il contrario, visto che per la 3N c'è un meccanismo in più di promozione (i tornei di promozione) con possibilità di sovrastima. È vero anche però che per la promozione da 3N a 2N c'è la possibilità di promozione ”a salto” (ai 3N che realizzano la percentuale di promozione viene assegnato direttamente 1600 come nuovo punteggio, superiore al punteggio che si avrebbe facendo l' aggiornamento normale).

Per le categorie 1N e CM i ”retrocessi” sono relativamente molto meno, e ciò è spiegabile con fatto che la promozione a queste categorie viene ottenuta aumentando gradualmente il proprio punteggio. Quindi, quando un giocatore arriva alla soglia della categoria, è abbastanza certo che quel valore corrisponde alla sua forza. Per queste due categorie è evidente dalle distribuzioni una caratteristica, presente anche per la 3N e la 2N ma in misura minore, vale a dire la discontinuità a ”gradino” ai valori di soglia delle categorie. Una possibile spiegazione di questo fatto potrebbe essere che i giocatori, una volta varcata la soglia della promozione, rimangono inattivi per un certo tempo (almeno un semestre) per evitare di retrocedere.

Per i Maestri la percentuale di giocatori sotto al valore teorico (2100) è il 26.6% del totale, e 38.9% se riferita a quelli che hanno meno di 2200 punti. Però le considerazioni che abbiamo fatto sono difficilmente applicabili ai Maestri, in quanto il meccanismo di promozione ha subito vari cambiamenti.

Nella graduatoria Elo FSI del gennaio 2000 ci sono 651 giocatori in possesso anche dell’ Elo FIDE. La figura 15 mostra la distribuzione dell’ Elo FIDE per questi giocatori. Come si può vedere, l’ Elo FIDE parte da 2000 punti.

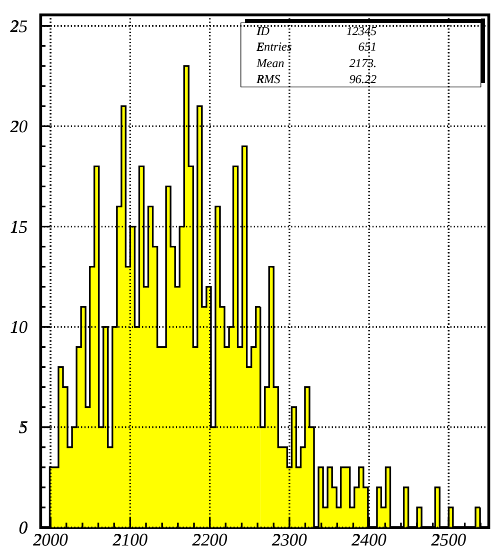

Figure 15: Distribuzione dell’ Elo FIDE.

La figura 16, invece, mostra la distribuzione dell’ Elo Italia per i giocatori che hanno anche l’ Elo FIDE, ed ha un andamento piuttosto differente.

La relazione tra Elo Italia e Elo FIDE è mostrata nella figura 17. Sull’ asse delle ascisse è rappresentato l’ Elo Italia e sull’ asse delle ordinate è rappresentato l’ Elo FIDE. È abbastanza evidente che sebbene la relazione sia con buona approssimazione lineare, il coefficiente angolare della retta è maggiore di 1.

Questo può essere visto meglio nella figura 18, che mostra la differenza tra Elo FIDE e Elo Italia in funzione dell’ Elo Italia. L’ Elo FIDE medio risulta sistematicamente maggiore dell’ Elo Italia, specialmente sotto i 2200, e per un buon numero

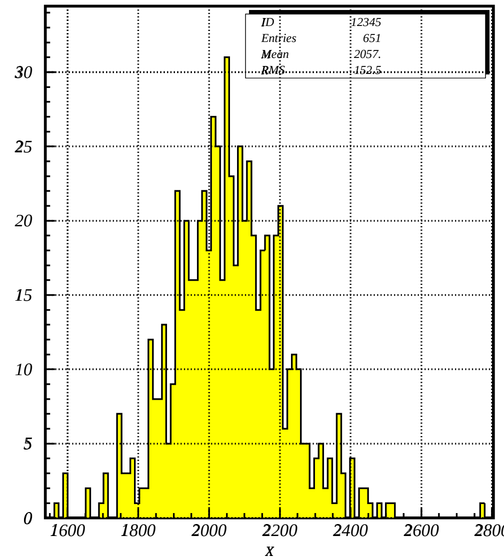

Figure 16: Distribuzione dell’ Elo Italia per i giocatori con Elo FIDE.

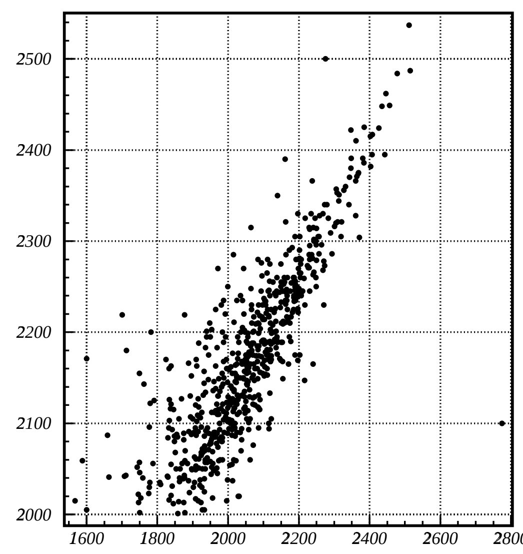

Figure 17: Elo FIDE in funzione dell’ Elo Italia.

di giocatori tale differenza è superiore ai 200 punti.

È probabile che un comportamento del genere sia dovuto ad una sovrastima dell’ Elo FIDE nella zona sotto i 2200 punti, legata al meccanismo di acquisizione dell’ Elo FIDE.

Relativamente all’ età dei giocatori in graduatoria, la figura 19 mostra la distribuzione dei giocatori per anno di nascita (tutti nati tra il 1900 e il 200) nei casi in cui questo dato è noto (6176 su 7481). L’ età media risulta 40 anni.

La figura 20 mostra la distribuzione del punteggio Elo per i nati dopo il 1980.

Il numero dei giocatori nella graduatoria Elo è riportato

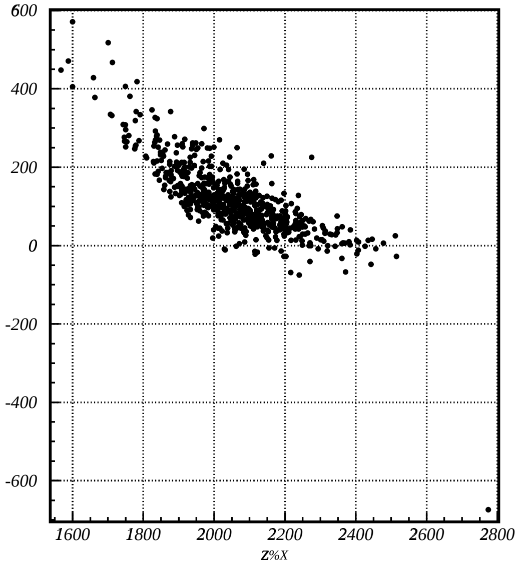

Figure 18: Differenza tra Elo FIDE e Elo Italia in funzione dell' Elo Italia.

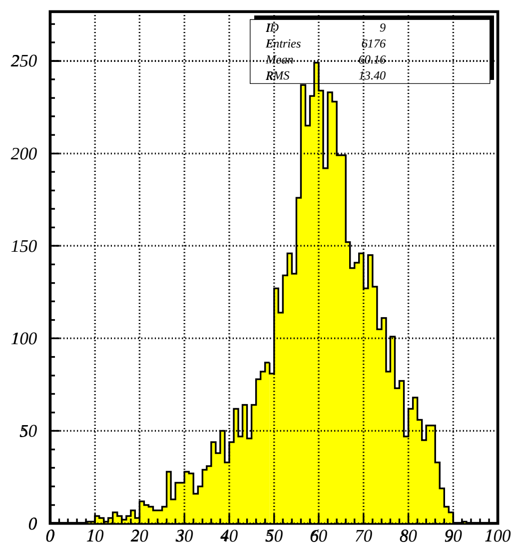

Figure 19: Distribuzione dei giocatori per anno di nascita

in figura 21, a partire dal secondo semestre del 1997 fino al secondo semestre del 2000 (7 semestri in tutto).

La variazione del numero di giocatori può essere valutata meglio se guardiamo la variazione del numero dei giocatori rispetto al semestre precedente, riportata in figura 22.

La figura 23 riporta la media dei punteggi Elo dei giocatori presenti in graduatoria per gli stessi semestri. Vengono anche riportate le stime degli errori statistici sulle medie, calcolati dividendo la deviazione standard delle distribuzioni per le radici quadrate dei numeri dei giocatori. Tali errori vanno da

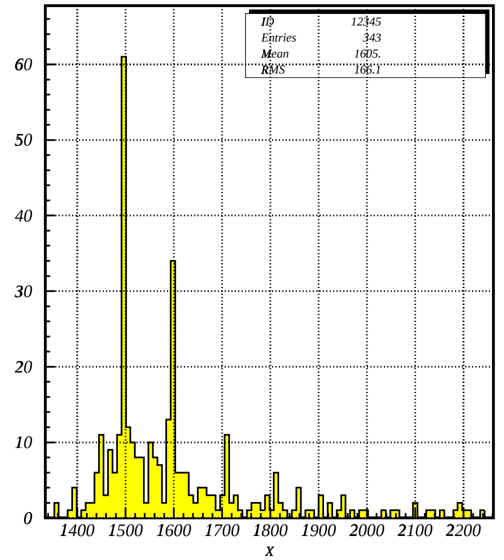

Figure 20: Distribuzione dell' Elo dei nati dopo il 1980

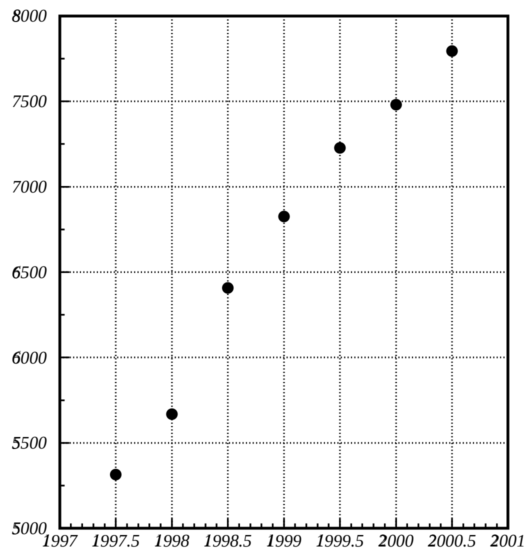

Figure 21: Numero dei giocatori nella graduatoria Elo Italia

un massimo di 2.6 per il punto più a sinistra e diminuiscono progressivamente fino a 2.2 per il punto più a destra.

Il salto tra il primo e il secondo semestre del 1998 è dovuto al fatto che in quel periodo il sistema Elo è stato esteso anche alla 3^a categoria nazionale, per cui si è avuto un aumento anomalo del numero dei giocatori in graduatoria, e una diminuzione altrettanto anomala dell' Elo medio, in quanto i 3N sono entrati col punteggio convenzionale di 1500.

A parte questo, è evidente un andamento deflattivo piuttosto significativo, ben oltre l' errore statistico.

Per concludere questo capitolo, torniamo ai meccanismi di

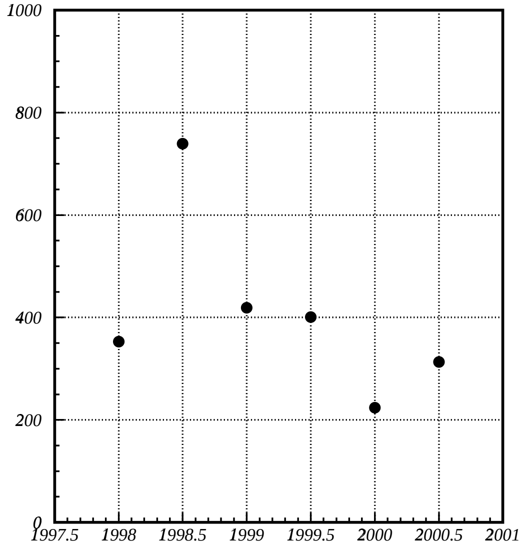

Figure 22: Aumento del numero dei giocatori nella graduatoria Elo Italia

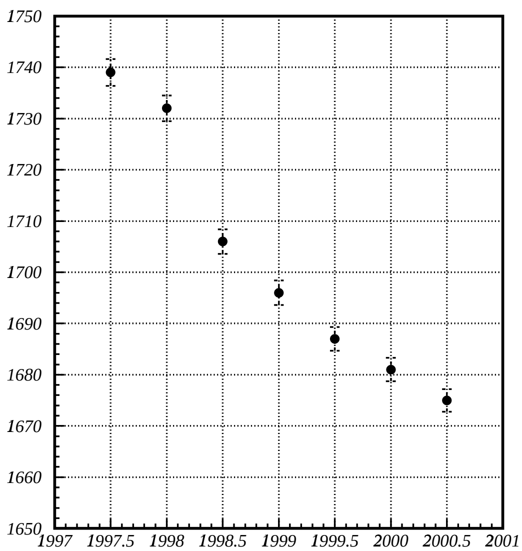

Figure 23: Media Elo dei giocatori in graduatoria

promozione alla 3^a e alla 2^a categoria nazionale. Abbiamo già osservato come questi meccanismi siano approssimativi, in quanto consentono l' entrata di un giocatore nella graduatoria Elo nazionale con un certo punteggio ma rinviano l' aggiustamento di esso a quando il giocatore ha accumulato un numero sufficiente di partite. Per avere un' idea dell' entità dell' approssimazione consideriamo un esempio pratico, un torneo svoltosi realmente, i cui risultati si trovano nella tabella 3. Come si vede, su 9 partecipanti ci sono 5 NC, 3 giocatori di 3^a nazionale, 1 giocatore di 2^a nazionale e 1 giocatore di 1^a nazionale. In base alle tabelle di promozione della FSI, si sono avute le seguenti promozioni:

- Il giocatore n. 3 (3^a naz.) promosso alla 2^a nazionale;
- Il giocatore n. 9 (NC) promosso alla 2^a nazionale;
- Il giocatore n. 6 (NC) promosso alla 3^a nazionale.

Ai fini del calcolo dell' Elo i giocatori NC sono considerati a 1300 punti. Dopo il torneo i punteggi dei giocatori 3, 9 e6 saranno quindi 1600, 1600 e 1500 rispettivamte.

C'è modo di sapere se questi valori riflettono realmente la forza di questi giocatori? Il modo più semplice è quello di calcolare il PR dei vari giocatori. Ovviamente dobbiamo basarci sui punteggi iniziali dei giocatori, e per i NC non abbiamo di meglio che prendere il punteggio convenzionale di 1300.

Table 3
Torneo Open

Giocatore	Elo	Turno 1	2	3	4	5
1	1872	+5	+4	+9	+3	=2
3	1500	+7	+2	+4	-1	+8
2	1600	+6	-3	+5	+9	=1
9	NC	+10	+6	-1	-2	+7
6	NC	-2	-9	+8	+10	+4
4	1467	+8	-1	-3	+5	-6
5	1440	-1	+7	-2	-4	+10
7	NC	-3	-5	+10	+8	-9
8	NC	-4	+10	-6	-7	-3
10	NC	-9	-8	-7	-6	-5

Ottenuti i valori del PR per i giocatori NC, questi possono essere usati come punteggi iniziali per una nuova iterazione, al posto dei 1300 convenzionali. Si procede cosi in modo iterativo finchè i valori non si stabilizzano. La tabella 4 mostra i risultati fino alla quarta iterazione per i giocatori NC. Dopo la quarta iterazione i valori cambiano di meno di 5 punti e possiamo considerarli stabili.

Table 4
Performances giocatori NC

Iterazione	n. 6	n. 7	n. 8	n. 9	n. 10
1	1473	1288	1133	1554	928
2	1416	1231	1091	1512	978
3	1410	1224	1078	1499	938
4	1396	1209	1068	1489	930

Per i giocatori 6 e 9 la discrepanza tra punteggio ottenuto e performance è superiore ai 100 punti.

La performance dell' altro giocatore promosso, il n. 3, è invece 1683.

CONCLUSIONI

Il sistema Elo deve una grossa parte del suo successo al fatto che ha delle basi statistiche ben fondate e che alla fine può essere applicato utilizzando poche e semplici formule matematiche. I modi di applicazione del sistema sono molti (per esempio per quanto riguarda l’ aggiornamento del punteggio), ma in linea di massima possono essere interpretati in termini di diversi valori dei parametri fondamentali.

Sistemi modificati, più precisi e più fondati statisticamente sono stati sviluppati (per esempio il sistema Glicko), perdendo però la semplicità matematica originaria.

In generale il sistema Elo ha dimostrato di funzionare abbastanza bene. Ci sono stati tuttavia casi anche eclatanti in cui il sistema ha rivelato delle carenze. Il ”caso Ricca” è il caso più famoso, ma esiste evidenza di altri casi, specialmente di promozioni ottenute allo stesso modo. Anche a livello internazionale alcuni giocatori praticamente sconosciuti sono riusciti ad inserirsi nei top 100.

Con la panoramica sul sistema Elo che abbiamo presentato in questo articolo ci auguriamo di aver chiarito un certo numero di aspetti anche in vista di ulteriori discussioni e approfondimenti.

BIBLIOGRAFIA

1. *The Rating of Chess Players Past and Present*, Arpad Elo, Arco Publishing, New York, 1978.

2. *Paired Comparison Models with Time-Varying Parameters*, Mark Glickman, Department os Statistics, Harvard University, Cambridge, Massachusetts, May 1993. PhD Dissertation.

3. *Valutazione Bayesiana dell’ abilità dei giocatori di scacchi*, Barbara Fontani, Università degli Studi di Siena, Facoltà di Scienze Matematiche Fisiche e Naturali, Corso di Laurea in Matematica, Tesi di Laurea, anno accademico 1997-98.

4. *Report of the USCF Ratings Committee*, Mark Glickman, Chairman, USCF Ratings Committee, august 1994.

5. ”A Comprehensive Guide to Chess Ratings”, Mark Glickman, *American Chess Journal* 3, pp. 59-102 (1995).

6. *Becoming a Chess Master - The Development of a Rating System for Tournament Chess Players*, Mark Glickman, Boston University, USA, Proceedings of the 1995 Joint Statistical Meetings: Sports Section, 6-15.

7. ”Paramenter Estimation in Large Dynamic Paired Comparison Experiments”, Mark Glickman, *Applied Statistics* 48, pp. 377-394 (1999).

8. *Applicazione del Sistema Elo Italia*, Luigi Troso, Maurizio Mascheroni, Francesco Rinaldi, Roberto Carosi, Contributo alla Tavola Rotonda, Saint Vincent, novembre 1999.

9. *Atti della tavola rotonda sui metodi di classificazione di coppie comparabili e applicazione del sistema Elo Italia*, Giuseppe Campioli, Saint Vincent, novembre 1999.