A simple approximation to the Elo chess ratings formula

Gary A. Mamon

26 August 2012

Around the time of the Fischer-Spassky world chess championship match in 1972 (I was then 14 years old), I tried to understand how the Elo chess rating system was designed. More precisely, I wondered how the chess world estimated the chess rating difference as a function of a given match score. I assumed that if player A scores k times as many points (with wins counting for 1, draws counting for 1/2) against player B, and player B scores k times as many points than his opponent C, then in a match between A and C, A should score k^2 times as many points as C. This led me to adopt a logarithmic formula for the difference of ratings Δ:

$$\Delta = A\,\log\left(\frac{\mathrm{Wins}}{\mathrm{Losses}}\right) = A\,\log\left(\frac{x}{1-x}\right)\ , \tag{1}$$

where 'Wins' and 'Losses' correspond to points and not games, so that they both include (here and below) half of the drawn games, x is the fraction of points scored and where 'log' is the decimal logarithm.

The 1972 Fischer-Spassky match ended with a score of 12.5 to 8.5 in favor of Bobby Fischer. Given that Fischer had forfeited game 2, the actual score among games actually played was 12.5 to 7.5, i.e. Fischer scored a fraction of 0.625 of possible points among games actually played, while Boris Spassky scored a fraction of 0.375. I had read in a chess magazine that this fractional score of 0.625 corresponded to a difference of Elo ratings of 90 points. Then, solving for A in equation (1) with $x = 0.625$ and $\Delta = 90$, one gets $A = 405.7$. Suspecting a roundoff error, I finally adopted $A = 400$, hence equation (1) became

$$\Delta_{\mathrm{log}} = 400\,\log\left(\frac{x}{1-x}\right) = 400\,\log\left(\frac{\mathrm{Wins}}{\mathrm{Losses}}\right) . \tag{2}$$

Although, equation (2) made sense, was simple enough to be computable with a hand-held calculator, and made good approximations to the performances of players in tournaments, it predicted a ratings difference of $\Delta = 88.7$ points, instead of 90 for the Fischer-Spassky match, so I suspected that something must be slightly wrong with equation (2).

As I discovered a few years later, the Elo formula is indeed different. Arpad Elo, who was a professor of Physics at Marquette University in Milwaukee (Wisconsin), assumed that, at any given time, every chess player has a normal (Gaussian) distribution of chess levels (i.e. ratings), all with the same standard deviation, $\sigma = 200$ points, but each with his/her specific mean level. Elementary Statistics informs us that the distribution of the the difference of two Gaussian-distributed variables of the same standard deviation is itself a Gaussian, whose mean is the difference of the two means, and whose standard deviation is $\sqrt{2}$ times the original one. Therefore, the normalized distribution (i.e. probability distribution function) of differences of levels of two players with means y_{A} and y_{B} is a Gaussian distribution of mean $\Delta = y_{\mathrm{A}} - y_{\mathrm{B}}$ and a standard deviation of $\sqrt{2}\,\sigma \simeq 282$ points:

$$\frac{\mathrm{d}N}{\mathrm{d}\delta} = \frac{1}{\sqrt{4\pi\sigma^2}}\,\exp\left[-\frac{(\delta-\Delta)^2}{4\,\sigma^2}\right]\ , \tag{3}$$

(which is normalized to unity: $\int_{-\infty}^{+\infty} (\mathrm{d}N/\mathrm{d}\delta)\,\mathrm{d}\delta = 1$). Then, if one disregards draws, the expected performance (now, probability of wins) of player A against player B is the probability for $\delta > 0$, which is found by integrating over the Gaussian distribution of equation (3):

$$x = \int_0^\infty \frac{\mathrm{d}N}{\mathrm{d}\delta}\,\mathrm{d}\delta = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\Delta}{2\,\sigma}\right)\right] \ , \tag{4}$$

where 'erf' is the error function. Solving for Δ in equation (4) yields

$$\Delta_{\mathrm{Elo}} = 2\,\sigma\,\mathrm{erf}^{-1}(2x-1) = 400\,\mathrm{erf}^{-1}(2x-1) = 400\,\mathrm{erf}^{-1}\left(\frac{\mathrm{Wins}-\mathrm{Losses}}{\mathrm{Wins}+\mathrm{Losses}}\right) \ , \tag{5}$$

where erf^{-1} is the inverse error function, reciprocal of the error function, such that if $v = \mathrm{erf}\,u$, then $u = \mathrm{erf}^{-1} v$. According to equation (5), Fischer's performance of $x = 0.625$ against Spassky translates to a rating difference of $\Delta_{\mathrm{Elo}} = 90.1$ points (consistent with $\Delta = 90$ given in the chess magazine).

For close matches ($x \approx 1/2$), Taylor series expansions to first order of formulae (2) and (5) yield

$$\Delta_{\log} \approx \frac{1600}{\ln 10}\left(x - \frac{1}{2}\right) \simeq 694.9\left(x - \frac{1}{2}\right) \ , \tag{6}$$

$$\Delta_{\mathrm{Elo}} \approx 400\,\sqrt{\pi}\left(x - \frac{1}{2}\right) \simeq 709.0\left(x - \frac{1}{2}\right) \ . \tag{7}$$

Hence, the first order terms differ by only 2%. As seen in Figure 1, the two models (eqs. [2] and

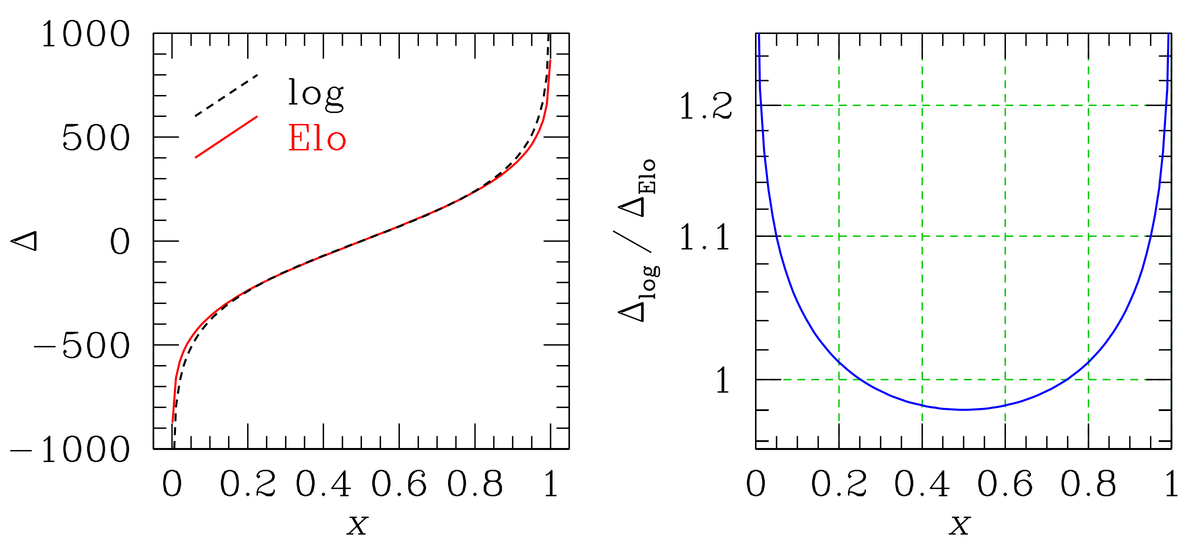

Figure 1: *Left*: Rating difference for given fractional score for the logarithmic (*dashed black*) and Elo (*solid red*) models. *Right*: Ratio of the two rating differences for given fractional score.

[5]) appear indeed very similar. A closer look (right panel) reveals differences that are greater than 10% for players of very unequal strength. But for players with ratings within 268 points,

i.e. where the stronger player is expected to score less than 83% of the points (see left panel of Figure 1), then the approximation of equation (2) is good to 2% relative accuracy (see right panel of Figure 1).

Comparing the first-order terms in equations (6) and (7), one can achieve a better approximation to the Elo formula for close scores ($x \approx 1/2$) by multiplying the logarithmic formula by $(\sqrt{\pi}/4) \ln 10 \simeq 1.020$ to have

$$\Delta_{\log,2} = 100\sqrt{\pi}\, \ln\left(\frac{x}{1-x}\right) \simeq 408.1\, \log\left(\frac{x}{1-x}\right) \ . \tag{8}$$

The third-order Taylor series expansions of the Elo and second logarithmic formula are

$$\Delta_{\rm Elo} \simeq 400\sqrt{\pi}\left(x-\frac{1}{2}\right)\left[1+\frac{\pi}{3}\left(x-\frac{1}{2}\right)\right] \ , \tag{9}$$

$$\Delta_{\log,2} \simeq 400\sqrt{\pi}\left(x-\frac{1}{2}\right)\left[1+\frac{4}{3}\left(x-\frac{1}{2}\right)\right] \ . \tag{10}$$

Comparing the terms in brackets of equations (9) and (10), the third-order term in the second logarithmic formula is therefore only $4/\pi$ times that of the Elo formula, i.e. only 27% larger. But this 2% increase in the logarithmic formula will push up the curve of the right panel of Figure 1 and thus worsen the disagreement with the Elo formula for lopsided scores.

In summary, the simple logarithmic formula of equation (2) is an excellent compromise for typical situations where the ratings of the two players differ by less than 268 points. For very closely rated players, the similar logarithmic formula of equation (8) is even more accurate.