FROM THE DRAW RATIO TO NORMALIZED ELO

MICHEL VAN DEN BERGH

Recall that normalized Elo [1] is defined in such a way that the number of games required to prove that one engine is stronger than another is proportional to $1/(\text{n.e.})^2$. In other words normalized Elo is an important quantity for understanding the resource consumption of tests. This is why for example in Fishtest [2] SPRT bounds are expressed in normalized Elo.

However normalized Elo depends on the testing environment (e.g. the opening book). To transfer Elo values from one testing environment to another one needs an Elo model. The BayesElo (Rao-Kupper) model (see [3]), is such a model.[1] In the BayesElo model the (w, d, l) probabilities of a game between two engines are given by

$$\begin{aligned} w &= L(e + \delta - \gamma) \\ l &= L(-e - \delta - \gamma) \\ d &= 1 - w - l \end{aligned}$$

where

(1) L is the logistic function
$$L(x) = \frac{1}{1 + \exp(-x)}$$
(2) e is the Elo difference, expressed in (scaled) BayesElo units.
(3) δ is the bias of the starting position (presumably taken from an opening book).
(4) γ is a constant called "draw Elo".

The following proposition describes the conversion between normalized Elo and BayesElo.

Proposition 0.1. *For small Elo differences we have approximately*

$$\frac{\text{normalized Elo}}{\text{BayesElo}} \cong C\sqrt{d(1-d)}$$

where C is a universal constant.

Remark 0.2. Since the function $\sqrt{d(1-d)}$ reaches a maximum for $d = 1/2$ this says that if we accept the BayesElo model then for maximum efficiency we should choose our opening book in such a way that the draw ratio is 50%.

[1] In this note we make no claims about the validity of the BayesElo model. We accept it as an axiom.

Example 0.3. Assume we use a book which reduces the draw ratio from 94.3% to 50%. Then according to the BayesElo model, this would magnify normalized Elo differences (for engines which are close together in strength) with a factor

$$\frac{\sqrt{0.5(1-0.5)}}{\sqrt{0.943(1-0.943)}} = 2.16$$

Since $2.16^2 = 4.66$ this means that in this example we would get an almost 80% reduction of games needed to establish the superiority of one engine over another (for engines that are close in strength).

Proof of Proposition 0.1. Below we write $w(\delta) = L(\delta - \gamma)$. The outcome of a game pair then satisfies the probabilities

$$\begin{aligned}(w_1, d_1, l_1) &= (w(e+\delta), 1 - w(e+\delta) - w(-e-\delta), w(-e-\delta))\\ (w_2, d_2, l_2) &= (w(e-\delta), 1 - w(e-\delta) - w(-e+\delta), w(-e+\delta))\end{aligned}$$

We will score games as $-1, 0, 1$. Then the expected score for small e will be

$$\begin{aligned}&w(e+\delta) - w(-e-\delta) + w(e-\delta) - w(-e+\delta)\\ &\qquad \cong (w(\delta) + w'(\delta)e) - (w(-\delta) - ew'(-\delta)) + (w(-\delta) + w'(\delta)e) - (w(\delta) - ew'(\delta))\\ &\qquad = 2e(w'(\delta) + w'(-\delta))\end{aligned}$$

Taken into account that L satisfies the functional equation,

$$L'(x) = L(x)(1 - L(x))$$

we also obtain

$$w'(\delta) + w'(-\delta) = w(\delta) - w(\delta)^2 + w(-\delta) - w(-\delta)^2$$

On the other hand the variance for $e = 0$ will be

$$2((w(\delta) + w(-\delta)) - (w(\delta) - w(-\delta))^2)$$

Hence up to an irrelevant scalar factor the ratio between normalized Elo and BayesElo will be

$$\frac{w(\delta) - w(\delta)^2 + w(-\delta) - w(-\delta)^2}{\sqrt{(w(\delta) + w(-\delta)) - (w(\delta) - w(-\delta))^2}} \tag{1}$$

$$\begin{aligned}d(1-d) &= (1 - w(\delta) - w(-\delta))(w(\delta) + w(-\delta))\\ &= w(\delta) + w(-\delta) - w(\delta)^2 - w(-\delta)^2 - 2w(\delta)w(-\delta)\end{aligned}$$

We may then further approximate (1) as

$$\begin{aligned}\frac{d(1-d) + 2w(\delta)w(-\delta)}{\sqrt{d(1-d) + 4w(\delta)w(-\delta)}} &= \sqrt{d(1-d)}\frac{1 + 2w(\delta)w(-\delta)/(d(1-d))}{\sqrt{1 + 4w(\delta)w(-\delta)/(d(1-d))}}\\ &\cong \sqrt{d(1-d)}\left(1 + 2\left(\frac{w(\delta)w(-\delta)}{d(1-d)}\right)^2\right)\end{aligned}$$

where we have used $(1+2x)/\sqrt{1+4x} \cong 1 + 2x^2$. Let d_0 be the draw ratio corresponding to zero bias. I.e. $d_0 = 1 - 2w(0)$. It is easy to see that in the BayesElo model wl/d is only depends on draw Elo. Hence

$$w(\delta)w(-\delta) = w(0)^2 d/d_0 = (1/4)(1-d_0)^2 d/d_0$$

So we get

$$\frac{w(\delta)w(-\delta)}{d(1-d)} = \frac{(1-d_0)^2}{4d_0(1-d)}$$

and hence

$$\frac{\text{normalized Elo}}{\text{BayesElo}} \sim \sqrt{d(1-d)}\left(1 + \frac{(1-d_0)^4}{8d_0^2(1-d)^2}\right)$$

Under the assumption that $1 - d_0$ is small, this yields our claim. □

Remark 0.4. The term $(1-d_0)^4/(8d_0^2(1-d)^2)$ is a measure for the relative error of the approximation $\sqrt{d(1-d)}$. In the above example $d_0 = 0.943$, $d = 0.5$. We get

$$\frac{(1-d_0)^4}{8d_0^2(1-d)^2} = 6 \times 10^{-6}, \qquad \frac{(1-d_0)^4}{8d_0^2(1-d_0)^2} = 4 \times 10^{-4}$$

which is negligeable. If we take $d_0 = 0.6$, $d = 0.5$ then we get

$$\frac{(1-d_0)^4}{8d_0^2(1-d)^2} = 0.036, \qquad \frac{(1-d_0)^4}{8d_0^2(1-d_0)^2} = 0.056$$

which is still small (and note that the relative error of the quotient $\sqrt{d(1-d)}/\sqrt{d_0(1-d_0)}$ is $0.056 - 0.036 = 0.02$).

References

1. Michel Van den Bergh, *Comments on Normalized Elo*, http://hardy.uhasselt.be/Fishtest/normalized_elo_practical.pdf.
2. Gary Linscott (original author), *The Fishtest framework*, https://github.com/glinscott/fishtest.
3. D. Shawul and R. Coulom, *Paired Comparisons with Ties: Modeling Game Outcomes in Chess* , http://www.grappa.univ-lille3.fr/~coulom/ChessOutcomes.pdf.