

# A Note on Estimating Elo

Isak Ellmer

December 2023

## 1 Introduction

Suppose that two players are playing a game of chess. We want to make some inference on the difference in strength between the two players. Let  $l$ ,  $d$  and  $w$  be the probabilities that player 1 loses, draws or wins respectively. It is not hard to see that player 1 is stronger than player 2 if and only if  $w > l$ . In fact,  $d$  does not matter at all, if we only want to know which of the two players is stronger.

Consider however the two following cases. In the first case,  $l = d = 0$  and  $w = 1$ . And in the second case,  $l = 0$ ,  $d = 0.99$  and  $w = 0.01$ . We see that player 1 is stronger than player 2 in both of these cases, but much more so in the first.

To quantify this strength difference we naturally look at the expected score of a game. If a loss is worth 0 points, a draw  $\frac{1}{2}$  and a win 1, the expected score is given by  $s = \frac{1}{2}d + w$ . The closer  $s$  is to  $\frac{1}{2}$  the more even the games are. And if  $s$  is close to 1 then player 1 wins always every game. We see that in the first case  $s = \frac{1}{2} \cdot 0 + 1 \cdot 1 = 1$ , and in the second case  $s = \frac{1}{2} \cdot 0.99 + 1 \cdot 0.01 = 0.505$ .

We see that player 1 is stronger than player 2 if and only if  $s > \frac{1}{2}$ . This is because  $l + d + w = 1$  and we then have

$$s = \frac{1}{2}d + w = \frac{1}{2}(d + w) + \frac{1}{2}w = \frac{1}{2} + \frac{1}{2}(w - l).$$

The relationship between expected score and Elo difference [2] is given by

$$\Delta E = S^{-1}(s)$$

where

$$S(x) = \frac{1}{1 + 10^{-x/400}}.$$

The inverse of  $S$  is given by

$$S^{-1}(s) = -\frac{400}{\log 10} \log \left( \frac{1}{s} - 1 \right).$$

## 2 Estimating Elo

We make the language a little more rigorous. Let  $X_1, X_2, \dots$  be independent random variables taking the values 0,  $\frac{1}{2}$  and 1 with probabilities  $l$ ,  $d$  and  $w$ . Let  $\hat{s}$  be the random variable given by

$$\hat{s} = \frac{1}{N} \sum_{n=1}^N X_n.$$

The central limit theorem [1] then asserts that

$$\frac{\hat{s} - s}{\sigma/\sqrt{N}}$$

where  $\sigma^2$  is the variance of the variables  $X_n$ , approaches  $N(0, 1)$  in distribution.

We want to construct a confidence interval for  $s$  of confidence grade  $1 - \alpha$ . Let  $\lambda_{\alpha/2}$  be the upper  $\alpha/2$  quantile of  $N(0, 1)$ . Then

$$P\left(-\lambda_{\alpha/2} < \frac{\hat{s} - s}{\sigma/\sqrt{N}} < \lambda_{\alpha/2}\right) \approx 1 - \alpha$$

or equivalently

$$P\left(\hat{s} - \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}} < s < \hat{s} + \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right) \approx 1 - \alpha.$$

Since  $S$  is an increasing function we easily derive the confidence interval

$$P\left(S^{-1}\left(\hat{s} - \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right) < \Delta E < S^{-1}\left(\hat{s} + \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right)\right) \approx 1 - \alpha$$

for  $\Delta E = S^{-1}(s)$ . Note however that unless  $\hat{s} = 0$ , this interval will not be centered around  $\Delta\hat{E} = S^{-1}(\hat{s})$ .

### 2.1 Centered Approximate Confidence Intervals

We will take two approaches. The first approach assumes that the Elo difference is small and that  $N$  is large. The second approach assumes only that  $N$  is large.

We begin with the first approach. We have  $s = S(\Delta E)$ . Expanding  $S$  in its first order Taylor polynomial around 0 implies

$$s = S(\Delta E) \approx S(0) + \Delta E S'(0) = \frac{1}{2} + \Delta E \frac{\log 10}{1600}.$$

Solving for  $\Delta E$  then gives

$$\Delta E \approx \frac{1600}{\log 10} \left(s - \frac{1}{2}\right).$$

Starting with the confidence interval

$$P\left(\hat{s} - \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}} < s < \hat{s} + \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right) \approx 1 - \alpha$$

from Section 2, we subtract  $\frac{1}{2}$  and multiply by the constant  $\beta = \frac{1600}{\log 10}$ . This gives

$$P\left(\beta\left(\hat{s} - \frac{1}{2}\right) - \frac{\beta\lambda_{\alpha/2}\sigma}{\sqrt{N}} < \beta\left(s - \frac{1}{2}\right) < \beta\left(\hat{s} - \frac{1}{2}\right) + \frac{\beta\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right) \approx 1 - \alpha.$$

The assumption that  $\Delta E$  is small now gives

$$P\left(\Delta\hat{E} - \frac{\beta\lambda_{\alpha/2}\sigma}{\sqrt{N}} < \Delta E < \Delta\hat{E} + \frac{\beta\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right) \approx 1 - \alpha.$$

For the second approach we start from

$$P\left(S^{-1}\left(\hat{s} - \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right) < \Delta E < S^{-1}\left(\hat{s} + \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right)\right) \approx 1 - \alpha.$$

We now instead expand  $S^{-1}$  in its first order Taylor polynomial around  $s = \hat{s}$ . We simply get

$$\begin{aligned} & S^{-1}\left(\hat{s} \pm \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}}\right) \\ & \approx S^{-1}(\hat{s}) + \left(\hat{s} \pm \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}} - \hat{s}\right) S^{-1'}(\hat{s}) \\ & = \Delta\hat{E} \pm \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}} \frac{1}{S'(\Delta\hat{E})}. \end{aligned}$$

This implies

$$P\left(\Delta\hat{E} - \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}} \frac{1}{S'(\Delta\hat{E})} < \Delta E < \Delta\hat{E} + \frac{\lambda_{\alpha/2}\sigma}{\sqrt{N}} \frac{1}{S'(\Delta\hat{E})}\right) \approx 1 - \alpha.$$

## References

- [1] Wikipedia contributors. Central limit theorem — Wikipedia, the free encyclopedia, 2023. [https://en.wikipedia.org/w/index.php?title=Central\\_limit\\_theorem&oldid=1188038646](https://en.wikipedia.org/w/index.php?title=Central_limit_theorem&oldid=1188038646).
- [2] Wikipedia contributors. Elo rating system — Wikipedia, the free encyclopedia, 2023. [https://en.wikipedia.org/w/index.php?title=Elo\\_rating\\_system&oldid=1185663969](https://en.wikipedia.org/w/index.php?title=Elo_rating_system&oldid=1185663969).