The plan:

It's basically Elo; we're using the Gaussian rather than Logistic version for convenience.

In the Elo system, the "performance" of players is modelled as a random variable with mean s_i (their skills) and fixed variance:

$$X_i \sim N(s_i, \sigma^2).$$

Then, the probability that player 1 wins is the probability that their performance exceeds that of the other player

$$\begin{aligned}
E_1 &= P(X_1 > X_2) = P\left(\underbrace{\frac{1}{\sigma}\left(X_1 - X_2 - (s_1 - s_2)\right)}_{\sim N(0,1)} > -\frac{1}{\sigma}(s_1 - s_2)\right) = \varphi\left(\frac{s_2 - s_1}{\sigma}\right) \\
E_2 &= P(X_2 > X_1) = \varphi\left(\frac{s_1 - s_2}{\sigma}\right) = 1 - E_1,
\end{aligned}$$

and then (supposing player 1 wins) skills are updated:

$$\begin{aligned}
s_1 &\leftarrow s_1 + k(1 - E_A) \\
s_2 &\leftarrow s_2 + k(\underbrace{0 - E_B}_{\text{actual performance}-\text{expected performance}}),
\end{aligned}$$

where k is some magic constant. A draw is treated as actual performance $= \frac{1}{2}$ and therefore a draw between two equally matched people (so $E_A = \frac{1}{2}$) results in no change in scores.

We need to modify this to handle multiple teams, take into account the margin by which each team won. The key points are that:

- We can calculate a "probability of winning something" from the "skills".
- The size of the score update is proportional to the probability of the result not occuring.

The Setup:

Teams	$I, J \quad$ (index sets)
Skills	$s_i, s_j \quad (i \in I, j \in J)$
Player performance	$U_i \sim N(s_i, \sigma^2)$
	$V_i \sim N(s_j, \sigma^2)$
Team performance	$X_I = \frac{1}{\lvert I\rvert}\sum_{i\in I} U_i + K_{\lvert I\rvert} \sim N\left(\frac{1}{\lvert I\rvert}\sum_{i\in I} s_i + K_{\lvert I\rvert},\quad \frac{1}{\lvert I\rvert}\sigma^2\right)$
	$X_J = \frac{1}{\lvert J\rvert}\sum_{i\in J} V_i + K_{\lvert J\rvert} \sim N\left(\frac{1}{\lvert J\rvert}\sum_{j\in J} s_j + K_{\lvert J\rvert},\quad \frac{1}{\lvert J\rvert}\sigma^2\right)$

where $K_{\{1,2,3,4\}}$ is a magic constant.

The magic constant K is to handle the case where the teams are unbalanced.

$$K_1 = -0.6\sigma \qquad K_2 = 0 \qquad K_3 = 0.2\sigma \qquad K_4 = -0.1\sigma$$

If the teams are the same size, the presence of K has no effect on $X_I - X_J$. However, when they are mismatched, it serves to penalise one of the teams. In words, the above choices assert that given people of equal skill, a three player team is best, followed by two, then four, then one. As an example, those numbers chosen correspond (rougly) to expecting a 10–4 loss when a three player team plays someone on their own.

So, the probability that team I wins any point is:

$$E = P(X_I > X_J) = P\left(\underbrace{\frac{1}{\varphi}(X_I - X_J - \tau)}_{\sim N(0,1)} > \frac{-\tau}{\varphi}\right) = \varphi\left(\frac{\tau}{\varphi}\right)$$

where

$$\tau = \frac{1}{|I|}\sum_{i\in I} s_i - \frac{1}{|J|}\sum_{j\in J} s_j + K_{|I|} - K_{|J|}$$
$$\varphi = \sqrt{\sigma^2\left(\frac{1}{|I|} + \frac{1}{|J|}\right)}$$

We reward team I with $(1-E)m$ points every time they score (and take the same amount from team J; Em points change hands when team J scores (where m is some magic constant).

So if the score were a (team I) : b (team J), the score updates would be

$$s_i \leftarrow s_i + ((1-E)a - Eb)m \qquad \text{each } i \in I$$
$$s_i \leftarrow s_i + (Eb - (1-E)a)m \qquad \text{each } j \in J.$$

Because $(1-E)a - Eb = a - (a+b)E$, you can alternatively, and equivalently, think of this as “update is proportional to difference between actual, and expected score” (if score is $a{:}b$, $a+b$ points were played, and we expected I to win $(a+b)E$ of them).

Oh, and as a bit of fun, we can also “predict” scores:

If $E > \frac{1}{2}$ we predict a score of

$$10 : 10\frac{1-E}{E}$$

and if $E < \frac{1}{2}$

$$10\frac{E}{1-E} : 10$$

that is, multiplying $E : (1-E)$ up so that the larger one is equal to ten.