THE ACCOUNTING IDENTITY

MICHEL VAN DEN BERGH

Let $X_1, Y_1, X_2, Y_2, \ldots, X_n, Y_n \in \{0, 1/2, 1\}$ be the outcomes of a match between two chess engines. The openings of $X_1, \ldots, X_n$ are picked from an opening book and $Y_1, \ldots, Y_n$ are played with the same openings but with reversed colors. We put

$$\mu_i := E(X_i),$$
$$\mu_i^\circ := E(Y_i).$$

The *match score* is defined as

$$S := \frac{X_1 + Y_1 + \cdots + X_n + Y_n}{2n}.$$

In the trinomial model the $(X_i)_i$, $(Y_i)_i$ are (incorrectly) considered as i.i.d random variables. The trinomial model leads to the following (also incorrect) estimate for the variance of the match score

$$\mathrm{Var}_3(S) := \frac{1}{2n} \frac{(X_1 - S)^2 + (Y_1 - S)^2 + \cdots + (X_n - S)^2 + (Y_n - S)^2}{2n}.$$

In the (more correct) pentanomial model we put $Z_i = (X_i + Y_i)/2$. We have

$$S = \frac{Z_1 + \cdots + Z_n}{n}$$

and, assuming the $(Z_i)_i$ are i.i.d. random variables, the pentanomial estimate for the variance of S is

$$\mathrm{Var}_5(S) := \frac{1}{n} \frac{(Z_1 - S)^2 + \cdots + (Z_n - S)^2}{n}.$$

Hence the trinomial variance, *normalized per game* is

$$V_3 := 2n \, \mathrm{Var}_3(S) = \frac{(X_1 - S)^2 + (Y_1 - S)^2 + \cdots + (X_n - S)^2 + (Y_n - S)^2}{2n}$$

whereas the pentanomial variance, *normalized per game* is

$$V_5 := 2n \, \mathrm{Var}_5(S) = 2 \frac{(Z_1 - S)^2 + \cdots + (Z_n - S)^2}{n}.$$

Assuming small elo differences, and additivity of elo, we have

$$\mu_i = s + b_i, \qquad (0.1)$$
$$\mu_i^\circ = s - b_i.$$

where b_i is the *bias* of the i'th opening position and s is de expected match score for balanced positions between the given engines.

Theorem 0.1 (The accounting identity)**.** *Assume that the* $2n$ *random variables* $(X_i)_i, (Y_i)_i$ *are independent and that additivity of elo in the sense of* (0.1) *holds. Then under reasonable regularity conditions on* $(\mu_i)_i$, $(\mu_i^\circ)_i$ *we have*

$$V_3 - V_5 = \frac{1}{n}\sum_{i=1}^{n} b_i^2 + O(n^{-1/2}). \tag{0.2}$$

We call (0.2) an *accounting identity* since it basically amounts to rewriting some sums.

Assuming that the opening positions are picked randomly and b is the random variable representing the bias of a position, the formula (0.2) may be rewritten as

$$V_3 - V_5 = E(b^2) + O(n^{-1/2}).$$

We call

$$\sqrt{E(b^2)}$$

(possibly converted to Elo for clarity) the *RMS bias* of an opening book.

Proof of Theorem 0.1. We have

$$\begin{aligned} V_5 = 2n\,\mathrm{Var}_5(S) &= 2\frac{(Z_1 - S)^2 + \cdots + (Z_n - S)^2}{n} \\ &= \frac{(X_1 + Y_1 - 2S)^2 + \cdots + (X_n + Y_n - 2S)^2}{2n}. \end{aligned}$$

Hence

$$\begin{aligned} V_3 - V_5 &= \frac{(X_1 - S)^2 + (Y_1 - S)^2 - (X_1 + Y_1 - 2S)^2 + \cdots}{2n} \\ &= -\frac{(X_1 - S)(Y_1 - S) + \cdots + (X_n - S)(Y_n - S)}{n} \end{aligned}$$

We have

$$(X_i - S)(Y_i - S) = X_iY_i - (X_i + Y_i)S + S^2$$

$$\begin{aligned} \sum_i (X_i - S)(Y_i - S) &= \sum_i X_iY_i - S\sum_i (X_i + Y_i) + nS^2 \\ &= \sum_i X_iY_i - 2nS^2 + nS^2 \\ &= \sum_i X_iY_i - nS^2 \end{aligned}$$

and hence

$$V_3 - V_5 = -\frac{\sum_i X_iY_i}{n} + S^2.$$

By the central limit theorem we have

$$S = E(S) + O(n^{-1/2})$$

so that (assuming reasonable $(\mu_i)_i$, $(\mu_i^\circ)_i$)

$$S^2 = E(S)^2 + O(n^{-1/2}).$$

The random variables X_iY_i are independent but not identically distributed. Nonetheless, under very weak hypothesis [1], we may assume that the central limit theorem applies, so that

$$\frac{\sum_i X_iY_i}{n} = \frac{\sum_i E(X_iY_i)}{n} + O(n^{-1/2})$$

Thus

$$\begin{aligned} V_3 - V_5 &= -\frac{\sum_i E(X_i)E(Y_i)}{n} + E(S)^2 + O(n^{-1/2}) \\ &= -\frac{\sum (E(X_i) - E(S))(E(Y_i) - E(S))}{n} + O(n^{-1/2}) \\ &= \frac{\sum_i b_i^2}{n} + O(n^{-1/2}) \end{aligned}$$

□

References

1. *Lyapunov theorem*, `https://www.encyclopediaofmath.org/index.php/Lyapunov_theorem`.