Probability generating functions for soccer

Mark A. van de Wiel[*]

The probability generating function (PGF) is a multivariate polynomial consisting of terms $pw^a x^b y^c z^d$, which means the following: the probability that $W = a, X = b, Y = c$ and $Z = 6$ is equal to p. Example. Denote the number of points after round 1 by W_1, X_1, Y_1, and Z_1. Then, after one round (one match team t_W against t_X, $t_W - t_X$, one match $t_Y - t_Z$) with probabilities (0.7,0.1,0.2) and (0.5,0.2,0.3) we have, by enumerating all nine combinations of wins, draws and losses,

$$\begin{aligned}
&\mathrm{PGF}(W_1, X_1, Y_1, Z_1)(x, w, y, z) \\
&\quad = 0.7 * 0.5w^3x^0y^3z^0 + 0.7 * 0.2w^3x^0y^1z^1 + 0.7 * 0.3w^3x^0y^0z^3 \\
&\quad + 0.1 * 0.5w^1x^1y^3z^0 + 0.1 * 0.2w^1x^1y^1z^1 + 0.1 * 0.3w^1x^1y^0z^3 \\
&\quad + 0.2 * 0.5w^0x^3y^3z^0 + 0.2 * 0.2w^0x^3y^1z^1 + 0.2 * 0.3w^0x^3y^0z^3.
\end{aligned}$$

A very nice property of PGFs is that we may multiply these for summing independent events. That is denote the total number of point after all six rounds $W = \sum_{i=1}^{6} W_i$, where W_i is the number of points for team t_W in round i. Likewise for X, Y and Z. Then, we simply have:

$$\mathrm{PGF}(W, X, Y, Z) = \prod_{i=1}^{6} \mathrm{PGF}(W_i, X_i, Y_i, Z_i)(w, x, y, z).$$

Hence, all we need to do to obtain all possible end rankings (= all combinations of number of points for all four teams) is to multiply four-variate polynomials.

The R-script makes use of the R-package `mpoly` which implements multiplication of multivariate polynomials. Finally, to determine the probability that team t_W ends up in the top 2 we simply add all probabilities p of terms $pw^a x^b y^c z^d$ for which a is larger than at least two other exponents. Ties are broken by spreading the probability equally. E.g. the term $pw^{10}x^{10}y^{10}z^4$, contributes $2/3p$, as there are two spots for the first three teams, which have the same number of points.

[*]Email: mark.vdwiel@amsterdamumc.nl