EURO 2020 predictions: final

Leonardo Egidi - DEAMS, University of Trieste, Italy. Mail: legidi@units.it

8 July 2021

Contents

The statistical model (in brief)

We use a **double Poisson model with dynamic team-specific abilities** for the attack and the defence. Let (X_i, Y_i) denote the random number of goals scored by the home and the away team in the i-th game, $i = 1, \dots, n$, respectively. ranking denotes the Coca-Cola FIFA ranking at May 27th, 2021, whereas att and def denote the attack and the defence abilities, respectively.

$$
\begin{aligned}
X_i|\lambda_{1i} &\sim \text{Poisson}(\lambda_{1i}), && (1)\\
Y_i|\lambda_{2i} &\sim \text{Poisson}(\lambda_{2i}), && (2)\\
\log(\lambda_{1i}) &= \text{home} + \text{att}_{h_i,t} + \text{def}_{a_i,t} + \frac{\gamma}{2}(\mathsf{ranking}_{h_i} - \mathsf{ranking}_{a_i}) && (3)\\
\log(\lambda_{2i}) &= \text{att}_{a_i,t} + \text{def}_{h_i,t} - \frac{\gamma}{2}(\mathsf{ranking}_{h_i} - \mathsf{ranking}_{a_i}), \quad i = 1, \dots, n \text{ (matches)}, && (4)\\
\text{att}_{k,t} &\sim \mathcal{N}(\text{att}_{k,t-1}, \sigma^2), && (5)\\
\text{def}_{k,t} &\sim \mathcal{N}(\text{def}_{k,t-1}, \sigma^2), && (6)\\
\sum_{k=1}^{n_t} \text{att}_{k,} = 0,\ &\sum_{k=1}^{n_t} \text{def}_{k,} = 0, \quad k = 1, \dots n_t \text{ (teams)},\ t = 1, \dots, T \text{ (times)}. && (7)
\end{aligned}
$$

Lines (1)-(2) display the likelihood's equations (two Poisson distributions); lines (3)-(4) display the log-linear models for the scoring rates λ_1, λ_2; lines (5)-(6) display the dynamic prior distributions for the attack and the defence parameters, respectively; line (7) displays the sum-to-zero identifiability constraints. Model fitting has been obtained through the Hamiltonian Monte Carlo sampling, 2000 iterations, 4 chains (`rstan` package). The historical data used to fit the models come from: **Nations' League** (2019-2020), **Euro UEFA Qualifiers** (2020-2021), **World Cup UEFA Qualifiers** (2021), **UEFA Euro 2020** (groupstage + round of 16 + quarter of finals + semifinals matches).

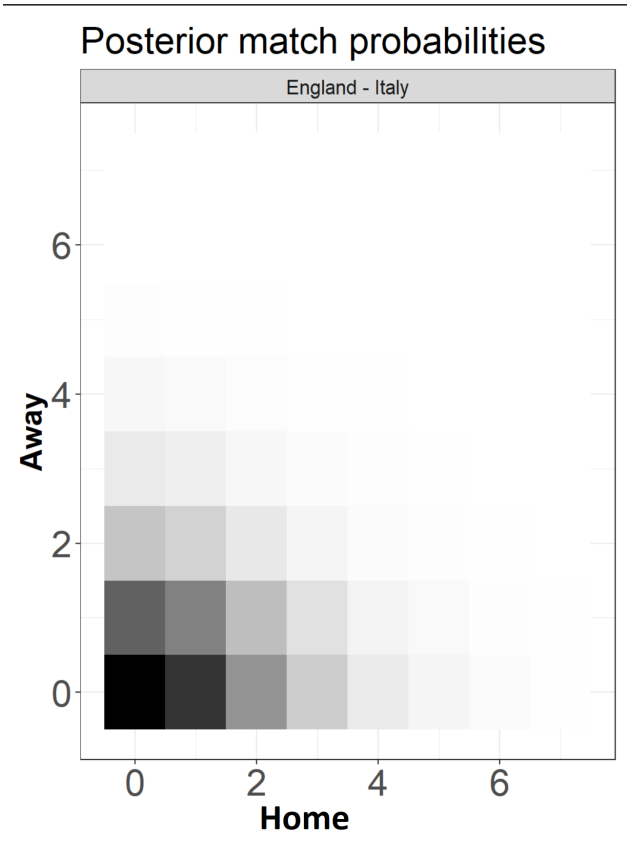
The idea is to provide a dynamic predictive scenario: at the end of each match-day, the model will be refitted to predict the remaining matches.

Final prediction (11 July)

Posterior matches probabilities from the posterior predictive distribution of the model above are displayed in the table below. **mlo** denotes the most likely exact outcome (in parenthesis, the corresponding posterior probability). Darker regions in the plots below denote more likely outcomes: on the x-axis the home goals, on the y-axis the away goals.

Attention: the matches probabilities below refer to the results within the regular 90 minutes.

home	away	home win	draw	away win	mlo
England	Italy	0.415	0.331	0.254	0-0 (0.213)



Expected number of goals

We compute also the **expected number of goals** λ_1, λ_2 for each match, obtained by computing the median values from the MCMC sampling for the scoring rates.

home	away	exp_home	exp_away
England	Italy	0.91	0.62

Estimated attack/defence abilities

In the plot below we display the posterior intervals for the **attack** (red) and **defence** (blue) abilities estimated through the training set matches, from **October 2019** until the **semifinals**: the higher the attack and the lower the defence values for a given team, and the better is the estimated overall team's ability.