World Cup Qatar 2022 groupstage predictions: 1st match day

Leonardo Egidi, Vasilis Palaskas - Mail: legidi@units.it, vasilis.palaskas94@gmail.com

19 November 2022

Contents

The statistical model (in brief)

We use a **diagonal-inflated Bivariate-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. $\mathsf{ranking}$ denotes the Coca-Cola FIFA ranking at October 6th, 2022, whereas att and def denote the attack and the defence abilities, respectively.

$$
\begin{aligned}
(X_i, Y_i) \ &\sim \ \begin{cases} (1-p)\text{BP}(x_i, y_i|\lambda_1, \lambda_2, \lambda_3) & \text{if } x \neq y \\ (1-p)\text{BP}(x_i, y_i|\lambda_1, \lambda_2, \lambda_3) + pD(x, \eta) & \text{if } x = y, \end{cases}, && (1)\\
\log(\lambda_{1i}) =& \ \text{att}_{h_i,t} + \text{def}_{a_i,t} + \frac{\gamma}{2}(\mathsf{ranking}_{h_i} - \mathsf{ranking}_{a_i}) && (2)\\
\log(\lambda_{2i}) =& \ \text{att}_{a_i,t} + \text{def}_{h_i,t} - \frac{\gamma}{2}(\mathsf{ranking}_{h_i} - \mathsf{ranking}_{a_i}), \quad i = 1, \dots, n \text{ (matches)}, && (3)\\
\log(\lambda_{3i}) =& \ \eta, && (4)\\
\text{att}_{k,t} \sim& \ \mathcal{N}(\text{att}_{k,t-1}, \sigma^2), && (5)\\
\text{def}_{k,t} \sim& \ \mathcal{N}(\text{def}_{k,t-1}, \sigma^2), && (6)\\
\eta, \ \gamma \sim& \ \mathcal{N}(0, 1) && (7)\\
p \sim& \ \text{Uniform}(0, 1) && (8)\\
\sum_{k=1}^{n_t} \text{att}_{k,} = 0,& \ \sum_{k=1}^{n_t} \text{def}_{k,} = 0, \quad k = 1, \dots n_t \text{ (teams)}, \ t = 1, \dots, T \text{ (times)}. && (9)
\end{aligned}
$$

Lines (1) displays the likelihood's equations (diagonal inflated bivariate Poisson); lines (2)-(4) display the log-linear models for the scoring rates λ_1, λ_2 and the covariance parameter λ_3; lines (5)-(6) display the dynamic prior distributions for the attack and the defence parameters, respectively; lines (7)-(8) display prior distributions for the other model parameters; line (9) displays the sum-to-zero identifiability constraints. Model fitting has been obtained through the Hamiltonian Monte Carlo sampling, 2000 iterations, 4 chains using the `footBayes` `R` package (with the underlying `rstan` package). The historical data used to fit the models come from *all the international matches played during the years' range 2018-2022.*

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.

Groupstage predictions: 1st day (20-24 November)

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.

home	away	home win	draw	away win	mlo
Qatar	Ecuador	0.228	0.255	0.516	0-1 (0.134)
England	Iran	0.479	0.310	0.210	1-0 (0.183)
Senegal	Netherlands	0.124	0.241	0.635	0-1 (0.182)
United States	Wales	0.461	0.300	0.239	1-0 (0.169)
Argentina	Saudi Arabia	0.715	0.223	0.061	1-0 (0.213)
Denmark	Tunisia	0.598	0.247	0.156	1-0 (0.166)
Mexico	Poland	0.410	0.282	0.308	1-0 (0.131)
France	Australia	0.619	0.254	0.127	1-0 (0.192)
Morocco	Croatia	0.285	0.301	0.414	0-0 (0.147)
Germany	Japan	0.414	0.266	0.319	1-0 (0.12)
Spain	Costa Rica	0.701	0.219	0.079	1-0 (0.191)
Belgium	Canada	0.625	0.227	0.148	1-0 (0.144)
Switzerland	Cameroon	0.509	0.283	0.208	1-0 (0.171)
Uruguay	South Korea	0.482	0.282	0.236	1-0 (0.15)
Portugal	Ghana	0.828	0.130	0.042	2-0 (0.157)
Brazil	Serbia	0.762	0.167	0.071	2-0 (0.148)