World Cup Qatar 2022 groupstage predictions: 2nd match day

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

24 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. ranking denotes the Coca-Cola FIFA ranking at October 6th, 2022, whereas att and def denote the attack and the defence abilities, respectively.

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

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: 2nd match-day day (25-28 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
Wales	Iran	0.324	0.307	0.369	0-0 (0.145)
Qatar	Senegal	0.244	0.294	0.462	0-1 (0.159)
Netherlands	Ecuador	0.603	0.243	0.154	1-0 (0.153)
England	United States	0.493	0.270	0.237	1-0 (0.142)
Tunisia	Australia	0.363	0.317	0.320	0-0 (0.164)
Poland	Saudi Arabia	0.400	0.320	0.280	0-0 (0.167)
France	Denmark	0.436	0.291	0.273	1-0 (0.144)
Argentina	Mexico	0.616	0.261	0.123	1-0 (0.197)
Japan	Costa Rica	0.644	0.237	0.119	1-0 (0.174)
Germany	Spain	0.218	0.244	0.538	0-1 (0.119)
Belgium	Morocco	0.571	0.267	0.163	1-0 (0.173)
Croatia	Canada	0.435	0.290	0.275	1-0 (0.146)
Cameroon	Serbia	0.208	0.277	0.515	0-1 (0.154)
Brazil	Switzerland	0.712	0.208	0.081	1-0 (0.177)
Portugal	Uruguay	0.413	0.289	0.298	1-0 (0.135)
South Korea	Ghana	0.602	0.254	0.144	1-0 (0.172)