World Cup Qatar 2022 groupstage predictions: 3rd match day

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

28 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: 3rd match-day day (29 November-2nd December)

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 favorite team goals, on the y-axis the underdog team goals.

favorite	underdog	favorite win	draw	underdog win	mlo
Ecuador	Senegal	0.389	0.302	0.309	1-0 (0.145)
Netherlands	Qatar	0.767	0.160	0.073	2-0 (0.133)
United States	Iran	0.402	0.297	0.301	1-0 (0.139)
England	Wales	0.624	0.236	0.140	1-0 (0.158)
France	Tunisia	0.643	0.236	0.122	1-0 (0.159)
Denmark	Australia	0.546	0.274	0.180	1-0 (0.165)
Argentina	Poland	0.658	0.229	0.112	1-0 (0.165)
Mexico	Saudi Arabia	0.436	0.322	0.243	0-0 (0.179)
Belgium	Croatia	0.470	0.260	0.270	1-0 (0.119)
Morocco	Canada	0.428	0.313	0.259	0-0 (0.166)
Spain	Japan	0.575	0.251	0.175	1-0 (0.148)
Germany	Costa Rica	0.646	0.224	0.130	1-0 (0.148)
Portugal	South Korea	0.611	0.236	0.153	1-0 (0.145)
Uruguay	Ghana	0.681	0.208	0.111	1-0 (0.145)
Switzerland	Serbia	0.393	0.284	0.322	1-0 (0.121)
Brazil	Cameroon	0.796	0.162	0.041	2-0 (0.186)