

Predictive bookmaker consensus for FIFA World Cup 2018

Andreas Wittmann

17.05.2018

Motivation

Football is a simple game. Twenty-two men chase a ball for 90 minutes and at the end, the Germans always win.

– Gary Lineker

- Prediction of the FIFA World Cup 2018 based on bookmakers odds.
- Other approaches use historical data based on previous games.
- Bookmakers odds are an assessment of the expected chance to win a specific tournament.
- Bookmakers have try to rate the teams correctly to earn their money.

Steps to forecast the winner of the 2018 FIFA World Cup:

1. Obtain the long-term winning odds from online bookmakers
(<https://www.oddschecker.com/football/world-cup/winner>).
2. Transform the odds to winning probabilities.
3. The odds infer the ability of a team. Using these abilities pairwise winning probabilities can be calculated.
4. Using the pairwise winning probabilities the whole tournament can be simulated 100,000 times.
5. Therefore we can obtain relative frequencies for each team winning the tournament.

Online Bookmakers (sample)

group	team	code	Bet 365	Sky Bet	Ladbrokes	William Hill	Marathon Bet
A	Egypt	EGY	150	150	150	150	200
A	Russia	RUS	40	40	40	50	40
A	Saudi Arabia	KSA	1000	1000	1000	1000	1000
A	Uruguay	URU	33	28	25	25	33
B	Iran	IRN	500	750	500	500	500
B	Morocco	MAR	500	250	250	300	400

Bookmaker quoted odds

The quoted odds for Team i are calculated as:

$$\textit{quoted odds}_i = \textit{odds}_i \cdot \delta + 1$$

- $+1$ represents the *sake*, the bettor always receives his original stake back.
- Odds 4/1 (four-to-one), the bettor can make 400 € profit on 100 € stake.
- One would receive a total of 500 € in return upon winning.
- $\delta < 1$ is the proportion of bets that is actually paid out by the bookmakers.
- The *overround* is the remaining proportion $1 - \delta$, which is the main basis of the bookmakers profits.

Odds

Calculate the odds for Team i as:

$$odds_i = (quoted\ odds_i - 1)/\delta$$

Using this we can calculate the probabilities as

$$p_i = 1 - \frac{odds_i}{1+odds_i}$$

- It is assumed that each bookmakers δ is constant across the various teams in the tournament.
- The δ can be chosen such that $\sum_i p_i = 1$.
- We have to find δ and remove the overround.

Overround Example

```
quotedOdds <- c(1.6, 2.3, 16, 30)  
delta <- calcDelta(quotedOdds)  
delta
```

```
## [1] 0.7579091
```

```
calcOdds(quotedOdds, delta)
```

```
## [1] 0.7916516 1.7152452 19.7912904 38.2631615
```

```
calcProbs(quotedOdds, delta)
```

```
## [1] 0.55814422 0.36829087 0.04809706 0.02546917
```

Aggregation

- Aggregate the overround-adjusted odds across the bookmakers
- Transformation to the logit scale for averaging:

$$\text{logit}(p_i) = \frac{1}{n_b} \sum_{b=1}^{n_b} \text{logit}(p_{i,b})$$

- There are $b = 1, \dots, n_b$ bookmakers.
- $p_{i,b}$ represents the winning probability of team i from bookmaker b .
- logit represents the logit-function $\log(x/(1 - x))$.
- $\text{logit}(p_i)$ can be transferred back using the inverse logit-function $\exp(p)/(1 + \exp(p))$.

Pairwise comparisons

- The odds without overround infer the ability of a team.
- Using these abilities pairwise winning probabilities can be calculated.
- The *Bradley-Terry-Modell* can be used to calculate the winning probability that team A strikes team B as

$$P(A \text{ strikes } B) = \frac{\text{ability}(A)}{(\text{ability}(A) + \text{ability}(B))}$$

Inverse Tournament

Simulation based on team-specific abilities is used:

1. Using the team abilities, pairwise winning probabilities can be derived for each possible match.
2. The whole tournament can be simulated to see which teams proceeds to which stage and which team finally wins.
3. Run the tournament simulation sufficiently often (about 100.000 times) to obtain the relative frequencies for each team winning the tournament.

An iterative approach is used to find team abilities which closely match the boomaker consensus probabilities.

Aggregated Input-Data

code	group	logOdds	probabilities
EGY	A	-5.363374	0.0046632
RUS	A	-3.970998	0.0185057
KSA	A	-7.151055	0.0007834
URU	A	-3.692737	0.0242986
IRN	B	-6.531814	0.0014542
MAR	B	-6.268229	0.0018920
POR	B	-3.446987	0.0308588
ESP	B	-1.912633	0.1286854
AUS	C	-6.289958	0.0018514
DEN	C	-4.828222	0.0079372
FRA	C	-1.952347	0.1242976
PER	C	-5.557156	0.0038449

Implementation Details (1)

```
public static final int N_SIMULATIONS = 100000;  
public static final int N_ITERATIONS = 500;  
  
for (int i = 0; i < Util.N_ITERATIONS; i++) {  
    for (int j = 0; j < Util.N_SIMULATIONS; j++) {  
        simulateGroupStage();  
        simulateKnockoutStage();  
    }  
  
    if (checkTeamAbility(i)) {  
        break;  
    }  
}
```

Implementation Details (2)

```
// checkTeamAbility(i)

final List<Team> teams

teams.forEach(team -> {
    final double calculatedProbability =
        team.knockoutStage.countWinner.doubleValue() /
        Util.N_SIMULATIONS;
    final double diff = team.probability - calculatedProbability;
    team.ability += diff / 2.0;

    this.errorSum += Math.abs(team.probability -
        calculatedProbability);
});
```

Implementation Details (3)

- Use $ability = \exp(probability)$ as initial ability value.
- Use some relative target error (relative target error = 0.002 takes 141 iterations)
- Count final group standings, how often which teams reaches which round, count winners, ...
- Calculate relative frequencies.

Literature

- Predictive bookmaker consensus model for the UEFA Euro 2016 (<https://www2.uibk.ac.at/downloads/c4041030/wpaper/2016-15.pdf>)