**Statistical predictions for Euro 2020 and betting scenarios**

*Statistical models for football: an historical overview*

The art of statistical modelling has precious implications in the whole field of sport, with a particular attention to the ability of obtaining reliable predictions for future matches/games.

The use of statistics for football dates back to the mid of the XX century, with the seminal work from Charles Reep and Bernard Benjamin in 1968; a great advancement has been achieved in the ‘80s with the work of Micheal Maher, whereas the work of Stuart Coles and Mark Dixon in 1997 and Ioannis Ntzoufras and Dimitris Karlis in 2003 proposed some innovative ideas about goals’ correlation.

The basic idea of the current statistical models for football is to use a well-known probability distribution---the Poisson---to model the number of goals scored by the home team and the away team, respectively. We explain here how we set up a predictive model for the Euro Cup, with the aim to provide statistical predictions at least as much accurate as the bookmakers odds.

(Provide here, if you want, my website for more technical details:

https://www.leonardoegidi.com/euro-2020)

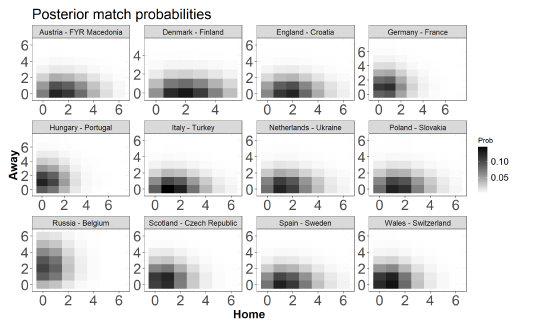
*A model for the Euro 2020 Cup*

We are all carefully following the football Uefa Euro Cup 2020 and, as amateurs, we would like to be able to predict in advance some of the matches, mixing the Apollo oracle capabilities and our personal prediction skills. However, without data and statistics our prediction are just rude, subjective, possibly driven by our personal feelings and, as such, not relying on scientific grounds.

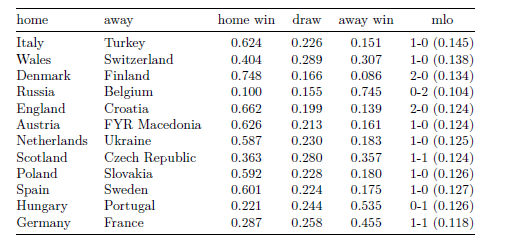
Once we collected historical data---the match results from the Nations’ League, the Euro Cup qualifiers, and the World Cup qualifiers for UEFA national teams---we may use this bag of past results to “train” the model and then obtain the predictions for the first matches from the Euro 2020. This means that the model uses all the past results and should be able to generalize the teams’ strengths by extrapolating what is more likely to happen in the future.

Should I say something more about the model? Something more technical? Maybe the fact to use attack/defensive skills?

Of course, football is a rare-events sport and, as such, is hardly predictable, even with the best statistical model. As a final output, we provide the probabilities for the home win, the draw and the away win, along with some probability plots for highlighting the most likely exact results for the first 12 matches from the Euro 2020 group-stage, as shown below.



These plots can be very useful to grasp the uncertainty surrounding the exact match results in a football match. For illustration purposes only, let us focus on the most fascinating match from the first match-day of the group-stage, Germany against France. Darker regions denote more likely results: the black square is in correspondence of the draw 1-1, with an associated probability of 0.1 (10%), whereas dark grey squares are in correspondence of results such as 1-0, 0-1 and 0-0, with associated probabilities of about 0.06 (6%) each. The model acknowledges a non-negligible balance in this match, neither of the two teams appears to be favorite over the other. Moreover, the model seems to predict a low final number of goals: both the teams exhibit high defensive performance in the recent past, for such a reason the predicted number of goals for the current match is not much high. However, if we look at the table below, we have some more precise suggestions: the table reports the model’s probabilities for the home win, the draw and the away win, along with the most likely exact outcome (mlo) among the exact results. So, according to these probabilities, under the model France is favorite to win the match (prob. 0.455), even though the mlo for this match is 1-1. Actually, France won the match 1-0, with a final result quite plausible under the model (the 1-0 was the second mlo for this match). But what can we say about model’s probability accuracy? And which were the bookmakers feeling about this match?



*A betting experiment*

The majority of the bookmakers indicated the Germany as the match favorite: in other way said, the odds for Germany were lower than those for France. Betting 1 euro on Germany’s win would have yielded, say, 2.5 euro, whereas betting 1 euro on France would have given, say, 2.7 euro. So, according to the bookmakers, France’s win was less likely than Germany’s win. However, the model said something different: for such a reason, betting on France’s win could be advantageous here, since the model probability exceeds the bookmakers probability, and, as such, yields a positive expected net profit. As statisticians, we should always look for these mis-matches between the probabilities implied by our model/algorithm and those derived from the bookmakers: when our model’s probability for a given event is greater than that implied by the bookmaker for the same event, there is in fact a profit’s chance.

We applied this way of reasoning for all the matches from the group-stages, by proposing a betting experiment driven by mathematical and probabilistic rules. Regarding this match, Germany-France, we bet 20 euros on the France’s win, winning 54 euros. Beyond this, we also bet on some “low-goals” scenarios, such as the 0-0, 0-1, 1-1, 1-0, that payed about 8.5, 9, 7 and 6.5 times the 1 euro bet and that were the most likely outcomes from the model. At the end of the match, our final profit operation was:

20 euros on France’s win 🡪 + 54 euros

10 euros on 1-1 🡪 - 10 euros

10 euros on 0-1 France 🡪 + 90 euro

10 euros on 1-0 Germany 🡪 -10 euro

Total: net profit of 124 euro.

This is a small example on how obtaining good probabilities from a model and how use them to set up a wise betting strategy. But remind that betting can be dangerous, and even smart statisticians can loose a lot of money.