

---

# Sports betting: A source for empirical Bayes

Mateo Graciano Londoño\*  
Andrés Ramírez Hassan,†

EAFIT University  
Medellín, Colombia

August 21, 2015

## 1 Problem statement

Sports forecasting is a topic which have been widely studied in the literature (Stekler et al., 2010), where betting odds have played a remarkable role in this aim. Some methodologies have been based on a frequentist approach (Leitner et al., 2010), whereas others on a Bayesian framework (Baio and Blangiardo, 2010).

Regarding the latter, one of the main concerns is related to building the prior distribution, we propose an empirical Bayes approach based on betting odds to obtain the hyper-parameters of some simple conjugate families, as in Andrés Ramírez (2014), to forecast the outcomes in some selected tournaments such as the US open (tennis), 2013-14 premier league (football) and 2014-15 NBA. This approach allows to incorporate prior information based on betting odds and historical information to forecast outcomes, and design a strategy to obtain profits in betting markets

Using a Dirichlet categorized model, we obtain the following posterior(Andrés Ramírez, 2014) :

$$\pi(\mathbf{p}|Data) = \frac{\Gamma(\sum_{i=1}^k \alpha_i + c_i)}{\prod_{i=1}^k \Gamma(\alpha_i + c_i)} \prod_{i=1}^k p_i^{\alpha_i + c_i - 1} \quad (1)$$

where  $k$  is the number of possible outcomes,  $c_i$  is the number of times the event  $i$  had happen,  $p_i$  is the probability for the event  $i$  and  $\alpha$  is the vector of hyper-parameters of the prior.

To estimate  $\alpha$  we must consider that is each book-marker overestimate the associated probabilities to their betting odds, considering  $o_i$  as the betting odd for the event  $i$ ,  $\sum_{i=1}^k \frac{1}{o_i} \geq 1$  so is necessary to obtain the associated "fair" probabilities, using Shin (1993) methodology:

$$p(z)_i^{bettingodds} = \frac{\sqrt{z^2 + 4(1-z) \frac{(1/o_i)^2}{\sum_l (1/o_l)}} - z}{2(1-z)}$$

---

\*Student of Mathematical Engineer, EAFIT University, Medellín, Colombia

†Tutor Professor, Department of Economics, EAFIT University, Medellín, Colombia

---

So that:

$$\sum_{i=1}^k p_i^{\text{betting odds}} = 1$$

So after the process of estimation of  $\alpha$  there is everything that is needed for drawing the probabilities from (1) and then simulate the event and finally obtain the probabilities of each possible outcome.

## 2 Objectives

### 2.1 General Objective

To propose a methodology to forecast and design a betting strategy in sports competitions based on an empirical Bayes approach.

### 2.2 Specific Objectives

- Review literature regarding to the topic of sports forecasting
- Gather results and betting odds from US open 2014, premier league 13-14, FIFA world cup 2014 and NBA 14-15
- Check performance of a specific model at several sports
- Check a betting strategy in order to conclude if either is profitable or not

## 3 Background

Sports forecasting have had a lot of attention, there are different ways for approaching this matter, first, we mention the statistical models which uses sport-related inputs, then the expert tipsters ([Štrumbelj, 2014](#)), also there are the ones which are based on prediction markets ([Spann and Skiera, 2009](#)) and finally the models that are based on betting odds ([Vlastakis et al., 2009](#)).

Bets have been used for different issues and there is also empirical evidence suggesting that they are the best (most accurate) publicly source of information for sports forecasting ([Štrumbelj, 2014](#)), on their work they also conclude that forecasting from betting odds are better than any other approach, or at least as good.

In [Stekler et al. \(2010\)](#) there is a great work on the analysis of the results on the forecasting at different sports, with three different approaches, they also show that each sport has a different way of betting, and that leads to different ways of forecasting being the best approach for each sport. That paper is focused on the forecasting power, not on a profitable betting strategy which is what matter in most other papers.

[Spann and Skiera \(2009\)](#) conclude on their work that betting odds are the best source for sport forecasting but we can see at [Štrumbelj \(2014\)](#) that there are different ways to obtain the associated probability from a betting odd and they conclude that the [Shin \(1993\)](#) methodology is the most accurate for calculate the probabilities among the different methods that had been used before.

---

## 4 Justification

Today, on our world where there are lot of fans of every kind of sports and also a growing number of bets on-line there is a important issue in the area of sport forecasting (Stekler et al., 2010). In fact, there are different kind of source for forecasting such as predictions from the experts, the prediction markets and also from betting bookmarkers. On the literature there are a lot of works regarding to sports forecasting, but different approaches are used for different kind of sports. For instance, there is a model for basketball using a Markov model (Štrumbelj and Vračar, 2012) and in the other hand there are different approaches for football such as one based on betting odds (Strumbelj and Sikonja, 2010).

To sum up, the main idea is to use a Bayesian model to include betting odds and historical data in order to generalize a model so that it can be used for different kind of sports with no significant differences in the model. That could lead us to create a betting strategy towards a profitable one.

## 5 Scope

The project focuses on the development of a model which can be used, with a few modifications, to forecast the results of different kind of sports. It would be interesting if the model was tested at every different sport but the main objective is to test it just in football, basketball and tennis because those are representatives sports, tennis only has two possible outcomes and it is individual, basketball is a team sport and has two possible outcomes while football is a team sport as well but has three possible outcomes.

Furthermore, after checking the model performance, the idea is to focus on a betting strategy and check how it would have worked in previous tournaments and conclude if either it would had been profitable or not.

## 6 Methodology

The first stage is understanding the basics of Bayesian statistics and implement some models with this characteristics, then the goal is a literature review regarding to sports forecasting and also previous works based on betting odds.

The second stage is to estimate the probabilities of each possible outcome at several sports, the first sub-stage is implementing the model from Andrés Ramírez (2014) for the case of tennis, for it is needed historic information and also information for betting odds (both will be gather), the second sub-stage is to proceed in the same way but for the football case. The last sub-stage is to check the model prediction power.

The last stage is the development of the betting strategy and check if it would have been profitable in the past.

---

## 7 Schedule

Dates	Activity
July 1st - 31th	Study Bayesian methods
July 20st - August 2nd	Literature Review
August 1st - 6th	pre-project
August 14th	Proposal presentation
August 14th - september 18th	Model development
September 25th	Oral progress report
September 28th - October 5th	Betting strategy development
October 5th - 12th	Check performace
October 12th - 30th	Write the final report
November 6th	Final project report
November 1st - 19th	Preparation of final project presentation
November 20th	Final project presentation

## 8 Budget

This research will not required any budget, because EAFIT University provides data bases for the literature review, software licenses to implement the computer model and the tutor professor.

## 9 Intellectual property

According to the internal regulation on intellectual property within EAFIT University, the results of this investigation practice are product of Mateo Graciano Londoño as student and Andrés Ramírez Hassan as tutor professor.

In case further products, beside academic articles, should be generated from this work, the intellectual property distribution related to them will be directed under the current regulation of this matter determined by EAFIT University ([EAFIT, 2009](#)).

## References

- Andrés Ramírez, J. C. (2014). Which Team Will Win the 2014 FIFA World Cup? A Bayesian Approach for Dummies. <http://ssrn.com/abstract=2396259>. [Online; accessed 18-August -2015].
- Baio, G. and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37(2):253–264.
- EAFIT, U. (2009). Reglamento de propiedad intelectual.
- Leitner, C., Zeileis, A., and Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EUROÁĆ 2008. *International Journal of Forecasting*, 26(3):471–481.

- 
- Shin, H. (1993). Measuring the incidence of insiders trading in a market for state-contingent claims. *The Economic Journal*, 103(420):1141–1153.
- Spann, M. and Skiera, B. (2009). Sports forecasting: a comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, 28(1):55–72.
- Stekler, H., Sendor, D., and Verlander, R. (2010). Issues in sports forecasting. *International Journal of Forecasting*, 26(3):606–621.
- Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4):934 – 943.
- Strumbelj, E. and Sikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*, 26(3):482–488.
- Štrumbelj, E. and Vračar, P. (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting*, 28(2):532–542.
- Vlastakis, N., Dotsis, G., and Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, 28(5):426–444.