Sports betting: A source for empirical Bayes

Research practice 1: Progress report

Mateo Graciano-Londoño
Mathematical Engineering
Student

Andrés Ramírez-Hassan
Department of Economics
Tutor

EAFIT University, Medellín
Colombia

October 2^{nd}, 2015

Previous works

Sports forecasting is a topic which has 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].

Previous works

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 to calculate the probabilities among the different methods that had been used before.

Obtaining the probabilities

The idea is to use a Dirichlet categorized model, so we obtain the following posterior [Ramírez and Cardona 2014]:

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

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

Betting strategy

For the betting strategy we will proceed since we already have a good estimation of the probability ($\boldsymbol{p}$) of each outcome from the posterior. So now, given the odds (θ_i) for each outcome, how are we suppose to bet?

- Bet for the most likely event.
- Bet if $\theta_i p_i > 1$.
- Bet for the greatest $\theta_i p_i$ for each bookmarker.

Betting strategy

It is easy to show that the first (which is the most common idea), is statistically the worst, but with the second and third one the gambler is supposed to win.

The question is how much to bet? For that matter is a good idea to use the Kelly [1956] criteria which is based on the following equation:

$$f^* = \frac{\theta p - 1}{\theta - 1} \tag{1}$$

From (1) we calculate which fraction of the bankroll we are supposed to bet.

Dirichlet distribution

- Random numbers generation from the Dirichlet distribution.
- Estimation of the Dirichlet distribution parameters with the moments method for the tennis case.
- Estimation of the Dirichlet distribution parameters with maximum likelihood estimation method for the football case.

Posterior and simulation exercises

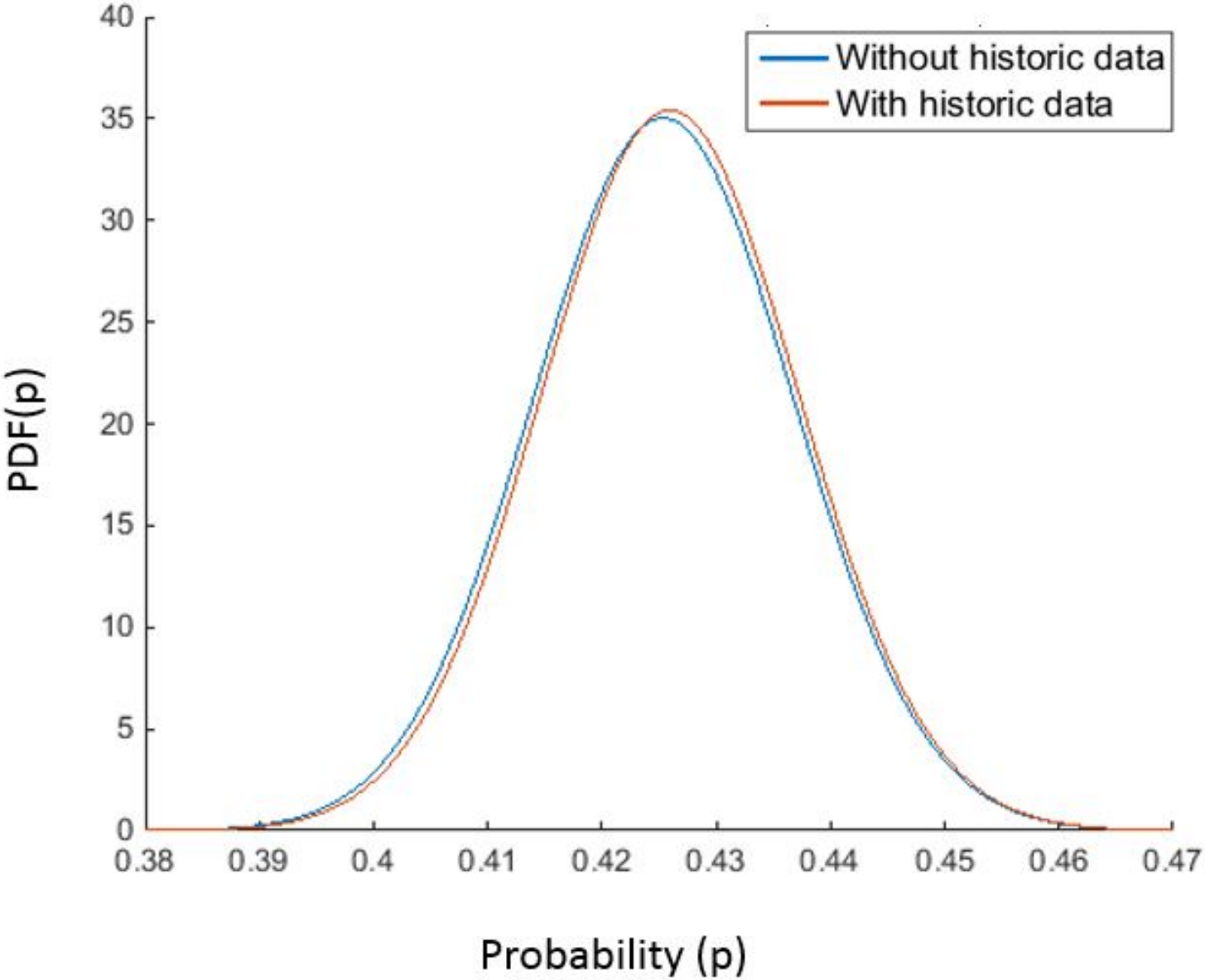
Once we obtained the posterior distributions of the parameters, we simulated our model 500.000 times for each game, and then, we calculate the mean for our estimation of $\boldsymbol{p}$.

Tennis results

	Shin	**Posterior**
Djokovic Vs Nadal	0.4254	0.42885
Djokovic Vs Wawrinka	0.8162	0.81421
Gasquet Vs Nadal	0.0728	0.07077
Djokovic Vs Youzhny	0.9403	0.93537
Murray Vs Wawrinka	0.7588	0.75760
Gasquet Vs Ferrer	0.2964	0.29595
Robledo Vs Nadal	0.0608	0.05985
Djokovic Vs Granollers	0.9634	0.9642

Table: Estimated probabilities for the win of the first player.

Tennis results

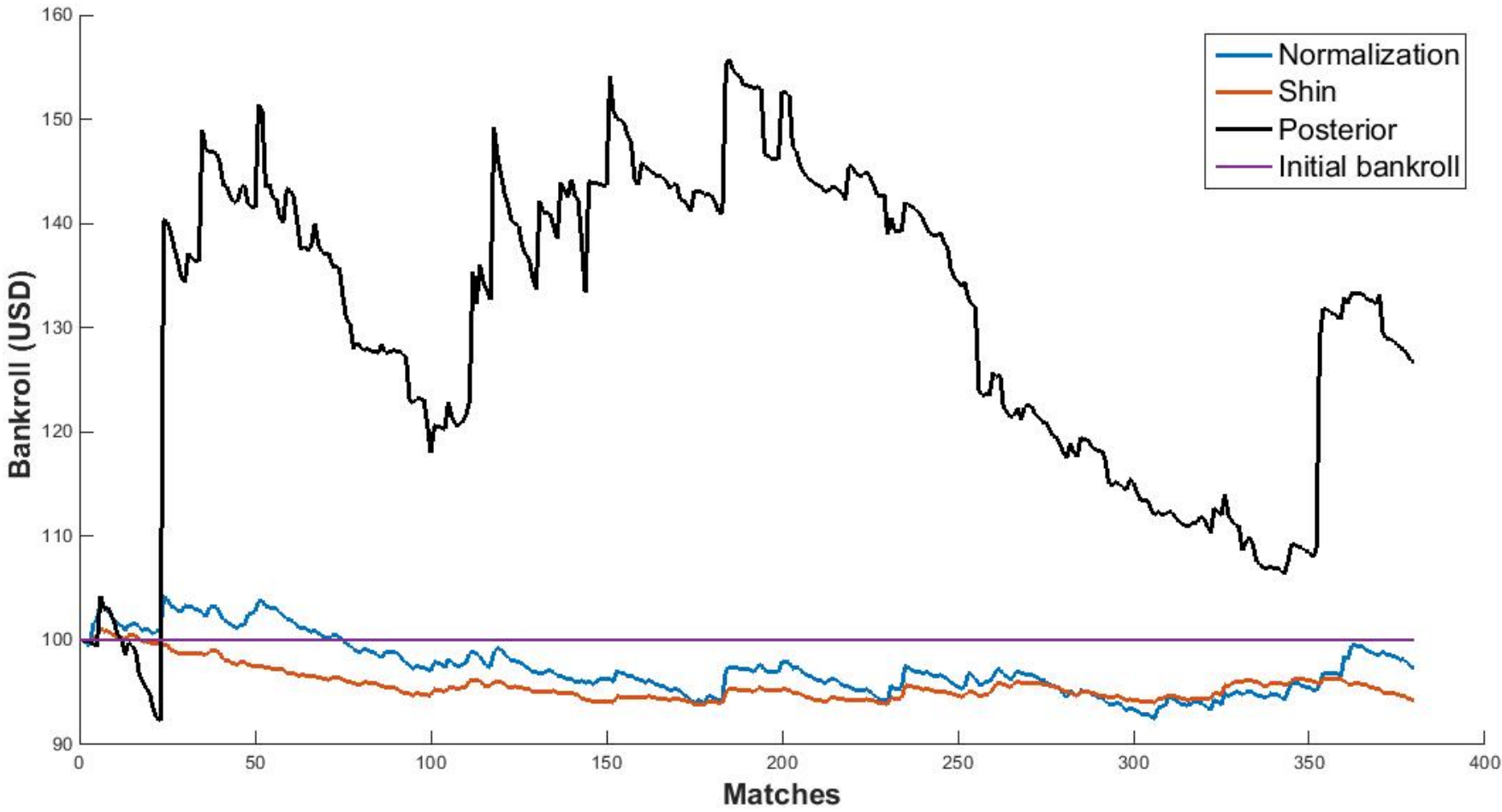


Premier league results

	Shin			Posterior		
Match	Home	Draw	Away	Home	Draw	Away
Ars Vs Cry	0.7740	0.1562	0.0696	0.7733	0.1559	0.0708
Lei Vs Eve	0.3076	0.2871	0.4052	0.3122	0.2915	0.3963
ManU Vs Swan	0.7166	0.1906	0.0927	0.7142	0.2009	0.0849
QPR Vs Hull	0.3877	0.2972	0.3150	0.3786	0.3086	0.3128
Stoke Vs Villa	0.4941	0.2838	0.2219	0.4607	0.2979	0.2415
Brom Vs Sunder	0.4272	0.2904	0.2822	0.4089	0.2936	0.2975
WHm Vs Tott	0.2584	0.2769	0.4646	0.2630	0.2608	0.4762
Liv Vs Soton	0.7195	0.1846	0.0958	0.7013	0.1906	0.1081
New Vs ManC	0.1802	0.2348	0.5848	0.1921	0.2460	0.5619

Table: Estimated probabilities for the football case.

Better than Shin?



Further work

- Check the betting strategy with PL another season.
- Consider an strategy which involves only a fraction of the Kelly fraction.
- Gather information for the NBA case.

References I

H.O. Stekler, David Sendor, and Richard Verlander. Issues in sports forecasting. *International Journal of Forecasting*, 26(3):606–621, July 2010.

Christoph Leitner, Achim Zeileis, and Kurt Hornik. Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. *International Journal of Forecasting*, 26(3):471–481, July 2010.

Gianluca Baio and Marta Blangiardo. Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37 (2):253–264, 2010.

Martin Spann and Bernd Skiera. Sports forecasting: a comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, 28(1):55–72, 2009.

References II

Erik Štrumbelj. On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4):934–943, 2014.

H. Shin. Measuring the incidence of insiders trading in a market for state-contingent claims. *The Economic Journal*, 103(420): 1141–1153, 1993.

Andrés Ramírez and Jhonatan Cardona. Which Team Will Win the 2014 FIFA World Cup? A Bayesian Approach for Dummies. `http://ssrn.com/abstract=2396259`, 2014. [Online; accessed 18-August -2015].

J. L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926, 1956.

Any questions?