DE GRUYTER J. Quant. Anal. Sports 2015; 11(1): 39–52

Francisco J. R. Ruiz* and Fernando Perez-Cruz

A generative model for predicting outcomes in college basketball

Abstract: We show that a classical model for soccer can also provide competitive results in predicting basketball outcomes. We modify the classical model in two ways in order to capture both the specific behavior of each National collegiate athletic association (NCAA) conference and different strategies of teams and conferences. Through simulated bets on six online betting houses, we show that this extension leads to better predictive performance in terms of profit we make. We compare our estimates with the probabilities predicted by the winner of the recent Kaggle competition on the 2014 NCAA tournament, and conclude that our model tends to provide results that differ more from the implicit probabilities of the betting houses and, therefore, has the potential to provide higher benefits.

Keywords: NCAA tournament; Poisson factorization; Probabilistic modeling; variational inference.

DOI 10.1515/jqas-2014-0055

1 Introduction

In this paper, we aim at estimating probabilities in sports. Specifically, we focus on the March Madness Tournament in college basketball,[1] although the model is general enough to model nearly any team sport for regular season and play-off games (assuming that both teams are willing to win). Estimating probabilities in sport events is challenging, because it is unclear what variables affect the outcome and what information is publicly known before the games begin. In team sports, it is even more complicated, because the information about individual players becomes relevant. Although there has been some attempts to model individual players (Miller et al. 2014), there is no standard method to evaluate the importance of individual players and remove their contribution to the team when players do not play or get injured or suspended. It is also unclear if considering individual player information can improve predictions with no overfit. For college basketball, even more variables come into play, because there are 351 teams divided in 32 conferences, they only play about 30 regular games and the match-ups are not random, so the results do not directly show the level of each team.

In the literature, we can find several variants of a simple model for soccer that identifies each team by its attack and defense coefficients (Baio and Blangiardo 2010; Crowder et al. 2002; Dixon and Coles 1997; Heuer, Muller, and Rubner 2010; Maher 1982). In all these works, the score for the home team is drawn from a Poisson distribution, whose mean is the multiplicative contribution of the home team attack coefficient and the away team defense coefficient. The score of the visitor team is an independent Poisson random variable, whose mean is the visitor attack coefficient multiplied by the home team defense coefficient. These coefficients are estimated by maximum likelihood using the past results and used to predict future outcomes.

A similar model can be found in the literature of Poisson factorization (Canny 2004; Cemgil 2009; Dunson and Herring 2005), where the elements of a matrix are assumed to be independent Poisson random variables given some latent attributes. For instance, in Poisson factorization for recommendation systems (Gopalan, Hofman, and Blei 2013), where the input is a user/item matrix of ratings, each user and each item is represented by a K-dimensional latent vector of positive weights. Each rating is modeled by a Poisson distribution parameterized by the inner product of the user's and item's weights.

We build a model that combines these two ideas (Poisson factorization and the model for soccer) and takes into account the structure of the Men's Division I Basketball of the National collegiate athletic association (NCAA). In order to estimate the mean of the Poisson distributions, we define an attack and defense vector for each team and for each NCAA conference. The conference-specific

1 http://www.ncaa.com/march-madness

***Corresponding author: Francisco J. R. Ruiz,** University Carlos III in Madrid – Signal Theory and Communications Department. Avda. de la Universidad, 30. Lab 4.3.A03, Leganes, Madrid 28911, Spain, e-mail: franrruiz@tsc.uc3m.es. http://orcid.org/0000-0002-2200-901X
Fernando Perez-Cruz: University Carlos III in Madrid – Signal Theory and Communications Department. Avda. de la Universidad, 30, Leganes, Madrid 28911, Spain; and Bell Labs, Alcatel-Lucent, New Providence, NJ 07974 USA, e-mail: fernando@tsc.uc3m.es, fernando.perez-cruz@alcatel-lucent.com

coefficients model the overall behavior of each conference, while the team-specific coefficients capture differences within each conference. To estimate the coefficients, we apply a variational inference algorithm. For comparisons, we adhere to the rules in the recent Kaggle competition,[2] in which all the predictions have to be in place before the actual tournament starts, i.e., we do not use the results in the first rounds of the tournament to improve the predictions in the subsequent rounds.

We use two metrics to validate the model. First, we compute the average negative log-likelihood of the predicted probabilities for the winning teams. This metric is used to determine the winners of the Kaggle competition. Unfortunately, the test sample size (63 games) is too small and almost all reasonable participants are statistically indistinguishable from the winner. Compared to the winner's probability estimates, we could not reject the null hypothesis of a Wilcoxon signed-rank test (Wilcoxon 1945) for 198 out of the remaining 247 participants. With so few test cases, it is unclear if the winners had a better model or were just lucky. This serves as an excuse for sore losers (we were ranked #39 in the competition), but more importantly as a word of advice for these competitions, in which metrics should be able to tell without doubt that some participants did significantly better (making use of statistical tests to tell them apart). Second, we compute the profit we would make after betting on six on-line betting houses using Kelly's criterion (Kelly 1956). Kelly's criterion assumes that our estimates are the true underlying probabilities, and the betting house odds are only an estimate. It provides the fraction of our bankroll that we should stake on each bet in order to maximize the long term growth rate of our fortune and make sure that we do not lose it all. This metric tells us how good our probability estimates are when compared to those of the betting houses. Our model outperforms the considered betting houses and the Kaggle competition winner.

2 Model description

We develop a statistical model for count data, corresponding to the outcomes of each basketball game. For each game $m=1,\ldots,M$, we observe the pair (y_m^H, y_m^A), which are the points scored by the home and away teams, respectively.

The soccer model by Maher (1982) or Dixon and Coles (1997) introduces an attack and defense coefficient for each team $t=1,\ldots,T$, denoted, respectively, by α_t and β_t. Given these coefficients, the number of scores obtained by the home and away sides at game m are independently distributed as

$$\begin{aligned} y_m^H &\sim \text{Poisson}(\gamma\alpha_{h(m)}\beta_{a(m)}), \\ y_m^A &\sim \text{Poisson}(\alpha_{a(m)}\beta_{h(m)}), \end{aligned} \tag{1}$$

respectively. Here, the index $h(m)\in\{1,\ldots, T\}$ identifies the team that is playing at home in the m -th game and, similarly, $a(m)$ identifies the team that is playing away. The parameter γ is the home coefficient and represents the advantage for the team hosting the game. This effect is assumed to be constant for all the teams and throughout the season. Note also that β_t is actually a "inverse defense" coefficient, in the sense that smaller values represent better defense capabilities.

For the NCAA Tournament, we modify the model in Eq. 1 in two ways. First, we represent each team with K_1 attack coefficients and K_1 defense coefficients, which are grouped for each team in vectors $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$, respectively. Each coefficient may represent a particular tactic or strategy, so that teams can be good at defending some tactics but worse at defending others (the same applies for attacking). Second, we also take into account the conference to which each team belongs.[3] For that purpose, we introduce conference-specific attack and defense coefficients, allowing us to capture the overall behavior of each conference. We denote by η_l and ρ_l the K_2-dimensional attack and defense coefficient vectors of conference l, respectively, and we introduce index $\ell(t)\in\{1,\ldots, L\}$ to represent the conference to which team t belongs. Hence, we model the outcome at game m as

$$\begin{aligned} y_m^H &\sim \text{Poisson}(\gamma\boldsymbol{\alpha}_{h(m)}^\top\boldsymbol{\beta}_{a(m)} + \gamma\boldsymbol{\eta}_{\ell(h(m))}^\top\boldsymbol{\rho}_{\ell(a(m))}), \\ y_m^A &\sim \text{Poisson}(\boldsymbol{\alpha}_{a(m)}^\top\boldsymbol{\beta}_{h(m)} + \boldsymbol{\eta}_{\ell(a(m))}^\top\boldsymbol{\rho}_{\ell(h(m))}). \end{aligned} \tag{2}$$

To complete the specification of the model, we place independent gamma priors over the elements of the attack and defense vectors, as well as a gamma prior over the home coefficient. Throughout the paper, we parametrize the gamma distribution with its shape and rate. Therefore, the generative model is as follows:

1. Draw the home coefficient $\gamma\sim\text{gamma}(s_\gamma, r_\gamma)$.
2. For each team $t=1,\ldots,T$:
   (a) Draw the attack coefficients $\alpha_{t,k}\sim\text{gamma}(s_\alpha, r_\alpha)$ for $k=1,\ldots,K_1$
   (b) Draw the defense coefficients $\beta_{t,k}\sim\text{gamma}(s_\beta, r_\beta)$ for $k=1,\ldots,K_1$.

2 https://www.kaggle.com/c/march-machine-learning-mania

3 We consider North and South Divisions of Big South Conference as two different conferences. The same applies to Mid-American Conference (East and West) and Ohio Valley (East and West).

3. For each conference $l=1, \dots, L$:
   (a) Draw the attack coefficients $\eta_{l,k} \sim \text{gamma}(s_\eta, r_\eta)$ for $k=1, \dots, K_2$.
   (b) Draw the defense coefficients $\rho_{l,k} \sim \text{gamma}(s_\rho, r_\rho)$ for $k=1, \dots, K_2$.
4. For each game $m=1, \dots, M$:
   (a) Draw the score
   $$y_m^H \sim \text{Poisson}(\gamma \boldsymbol{\alpha}_{h(m)}^\top \boldsymbol{\beta}_{a(m)} + \gamma \boldsymbol{\eta}_{\ell(h(m))}^\top \boldsymbol{\rho}_{\ell(a(m))}).$$
   (b) Draw the score
   $$y_m^A \sim \text{Poisson}(\boldsymbol{\alpha}_{a(m)}^\top \boldsymbol{\beta}_{h(m)} + \boldsymbol{\eta}_{\ell(a(m))}^\top \boldsymbol{\rho}_{\ell(h(m))}).$$

Thus, the shape and rate parameters of the *a priori* gamma distributions are hyperparameters of our model. The corresponding graphical model is shown in Figure 1, in which circles correspond to random variables and gray-shaded circles represent observations.

3 Inference

In this section, we describe a mean-field inference algorithm to approximate the posterior distribution of the attack and defense coefficients, as well as the home coefficient, which we need to predict the outcomes of the tournament games.

Variational inference provides an alternative to Markov chain Monte Carlo (MCMC) methods as a general source of approximation methods for inference in probabilistic models (Jordan et al. 1999). Variational algorithms turn inference into a non-convex optimization problem, but they are in general computationally less demanding compared to MCMC methods and do not suffer from limitations involving mixing of the Markov chains. In a general variational inference scenario, we have a set of hidden variables Φ whose posterior distribution given the observations $\boldsymbol{y}$ is intractable. In order to approximate the posterior $p(\Phi | \boldsymbol{y}, \mathscr{H})$, where $\mathscr{H}$ denotes the set of hyperparameters of the model, we first define a parametrized family of distributions over the hidden variables, $q(\Phi)$, and then fit their parameters to find

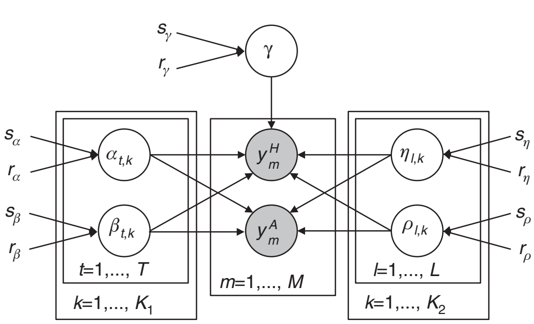

Figure 1: Graphical model representation for our generative model.

a distribution that is close to the true posterior. Closeness is measured in terms of Kullback-Leibler (KL) divergence between both distributions $D_{KL}(q||p)$. The computation of the KL divergence is intractable, but fortunately, minimizing $D_{KL}(q||p)$ is equivalent to maximizing the so-called evidence lower bound (ELBO) $\mathscr{L}$, since

$$\begin{aligned} \log p(\boldsymbol{y} | \mathscr{H}) &= \mathbb{E}[\log p(\boldsymbol{y}, \Phi | \mathscr{H})] + H[q] + D_{KL}(q \| p) \\ &\geq \mathbb{E}[\log p(\boldsymbol{y}, \Phi | \mathscr{H})] + H[q] \triangleq \mathscr{L}, \end{aligned} \tag{3}$$

where the expectations above are taken with respect to the variational distribution $q(\Phi)$, and H[q] denotes the entropy of the distribution $q(\Phi)$.

Typical variational inference methods maximize the ELBO $\mathscr{L}$ by coordinate ascent, iteratively optimizing each variational parameter. A closed-form expression for the corresponding updates can be easily found for conditionally conjugate variables, i.e., variables whose complete conditional is in the exponential family. We refer to (Ghahramani and Beal 2001; Hoffman et al. 2013) for further details. In order to obtain a conditionally conjugate model, and following (Dunson and Herring 2005; Gopalan et al. 2013, 2014; Zhou et al. 2012), we augment the representation by defining for each game the auxiliary latent variables

$$\begin{aligned} &z_{m,k}^{H1} \sim \text{Poisson}(\gamma \alpha_{h(m),k} \beta_{a(m),k}), \; z_{m,k}^{H2} \sim \text{Poisson}(\gamma \eta_{\ell(h(m)),k} \rho_{\ell(a(m)),k}), \\ &z_{m,k}^{A1} \sim \text{Poisson}(\alpha_{a(m),k} \beta_{h(m),k}), \;\; z_{m,k}^{A2} \sim \text{Poisson}(\eta_{\ell(a(m)),k} \rho_{\ell(h(m)),k}), \end{aligned} \tag{4}$$

so that the observations for the home and away scores can be, respectively, expressed as

$$y_m^H = \sum_{k=1}^{K_1} z_{m,k}^{H1} + \sum_{k=1}^{K_2} z_{m,k}^{H2}, \quad \text{and} \quad y_m^A = \sum_{k=1}^{K_1} z_{m,k}^{A1} + \sum_{k=1}^{K_2} z_{m,k}^{A2}, \tag{5}$$

due to the additive property of Poisson random variables. Thus, the auxiliary variables preserve the marginal Poisson distribution of the observations. Furthermore, the complete conditional distribution over the auxiliary variables, given the observations and the rest of latent variables, is a Multinomial. Using the auxiliary variables, and denoting $\boldsymbol{\alpha}=\{\boldsymbol{\alpha}_t\}$, $\boldsymbol{\beta}=\{\boldsymbol{\beta}_t\}$, $\boldsymbol{\eta}=\{\boldsymbol{\eta}_l\}$, $\boldsymbol{\rho}=\{\boldsymbol{\rho}_l\}$ and $\boldsymbol{z}=\{z_{mk}^{H1}, z_{mk}^{H2}, z_{mk}^{A1}, z_{mk}^{A2}\}$, the joint distribution over the hidden variables can be written as

$$\begin{aligned} p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\rho}, \gamma, \boldsymbol{z} | \mathscr{H}) &= \prod_{t=1}^{T} \prod_{k=1}^{K_1} p(\alpha_{t,k} | s_\alpha, r_\alpha) p(\beta_{t,k} | s_\beta, r_\beta) \\ &\times p(\gamma | s_\gamma, r_\gamma) \prod_{l=1}^{L} \prod_{k=1}^{K_2} p(\eta_{l,k} | s_\eta, r_\eta) p(\rho_{l,k} | s_\rho, r_\rho) \\ &\times \prod_{m=1}^{M} \prod_{k=1}^{K_1} p(z_{m,k}^{H1} | \gamma, \alpha_{h(m),k}, \beta_{a(m),k}) p(z_{m,k}^{A1} | \alpha_{a(m),k}, \beta_{h(m),k}) \\ &\times \prod_{m=1}^{M} \prod_{k=1}^{K_2} p(z_{m,k}^{H2} | \gamma, \eta_{\ell(h(m)),k}, \rho_{\ell(a(m)),k}) p(z_{m,k}^{A2} | \eta_{\ell(a(m)),k}, \rho_{\ell(h(m)),k}), \end{aligned} \tag{6}$$

and the observations are generated according to Eq. 5. In mean-field inference, the posterior distribution is approximated with a completely factorized variational distribution, i.e., q is chosen as

$$q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\rho},\gamma,\boldsymbol{z})=q(\gamma)\prod_{t=1}^{T}\prod_{k=1}^{K_1}q(\alpha_{t,k})q(\beta_{t,k})$$
$$\prod_{l=1}^{L}\prod_{k=1}^{K_2}q(\eta_{l,k})q(\rho_{l,k})\prod_{m=1}^{M}q(\boldsymbol{z}_m^H)q(\boldsymbol{z}_m^A), \tag{7}$$

being $\boldsymbol{z}_m^H$ the vector containing the variables $\{z_{mk}^{H1}, z_{mk}^{H2}\}$ for game m (and similarly for $\boldsymbol{z}_m^A$ and $\{z_{mk}^{A1}, z_{mk}^{A2}\}$). For conciseness, we have removed the dependency on the variational parameters in Eq. 7. We set the variational distribution for each variable in the same exponential family as the corresponding complete conditional, therefore yielding

$$\begin{aligned}
&q(\gamma)=\text{gamma}(\gamma\,|\,\gamma^{\text{shp}},\gamma^{\text{rte}}),\\
&q(\alpha_{t,k})=\text{gamma}(\alpha_{t,k}\,|\,\alpha_{t,k}^{\text{shp}},\alpha_{t,k}^{\text{rte}}),\\
&q(\beta_{t,k})=\text{gamma}(\beta_{t,k}\,|\,\beta_{t,k}^{\text{shp}},\beta_{t,k}^{\text{rte}}),\\
&q(\eta_{l,k})=\text{gamma}(\eta_{l,k}\,|\,\eta_{l,k}^{\text{shp}},\eta_{l,k}^{\text{rte}}),\\
&q(\rho_{l,k})=\text{gamma}(\rho_{l,k}\,|\,\rho_{l,k}^{\text{shp}},\rho_{l,k}^{\text{rte}}),\\
&q(\boldsymbol{z}_m^H)=\text{multinomial}(\boldsymbol{z}_m^H\,|\,y_m^H,\boldsymbol{\phi}_m^H),\\
&q(\boldsymbol{z}_m^A)=\text{multinomial}(\boldsymbol{z}_m^A\,|\,y_m^A,\boldsymbol{\phi}_m^A).
\end{aligned} \tag{8}$$

Then, the set of variational parameters is composed of the shape and rate for each gamma distribution, as well as the probability vectors $\boldsymbol{\phi}_m^H$ and $\boldsymbol{\phi}_m^A$ for the multinomial distributions. Note that $\boldsymbol{\phi}_m^H$ and $\boldsymbol{\phi}_m^A$ are both (K_1+K_2)-dimensional vectors. To minimize the KL divergence and obtain an approximation of the posterior, we apply a coordinate ascent algorithm (the update equations of the variational parameters are given in Appendix A).

4 Experiments

4.1 Experimental setup

We apply our variational algorithm to last 4 years of NCAA Men's Division I Basketball Tournament. Here, we focus on 2014 tournament, while results for previous years can be found in Appendix B. Following the recent Kaggle competition procedure, we fit the model using the regular season results of over 5000 games to predict the outcome of the 63 tournament games.[4] As in Kaggle competition, we do not predict the "first four" games (they are not considered in the learning stage either). We apply the algorithm described in Section 3 independently for each season, because teams exhibit different strength even at consecutive seasons, probably due to the high turnaround of players. Note that the data include a variable which indicates whether one of the teams was hosting the game, or it was played on a neutral court. We include this variable in our formulation of the problem, and therefore we remove the home coefficient γ for games in which the site was considered neutral. We use the output of our algorithm, i.e., the parameters for the approximate posterior distribution over the hidden coefficients, to estimate the probability of teams winning in each Tournament game.[5] To test the model, we simulate betting on the Tournament games using data from several betting houses[6] (missing entries in the bookmaker betting odd matrices were not taken into account).

For hyperparameter selection, we carried out an exhaustive grid search, but did not find significant differences in our results as a consequence of the shape and rate values of the *a priori* gamma distributions. The experiments that we describe in this section were run with shape 1 and rate 0.1, except for the home coefficient, for which we use unit shape and rate.

For the training stage, we initialize our algorithm by randomly setting all the variational parameters. Every 10 iterations, we compute the ELBO as $\mathscr{L}=\mathbb{E}[\log p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\rho},\gamma,\boldsymbol{z}\,|\,\mathscr{H})]+\mathbb{E}[\log q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\rho},\gamma,\boldsymbol{z})]$, where the expectations are taken with respect to the variational distribution q. The training process stops when the relative change in the ELBO is $<10^{-8}$, or when 10^6 iterations are reached (whatever happens first).

After convergence, we estimate the probabilities of each team winning for the 63 games in the tournament. We estimate them for each game m by computing the expected Poisson means as $\mathbb{E}[y_m^H]=\mathbb{E}[\boldsymbol{\alpha}_{h(m)}^\top\boldsymbol{\beta}_{a(m)}+\boldsymbol{\eta}_{\ell(h(m))}^\top\boldsymbol{\rho}_{\ell(a(m))}]$ and $\mathbb{E}[y_m^A]=\mathbb{E}[\boldsymbol{\alpha}_{a(m)}^\top\boldsymbol{\beta}_{h(m)}+\boldsymbol{\eta}_{\ell(a(m))}^\top\boldsymbol{\rho}_{\ell(h(m))}]$. Holding both means fixed, the difference $y_m^H-y_m^A$ follows a Skellam distribution (Skellam 1946) with parameters $\mathbb{E}[y_m^H]$ and $\mathbb{E}[y_m^A]$. We compute the probability of team $h(m)$ winning the game as $\text{Prob}(y_m^H-y_m^A>0\,|\,y_m^H-y_m^A\neq 0)$. Alternatively, we can estimate probabilities by sampling from the approximate posterior distribution, with no significant difference in the predictions. We average the predicted probabilities for 100 independent runs of the variational algorithm,

4 Data was collected from https://www.kaggle.com/c/march-machine-learning-mania/data

5 We also remove γ for predictions, as all tournament games are played in a neutral court.

6 Bookmaker betting odds were extracted from http://www.oddsportal.com/basketball/usa/ncaa-i-a/results

under different initializations to alleviate the sensibility of the variational algorithm to its starting point.

4.2 Results for 2014 tournament

Exploratory analysis. One of the benefits of a generative model is that, instead of a black-box approach, it provides an explanation of the results. Furthermore, generative models allow integrating the information from experts in sports as prior knowledge in the Bayesian generative model. This would constrain the statistical model and may provide more accurate predictions and usable information to help understand the teams performance.

We found that the expected value of the home coefficient is $\mathbb{E}[\gamma]=1.03$ (we obtained this value after averaging the results for 100 independent runs for a model with $K_1=K_2=10$, being the standard deviation around 5×10^{-4}). This indicates that playing at home provides some advantage, but this advantage is not as relevant as in soccer, where the home coefficient is typically around 1.4 (Dixon and Coles 1997).

We can also use our generative model to rank conferences and provide a qualitative measure on how well it follows general appreciation. Although there are several ways for ranking, we have taken a simple approach. For a model with $K_1=10$ and $K_2=10$ we have ranked the conferences according to $\sum_{k=1}^{K_2}(\mathbb{E}[\eta_{\ell,k}]-\mathbb{E}[\rho_{\ell,k}])$, with expectations taken with respect to the variational distribution. In Table 1 we show the obtained ranking, together with the number of teams for each conference that entered the March Madness Tournament. The top-5 conferences (Pac-12, Big Ten, ACC, Big 12 and Atlantic 10) are the stronger ones, as they contribute with six or seven teams to the Tournament. There are two conferences that contribute with four teams (Big East and American) and they are ranked 7th and 8th. There are three conferences (Mountain West, West Coast and SEC) that contribute with two or three teams and they are ranked 11th–13th. There are only three conferences that contribute with only the conference winner and that are stronger than the second tier conferences (those with two to four teams in the Tournament). There are also three conferences (Big South, Mid-American and Ohio Valley) that we divide into two sub-conferences, but they only contribute with one team to the tournament. The sub-conference that contributed with a team to the tournament is always ranked higher with our score.

We also provide some qualitative results about the team-level parameters. For the model above with $K_1=K_2=10$, we rank teams according to the value of $\sum_{k=1}^{K_1}\mathbb{E}[\alpha_{t,k}-\beta_{t,k}]+\sum_{k=1}^{K_2}\mathbb{E}[\eta_{\ell(t),k}-\rho_{\ell(t),k}]$. We show in Table 2 the top-64 teams of the obtained ranking. Out of the 36 teams that entered the tournament as "at large" bids, 34 of them are placed on the top-60 positions of the ranking. The two other teams are Tennessee, which is

Table 1: Ranking of conferences provided by our model.

#	Value	Conference	# Teams	#	Value	Conference	# Teams
1	8.2	Pac-12	6	19	−0.6	Big Sky	1
2	8.0	Big Ten	6	20	−1.3	Sun Belt	1
3	4.7	ACC	6	21	−1.3	Southern	1
4	4.2	Big 12	7	22	−1.9	Ivy League	1
5	4.1	Atlantic 10	6	23	−1.9	Ohio Valley (E)	1
6	3.9	Colonial Athletic Association	1	24	−2.4	Northeast	1
7	3.8	Big East	4	25	−2.5	Summit League	1
8	3.3	American	4	26	−2.9	Mid-American (E)	0
9	3.3	Conference USA	1	27	−3.0	SWAC	1
10	3.2	Big South (S)	1	28	−3.0	Ohio Valley (W)	0
11	2.6	Mountain West	2	29	−3.2	Southland	1
12	2.2	West Coast	2	30	−3.4	MEAC	1
13	0.9	SEC	3	31	−3.5	Patriot League	1
14	0.8	Horizon League	1	32	−4.0	WAC	1
15	0.5	Missouri Valley	1	33	−4.2	MAAC	1
16	0.3	Mid-American (W)	1	34	−4.4	Atlantic Sun	1
17	0.3	Big South (N)	0	35	−7.7	America East	1
18	0.2	Big West	1	–	–	–	–

For each conference, we show the value of the metric that we use to produce the ranking, and the number of teams that entered the March Madness Tournament for that conference.

Table 2: Ranking of teams provided by our model (only shown top-64 teams).

#	Value	Team	#	Value	Team	#	Value	Team	#	Value	Team
1	12.7	Arizona	17	8.3	Arizona State	33	7.3	Connecticut	49	6.2	Butler
2	12.0	Iowa	18	8.3	Dayton	34	7.2	St Louis	50	6.1	Cincinnati
3	11.4	Michigan State	19	8.2	St Bonaventure	35	7.2	Wichita State	51	6.0	Texas Tech
4	10.5	Ohio State	20	8.1	Colorado	36	7.1	Indiana	52	5.9	Massachusetts
5	10.3	Louisville	21	8.1	North Carolina	37	7.1	Nebraska	53	5.9	Southern Methodist
6	9.9	Michigan	22	8.0	Gonzaga	38	7.1	Texas	54	5.8	Georgetown
7	9.5	UCLA	23	7.9	Oklahoma	39	7.1	Iowa State	55	5.8	Xavier
8	9.5	Villanova	24	7.9	Syracuse	40	7.0	Memphis	56	5.8	Clemson
9	9.4	Utah	25	7.9	Baylor	41	7.0	Virginia Commonwealth	57	5.7	New Mexico
10	9.4	Pittsburgh	26	7.8	Purdue	42	6.8	Arkansas	58	5.6	San Diego State
11	9.2	Creighton	27	7.7	Providence	43	6.5	Florida State	59	5.6	Kansas State
12	9.0	Wisconsin	28	7.7	Stanford	44	6.5	Florida	60	5.6	St John's
13	8.7	Minnesota	29	7.6	Duke	45	6.4	Kentucky	61	5.3	Tennessee
14	8.6	Kansas	30	7.6	Oregon	46	6.4	Brigham Young	62	5.2	Boise State
15	8.6	California	31	7.4	Middle Tennessee State	47	6.3	George Washington	63	5.2	Virginia
16	8.5	Oklahoma State	32	7.4	Illinois	48	6.2	Tulsa	64	5.2	Maryland

The column "value" corresponds to the metric that we use to produce the ranking.

ranked #61, and North Carolina State, ranked #78. Out of the 32 teams that entered the Tournament as "automatic bids" (i.e., teams winning their conference tournaments), half of them are placed on the top-100 positions, while the rest are ranked up to position #280 (for Texas Southern). In addition, for nine out of the 10 conferences that contribute with more than one team to the March Madness competition, the conference winner is also listed in Table 2 (top-64 positions), and 44 out of the 46 teams of these 10 conferences that entered the Tournament are also in that list. The two teams that do not appear in the top-64 positions are St Joseph's (winner of the Atlantic 10 conference) and North Carolina State, which entered the competition in the pre-round. St Joseph's was the play-off winner at the Atlantic 10 conference, but it had a poor record in that conference, which explains why its rating is not that high with out score. Regarding the teams in the March Madness competition belonging to the weaker conferences (i.e., those conferences that only contribute with one team to the Tournament), only two out of 22 teams are in the top-64 positions. Qualitatively, our results coincide with the way teams are selected for the tournament.

If we focus on Pac-12 conference, the six teams that entered the competition are placed in positions #1, #7, #17, #20, #28 and #30 of Table 2 (for Arizona, UCLA, Arizona State, Colorado, Stanford and Oregon, respectively), and the conference winner was UCLA, which is the second of the six teams. This is not a contradiction, because it is the number of won games in the conference tournament what determines the conference winner, while our model takes into account all the games and the score of each game as input. Under our ranking, a team that loses a few games by a small difference and win many games by a large difference will be better placed than a team that wins all the games by a small margin.

Finally, our model has the ability to provide predictions for the expected results in each game, since we directly model the number of points. We include in Table 3 a subset of the games in the March Madness competition, together with their results, our predictions, and the 90% credible intervals (the rest of the games of the Tournament are shown in Appendix C). The predictions have been obtained after averaging the expected Poisson means $\mathbb{E}[y_m^H]$ and $\mathbb{E}[y_m^A]$ for 100 independent runs, using a model with $K_1=K_2=10$. Out of the 126 scores, 21 of them are outside the 90% credible interval, which is a bit high but not unheard of. What might be more surprising is that 17 out of these 21 scores are below the credible interval and only four of them above the credible interval. There are several explanations for this effect. The most plausible is that we train the model with the regular season results but predict Tournament games instead. In regular season games, losing a game is not the end of the world, but losing a Tournament game has greater importance. Hence, players, which are young college students, might feel some additional pressure and it should be unsurprising that teams tend to score less than in the regular season. Nevertheless, we can still say that this is only a minor effect and that the loss of performance due to pressure is not significant enough to make us state that a model trained for the regular season cannot be used for predicting the March Madness Tournament.

Table 3: List of a subset of the games in the 2014 tournament.

#	Team 1			Team 2		
	Team	**Result**	**Prediction (CI)**	**Team**	**Result**	**Prediction (CI)**
36	North Dakota State	44	61.4 (49–75)	San Diego State	63	71.7 (58–86)
37	Dayton	55	63.1 (50–76)	Syracuse	53	71.8 (58–86)
38	Oregon	77	76.9 (63–92)	Wisconsin	85	78.8 (64–94)
39	Harvard	73	67.7 (54–81)	Michigan State	80	76.1 (62–91)
40	Connecticut	77	62.5 (50–76)	Villanova	65	72.0 (58–86)
41	Kansas	57	80.8 (66–96)	Stanford	60	71.5 (58–86)
42	Wichita State	76	69.9 (56–84)	Kentucky	78	70.3 (57–84)
43	Iowa State	85	87.8 (73–103)	North Carolina	83	85.7 (71–101)
44	Tennessee	83	72.6 (59–87)	Mercer	63	65.0 (52–79)
45	UCLA	77	80.7 (66–96)	Stephen F. Austin	60	71.6 (58–86)
46	Creighton	55	70.9 (57–85)	Baylor	85	66.9 (54–81)
47	Virginia	78	66.8 (54–81)	Memphis	60	67.2 (54–81)
48	Arizona	84	75.4 (61–90)	Gonzaga	61	59.2 (47–72)
49	Stanford	72	71.9 (58–86)	Dayton	82	76.7 (63–91)
50	Wisconsin	69	68.3 (55–82)	Baylor	52	65.6 (53–79)
51	Florida	79	74.5 (61–89)	UCLA	68	74.8 (61–89)
52	Arizona	70	66.5 (53–80)	San Diego State	64	57.3 (45–70)
53	Michigan	73	70.8 (57–85)	Tennessee	71	63.1 (50–76)
54	Iowa State	76	79.8 (65–95)	Connecticut	81	73.5 (60–88)
55	Louisville	69	68.5 (55–82)	Kentucky	74	70.7 (57–85)
56	Virginia	59	64.9 (52–78)	Michigan State	61	73.1 (59–87)
57	Florida	62	69.2 (56–83)	Dayton	52	62.1 (49–75)
58	Arizona	63	70.0 (57–84)	Wisconsin	64	63.6 (51–77)
59	Michigan State	54	73.1 (59–87)	Connecticut	60	64.7 (52–78)
60	Michigan	72	73.3 (60–88)	Kentucky	75	70.8 (57–85)
61	Florida	53	64.7 (52–78)	Connecticut	63	62.4 (50–76)
62	Wisconsin	73	70.2 (57–84)	Kentucky	74	69.0 (56–83)
63	Connecticut	60	68.5 (55–82)	Kentucky	54	72.3 (59–87)

For each game and team, we show the actual outcome of the game, as well as the predicted mean value, together with the 90% credible interval (labeled as "CI" in the Table).

Quantitative analysis. To quantitatively evaluate our proposal, we report five solutions and compare them with the Kaggle competition winner and the implicit probabilities of six online betting houses.[7] We use four models with a fixed value of K_1 and K_2, but we also report the probabilities obtained as the average of the predictions for 10 different models, with K_1 ranging between 4 and 10 and K_2 ranging between 10 and 15. In Kaggle competition, our 10-model average predictions led us to position =39 out of 248. We first report the negative logarithmic loss, which is computed as in Kaggle competition as

$$\text{LogLoss}=-\frac{1}{63}\sum_{m=1}^{63}(\nu_m\log(\hat{\nu}_m)+(1-\nu_m)\log(1-\hat{\nu}_m)), \qquad (9)$$

where $\nu_m\in\{0,1\}$ indicates whether team $h(m)$ beats team $a(m)$, and $\hat{\nu}_m\in[0,1]$ is the predicted probability of team $h(m)$ beating team $a(m)$. To be able to understand the variability in these results, we take 500 bootstrap samples (Efron 1979) and show the boxplot for these samples in Figure 2. We report the mean, the median, the 25/75% and the 10/90% percentiles, as well as the extreme values in the standard format. Note that K_1=1, K_2=0 corresponds to the classical model for soccer. We have included some markers for comparison: the best and 100th best results in the Kaggle competition, the median probability prediction for all Kaggle participants, the Kaggle seed benchmark (in which the winning probability predicted for the stronger team is 0.5+0.03*seed difference) and the 0.5-benchmark for all games and teams. In this figure, the boxplot for the winner of the Kaggle competition is lower than the boxplot for our models and the online betting houses. However, we found that the predictions of the Kaggle winner are not statistically different from our predictions, as reported by a Wilcoxon signed-rank test (Wilcoxon 1945) with a

7 The implicit probabilities are computed from the published payoffs using a linear system of equations in which the implicit probability for each team is $p_i=Q/\text{payoff}_i$ for $i=\{1, 2\}$ and $1-Q$ represents the margin the betting house keeps for risk management and profit. For example, if for a given game the payoff for both teams were \$2.85 and \$1.425, the implicit probabilities would, respectively, be 1/3 and 2/3 and Q=0.95.

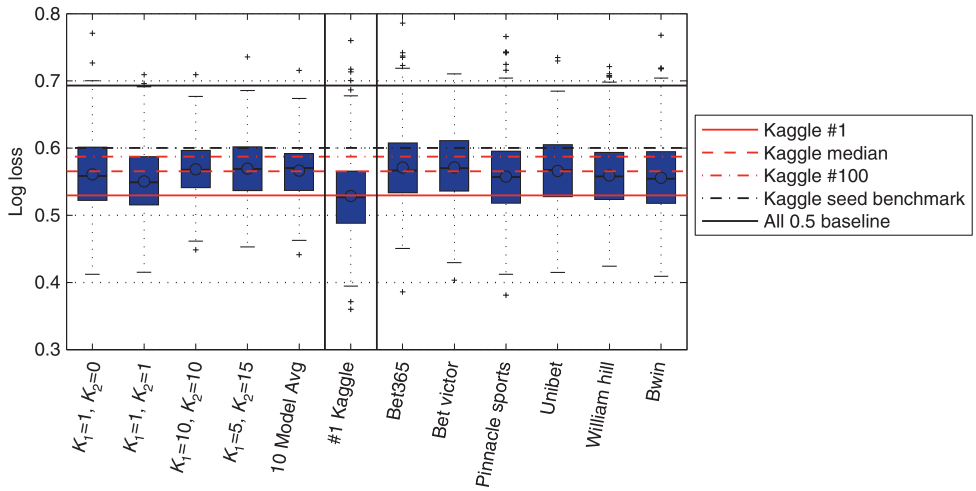

Figure 2: Boxplot representation of logarithmic loss after bootstrap. From left to right, we depict results for the considered models, Kaggle winner's estimates, and the six betting houses.

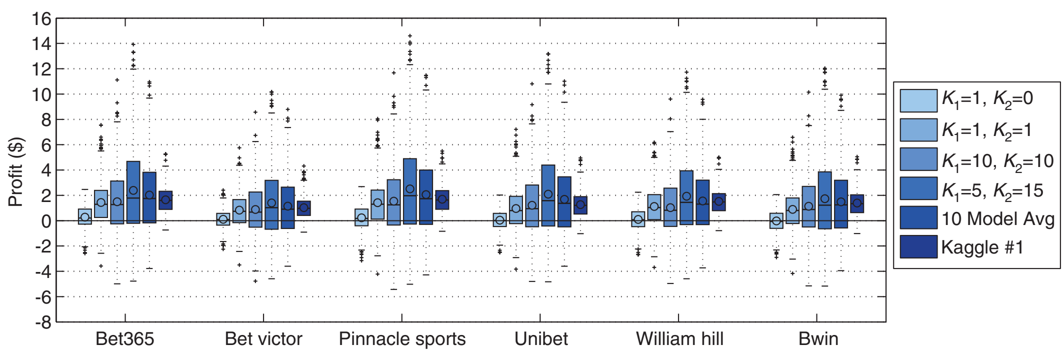

Figure 3: Boxplot representation of profit after bootstrap, broken down by betting house.

significance level of 1%. Specifically, we found that the predictions by Kaggle winner are not statistically different when compared to our 10-model average predictions. Furthermore, for 198 (out of 248) participants in the Kaggle competition, the Wilcoxon test failed to reject the null hypothesis (which corresponds to the median between the winner and the other participants being the same). This just indicates that the sample size is too small and we would need a larger test set to measure the goodness of fit of each proposal.

We now turn to a monetary metric that allows comparing our results with respect to the different betting houses. We assume that our probability estimates are the true ones and use Kelly's criterion (Kelly 1956) to decide how much we should bet (and for which team). Roughly, Kelly's criterion tells that the amount that we should bet grows with the difference between our probabilities and the implicit probabilities of the betting houses, and that we should bet for the team for which this difference is positive.[8] If the probabilities are very similar or Q is very large then Kelly's criterion might recommend not to bet. We have applied Kelly's criterion for the 63 games in the Tournament assuming that we have $1 per game. We could have aggregated the bankroll after each day or each weekend and bet more aggressively in the latter stages of the Tournament, but we believe that results with $1 per game are easier to follow. In Figure 3, we show the boxplot representation of our profit in the six considered betting houses, as well as the profit of the Kaggle competition winner (again, we use 500 bootstrap samples). For all the methods, the mean and the median are positive and away from zero, but the differences are not significant and are

8 Kelly's criterion does not tell us to bet in favor of the team that we believe will win (higher predicted probability), but it tells us to bet for the team for which we will make more money in average, making sure that we do not bankrupt. For example, if the betting house implicit probability is 0.8 for the stronger team and our model considers this probability is 0.7 Kelly's criterion will say that we should bet for the weaker team, because in a repeated game this strategy delivers the largest growth of the bankroll.

not significant amongst them (according to a Wilcoxon signed-rank test). The mean of our 10-model average and the mean of the K_1=K_2=10 model are larger than the mean of the Kaggle competition winner for all the betting houses. Our variance is larger because our model points towards a high variance strategy, in which we tend to bet for the underdog (see next paragraph). Also, the probabilities given by our model are more dissimilar than Kaggle winner's when compared to the betting houses and, as a consequence, we bet in more games and larger quantities, as detailed below. However, we would require a (much) larger number of test games to properly analyze the differences between both models, if they actually exist. Over the 63 tournament games we can state that the Kaggle competition winner follows a lower risk strategy, while our model points towards a higher risk strategy.

The contradiction between this monetary metric and the negative logarithmic loss can be easily explained, because in betting it typically pays off to bet in favor of the underdog (if it is undervalued), and our model tends to provide less extreme probabilities compared to the probabilities submitted by the winner of the Kaggle competition and the implicit probabilities of the betting houses. We end up betting in favor of the team with larger odds and we lose most of the bets, but when we win we recover from the losses. To illustrate this, we include Table 4, where we show the number of games in which we have won the bets that we have placed. For instance, for Pinnacle Sports we decide to bet on 60 games out of 63 (under our 10-model average predictions) and win 21 bets (about a third), while the winner of the Kaggle competition wins 29 bets out of 44 (about two thirds). The winner of the Kaggle competition tends to bet for the favorite, winning a small amount that compensates the few losses. Additionally, we tend to bet more in each game: in average, we stake 14 cents per bet, while the average bet for the Kaggle winner is 9 cents (for Pinnacle Sports). This means that our probabilities are further than those of the winner of the Kaggle competition when compared to the betting houses probabilities. This is not a bad thing for betting, since we need a model that is not only accurate, but also provides different predictions than the implicit probabilities of the betting houses. The betting houses do not necessary need to predict the true probabilities of the event, but they need to predict what people think are the true probabilities (and are willing to bet on). A model that identifies weaker but undervalued teams has the potential to provide huge benefits.

Finally, we show in Figure 4 the profit for each of the Kaggle participants after betting using Kelly's criterion on the 63 games in the Tournament. The results are ordered

Table 4: Number of games in which we win, number of games in which we bet, and number of games for which we have available the bookmaker odds (#Wins/#Bets/#Total).

	(1,0)	(1,1)	(10,10)	(5,15)	10 M. Avg	Kaggle #1
bet365	19/45/62	22/48/62	20/56/62	21/59/62	21/59/62	25/38/62
BetVictor	11/30/61	16/39/61	16/47/61	18/50/61	17/50/61	15/25/61
Pinnacle Sports	23/52/63	24/52/63	20/59/63	21/63/63	21/60/63	29/44/63
Unibet	15/37/62	19/44/62	19/53/62	20/57/62	20/56/62	23/38/62
William Hill	20/47/63	23/50/63	19/54/63	21/58/63	21/58/63	26/42/63
bwin	17/40/61	18/45/61	20/54/61	19/56/61	20/59/61	25/40/61

Rows contain the results for different bet houses, while columns represent different models or specific (K_1, K_2) configurations.

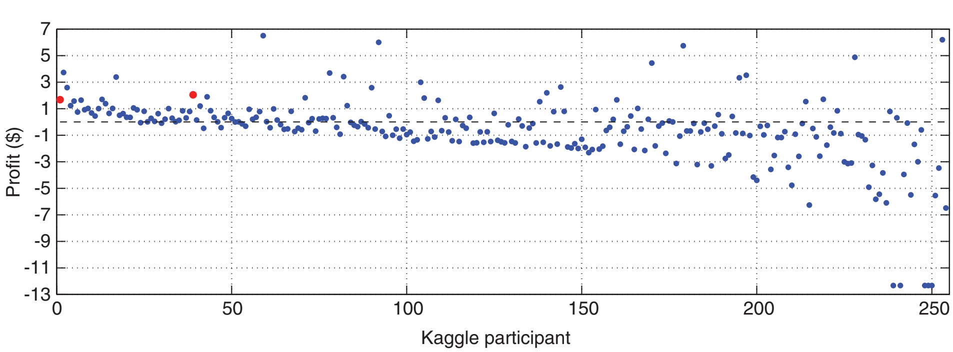

Figure 4: Profit on Pinnacle Sports for all Kaggle participants (ordered according to their final score in Kaggle competition). Red markers show the results by Kaggle winner and our 10-model average.

according to Kaggle leaderboard. The winner is represented by the first red dot and we are represented by the second red dot (the 39th dot overall). From this figure, we can see that the log-loss and the betting profits are related, but they are not a one to one mapping: 46 out of the first 50 participants have positive returns, and so do 23 out of the second 50 participants, 13 out of the third 50 participants, 13 out of the fourth 50 participants and only 7 of the last group of 54. This is easy to understand if we focus on participants with a positive return and a low negative log-loss score. These participants typically post over-confident predictions (close to 100% sure that a certain team will win), these predictions when wrong only give limited betting losses (at most $1 in our comparison), but a nearly unbounded log-loss. We can see that some of these participants would have obtained big wins even though their predictions are over-confident. If this would have been the error measured in Kaggle,[9] we would have been ranked #17.

5 Conclusions

In this paper, we have extended a simple soccer model for college basketball. Outcomes at each game are modeled as independent Poisson random variables whose means depend on the attack and defense coefficients of teams and conferences. Our conference-specific coefficients account for the overall behavior of each conference, while the per-team coefficients provide more specific information about each team. Our vector-valued coefficients can capture different strategies of both teams and conferences. We have derived a variational inference algorithm to learn the attack and defense coefficients, and have applied this algorithm to four March Madness Tournaments. We compare our predictions for the 2014 Tournament to the recent Kaggle competition results and six online betting houses. Simulations show that our model identifies weaker but undervalued teams, which results in a positive mean profit in all the considered betting houses. We also outperform the Kaggle competition winner in terms of mean profit.

Acknowledgments: We thank Kaggle competition organizers for providing us with the individual submissions of all the participants. Francisco J. R. Ruiz is supported by an FPU fellowship from the Spanish Ministry of Education (AP2010-5333). This work is also partially supported by Ministerio de Economía of Spain (projects 'COMONSENS', id. CSD2008-00010, and 'ALCIT', id. TEC2012-38800-C03-01), by Comunidad de Madrid (project 'CASI-CAM-CM', id. S2013/ICE-2845), and by the European Union 7th Framework Programme through the Marie Curie Initial Training Network 'Machine Learning for Personalized Medicine' (MLPM2012, Grant No. 316861).

9 We do not advocate for a change in error measure. Log-loss is more robust and a better indicator, but for a competition we either need more test cases or a statistical test to tell if the winner is significantly better than the other participants.

Appendix A

A Variational Update Equations

In this section, we provide further details on the variational inference algorithm detailed in Section 3. Here, we denote by $\boldsymbol{\phi}_m^{H1}$ ($\boldsymbol{\phi}_m^{A1}$) the K_1-vector composed of the first K_1 elements of $\boldsymbol{\phi}_m^{H}$ ($\boldsymbol{\phi}_m^{A}$), and by $\boldsymbol{\phi}_m^{H2}$ ($\boldsymbol{\phi}_m^{A2}$) the K_2-vector composed of the remaining K_2 elements. We show below the update equations for all the variational parameters, which are needed for the coordinate ascent algorithm:

1. For the home coefficient γ, the updates are given by

$$\gamma^{\mathrm{shp}} = s_\gamma + \sum_{m=1}^{M} y_m^H, \tag{10}$$

$$\gamma^{\mathrm{rte}} = r_\gamma + \sum_{m=1}^{M} \mathbb{E}[\boldsymbol{\alpha}_{h(m)}^\top \boldsymbol{\beta}_{a(m)} + \boldsymbol{\eta}_{\ell(h(m))}^\top \boldsymbol{\rho}_{\ell(a(m))}], \tag{11}$$

   where we denote by $\mathbb{E}[\cdot]$ the expectation with respect to the distribution q.

2. For the team attack and defense parameters $\alpha_{t,k}$ and $\beta_{t,k}$, we obtain

$$\alpha_{t,k}^{\mathrm{shp}} = s_\alpha + \sum_{m:h(m)=t} \phi_{m,k}^{H1} y_m^H + \sum_{m:a(m)=t} \phi_{m,k}^{A1} y_m^A, \tag{12}$$

$$\alpha_{t,k}^{\mathrm{rte}} = r_\alpha + \sum_{m:h(m)=t} \mathbb{E}[\gamma]\mathbb{E}[\beta_{a(m),k}] + \sum_{m:a(m)=t} \mathbb{E}[\beta_{h(m),k}], \tag{13}$$

$$\beta_{t,k}^{\mathrm{shp}} = s_\beta + \sum_{m:a(m)=t} \phi_{m,k}^{H1} y_m^H + \sum_{m:h(m)=t} \phi_{m,k}^{A1} y_m^A, \tag{14}$$

$$\beta_{t,k}^{rte} = r_\beta + \sum_{m:a(m)=t} \mathbb{E}[\gamma]\mathbb{E}[\alpha_{h(m),k}] + \sum_{m:h(m)=t} \mathbb{E}[\alpha_{a(m),k}]. \tag{15}$$

3. For the conference attack and defense parameters $\eta_{l,k}$ and $\rho_{l,k}$, the updates are

$$\eta_{l,k}^{\mathrm{shp}} = s_\eta + \sum_{m:\ell(h(m))=l} \phi_{m,k}^{H2} y_m^H + \sum_{m:\ell(a(m))=l} \phi_{m,k}^{A2} y_m^A, \tag{16}$$

$$\eta_{l,k}^{rte} = r_\eta + \sum_{m:\ell(h(m))=l} \mathbb{E}[\gamma]\mathbb{E}[\rho_{\ell(a(m)),k}] + \sum_{m:\ell(a(m))=l} \mathbb{E}[\rho_{\ell(h(m)),k}], \tag{17}$$

$$\rho^{\text{shp}}_{l,k} = s_\rho + \sum_{m:\ell(a(m))=l} \phi^{H2}_{m,k} y^H_m + \sum_{m:\ell(h(m))=l} \phi^{A2}_{m,k} y^A_m, \tag{18}$$

$$\rho^{\text{rte}}_{l,k} = r_\rho + \sum_{m:\ell(a(m))=l} \mathbb{E}[\gamma]\mathbb{E}[\eta_{\ell(h(m)),k}] + \sum_{m:\ell(h(m))=l} \mathbb{E}[\eta_{\ell(a(m)),k}], \tag{19}$$

4. For the multinomial probabilities of the auxiliary variables, we obtain

$$\phi^{H1}_{m,k} \propto \exp\{\mathbb{E}[\log\gamma] + \mathbb{E}[\log\alpha_{h(m),k}] + \mathbb{E}[\log\beta_{a(m),k}]\}, \tag{20}$$

$$\phi^{H2}_{m,k} \propto \exp\{\mathbb{E}[\log\gamma] + \mathbb{E}[\log\eta_{\ell(h(m)),k}] + \mathbb{E}[\log\rho_{\ell(a(m)),k}]\}, \tag{21}$$

$$\phi^{A1}_{m,k} \propto \exp\{\mathbb{E}[\log\alpha_{a(m),k}] + \mathbb{E}[\log\beta_{h(m),k}]\}, \tag{22}$$

$$\phi^{A2}_{m,k} \propto \exp\{\mathbb{E}[\log\eta_{\ell(a(m)),k}] + \mathbb{E}[\log\rho_{\ell(h(m)),k}]\}, \tag{23}$$

where the proportionality constants ensure that $\boldsymbol{\phi}^H_m$ and $\boldsymbol{\phi}^A_m$ are probability vectors.

All expectations above can be written in closed form, since for a random variable $X \sim \text{gamma}(s, r)$, we have $\mathbb{E}[X] = s/r$ and $\mathbb{E}[\log X] = \psi(s) - \log(r)$, being $\psi(\cdot)$ the digamma function (Abramowitz and Stegun 1972).

B Results for 2011–2014 Tournaments

We now provide some additional results including the 2011–2014 tournaments. In Figure 5, we plot the normalized histogram corresponding to the proportion of observed events for which the predicted probabilities are comprised between the values in the x-axis, across the four considered Tournaments, for several models. The legend indicates the corresponding values of K_1 and K_2. In the figure, we can see that, as the predicted probability increases, so does the proportion of observed events.

Figure 6 shows a boxplot representation (after 500 bootstrap samples) of the negative logarithmic loss for each of the considered season. Here, we can see that 2011 Tournament yielded more unexpected results than in 2014.

Figure 7 shows the average profit we make in all the bet houses after adding together the profit (or loss) for each individual season. Figures 8–10 show the boxplot representation (after 500 bootstrap samples) for each

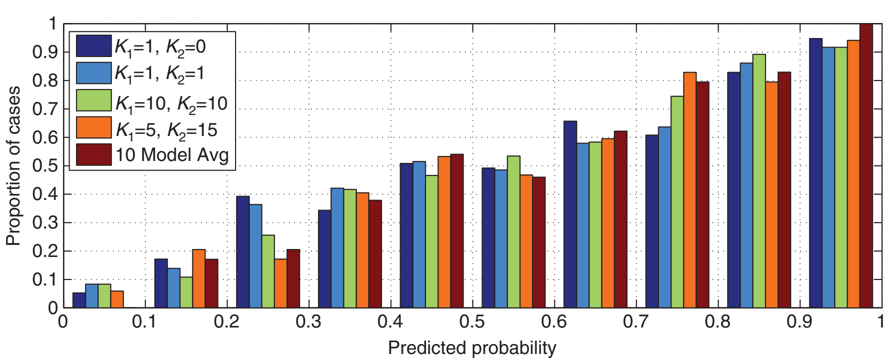

Figure 5: Proportion of observed events for which the predicted probabilities are comprised between the values in the x-axis, across the four considered seasons.

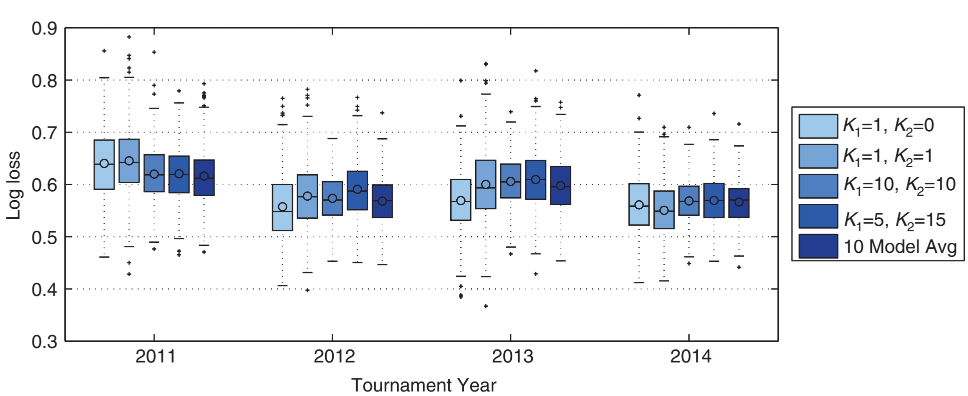

Figure 6: Boxplot representation of logarithmic loss after bootstrap, broken down by season.

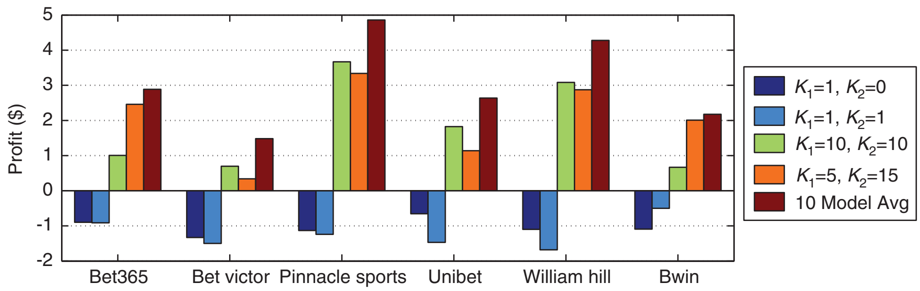

Figure 7: Profit broken down by house, across the four considered seasons.

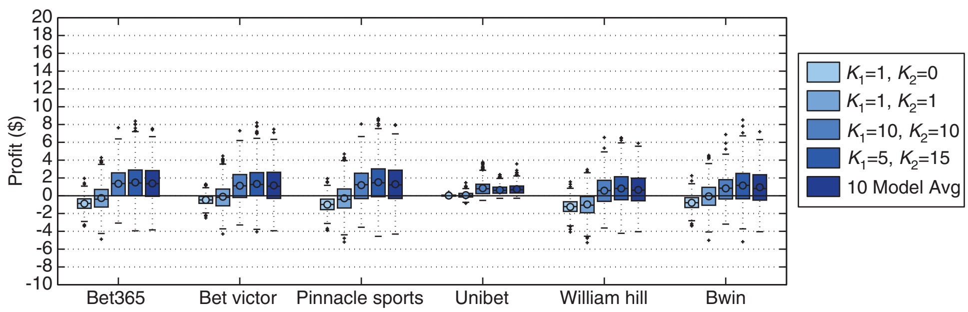

Figure 8: Boxplot representation of profit after bootstrap for season 2010/2011, broken down by house.

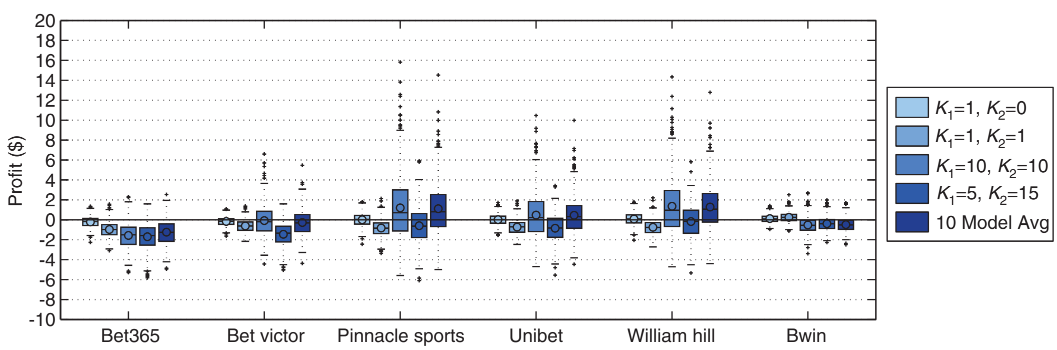

Figure 9: Boxplot representation of profit after bootstrap for season 2011/2012, broken down by house.

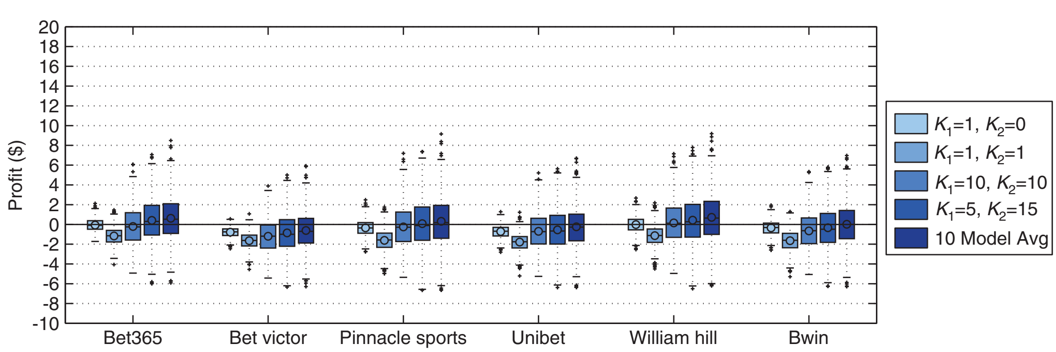

Figure 10: Boxplot representation of profit after bootstrap for season 2012/2013, broken down by house.

Table 5: List of the first 35 games in the 2014 tournament.

#	Team 1			Team 2		
	Team	**Result**	**Prediction (CI)**	**Team**	**Result**	**Prediction (CI)**
1	Ohio State	59	70.9 (57–85)	Dayton	60	60.7 (48–74)
2	Wisconsin	75	69.4 (56–83)	American	35	57.8 (46–71)
3	Colorado	48	64.9 (52–78)	Pittsburgh	77	71.2 (58–85)
4	Cincinnati	57	60.8 (48–74)	Harvard	61	57.2 (45–70)
5	Syracuse	77	74.7 (61–89)	Western Michigan	53	59.3 (47–72)
6	Oregon	87	95.7 (80–112)	Brigham Young	68	95.3 (80–112)
7	Florida	67	71.0 (57–85)	Albany NY	55	53.3 (42–66)
8	Michigan State	93	87.6 (73–103)	Delaware	78	73.9 (60–88)
9	Connecticut	89	78.0 (64–93)	St Joseph's	81	63.4 (51–77)
10	Michigan	57	73.1 (59–87)	Wofford	40	54.1 (42–66)
11	St Louis	83	75.6 (62–90)	North Carolina State	80	66.0 (53–80)
12	Oklahoma	75	81.7 (67–97)	North Dakota State	80	77.5 (63–92)
13	Villanova	73	88.6 (73–104)	Milwaukee	53	71.6 (58–86)
14	Texas	87	71.8 (58–86)	Arizona State	85	74.9 (61–89)
15	Louisville	71	80.2 (66–95)	Manhattan	64	64.0 (51–77)
16	San Diego State	73	70.6 (57–85)	New Mexico State	69	63.6 (51–77)
17	Duke	71	79.6 (65–94)	Mercer	78	72.8 (59–87)
18	Baylor	74	68.6 (55–82)	Nebraska	60	66.7 (54–80)
19	New Mexico	53	72.4 (59–87)	Stanford	58	70.7 (57–85)
20	Arizona	68	75.2 (61–90)	Weber State	59	53.2 (42–66)
21	Massachusetts	67	71.7 (58–86)	Tennessee	86	75.8 (62–90)
22	Creighton	76	86.3 (71–102)	Louisiana-Lafayette	66	69.4 (56–83)
23	Kansas	80	88.0 (73–104)	Eastern Kentucky	69	68.4 (55–82)
24	Gonzaga	85	74.7 (61–89)	Oklahoma State	77	79.2 (65–94)
25	Memphis	71	79.4 (65–94)	George Washington	66	69.7 (56–84)
26	Wichita State	64	64.6 (52–78)	Cal Poly	37	53.8 (42–66)
27	North Carolina	79	75.7 (62–90)	Providence	77	71.6 (58–86)
28	Virginia Commonwealth	75	71.5 (58–86)	Stephen F. Austin	77	67.7 (54–81)
29	Virginia	70	66.6 (53–80)	Coastal Carolina	59	49.5 (38–61)
30	Kentucky	56	67.7 (54–82)	Kansas State	49	60.4 (48–73)
31	Iowa State	93	80.1 (66–95)	North Carolina Central	75	68.3 (55–82)
32	UCLA	76	81.3 (67–96)	Tulsa	59	70.1 (57–84)
33	Florida	61	65.7 (53–79)	Pittsburgh	45	61.6 (49–75)
34	Louisville	66	74.2 (60–89)	St Louis	51	64.5 (52–78)
35	Texas	65	66.4 (53–80)	Michigan	79	72.8 (59–87)

For each game and team, we show the actual outcome of the game, as well as the predicted mean value, together with the 90% credible interval (labelled as "CI" in the Table).

individual season (the plot for 2014 tournament is included in the main text).

C List of 2014 Tournament Games

We show in Table 5 the list corresponding to the 35 games in the 2014 March Madness Tournament not shown in Table 3. For each game, we show the actual outcome of the game, as well as the predicted mean values and the 90% credible intervals.

References

Abramowitz, M. and I. A. Stegun. 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dover Publications.

Baio, G. and M. A. Blangiardo. 2010. "Bayesian Hierarchical Model for the Prediction of Football Results." *Journal of Applied Statistics* 37:253–264.

Canny, J. 2004. "GaP: A Factor Model for Discrete Data." In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, pp. 122–129.

Cemgil, A. T. 2009. “Bayesian Inference for Nonnegative Matrix Factorisation Models.” *Computational Intelligence and Neuroscience* 2009: 17.

Crowder, M., M. Dixon, A. Ledford, and M. Robinson. 2002. “Dynamic Modelling and Prediction of English Football League Matches for Betting.” *Journal of the Royal Statistical Society: Series D (The Statistician)* 51:157–168.

Dixon, M. J. and S. G. Coles. 1997. “Modelling Association Football Scores and Inefficiencies in the Football Betting Market.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 46:265–280.

Dunson, D. B. and A. H. Herring. 2005. “Bayesian Latent Variable Models for Mixed Discrete Outcomes.” *Biostatistics* 6:11–25.

Efron, B. 1979. “Bootstrap Methods: Another Look at the Jackknife.” *The Annals of Statistics* 7:1–26.

Ghahramani, Z. and M. J. Beal. 2000. “Propagation Algorithms for Variational Bayesian Learning.” In Advances in Neural Information Processing Systems 13, pp. 507–513.

Gopalan, P., J. M. Hofman, and D. M. Blei. 2013. “Scalable Recommendation with Poisson Factorization.” *arXiv preprint arXiv:1311.1704*.

Gopalan, P., F. J. R. Ruiz, R. Ranganath, and D. M. Blei. 2014. “Bayesian Nonparametric Poisson Factorization for Recommendation Systems,” *Artificial Intelligence and Statistics (AISTATS)* 33:275–283.

Heuer, A., C. Müller, and O. Rubner. 2010. “Soccer: is Scoring Goals a Predictable Poissonian Process?” *arXiv preprint arXiv:1002.0797*.

Hoffman, M. D., D. M. Blei, C. Wang, and J. Paisley. 2013. “Stochastic Variational Inference.” *Journal of Machine Learning Research* 14:1303–1347.

Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. 1999. “An Introduction to Variational Methods for Graphical Models.” *Machine Learning* 37:183–233.

Kelly, J. L. 1956. “A New Interpretation of Information Rate.” *IRE Transactions on Information Theory* 2:185–189.

Maher, M. J. 1982. “Modelling Association Football Scores.” *Statistics Neerland* 36:109–118.

Miller, A., L. Bornn, R. Adams, and K. Goldsberry. 2014. “Factorized Point Process Intensities: A Spatial Analysis of Professional Basketball.” *arXiv preprint arXiv:1401.0942*.

Skellam, J. G. 1946. “The Frequency Distribution of the Difference between Two Poisson Variates Belonging to Different Populations.” *Journal of the Royal Statistical Society* 109:296+3.

Wilcoxon, F. 1945. “Individual Comparisons by Ranking Methods.” *Biometrics Bulletin* 1:80–83.

Zhou, M., L. Hannah, D. B. Dunson, and L. Carin. 2012. “Beta-Negative Binomial Process and Poisson Factor Analysis.” *Journal of Machine Learning Research – Proceedings Track* 22:1462–1471.