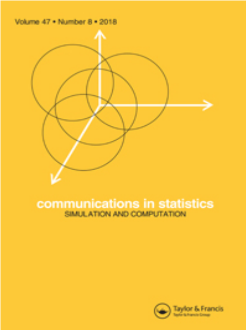

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

A discrete-time and finite-state Markov Chain model for association football matches

Qingrong Zou, Qi Li, Hao Guo & Jian Shi

To cite this article: Qingrong Zou, Qi Li, Hao Guo & Jian Shi (2018) A discrete-time and finite-state Markov Chain model for association football matches, Communications in Statistics - Simulation and Computation, 47:8, 2476-2485, DOI: 10.1080/03610918.2017.1348518

To link to this article: https://doi.org/10.1080/03610918.2017.1348518

Accepted author version posted online: 12 Jul 2017.
Published online: 10 Aug 2017.

COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2018, VOL. 47, NO. 8, 2476–2485
https://doi.org/10.1080/03610918.2017.1348518

A discrete-time and finite-state Markov Chain model for association football matches

Qingrong Zou[a], Qi Li[b], Hao Guo[a], and Jian Shi[a,c]

[a]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China; [b]Editorial Department of the Journal of Jilin University, Jilin University, Changchun, China; [c]School of Mathematical Sciences, University of the Chinese Academy of Sciences, Beijing, China

ABSTRACT
A birth process model proposed by Dixon and Robinson has been widely used in football spread betting market. However, multiple goals in a minute are permitted in the model, which does not conform to historical record. Moreover, it is difficult to calculate the outcome probability of the process accurately. The article presents a discrete-time and finite-state Markov chain model for real-time forecast of football matches and a recursive algorithm is derived to calculate the outcome probability accurately. The empirical study shows that the proposed model outperforms the models of Dixon and Robinson and Dixon and Coles.

ARTICLE HISTORY
Received 16 June 2016
Accepted 22 June 2017

KEYWORDS
Association football; Discrete-time and finite-state Markov chain; Football outcome forecast; Maximum likelihood; Recursive algorithm

MATHEMATICS SUBJECT CLASSIFICATION
65C20; 62M20; 60J10; 62P25

1. Introduction

Sports betting markets are becoming increasingly competitive (Hvattum and Arntzen, 2010). Association football is the most popular sport in the world and comprises the fastest growing gambling market (Constantinou et al., 2012). The outcome of association football match depends on many factors, such as home effect, the effect of injured players and various psychological effects (Rue and Salvesen, 2000).

Models for forecast of association football matches for betting have existed for a long time. Previously, people were in favor of the negative binomial model for association football scores and rejected the Poisson model (Moroney, 1951; Reep et al., 1971). Reep et al. (1971) showed that chance dominates the game. On the one hand, it is recognized that skill rather than chance dominates the game for many games, which is also proved by Hill (1974) and Fahrmeir and Tutz (1994). On the other hand, those researches failed to estimate probabilities on a team-specific basis. Maher (1982) first considered a model for football scores of a match between specific teams. Independent Poisson distributions for scores of home and away teams were assumed. Dixon and Coles (1997) modified the model to reflect dependence for low scores (0–0, 1–0, 0–1, and 1–1). Also, they introduced a time-dependence effect to make parameters locally constant through time rather than static. McHale and Scarf (2011) studied the dependence of goals scored. In addition, Bayesian and machine learning techniques are also applied

CONTACT Jian Shi jshi@iss.ac.cn Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.

to predict outcome of football matches, see Joseph et al. (2006), Constantinou et al. (2012), Baio and Blangiardo (2010), among others.

In spreading betting markets, it is crucial to predict the following score conditional on the current score at any time during the match between specified teams. To this end, Dixon and Robinson (1998) developed a birth process model. The processes of goal times of home and away teams are taken to be two nonhomogeneous Poisson processes, which indicates more than one goal in any time interval is permitted. Later, Volf (2009) proposed a random point process model similar to Dixon and Robinson (1998) by considering the effect of covariates. However, according to historical record, no more than one goal in a minute had happened, except for the time intervals (44, 45] and (89, 90] in consideration of injury. Goal times are generally recorded in minutes, therefore the birth process model can be seen as an approximation of the real data. In order to be more accordant with practical circumstances, we propose a discrete-time and finite-state Markov chain model on the basis of the Poisson processes. In addition, in view of the birth process formulation, it requires to calculate the probability of being in every state of $\{(x, y) : x, y = 0, 1, \ldots\}$ at the 90th minute, which is obtained by integrating over all possible times and all possible routes to arrive at a state (x, y). Therefore, the heavy computation makes direct calculation infeasible (Dixon and Robinson, 1998). Based on the discrete-time and finite-state Markov chain model, we derive a recursive algorithm that makes direct calculation feasible.

In Section 2, we develop the discrete-time and finite-state Markov chain model; in Section 3, we describe the data and present the results of empirical study, furthermore, we compare the proposed model with the models of Dixon and Robinson (1998) and Dixon and Coles (1997); finally, Section 4 presents the conclusion.

2. The discrete-time and finite-state Markov chain model

Before introducing our model for goal times, we summarize the model of Dixon and Robinson (1998) on which our model is based. The crucial point of the birth process model is the intensities $\lambda_k^*(t)$ and $\mu_k^*(t)$ of home and away teams. In the match k between two teams indexed by H and A, the intensities of the home and away teams are shown in Eqs. (1) and (2), respectively, where α_H and β_H measure the attack and defense abilities of home team H, respectively, α_A and β_A measure the attack and defense abilities of away team A respectively, γ is the home effect, ρ_1 and ρ_2 are the injury time effect parameters, λ_{xy} and μ_{xy} are score adjustment factor parameters, ξ_1 and ξ_2 measure the continuously increasing scoring rates. $\delta_1(t) = I(t = 45)$, $\delta_2(t) = I(t = 90)$, and $I(\cdot)$ is the indicator function.

$$\lambda_k^*(t) = \rho_1^{\delta_1(t)} \rho_2^{\delta_2(t)} \lambda_{xy} \gamma \alpha_H \beta_A + \xi_1 t, \tag{1}$$

$$\mu_k^*(t) = \rho_1^{\delta_1(t)} \rho_2^{\delta_2(t)} \mu_{xy} \alpha_A \beta_H + \xi_2 t, \tag{2}$$

Dixon and Robinson (1998) found in the best fitting-model that the home parameter λ_{xy} could be appropriately defined as

$$\lambda_{xy} = \begin{cases} \lambda_{10}, & \textit{if} \quad x = 1, y = 0; \\ \lambda_{01}, & \textit{if} \quad x = 0, y = 1; \\ \lambda_{21}, & \textit{if} \quad x + y > 1, x - y \geq 1; \\ \lambda_{12}, & \textit{if} \quad x + y < 1, x - y \leq -1; \\ 1, & \text{otherwise}, \end{cases}$$

and the away parameter μ_{xy} could also be defined similarly, where the score is $x : y$ at time $(t-1)$.

2.1. The proposed discrete-time and finite-state Markov chain model

Let $S_H^{(t)}$ and $S_A^{(t)}$ be the numbers of goals scored by the home and away sides at time t respectively for $t = 0, 1, 2, \dots, 90$. From the theory of Poisson processes, the probability of that home team scored x goals and away team scored y goals in the period of time $(t-1, t]$, on the condition that the score is $i : j$ at time $(t-1)$, is derived as the following:

$$\begin{aligned} f^{(t)}_{(i,j),(x,y)} &= Pr\left\{S_H^{(t)} = i+x, S_A^{(t)} = j+y \,|\, S_H^{(t-1)} = i, S_A^{(t-1)} = j\right\} \\ &= \frac{\left(\theta^{(t)}_{H,(i,j)}\right)^x}{x!} \cdot e^{-\theta^{(t)}_{H,(i,j)}} \cdot \frac{\left(\theta^{(t)}_{A,(i,j)}\right)^y}{y!} \cdot e^{-\theta^{(t)}_{A,(i,j)}}, \end{aligned}$$

where, $1 \le t \le 90$.

$$\begin{aligned} \theta^{(t)}_{H,(i,j)} &= \left(\lambda_{ij}^{(t-1)} \rho_1^{\delta_1(t)} \rho_2^{\delta_2(t)}\right) \cdot (\gamma \alpha_H \beta_A) + \frac{2t-1}{2} \cdot \xi_1, \\ \theta^{(t)}_{A,(i,j)} &= \left(\mu_{ij}^{(t-1)} \rho_1^{\delta_1(t)} \rho_2^{\delta_2(t)}\right) \cdot (\alpha_A \beta_H) + \frac{2t-1}{2} \cdot \xi_2. \end{aligned}$$

According to actual match situation, we can assume that the total number of goals scored by home and away teams is no more than 2 in the time interval (44, 45], no more than 3 in the time interval (89, 90], and no more than 1 in other one minute time intervals. Under the above assumptions, we have the following conditional probabilities:

$$p^{(t)}_{(i,j),(x,y)} = \begin{cases} \dfrac{f^{(t)}_{(i,j),(x,y)}}{\sum_{x+y\le 2} f^{(t)}_{(i,j),(x,y)}}, & x \ge 0, y \ge 0, x+y \le 2,\ t = 45; \\[2ex] \dfrac{f^{(t)}_{(i,j),(x,y)}}{\sum_{x+y\le 3} f^{(t)}_{(i,j),(x,y)}}, & x \ge 0, y \ge 0, x+y \le 3,\ t = 90; \\[2ex] \dfrac{f^{(t)}_{(i,j),(x,y)}}{\sum_{x+y\le 1} f^{(t)}_{(i,j),(x,y)}}, & x \ge 0, y \ge 0, x+y \le 1,\ \text{otherwise}. \end{cases} \tag{3}$$

We are now in a position to establish a discrete-time and finite-state Markov chain. Let

$$\mathcal{M} = \{(x, y) : \ x \text{ and } y \text{ are integers},\ 0 \le x \le 93, 0 \le y \le 93\}$$

be a two-dimensional finite-state space. Our Markov chain is defined to be the sequence of random vectors $\{(X(t), Y(t)) : t = 0, 1, 2, \dots, 90\}$ which satisfy the following conditions: (C.1)$(X(t), Y(t))$ takes values on $\mathcal{M}$ for $t = 0, 1, \dots, 90$; (C.2)$(0, 0)$ is the initial state, i.e., $(X(0), Y(0)) = (0, 0)$; (C.3) The transition probability for the chain to jump from the state (i, j) at time $(t-1)$ to the state $(i+x, j+y)$ at time t is

$$Pr\left\{(X(t), Y(t)) = (i+x, j+y) |\, (X(t-1), Y(t-1)) = (i, j)\right\} = p^{(t)}_{(i,j)(x,y)}$$

for $t = 1, 2, \dots, 90$.

2.2. The likelihood function of the Markov chain model

Having the transition probabilities of each state for home and away teams, in order to make predictions, we need to fit the model to data so that the parameters are estimated. This can be done by the method of maximum likelihood and the optimization of the likelihood function can be solved by the coordinate descent algorithm (see the Appendix).

The likelihood of the observed Markov chain, for a particular match k, is

$$L_k = \prod_{t=1}^{90} p^{(t)}_{(X(t-1),Y(t-1))(X(t)-X(t-1),Y(t)-Y(t-1))}. \tag{4}$$

For $t \neq 45$ and $t \neq 90$, denote $S_1^{(t)} = 1 + \theta^{(t)}_{H,(X(t-1),Y(t-1))} + \theta^{(t)}_{A,(X(t-1),Y(t-1))}$, according to the assumption and Eq. (3), the state transition probability is one of the following three cases:

$$\begin{aligned}
p^{(t)}_{(X(t-1),Y(t-1))(0,0)} &= \frac{1}{S_1^{(t)}};\\
p^{(t)}_{(X(t-1),Y(t-1))(1,0)} &= \frac{\theta^{(t)}_{H,(X(t-1),Y(t-1))}}{S_1^{(t)}};\\
p^{(t)}_{(X(t-1),Y(t-1))(0,1)} &= \frac{\theta^{(t)}_{A,(X(t-1),Y(t-1))}}{S_1^{(t)}}.
\end{aligned}$$

For $t = 45$, denote

$$\begin{aligned}
S_2^{(t)} = S_1^{(t)} &+ \theta^{(t)}_{H,(X(t-1),Y(t-1))}\theta^{(t)}_{A,(X(t-1),Y(t-1))}\\
&+\frac{1}{2}\left(\theta^{(t)}_{H,(X(t-1),Y(t-1))}\right)^2 + \frac{1}{2}\left(\theta^{(t)}_{A,(X(t-1),Y(t-1))}\right)^2
\end{aligned}$$

the state transition probability is one of the following six cases:

$$\begin{aligned}
p^{(t)}_{(X(t-1),Y(t-1))(0,0)} &= \frac{1}{S_2^{(t)}};\\
p^{(t)}_{(X(t-1),Y(t-1))(1,0)} &= \frac{\theta^{(t)}_{H,(X(t-1),Y(t-1))}}{S_2^{(t)}};\\
p^{(t)}_{(X(t-1),Y(t-1))(0,1)} &= \frac{\theta^{(t)}_{A,(X(t-1),Y(t-1))}}{S_2^{(t)}};\\
p^{(t)}_{(X(t-1),Y(t-1))(1,1)} &= \frac{\theta^{(t)}_{H,(X(t-1),Y(t-1))}\theta^{(t)}_{A,(X(t-1),Y(t-1))}}{S_2^{(t)}};\\
p^{(t)}_{(X(t-1),Y(t-1))(2,0)} &= \frac{\left(\theta^{(t)}_{H,(X(t-1),Y(t-1))}\right)^2}{2S_2^{(t)}};\\
p^{(t)}_{(X(t-1),Y(t-1))(0,2)} &= \frac{\left(\theta^{(t)}_{A,(X(t-1),Y(t-1))}\right)^2}{2S_2^{(t)}}.
\end{aligned}$$

For $t = 90$, denote

$$\begin{aligned}
S_3^{(t)} = S_2^{(t)} &+ \frac{1}{2}\left(\theta^{(t)}_{H,(X(t-1),Y(t-1))}\right)^2\theta^{(t)}_{A,(X(t-1),Y(t-1))}\\
&+\frac{1}{2}\theta^{(t)}_{H,(X(t-1),Y(t-1))}\left(\theta^{(t)}_{A,(X(t-1),Y(t-1))}\right)^2\\
&+\frac{1}{6}\left(\theta^{(t)}_{H,(X(t-1),Y(t-1))}\right)^3 + \frac{1}{6}\left(\theta^{(t)}_{A,(X(t-1),Y(t-1))}\right)^3,
\end{aligned}$$

the state transition probability is one of the following ten cases:

$$p^{(t)}_{(X(t-1),Y(t-1))(0,0)} = \frac{1}{S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(1,0)} = \frac{\theta^{(t)}_{H,(X(t-1),Y(t-1))}}{S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(0,1)} = \frac{\theta^{(t)}_{A,(X(t-1),Y(t-1))}}{S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(1,1)} = \frac{\theta^{(t)}_{H,(X(t-1),Y(t-1))}\theta^{(t)}_{A,(X(t-1),Y(t-1))}}{S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(2,0)} = \frac{\left(\theta^{(t)}_{H,(X(t-1),Y(t-1))}\right)^2}{2S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(0,2)} = \frac{\left(\theta^{(t)}_{A,(X(t-1),Y(t-1))}\right)^2}{2S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(2,1)} = \frac{\left(\theta^{(t)}_{H,(X(t-1),Y(t-1))}\right)^2\theta^{(t)}_{A,(X(t-1),Y(t-1))}}{2S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(1,2)} = \frac{\theta^{(t)}_{H,(X(t-1),Y(t-1))}\left(\theta^{(t)}_{A,(X(t-1),Y(t-1))}\right)^2}{2S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(3,0)} = \frac{\left(\theta^{(t)}_{H,(X(t-1),Y(t-1))}\right)^3}{6S_3^{(t)}};$$

$$p^{(t)}_{(X(t-1),Y(t-1))(3,0)} = \frac{\left(\theta^{(t)}_{A,(X(t-1),Y(t-1))}\right)^3}{6S_3^{(t)}}.$$

Scores between matches are independent, so the overall likelihood is obtained by taking the product over matches.

2.3. A recursive algorithm for prediction

In order to predict the outcome of the discrete-time and finite-state Markov chain model, we develop a recursive algorithm to calculate the probability of any future state given the current state for the above Markov chain $\{(X(t), Y(t))\}_{t=0}^{90}$.

In real situations, it is unnecessary to calculate the full probability matrix for $\{(X(t), Y(t))\}_{t=0}^{90}$ because maximum number of goals scored by a team in one match is far less than 93. A sub-matrix of probability is of interest in practice. Therefore, we assume $\max(X(90), Y(90)) \le M \le 93$, where M is an appropriate positive integer. According to historical record, M can be taken as 15.

Let $A^{(t_1)}_{t_0}$ be the conditional state probability matrix at time t_1 given the state information at time t_0, where $A^{(t_1)}_{t_0} = (a^{(t_1)}_{ij,t_0})_{(M+1)\times(M+1)}$ for $0 \le t_0 < t_1 \le 90$.

For simplicity of mathematical expression, we extend $p^{(t)}_{(i,j)(x,y)}$ as follows:

$$\widetilde{p}^{(t)}_{(i,j),(x,y)} = \begin{cases} p^{(t)}_{(i,j),(x,y)}, & \text{for } i \ge 0 \text{ and } j \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

Let $B^{(t)} = (b_{ij}^{(t)})_{(M+4)\times(M+4)}$ be an $(M+4)\times(M+4)$ matrix for $0 \le t \le 90$. Given $0 \le t_0 < t_1 \le 90$, suppose the state is (u, v) at time t_0. The recursive algorithm to calculate $B^{(t)}$ from $t = t_0$ to $t = t_1$ is as follows:

1. For $t = t_0 : B^{(t_0)} = (b_{ij}^{(t_0)})$, where $b_{ij}^{(t_0)} = \begin{cases} 1, & i = u+4,\ j = v+4, \\ 0, & \text{otherwise,} \end{cases}$ for $1 \le i, j \le (M+4)$.
2. For $t = (t_0+1), \ldots, t_1 : B^{(t)} = (b_{ij}^{(t)})$, where
   (1) for $1 \le i \le (u+3)$ or $1 \le j \le (v+3)$, $b_{ij}^{(t)} = b_{ij}^{(t-1)} = 0$,
   (2) for $(u+4) \le i \le (M+4)$ and $(v+4) \le j \le (M+4)$,
      i) if $t = 45$,

$$\begin{aligned} b_{ij}^{(t)} &= b_{ij}^{(t-1)}\widetilde{p}_{(i-4,j-4),(0,0)}^{(t)} + b_{(i-1)(j-1)}^{(t-1)}\widetilde{p}_{(i-5,j-5),(1,1)}^{(t)} \\ &\quad + b_{(i-1)j}^{(t-1)}\widetilde{p}_{(i-5,j-4),(1,0)}^{(t)} + b_{(i-2)j}^{(t-1)}\widetilde{p}_{(i-6,j-4),(2,0)}^{(t)} \\ &\quad + b_{i(j-1)}^{(t-1)}\widetilde{p}_{(i-4,j-5),(0,1)}^{(t)} + b_{i(j-2)}^{(t-1)}\widetilde{p}_{(i-4,j-6),(0,2)}^{(t)}; \end{aligned}$$

      ii) if $t = 90$,

$$\begin{aligned} b_{ij}^{(t)} &= b_{ij}^{(t-1)}\widetilde{p}_{(i-4,j-4),(0,0)}^{(t)} + b_{(i-1)(j-1)}^{(t-1)}\widetilde{p}_{(i-5,j-5),(1,1)}^{(t)} \\ &\quad + b_{(i-1)j}^{(t-1)}\widetilde{p}_{(i-5,j-4),(1,0)}^{(t)} + b_{(i-2)j}^{(t-1)}\widetilde{p}_{(i-6,j-4),(2,0)}^{(t)} + b_{(i-3)j}^{(t-1)}\widetilde{p}_{(i-7,j-4),(3,0)}^{(t)} \\ &\quad + b_{i(j-1)}^{(t-1)}\widetilde{p}_{(i-4,j-5),(0,1)}^{(t)} + b_{i(j-2)}^{(t-1)}\widetilde{p}_{(i-4,j-6),(0,2)}^{(t)} + b_{i(j-3)}^{(t-1)}\widetilde{p}_{(i-4,j-7),(0,3)}^{(t)} \\ &\quad + b_{(i-2)(j-1)}^{(t-1)}\widetilde{p}_{(i-6,j-5),(2,1)}^{(t)} + b_{(i-1)(j-2)}^{(t-1)}\widetilde{p}_{(i-5,j-6),(1,2)}^{(t)}; \end{aligned}$$

      iii) if $t \ne 45$ and $t \ne 90$,

$$b_{ij}^{(t)} = b_{ij}^{(t-1)}\widetilde{p}_{(i-4,j-4),(0,0)}^{(t)} + b_{(i-1)j}^{(t-1)}\widetilde{p}_{(i-5,j-4),(1,0)}^{(t)} + b_{i(j-1)}^{(t-1)}\widetilde{p}_{(i-4,j-5),(0,1)}^{(t)}.$$

Therefore, when the state at time t_0 is (u, v), the conditional state probability matrix at time t_1 is

$$A_{t_0}^{(t_1)} = \left(a_{ij,t_0}^{(t_1)}\right)_{(M+1)\times(M+1)} = \left(C \cdot b_{(i+3)(j+3)}^{(t_1)}\right)_{(M+1)\times(M+1)},$$

where $C = 1/\sum_{i=4}^{M+4}\sum_{j=4}^{M+4} b_{ij}^{(t_1)}$ is the normalizing constant. Particularly, when $t_0 = 0$. $A_0^{(t)}$ is the state probability matrix of the state transition model at time t for $t = 1, 2, \ldots, 90$.

The recursive algorithm is far faster than the Monte Carlo simulation, which greatly improves the efficiency of real-time forecast of football matches.

3. Empirical study

3.1. Data

Football match data generally contain the times of goals, the identity of team of each goal scored, and the full-time score result. The data to be analyzed consists of the English Premier League (denoted as EPL) and the German Bundesliga (denoted as GSL) football matches over 2009–2010, 2010–2011, 2011–2012, 2012–2013, and 2013–2014 seasons, which contains 11 semi-seasons. We use the first 8 semi-seasons as training dataset and the last 3 semi-seasons as test dataset.

In order to compare our model with the models of Dixon and Robinson (1998) and Dixon and Coles (1997), we delete the matches containing teams not appearing in the test dataset in order to avoid the impact of new team, for it is quite difficult to predict performance of new team. After data preprocessing, the EPL data comprise 2081 matches and the GSL data comprise 1720 matches.

3.2. Measure of prediction performance

Rue and Salvesen (2000) suggested to use the pseudo-likelihood statistic as a measure of prediction performance, which was first defined by Dixon and Coles (1997). Goddard (2005) adopted the criterion as well. The pseudo-likelihood is equivalent to the geometric mean of the predicted probabilities for the actual results of all matches in the test dataset, which is shown in Eq. (5) with N matches

$$PL = \left[\prod_{k=1}^{N} (p_k^H)^{\delta_k^H} (p_k^D)^{\delta_k^D} (p_k^A)^{\delta_k^A} \right]^{1/N}, \tag{5}$$

where, in the match k,

$$\left(\delta_k^H, \delta_k^D, \delta_k^A\right) = \begin{cases} (1, 0, 0), & \text{if home team win;} \\ (0, 1, 0), & \text{if draw;} \\ (0, 0, 1), & \text{if away team win,} \end{cases}$$

p_k^H, p_k^A, and p_k^D are the predicted probabilities for home win, away win and draw in the match k, respectively. If we randomly predict win–draw–lose match results, the PL is 1/3. The greater the PL value the better the models prediction ability. However, small numerical variation in the pseudo-likelihood statistic may indicate large variations in forecasting capability (Goddard, 2005).

3.3. Forecasting performance of the proposed model

For the model parameters $\alpha_i,\ \beta_j,\ \lambda_{xy},\ \mu_{xy},\ \rho_1,\ \rho_2,\ \xi_1,\ \xi_2,\ \gamma,\ (i, j = 1, 2, \ldots, d)$, we update them every two rounds, where d denotes the number of teams. To be specific, we fit the model on the training dataset and forecast outcomes of the following two rounds matches. Having made predictions, we enlarge the training dataset by taking into account the predicted matches and refit the model. We repeat this procedure until the last two rounds of games in the 2013/14 season have been forecasted.

Figure 1 shows the forecast performance of the discrete-time and finite-state Markov chain model and the birth process model on the EPL dataset and the GSL dataset, respectively. We can see from Fig. 1 that the prediction effects of the two models are basically the same for the EPL dataset, while the birth process model does better for the GSL dataset. For a clearer comparison, we report the pseudo-likelihoods of the last three semi-seasons of the EPL test dataset and the GSL test dataset in Tables 1 and 2, respectively. Denote test_1 as the second half season of 2012–2013, test_2 as the first half season of 2013–2014 and test_3 as the second half season of 2013–2014. For the English Premier League, the discrete-time and finite-state Markov chain model is better than the birth process model by relatively 0.26% higher, however, for the German Bundesliga, the Markov chain model performs worse than the birth process model by 0.68% lower.

Dixon and Robinson (1998) mentioned that the dependence was most noticeable when any of the two teams had a narrow lead. They conclude that the home scoring rate decreases

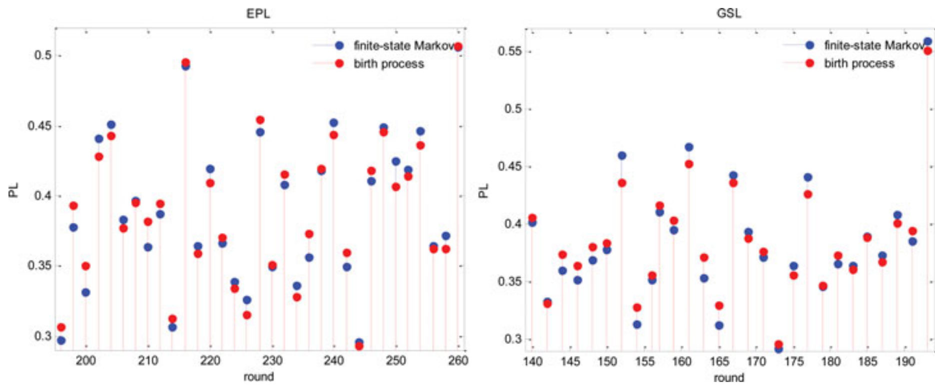

Figure 1. The pseudo-likelihood of test dataset for the discrete-time and finite-state Markov chain model and the birth process model.

Table 1. The pseudo-likelihoods of EPL.

	Number of matches	Markov chain	Markov chain*	Dixon & Robinso	Dixon & Coles
test_1	182	0.3811	0.3842	0.3812	0.3794
test_2	153	0.3723	0.3800	0.3727	0.3743
test_3	153	0.4003	0.4044	0.3965	0.395
Average	488	0.3842	0.3891	0.3832	0.3828

Note: Markov chain* denotes the discrete-time and finite-state Markov chain model without the score adjustment factor parameters.

and the away scoring rate increases significantly when the home team is leading, while both teams tend to increase scoring rates when the away team is leading. This is the reason they introduce the score adjustment factor parameters λ_{xy} and μ_{xy}. They also claim that there is no evidence for the immediate strike back. However, in our Markov chain model, the likelihood is the product of the transition probability of every one minute, the score adjustment factor parameters have to take effect immediately, which may distort the truth. Therefore, we have strong reason to remove the score adjustment factors from the Markov model. We call the discrete-time and finite-state Markov chain model without score adjustment factors the modified Markov chain model.

Figure 2 shows the stems of the pseudo-likelihood of the modified Markov chain model and the birth process model. We can see that the modified Markov chain model clearly performs better than the birth process model, especially for the German Bundesliga. Tables 1 and 2 also show the prediction performance of the modified Markov chain model on the last three semi-seasons. We can see that the prediction ability of the modified Markov chain model improves greatly 1.28% higher on the EPL dataset and 2.44% higher on the GSL dataset compared with the original Markov chain model. Moreover, the modified Markov chain model outperforms the birth process model, 1.54% higher on the EPL dataset and 1.74% higher on

Table 2. The pseudo-likelihoods of GSL.

	Number of matches	Markov chain	Markov chain*	Dixon & Robinso	Dixon & Coles
test_1	153	0.3676	0.3802	0.3728	0.3765
test_2	135	0.3774	0.3877	0.3816	0.3851
test_3	137	0.3890	0.3932	0.3869	0.3866
Average	425	0.3775	0.3867	0.3801	0.3825

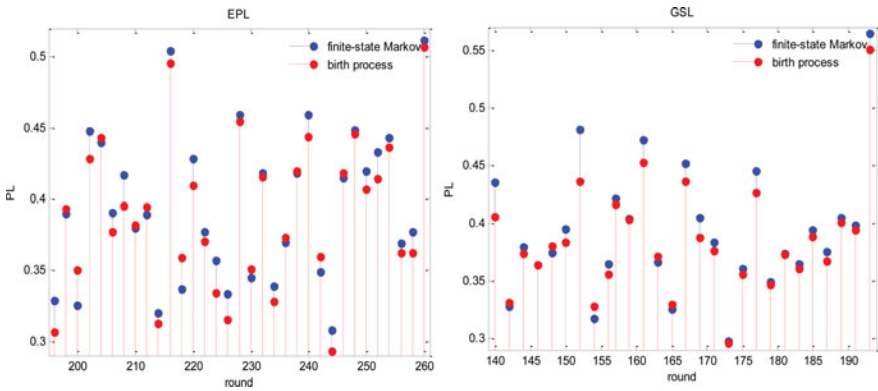

Figure 2. The pseudo-likelihood of test dataset for the modified Markov chain model and the birth process model.

the GSL dataset. It can also be seen that the birth process model performs worse than the model of Dixon and Coles (1997) on the GSL dataset with 0.63% lower and performs better on the EPL dataset with only 0.10% higher.

4. Conclusion

In this article, a discrete-time and finite-state Markov chain model is proposed to fit goal times of football matches. It can provide estimates of the following outcome conditional on the current score during a match between specified teams. The empirical study shows a significant improvement in match outcome prediction over the models of Dixon and Robinson (1998) and Dixon and Coles (1997). We find that it is better to remove the score adjustment factors from the Markov chain model, which corresponds to the conclusion of no evidence for the immediate strike back claimed by Dixon and Robinson (1998). A recursive algorithm is developed as well in order to calculate the probability of any future state given the current state for the proposed model. Comparing with the Monte Carlo method, this algorithm makes direct calculation of score probability feasible, which increases the efficiency of the model in application.

Acknowledgments

The authors would like to thank the editor and the two anonymous referees for their constructive comments and suggestions. We thank Beijing StausWin Lottery Operations Technology Ltd. for providing us the EPL data and the GSL data.

Appendix

A coordinate descent algorithm for parameter estimation

In the discrete-time and finite-state Markov chain model, we need to estimate $2d + 13$ parameters $(\alpha_1, \alpha_2, \ldots, \alpha_d, \beta_1, \beta_2, \ldots, \beta_d, \lambda_{10}, \lambda_{01}, \lambda_{21}, \lambda_{12}, \mu_{10}, \mu_{01}, \mu_{21}, \mu_{12}, \rho_1, \rho_2, \xi_1, \xi_2, \gamma)$ with d teams. Actually, this is a high-dimensional nonlinear optimization problem. The coordinate descent algorithm is a popular approach to pursue a solution of the problem.

The coordinate descent algorithm we used is listed as follows: denote

$$\eta = (\alpha_1, \ldots, \alpha_d, \beta_1, \ldots, \beta_d, \lambda_{10}, \lambda_{01}, \lambda_{21}, \lambda_{12}, \mu_{10}, \mu_{01}, \mu_{21}, \mu_{12}, \xi_1, \xi_2, \rho_1, \rho_2),$$
$$U_r = \frac{\partial l(\eta)}{\partial \eta_r}, \; V_r = \frac{\partial^2 l(\eta)}{\partial \eta_r^2}.$$

Step 1: Given the initial value of $\eta^{(0)}$, for example 1 for every element. Given the maximum number of iterations $n \in N$, first-order partial derivative tolerance $\tau > 0$, likelihood value tolerance $\tau_1 > 0$. Initialization the number of iterations $k = 0$.

Step 2: To calculate $U_r^{(k)}$, $r = 1, 2, \ldots, 2d + 13$, log-likelihood $L^{(k)}$ conditional on $\theta = \theta^{(k)}$.

Step 3: To calculate $r^* = \arg\max_r |U_r^{(k)}|$, $V_{r^*}^{(k)}$, let

$$\theta^{(k+1)} = \begin{cases} \theta_{r^*}^{(k+1)} = \theta_{r^*}^{(k+1)} - U_{r^*}^{(k)}/V_{r^*}^{(k)}, \\ \theta_r^{(k+1)} = \theta_r^{(k)}, \forall r \neq r^*. \end{cases}$$

Step 4: Conditional on $\theta = \theta^{(k+1)}$, calculating $U_r^{(k+1)}$, $r = 1, 2, \ldots, 2d + 13$, $L^{(k+1)}$ and $\Delta_1 = \max_r |U_r^{(k+1)}|$, $\Delta_2 = |L^{(k+1)} - L^{(k)}|$, updating $\Delta_2 = |L^{(k+1)} - L^{(k)}|$.

Step 5: If $k > n$ or $\Delta_1 < \tau$, or $\Delta_2 < \tau_1$, break, return $\theta^{(k)}$ as estimation; or else, go back to Step 3.

The crucial work of the algorithm is the calculation of the first order partial derivatives and the second-order partial derivatives. It can be done by differentiating the log-likelihood with respect to model parameters but is complicated. We do not report these results here but they are available upon request.

References

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

Constantinou, A. C., Fenton, N. E., Neil, M. (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems* 36:322–339.

Dixon, M. J., Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46(2):265–280.

Dixon, M., Robinson, M. (1998). A birth process model for association football matches. *Journal of the Royal Statistical Society: Series D (The Statistician)* 47(3):523–538.

Fahrmeir, L., Tutz, G. (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association* 89(428):1438–1449.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting* 21(2):331–340.

Hill, I. D. (1974). Association football and statistical inference. *Applied statistics* 23(2):213–227.

Hvattum, L. M., Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of forecasting* 26(3):460–470.

Joseph, A., Fenton, N.E., Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems* 19(7):544–553.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36(3):109–118.

McHale, I. G., Scarf, P. A. (2011). Modelling the dependence of goals scored by opposing teams in international soccer matches. *Statistical Modelling* 11(3):219–236.

Moroney, M. J. (1956). *Facts From Figures*. 3rd edition, reprinted 1969. United Kingdom:Penguin Books.

Reep, C., Pollard, R., Benjamin, B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society. Series A (General)* 134(4):623–629.

Rue, H., Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)* 49(3):399–418.

Volf, P. (2009). Random point process model for the score in sport matches. *Journal of Management Mathematics* 20:121–131.