IE Interfaces
Vol. 24, No. 4, pp. 323-329, December 2011.

마코비안 도착과정을 이용한 축구경기 득점결과의 예측

김남기[1] · **박현민**[2†]

[1]전남대학교 산업공학과 / [2]배재대학교 경영학과

Predicting the Score of a Soccer Match by Use of a Markovian Arrival Process

Namki Kim[1] · Hyunmin Park[2]

[1]Department of Industrial Engineering, Chonnam National University, 500-757, Korea
[2]Department of Business Administration, Pai Chai University, 302-735, Korea

We develop a stochastic model to predict the score of a soccer match. We describe the scoring process of the soccer match as a markovian arrival process (MAP). To do this, we define a two-state underlying Markov chain, in which the two states represent the offense and defense states of the two teams to play. Then, we derive the probability vector generating function of the final scores. Numerically inverting this generating function, we obtain the desired probability distribution of the scores. Sample numerical examples are given at the end to demonstrate how to utilize this result to predict the final score of the match.

Keyword: soccer, score, prediction, Markovian arrival process, probability model

1. 서 론

최근 여러 프로 스포츠 종목의 높은 인기와 더불어 올림픽, 월드컵 등의 스포츠 이벤트에 대한 높은 관심과 응원이 꾸준히 이어지고 있다. 이런 가운데 경기 참가 선수와 코칭스태프는 기록 향상 또는 경기 승리를 위해 과학적이고 체계적인 경기 분석의 도구를 적극 활용하고 있다. 선수 개개인의 능력 향상을 위한 과학적인 훈련 기법의 모색은 물론, 팀 스포츠에 있어 상대팀과의 이전 경기 데이터를 통계 수치화하여 승리를 위한 최적의 전략을 강구하고 경기 결과를 예측하는 다양한 과학적 분석 방법이 제시되고 있다. 이러한 스포츠과학 분야에 체계적 분석 수단으로써 경영과학의 다양한 모델링 기법과 분석 이론의 역할은 더욱 커질 것이다. 최고의 인기 종목인 축구 경기에 있어서도 이미 상대 전적의 통계 데이터를 활용한 최적의 경기 운영 전략 수립, 경기 결과 시뮬레이션 등이 널리 활용되고 있다.

특히, 축구 경기 득점(score)에 대한 연구는 합리적인 경기 배팅 정보 제공을 위해 득점 분포에 대한 다양한 통계 분석과 실증 데이터의 검증으로 활발히 전개되어 왔다. Dixon and Coles (1997)의 연구는 특정 팀 간의 상대전적 데이터를 이용한 Poisson regression model을 이용해 경기 결과와 스코어의 예측을 시도하였다. 곧이어 Dixon and Robinson(1998)은 home advantage, 현재 스코어, 잔여 경기 시간의 요소가 득점 분포에 미치는 영향을 추가 고려하여 경기 시간대별 스코어를 예측하는 non-homogeneous model을 제시하였다. 또한, Dyte and Clarke(2000)는 1998년 프랑스 월드컵 경기 결과를 실증 분석하여 FIFA 랭킹과 경기 장소가 득점 결과에 미치는 영향을 고려한 득점 예측의 Poisson model을 제시하였다.

이러한 이전의 연구들은 득점 상황에 영향을 미치는 요인들을 도출하여 이전 경기 결과의 통계 데이터를 이용한 스코어 예

†연락저자 : 박현민 교수, 302-735 대전광역시 서구 배재로 155-40 배재대학교 경영학과,
Fax : 042-520-5345, E-mail : hmpark12@pcu.ac.kr
투고일(2011년 08월 31일), 심사일(1차 : 2011년 10월 10일, 2차 : 2011년 10월 21일), 게재확정일(2011년 10월 27일).

측의 거시적 회귀분석 모델 개발에 치우친 반면, 본 연구는 실제 축구 경기에 있어 득점 과정(scoring process)을 마코비안 도착과정(Markovian arrival process; MAP)으로 모델링하여 공격과 수비가 반복되는 경기진행 과정 중 득점이 이루어지는 상황을 세밀하게 분석한다는 점에서 차이점을 갖는다. 마코비안 도착과정은 Neuts(1981)에 의해 개발된 Versatile Arrival Process를 Lucantoni (1991)가 확장한 것으로 상태에 따라 특정 현상의 발생률이 달라지는 확률적 현상의 모델링에 적합하다. 특히, 통신 공학 분야에서 높은 상관성(correlation)을 갖는 패킷의 도착 과정을 분석하는데 활발히 적용되고 있다(Lucantoni, 1993; Srinivas, 2001).

그런데, 공격권을 누가 갖고 있는가 하는 상태 정보를 바탕으로 공격 중인 상태에서 득점 성공 또는 득점 실패의 발생률을 분석하는 마코비안 도착과정 모형에 의해 한 경기 동안 양 팀의 득점분포를 보다 체계적으로 유도할 수 있다. 이를 통해, 경기 예측의 정확성을 높일 수 있고, 볼 점유율, 골 결정력의 주요 경기 지표의 민감도 분석을 통해 적절한 경기 운영 전략을 세울 수 있다. 또한, 본 연구는 마코비안 도착과정의 확률모형 분석 이론을 흥미로운 스포츠과학 분야에 새로이 적용한다는데 의의가 있다. 향후, 본 연구 결과의 활용을 통해 A매치를 비롯한 중요 경기 전략의 효율적 수립, 양질의 축구 배팅 정보의 제시 등 실질적 효과도 기대할 수 있다.

본 논문은 다음과 같이 구성된다. 제 2장에서는 마코비안 도착과정을 이용한 축구 경기 득점 과정의 모델링에 대해 설명하고, 제 3장에서는 득점 확률 분포의 유도 과정을 제시한다. 제 4장에서는 수치 예제를 통해 예상 스코어의 예측 방법을 설명하고, 제 5장에서 본 논문을 요약, 정리한다.

2. 마코비안 도착과정을 이용한 축구경기의 모델링

2.1 모형의 가정

A, B 두 팀이 펼치는 축구경기를 스케치 해 보자. 먼저 두 팀 중 한 팀(편의상 A팀)이 선축하여 공격을 펼치고, 다른 한 팀(편의상 B팀)이 수비를 한다. 공격을 펼치던 A팀이 실수하여 공격권을 B팀에게 넘긴다. 그러면 이제 B팀이 공격하게 되고 A팀이 수비를 한다. B팀의 공격이 득점으로 이어졌다. 그러면 공격권은 다시 A팀으로 주어지고 B팀은 수비를 한다. 이렇게 A, B 두 팀은 공격과 수비를 번갈아 가며 펼치다가 45분이 지나면 전반전 경기가 종료된다. 후반전에서는 B팀의 선축으로 시작하여 동일한 과정이 반복되고, 후반전 45분이 끝나면 그 때까지의 득점으로 승부를 가른다. 동점일 경우, 때론 연장전 전・후반 15분을 더 진행하여 승부를 가르기도 한다.

본 논문에서는 위와 같은 축구경기의 득점과정을 다음과 같이 모델링한다. 어느 한 팀이 공격하다가 상대팀에게 공격권을 넘겨주는 경우는, 득점과 함께 공격권을 넘겨주는 경우와 득점에 실패하고 공격권을 넘겨주는 경우 두 가지 경우이다. 각각의 경우에 대해 다음을 가정한다. 팀 $j(j=A,B)$가 공격을 시작하여 득점을 올리고 공격권을 넘겨주기까지 걸리는 시간이 서로 독립이고 동일한(i.i.d., independent and identically distributed) 지수분포 $\exp(\lambda_j)$를 따르고, 이 시간은 다른 모든 사건과 독립이라고 가정한다. 또한, 팀j가 공격을 시작하여 득점을 올리지 못하고 공격권을 넘겨주기까지 걸리는 시간이 서로 독립이고 동일한 지수분포 $\exp(\mu_j)$를 따르고, 이 시간 또한 다른 모든 사건과 독립이라고 가정한다. 이러한 가정은 마치, 공격중인 팀 j가 매 시점마다 $\lambda_j dt$의 확률로 득점을 올리거나, $\mu_j dt$의 확률로 득점 없이 공격권을 상대팀에게 넘겨주거나, $1-\lambda_j dt-\mu_j dt$의 확률로 공격을 계속 진행한다고 해석될 수 있다.

2.2 마코비안 도착과정을 사용한 모델링

임의시점에서 어느 팀이 공격중인가를 축구경기 진행과정의 상태(phase)로 정의하자. 상태 $j=A$이면 현재 A팀이 공격중(B팀 수비 중) 이라는 의미이며, 상태 $j=B$이면 그 반대를 의미한다. 분석의 편의를 위해 다른 상태는 고려하지 않는다(실제 축구경기에서는, 전・후반 축구 경기 시간 중에서, 반칙, 선수의 부상, 공이 필드 바깥으로 나감 등으로 인한 경기 일시정지(ball dead)시간들이 있다. 이 시간들은 전・후반 45분이 끝나면 추가시간(injury time)으로 (일부) 보충된다. 본 논문에서는 이러한 경기 일시정지시간들은 고려하지 않고 득점이 일어날 수 있는 인플레이(inplay)시간만을 분석 대상으로 삼는다. 그러면, 축구 경기의 진행 과정을 위와 같이 2개의 상태(phase)로만 정의할 수 있다).

이제, 득점 없이 상대팀에게 공격권을 넘기는 상황을 표현하는 상태전이율행렬 $\boldsymbol{C}$와 득점과 함께 상대팀에게 공격권을 넘기는 상황을 표현하는 상태전이율행렬 $\boldsymbol{D}$를 다음과 같이 정의하자.

$$\boldsymbol{C}=\begin{pmatrix}-\lambda_A-\mu_A & \mu_A \\ \mu_B & -\lambda_B-\mu_B\end{pmatrix},$$
$$\boldsymbol{D}=\begin{pmatrix}0 & \lambda_A \\ \lambda_B & 0\end{pmatrix}.$$

그러면, $\boldsymbol{C}+\boldsymbol{D}$는 연속시간 마코프과정(Continuous-Time Markov Chain)의 상태전이율행렬(transition rate matrix)이 된다.

$$\boldsymbol{C}+\boldsymbol{D}=\begin{pmatrix}-\lambda_A-\mu_A & \lambda_A+\mu_A \\ \lambda_B+\mu_B & -\lambda_B-\mu_B\end{pmatrix}.$$

더구나 상태전이율행렬 $\boldsymbol{C}$와 $\boldsymbol{D}$는 마코비안 도착과정(MAP)을 표현하는 모수행렬(parameter matrix) $(\boldsymbol{C},\boldsymbol{D})$에 해당한다. 즉, 축구경기에서 득점이 일어나는 과정이, 마치 두 개의 상태 $(j=A,B)$를 갖는 마코비안 도착과정에서 고객들이 도착하는 과정으로 표현된다는 것이다. 그러면, 전, 후반 90분 동안 A팀

(혹은 B팀)의 총 득점의 확률분포는, 마코비안 도착과정에서 90분 동안, 상태 A(혹은 상태 B)에서 일어난 고객도착의 총수의 확률분포에 해당한다. 이와 같이, 제 2.1절의 가정 하에서, 축구경기의 득점과정이 모수행렬$(\boldsymbol{C}, \boldsymbol{D})$를 갖는 마코비안 도착과정으로 모델링된다.

3. 축구경기 득점확률분포의 분석

3.1 상태확률의 정의

경기시작 후 $t \ge 0$ 시점에서 경기상태를 다음의 세변수로 정의한다.

$J(t)$: t시점에서의 경기 상태($J(t)=A, B$: A 또는 B 팀이 공격 중)
$N_A(t)$: t시점까지 A팀의 득점($N_A(t) = 0, 1, 2, \cdots$)
$N_B(t)$: t시점까지 B팀의 득점($N_B(t) = 0, 1, 2, \cdots$)

그러면, $t \ge 0$ 시점의 경기상태는 $\{N_A(t), N_B(t), J(t)\}$의 3차원 vector로 나타낼 수 있으며, 경기 상황에 대한 상태 확률을 다음과 같이 정의할 수 있다.

$$P_j(n_A, n_B; t) = \Pr\{N_A(t)=n_A, N_B(t)=n_B, J(t)=j\},$$
$$n_A, n_B = 0, 1, 2, \cdots, \ j=A, B.$$

위의 상태 확률을 2차원 vector로 정의하면 다음과 같다.

$$\boldsymbol{P}(n_A, n_B; t) = \{P_A(n_A, n_B; t), P_B(n_A, n_B; t)\}.$$

추가로, 경기 시작 직전, 동전 던지기를 통해 먼저 선축하는 팀을 편의상 A팀이라고 하면, $t=0$일 때의 초기 조건은

$$\Pr\{N_A(0)=0, N_B(0)=0, J(0)=A\} = 1. \tag{1}$$

이다.

3.2 득점확률분포의 확률생성함수

경기상태 확률과정$\{N_A(t), N_B(t), J(t)\}$가 연속시간 마코프과정을 따르므로, 다음과 같은 벡터미분방정식(vector differential equation)을 얻을 수 있다.

$$\frac{d}{dt}\boldsymbol{P}(n_A, n_B; t) = \boldsymbol{P}(n_A, n_B; t)\boldsymbol{C} + \boldsymbol{P}(n_A-1, n_B; t)\boldsymbol{D_A} + \boldsymbol{P}(n_A, n_B-1; t)\boldsymbol{D_B}, n_A, n_B = 0, 1, \cdots. \tag{2}$$

여기서 $\boldsymbol{D_A} = \begin{pmatrix} 0 & \lambda_A \\ 0 & 0 \end{pmatrix}$, $\boldsymbol{D_B} = \begin{pmatrix} 0 & 0 \\ \lambda_B & 0 \end{pmatrix}$이고, $\boldsymbol{D} = \boldsymbol{D_A} + \boldsymbol{D_B}$ 이다.

식 (2)는 연속시간 마코프과정에서 잘 알려진 콜모고로프 전진방정식(forward Kolmogorov equation)에 해당된다(Lee, 2006).

$t \ge 0$시점에서의 득점분포의 확률벡터생성함수(probability vector generating function)를 다음과 같이 정의하자.

$$\boldsymbol{P}^*(z_A, z_B; t) = \sum_{n_A=0}^{\infty}\sum_{n_B=0}^{\infty} \boldsymbol{P}(n_A, n_B; t)\, z_A^{n_A} z_B^{n_B}.$$

식 (2)의 양변에 $z_A^{n_A} z_B^{n_B}$를 곱하고 모든 n_A, n_B에 대하여 더하면 다음을 얻는다.

$$\frac{d}{dt}\boldsymbol{P}^*(z_A, z_B; t) = \boldsymbol{P}^*(z_A, z_B; t)\boldsymbol{C} + z_A\boldsymbol{P}(z_A, z_B; t)\boldsymbol{D_A} + z_B\boldsymbol{P}(z_A, z_B; t)\boldsymbol{D_B} \tag{3}$$

식 (3)의 해를 다음과 같이 행렬지수(matrix exponential)형태로 구할 수 있다.

$$\boldsymbol{P}^*(z_A, z_B; t) = \boldsymbol{P}^*(z_A, z_B; 0)e^{(\boldsymbol{C} + z_A\boldsymbol{D_A} + z_B\boldsymbol{D_B})t} = (1, 0)\, e^{(\boldsymbol{C} + z_A\boldsymbol{D_A} + z_B\boldsymbol{D_B})t}. \tag{4}$$

식 (4)에서, $\boldsymbol{P}^*(z_A, z_B; 0) = (1, 0)$은 초기조건 식 (1)를 사용하였다. 식 (4)는 마코비안 도착과정에서, $t \ge 0$시점까지의 고객도착수를 나타내는 계수함수(counting function)와 동일한 형태이다(Lee, 2006).

식 (4)로부터, 전반전($\bar{t} = 45$분) 종료 후의 득점분포의 확률생성함수(generating function)는 다음과 같다.

$$(1, 0)\, e^{(\boldsymbol{C} + z_A\boldsymbol{D_A} + z_B\boldsymbol{D_B})\bar{t}}\boldsymbol{e}.$$

여기서 $\boldsymbol{e} = (1, 1)$이다. 후반전에는 선축이 바뀜에 따라 후반 $\bar{t} = 45$분 동안의 득점분포의 확률생성함수는 다음과 같다.

$$(0, 1)\, e^{(\boldsymbol{C} + z_A\boldsymbol{D_A} + z_B\boldsymbol{D_B})\bar{t}}\boldsymbol{e}.$$

최종 득점분포는 모형의 가정에 의해 전반 득점과 후반 득점의 독립합(convolution)이므로, 전・후반 경기 종료 후 최종 득점분포의 확률생성함수는 다음과 같다.

$$\left\{(1, 0)\, e^{(\boldsymbol{C} + z_A\boldsymbol{D_A} + z_B\boldsymbol{D_B})\bar{t}}\boldsymbol{e}\right\} \cdot \left\{(0, 1)\, e^{(\boldsymbol{C} + z_A\boldsymbol{D_A} + z_B\boldsymbol{D_B})\bar{t}}\boldsymbol{e}\right\}. \tag{5}$$

식 (5)을 Mathematica, Maple과 같은 소프트웨어 패키지를 이용하여 역변환하면 양팀의 최종득점에 대한 확률분포를 수치적으로 얻을 수 있다. 이는 다음 장에서 수치예와 함께 보여진다.

특별히, 전・후반 경기 결과, 득점 없이 비길 확률은 식 (5)에

$z_A = z_B = 0$을 대입하여 다음과 같이 얻는다.

$$\left\{(1,0)\, e^{C\bar{t}} \boldsymbol{e}\right\} \cdot \left\{(0,1)\, e^{C\bar{t}} \boldsymbol{e}\right\}.$$

3.3 연장전 골든골 제도에서의 승부확률

골든골(golden goal)은 전후반 90분 경기에서 승부가 가려지지 않을 경우, 연장전에서 먼저 득점하는 팀이 승리를 하고 경기를 끝내는 방식으로 sudden death라고도 한다. 1998년 프랑스 월드컵에서부터 사용되었으나, 2004년 유럽 축구선수권(Euro 2004)을 끝으로 국제축구연맹(FIFA)은 골든골 제도를 폐지하였다. 이 절에서는 연장 전 · 후반 각 $\bar{t} = 15$동안 누가 먼저 득점하여 경기를 승리로 이끄는지의 확률을 유도해본다.

먼저, $\bar{t} = 15$분 이내에 골든골과 함께 경기가 종료되는 상태전이확률행렬은 다음과 같다.

$$\int_0^{\bar{t}} e^{Ct} \boldsymbol{D}\, dt = \left(\boldsymbol{I} - e^{C\bar{t}}\right)\left(-\boldsymbol{C}^{-1}\right)\boldsymbol{D}. \tag{6}$$

여기서 e^{Ct}는 시간 t동안 득점이 일어나지 않고 상태전이만 일어나는 상태전이확률을 나타내며, $e^{Ct}\boldsymbol{D}\,dt$는 시간 t동안 득점이 일어나지 않고 상태전이만 일어나다가 시점$(t,\, t+dt)$에 득점이 일어나는 상태전이확률행렬을 나타낸다. 식 (6)은 마코비안 도착과정에서, 도착간격분포행렬과 동일한 형태이다(Lee, 2006).

A팀이 골든골을 넣고 승리하는 확률을, 연장전반에서 A팀이 골든골을 넣는 경우와 연장전반에는 득점을 하지 못했지만 후반에서 골든골을 넣은 경우로 나누어 유도하면 아래와 같다 (B팀이 골든골을 넣고 승리하는 경우는 아래 식들의 $\boldsymbol{D_A}$대신 $\boldsymbol{D_B}$를 쓰면 된다).

(i) 연장 전반에 A팀이 골든골을 넣고 승리하는 경우
연장 전반에 A팀이 골든골을 넣고 승리할 확률은 다음과 같다.

$$\left(\frac{1}{2}, \frac{1}{2}\right)\int_0^{\bar{t}} e^{Ct} \boldsymbol{D_A}\, dt\ \boldsymbol{e} = \left(\frac{1}{2}, \frac{1}{2}\right)\left(\boldsymbol{I} - e^{C\bar{t}}\right)\left(-\boldsymbol{C}^{-1}\right)\boldsymbol{D_A}\, \boldsymbol{e}.$$

여기서 초기상태확률은 $\left(\frac{1}{2}, \frac{1}{2}\right)$이다. 이는 두 팀 중 어느 팀이 연장전반에서 선축할지가 랜덤하게 결정되기 때문이다.

(ii) 연장 후반에서 A팀이 골든골을 넣고 승리하는 경우
연장후반에 A팀이 골든골을 넣고 승리하는 경우는, A팀이 연장전반을 선축하는 경우와, A팀이 연장후반을 선축하는 경우로 나누어 계산한다. 먼저 A팀이 연장전반을 선축하는 경우에, 연장전반에 두 팀 모두 득점하지 못할 확률은 $(1,0)\, e^{C\bar{t}} \boldsymbol{e}$이고, 연장후반에 A팀이 골든골을 넣을 확률은 $(0,1)\int_0^{\bar{t}} e^{Ct}\boldsymbol{D_A}\, dt\ \boldsymbol{e}$이다. 마찬가지로, A팀이 연장후반을 선축하는 경우에, 연장전반에 두 팀 모두 득점하지 못할 확률은 $(0,1)\, e^{C\bar{t}} \boldsymbol{e}$ 이고, 연장후반에 A팀이 골든골을 넣을 확률은 $(1,0)\int_0^{\bar{t}} e^{Ct}\boldsymbol{D_A}\, dt\ \boldsymbol{e}$이다. 이를 종합하면, 연장후반에 A팀이 골든골을 넣고 승리할 확률은 다음과 같다.

$$\begin{aligned}
&\frac{1}{2}\times\left\{(1,0)\, e^{C\bar{t}}\boldsymbol{e}\right\}\left\{(0,1)\int_0^{\bar{t}} e^{Ct}\boldsymbol{D_A}\, dt\ \boldsymbol{e}\right\} \\
&+\frac{1}{2}\times\left\{(0,1)\, e^{C\bar{t}}\boldsymbol{e}\right\}\left\{(1,0)\int_0^{\bar{t}} e^{Ct}\boldsymbol{D_A}\, dt\ \boldsymbol{e}\right\} \\
&=\frac{1}{2}\times\left\{(1,0)\, e^{C\bar{t}}\boldsymbol{e}\right\}\left\{(0,1)\left(\boldsymbol{I}-e^{C\bar{t}}\right)\left(-\boldsymbol{C}^{-1}\right)\boldsymbol{D_A}\,\boldsymbol{e}\right\} \\
&+\frac{1}{2}\times\left\{(0,1)\, e^{C\bar{t}}\boldsymbol{e}\right\}\left\{(1,0)\left(\boldsymbol{I}-e^{C\bar{t}}\right)\left(-\boldsymbol{C}^{-1}\right)\boldsymbol{D_A}\,\boldsymbol{e}\right\}.
\end{aligned}$$

마지막으로 연장전후반 각 $\bar{t} = 15$동안 두 팀 모두 득점을 내지 못하고, 끝내 필드 플레이(field play)로 승부를 가르지 못할 확률은 다음과 같다.

$$\left\{(1,0)\, e^{C\bar{t}} \boldsymbol{e}\right\} \cdot \left\{(0,1)\, e^{C\bar{t}} \boldsymbol{e}\right\}.$$

4. 모수추정과 수치예

4.1 모수의 추정

축구경기를 모델링하고 득점결과를 예측하는데 필요한 모수는 $\lambda_j\ \mu_j,\ j = A, B$이다. 이 절에서는 어떻게 이들 모수들을 추정할 수 있는지를 논의한다.

팀 $j = A, B$가 공격을 시작하여 상대팀에게 공격권을 넘겨주는 경우는, 득점과 함께 공격권을 넘겨주는 경우와, 득점에 실패하고 공격권을 넘겨주는 경우 두 가지 경우이다. 모형의 가정에 의해 각각에 소요 되는 시간은 각각 서로 독립이고 동일한, 지수분포 $\exp(\lambda_j)$와 $\exp(\mu_j)$를 따른다. 따라서, 다음을 얻는다.

팀 j가 공격권을 갖은 시간부터 공격권을 상대팀에게 넘겨주기까지의 시간의 평균

$$= \frac{1}{\lambda_j + \mu_j}\ (j = A, B), \tag{7}$$

팀 j의 공격이 득점으로 마무리 되어 공격권을 넘겨주는 경

우의 확률

$$= \frac{\lambda_j}{\lambda_j + \mu_j} \ (j = A, B). \tag{8}$$

두 팀 A, B의 역대 경기결과 자료로부터, 각 팀이 (i) 한번 볼을 소유하면, 상대팀에게 공격권을 넘겨주기 전까지 볼을 소유하는 평균시간과, 각 팀이 (ii) 한번 볼을 소유하면, (득점실패가 아닌) 득점으로 공격권을 넘겨주는 경우의 상대빈도를 각각 실측하여 식 (7)과 식 (8)의 우변에 각각 대입하고 연립하여 풀면, 득점분포 예측에 필요한 모든 모수 $\lambda_j\ \mu_j,\ j = A, B$를 구할 수 있다.

참고로 쉽게 구할 수 있는 축구 경기자료 중의 하나인, A팀 대 B팀의 볼 점유율은

$$\frac{(\lambda_B + \mu_B)}{(\lambda_A + \mu_A) + (\lambda_B + \mu_B)} : \frac{(\lambda_A + \mu_A)}{(\lambda_A + \mu_A) + (\lambda_B + \mu_B)}$$

로 표현된다.

4.2 수치 예제

두 팀 A, B의 역대 경기결과 자료로부터 <표 1>을 얻었다고 가정하자. 축구 및 스포츠 전문 웹사이트인 goal.com 또는 sports data bank 등에서는 과거 주요 축구경기에 대한 주요통계수치(점유율, 득점 등)을 공개하지만 안타깝게도 식 (7)과 식 (8)의 자료와 같은 세부 자료는 공개적으로 제공하지 않는다. 여기서는 <표 1>과 같은 자료를 얻었다고 가정하자. 즉, 팀 A가 한번 공을 소유하면, 볼 소유권을 넘기기 전까지 평균 30초를 소유한다고 하자. 또한, 한번 공격을 펼치면 100번의 공격 중 평균 한 번의 비율로 득점에 성공한다고 하자(이를 편의상 골 결정력이라고 부르자). <표 1>에 따르면 B팀이 볼을 점유한 후, 소유하는 시간은 A팀에 비해 작고, 골 결정력은 A팀과 동일하다.

표 1. A, B 간의 축구 경기에 대한 통계 예

1회 공격시 평균 공격시간	A팀	30초(1/2분)
	B팀	20초(1/3분)
1회 공격시 득점 확률	A팀	1/100
	B팀	1/100

먼저 <표 1>의 자료를 식 (7)과 식 (8)의 우변에 놓고 연립하여 풀어서 λ_i, $\mu_i (i = A, B)$를 구한다. 이것을 식 (5)의 득점분포의 확률생성함수에 대입한 후, 이를 역변환하면 두 팀의 최종 득점분포를 수치적으로 구할 수 있다. Mathematica를 사용하여 역 변환한 결과는 다음 <표 2>와 같다.

<표 2>의 확률분포로부터 A팀과 B팀의 축구경기에 대한 득점확률을 제시할 수 있다. <표 2>에서 대각선에 위치한 셀의 값을 모두 합산하면 두 팀이 비길 확률 0.296을 구할 수 있고, 대각선 위 또는 아래의 확률 값을 합하면, A팀과 B팀이 승리할 확률이 각각 0.351, 0.353임을 구할 수 있다. 또한, <표 2>의 결과에서, 예상 스코어로써 1 : 1이 될 확률이 0.1351로 가장 크다는 것을 알 수 있다(이 결과에 따르면, 스포츠토토 혹은 다른 방법으로 내기에 참여하는 경우, 1 : 1에 내기를 걸어야 할 것이다). 마지막으로, A팀이 B팀을 2점차 이상으로 이길 확률은 0.1384이다(이러한 확률값은 2점차 이상으로 이겨야만 16강에 진출할 수 있는 상황이 벌어졌을 때 필요한 확률이다).

만일 훈련 A팀이 훈련을 통해 또는 유능한 스트라이커의 영입을 통해 골 결정력을 이전보다 10% 향상하였다고 가정하자. 즉, <표 1>에서 A팀의 1회 공격 시 득점확률이 1/100에서 11/1000로 커졌다. 이런 경우, 다시 λ_i, μ_i를 계산하여, 식 (5)에 대입한 후, 역변환하면 <표 3>과 같은 결과를 얻을 수 있다.

<표 3>으로부터 A팀이 골 결정력을 10% 향상시킴으로 인

표 2. 득점분포 : A, B팀의 골 결정력이 동일한 경우

B팀 \ A팀	0	1	2	3	4	5	6	7	total
0	0.1153	0.1244	0.0668	0.0238	0.0063	0.0013	0.0002	0.0000	0.3381
1	0.1246	**0.1351**	0.0729	0.0261	0.0070	0.0015	0.0003	0.0000	0.3675
2	0.0671	0.0730	0.0396	0.0142	0.0038	0.0008	0.0001	0.0000	0.1986
3	0.0239	0.0262	0.0143	0.0052	0.0014	0.0003	0.0001	0.0000	0.0714
4	0.0064	0.0070	0.0038	0.0014	0.0004	0.0001	0.0000	0.0000	0.0191
5	0.0014	0.0015	0.0008	0.0003	0.0001	0.0000	0.0000	0.0000	0.0041
6	0.0002	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0007
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
total	0.3389	0.3675	0.1983	0.0711	0.0190	0.0040	0.0007	0.0000	

표 3. 득점분포 : A팀 골 결정력이 10% 상승하였을 경우

B팀 \ A팀	0	1	2	3	4	5	6	7	total
0	0.1035	0.1229	0.0726	0.0285	0.0083	0.0019	0.0004	0.0001	0.3382
1	0.1118	**0.1334**	0.0792	0.0312	0.0092	0.0022	0.0004	0.0001	0.3675
2	0.0601	0.0721	0.0430	0.0170	0.0050	0.0012	0.0002	0.0000	0.1986
3	0.0215	0.0258	0.0155	0.0062	0.0018	0.0004	0.0001	0.0000	0.0713
4	0.0057	0.0069	0.0042	0.0017	0.0005	0.0001	0.0000	0.0000	0.0191
5	0.0012	0.0015	0.0009	0.0004	0.0001	0.0000	0.0000	0.0000	0.0041
6	0.0002	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0008
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
total	0.3040	0.3629	0.2156	0.0851	0.0249	0.0058	0.0011	0.0002	

표 4. 득점분포 : B팀 골 결정력이 10% 하락하였을 경우

B팀 \ A팀	0	1	2	3	4	5	6	7	total
0	0.1285	**0.1387**	0.0745	0.0266	0.0071	0.0015	0.0003	0.0000	0.3772
1	0.1249	0.1355	0.0731	0.0262	0.0070	0.0015	0.0003	0.0000	0.3685
2	0.0605	0.0659	0.0357	0.0129	0.0035	0.0007	0.0001	0.0000	0.1793
3	0.0194	0.0213	0.0116	0.0042	0.0011	0.0002	0.0000	0.0000	0.0578
4	0.0047	0.0051	0.0028	0.0010	0.0003	0.0001	0.0000	0.0000	0.0140
5	0.0009	0.0010	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0027
6	0.0001	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0004
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
total	0.3390	0.3677	0.1983	0.0711	0.0191	0.0040	0.0007	0.0000	

해 A팀이 승리할 확률은 0.351에서 0.383으로 높아진 것을 알 수 있다. B팀의 승리 확률은 0.353에서 0.330으로 떨어졌으며, 또한 무승부가 될 확률도 0.296에서 0.287로 떨어졌다. 가장 가능성이 높은 득점은 여전히 1 : 1이지만, 그 확률은 0.1351에서 0.1334로 줄어들었다. A팀이 2점차 이상으로 B팀을 이길 확률은 0.1384에서 0.1618로 늘었다.

반대로, B팀의 스트라이커의 부상, 경기력의 감퇴로 인해 골 결정력이 10% 감소한 경우를 가정하자. 이러한 경우에 대한 득점분포의 결과는 <표 4>와 같다.

이 경우, A팀이 승리할 확률이 2.4% 상승하여 0.375이다. 또한 가장 발생 가능성이 높은 예상 스코어는 1:0 이 됨을 알 수 있다.

5. 결 론

본 연구에서는 마코비안 도착과정의 분석 방법을 이용하여 체계적인 축구 득점 분포의 유도 과정을 제시하였다. 실제 경기 결과의 예측에 참고할 수 있도록 수치 예제도 함께 제공하였다. 득점이 일어나는 과정을 비교적 세밀하게 묘사한 모델링 분석을 통해 경기 결과 예측의 정확성을 높일 수 있고, 효율적인 경기 운영 전략을 수립하는데 도움이 될 것으로 기대된다. 또한, 마코비안 도착과정(MAP) 이론의 흥미로운 응용 분야를 발굴하여 스포츠과학 분야에서 경영 과학 응용의 좋은 사례로 남을 수 있을 것이다.

추후 연구에서는 수년간 누적된 실제 경기 데이터를 확보하여 제시한 확률 모형의 신뢰성을 검증해보고, 실제 경기 결과 예측 과정을 정교화 할 수 있다. 또한, 농구, 럭비와 같이 한 골당 다 득점인 구기 스포츠 종목에서의 득점과정을 집단도착 마코비안 도착과정(batch Markovian arrival process : BMAP)으로 모델링하여 득점 분포를 분석해 볼 수 있다.

Acknowledgement

축구를 확률적으로 모델링하는 것에 대한 영감을 주신 KAIST 산업 및 시스템공학과의 채경철 교수님과 Markovian Arrival Process(MAP)를 멀리 KAIST까지 오셔서 강의해 주신 성균관대학교 시스템경영공학과의 이호우 교수님께 특별한 감사를 드립니다.

참고문헌

Dixon, M. J. and 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. J. and Robinson, M. E. (1998), A Birth Process Model for Association Football Matches, *Journal of the Royal Statistical Society. Series D(The Statistician)*, 47(3), 523-538.

Dyte, D. and Clarke, S. R. (2000), A Rating Based Poisson Model for World Cup Soccer Simulation, *The journal of the Operational Research Society*, 51(8), 993-998.

Lee, H. W. (2006), Queueing Theory, Sigma Press, Seoul, Korea.

Lucantoni, D. M. (1991), New Results on the Single Sever Queue with BMAP, *Stochastic Models*, 7(10), 1-46.

Lucantoni, D. M. (1993), The BMAP/G/1 queue : A Tutorial, *Lecture Notes in Computer Science : Performance Evaluation of Computer and Communications Systems*, 729, 330-358.

Neuts, M. F.(1981), Matrix Geometric Solutions in Stochastic Models, The Johns hopkins University Press, Baltimore.

Srinivas R. C. (2001), The Batch Markovian Arrival Process : A Review and Future Work. In *Advances in Probability and Stochastic Processes*, Notable Publications, Inc., New Jersey, USA, 21-49.

김 남 기
한국과학기술원 경영과학과 학사
한국과학기술원 산업경영학과 석사
한국과학기술원 산업공학과 박사
현재 : 전남대학교 산업공학과 부교수
관심분야 : 확률모형, 대기행렬이론

박 현 민
연세대학교 경영학과 학사
한국과학기술원 산업공학과 석사
한국과학기술원 산업 및 시스템공학과 박사
현재 : 배재대학교 경영학과 전임강사
관심분야 : 확률모형, 대기행렬이론, 품질관리