The Korean Journal of Applied Statistics (2016) DOI: http://dx.doi.org/10.5351/KJAS.2016.29.2.321
29(2), 321–330

Run expectancy and win expectancy in the Korea Baseball Organization (KBO) League

Hyung Woo Moon[a] · Yong Tae Woo[b] · Yang Woo Shin[c,1]

[a]Institute of Industrial Technology Research Center, Changwon National University;
[b]Department of Computer Engineering, Changwon National University;
[c]Department of Statistics, Changwon National University

(Received November 4, 2015; Revised November 29, 2015; Accepted January 25, 2016)

Abstract

Run expectancy (RE) is the mean number of runs scored from a specific base runner/outs situation of an inning to the end of the inning. Win expectancy (WE) is the probability that a particular team will win the game at a specific game state such as half-inning, score difference, outs, and/or runners on base. In this paper, we derive RE and WE for the Korea Baseball Organization (KBO) League based on six-year data from 2007 to 2012 using a Markov chain model.

Keywords: baseball, run expectancy, win expectancy, run values, win values, Markov chain

1. 서론

야구경기에서 감독들은 승리확률을 높이기 위해 다양한 작전을 구사한다. 야구감독들은 개인적인 성향이나 과거의 경험 그리고 현재의 경기 상황에 따라 동일한 상황에서도 상반된 작전을 지시하기도 한다. 예를 들어 동일한 무사 주자 1루 상황에서 타율이 저조한 타자에게는 희생번트를 요구할 수도 있고, 타율이 좋은 타자에게는 치고달리기 작전을 지시할 수 있다. 또한 동일한 타자에 대해서도 이닝이나 점수차에 따라 감독이 요구하는 작전은 달라질 수 있다. 따라서 야구경기에서 현재의 상황이나 선수들의 기량을 적절하게 반영하여 효과적인 작전을 구사하는 것은 승패를 좌우할 수 있는 매우 중요한 문제이다.

타자에 대한 경기력은 주로 타율, 타점, 출루율, 장타율 등과 같이 선수별로 타격과 관련된 기록을 통계적으로 분석한 지표에 의해 평가를 받는다. 이중에서 타율은 타석수에 대한 안타의 비율을 단순하게 계산하는 관계로 경기 상황에 따른 타자의 활약과 이로 인한 득점이나 승리에 대한 기여도를 효과적으로 반영하기 어렵다. 예를 들어 동일한 안타라하더라도 9회말 투아웃에 나온 역전 끝내기 안타와 많은 점수차로 이기고 있는 상황에서 나온 단순 안타가 해당 경기의 승패에 미친 기여도는 상당한 차이가 있다.

이와같이 감독이 작전을 구사하는 상황이나 타자의 타격 가치를 평가하는데 유용하게 사용될 수 있는 지표로서 미국프로야구에 대해서는 기대득점(run expectancy; RE)과 기대승리확률(win expectancy,

This research is financially supported by hibrain.net in 2014 and Changwon National University in 2015−2016.

[1]Corresponding author: Department of Statistics, Changwon National University, 20 Changwondaehak-ro Uichang-gu Changwon-si, Gyeongsangnam-do 51140, Korea. E-mail:ywshin@changwon.ac.kr

WE), 기대승리확률을 바탕으로 타격의 득점가치(run value; RV), 타격의 승리가치(win value; WV) 등이 제시되었다 (Tango 등, 2006). 기대득점은 각각의 아웃카운트와 주자 상황에서 그 이닝이 끝날 때까지 얻는 점수의 기댓값이다. 예를 들어 2007년부터 2012년 6년 동안 한국프로야구경기에서 1사 주자 2루인 상황은 12,903건이 있었으며 그 상황에서 이닝이 끝날 때까지 얻은 점수의 총합은 8779점이었다. 이 경우 1사 주자 2루인 상황에서의 기대득점은 8799/12903 = 0.680이다. 기대승리확률은 이닝, 점수차, 아웃카운트, 주자상태가 주어진 상태에서 경기를 계속 한다고 할 때, 공격하고 있는 팀이 승리할 확률이다. 예를 들어 6년(2007−2012) 동안 5회말 공격에서 1점 뒤지고 있고 1사에 주자 없는 상태는 총 306회 발생하였으며 이 중 115회를 홈팀이 승리하였다. 이 상황에서 홈팀의 기대승리확률은 115/306 = 0.376이다. 반면 원정팀은 5회초에 310번의 같은 상황에서 84번 승리하였다. 이 상황에서 원정팀의 기대승리확률은 84/310 = 0.271이다. 타격의 득점가치는 타격전 상황의 기대득점과 타격결과에 의하여 변화된 상황의 기대득점 사이의 차이를 말한다. 예를 들어 1사 주자 1, 2루의 상황에서 안타를 쳤을 때, 1점 득점하고 주자 상태가 1, 3루가 되었다고 하자. 1사 주자 1, 2루와 1사 주자 1, 3루의 상황에서 기대득점이 각각 0.972점과 1.342점이면 이 경우 안타의 득점가치는 1점(득점수) + 1.342 − 0.972 = 1.370이다. 타격의 승리가치는 타격 전후 상황의 기대승리확률간의 차이로서 타격 결과가 승리에 미치는 영향을 나타낸다.

이와 같은 지표들은 미국 메이저리그의 통계사이트인 팬그래프스(www.fangraphs.com)과 베이스볼프로스펙터스(www.baseballprospectus.com) 등에서도 제시되고 있으나 한국프로야구에 대해서는 이에 대한 결과가 매우 드문 실정이다. Jeong (2014)은 한국프로야구 데이터를 이용하여 총 발생횟수에 대한 특정 상황의 상대돗수로 기대승리확률을 제시하였다. 그러나 이 연구에서는 이해하기 어려운 특이한 결과가 나타난다. 예를 들어 7회초 무사 주자 1, 3루에서 1점, 2점, 3점 앞선 상황은 각각 6회, 10회, 3회가 발생하였다. 이 중에서 원정팀이 승리한 횟수는 각각 5회, 10회, 2회로 기대승리확률은 0.833, 1.000, 0.667이 되어, 3점 앞선 상황보다 2점 앞선 상황의 승리확률이 오히려 높아지는 경우도 발생하였다. 이와 같은 현상은 표본 수가 적은 경우에 나타나는 것으로 여겨진다. 이와 같은 문제점을 극복하기 위하여 본 논문에서는 마르코프연쇄를 이용하여 한국프로야구에서 기대득점(RE), 기대승리확률(WE), 득점가치(RV)와 승리가치(WV)를 구하는 방법을 제시한다.

마르코프연쇄는 야구경기를 수학적으로 모형화하고 분석하는 도구로서 여러 학자들에 의하여 매우 유용하게 사용되어 왔으나 RE나 WE를 다루지 않고 있다 (Bukiet 등, 1997; Hirotsu와 Wright, 2003, 2005; Moon 등, 2013; Sokol, 2003). Tango 등 (2006)은 마르코프연쇄를 이용하여 미국프로야구에서 RV와 WV를 구했다고 하였으나 구체적인 방법은 밝히지 않고 결과만을 제시하고있다.

본 논문에서 제안한 마르코프연쇄를 이용하여 타격의 RV와 WV를 구하는 방법은 다음과 같다. 먼저, 경기 데이터로부터 상태변화를 나타내는 진루확률과 타자들의 평균 타격확률을 구하고 Moon 등 (2013) 등이 제안한 마르코프연쇄 모형을 이용하여 평균 타격확률을 갖는 가상선수로 구성된 팀이 특정 상황에서 한 이닝동안 얻는 득점분포를 구한다. 이 득점분포로부터 RE, WE, RV, WV를 구한다.

이 논문의 구성은 다음과 같다. 제 2절에서는 마르코프연쇄를 이용하여 기대득점과 타격의 기대득점가치를 제 3절에서는 기대승리확률과 각 상황에서 타격의 기대승리가치를 구한다. 제 4절에서는 이 논문의 결론 및 향후과제에 대하여 기술한다.

2. 기대득점

2.1. 마르코프연쇄 모형

야구경기의 진행상황은 진행중인 이닝(초반, 후반)과 아웃된 타자수, 주자상태 뿐만 아니라 두 팀간의

점수차 등에 의하여 기술된다. 이를 위하여 한 이닝에서 n명의 타자가 타석을 마친 상태를 나타내는 다음 확률변수를 정의하자.

- K_n은 현재의 이닝에서 3아웃으로 그 이닝의 공격을 마치기 전 공격팀 선수 중 아웃된 타자수를 타나내며 상태공간은 $\mathcal{S}_K = \{0, 1, 2\}$이다.
- B_n은 루상에 있는 주자상태를 타나낸다. 루상에 주자가 없는 경우는 '0'으로 나타내고 주자가 1루와 3루에 있고 나머지 루는 비어 있는 상태를 '13'과 같이 나타내면 B_n의 상태공간은 $\mathcal{S}_B = \{0, 1, 2, 3, 12, 13, 23, 123\}$이 된다.
- H_n은 n번째 타자의 타격 결과를 나타내며 그 결과가 1루타, 2루타, 3루타, 홈런, 볼넷 또는 사구(몸에 맞는 볼), 아웃일 때 각각 $1, 2, 3, 4, 5, 6$의 값을 갖는다고 하자. H_n의 상태공간은 $\mathcal{S}_H = \{1, 2, 3, 4, 5, 6\}$이 된다.
- R_n은 n번째 타자의 타격결과로 얻는 점수를 나타내며 상태공간은 $\mathcal{S}_R = \{0, 1, 2, 3, 4\}$가 된다.
- $X_n = (K_n, B_n)$은 3아웃이 발생하기 전 (아웃된 타자수, 주자상태)를 나타낸다. 3아웃으로 그 이닝의 공격을 마친 상태를 $X_n = \Delta$로 나타내면 X_n의 상태공간은 $\mathcal{S}_X = \mathcal{S}_K \times \mathcal{S}_B \cup \{\Delta\}$이다.

야구경기를 모형화하고 구체적으로 기술하기 위해서는 n번째 타자가 타석을 마쳤을 때의 상태가 직전의 상태와 그 타자의 타격 결과에만 의존한다는 가정뿐만 아니라 각 상태의 변화를 나타내는 확률(전이확률)이 필요하다. 전이확률은 아웃된 타자수, 주자상태 뿐만 아니라 진행중인 이닝, 두 팀간의 점수차 등 여러 상황에 따라 달라질 수 있다. 그러나 자료로부터 전이확률을 계산할 때 각 상태가 발생한 횟수가 매우 적은 경우에는 의미없는 전이확률이 될 수 있다. 따라서 여기서는 아웃된 타자수와 주자상태만을 고려한 확률과정 $\boldsymbol{X} = \{X_n, n = 1, 2, \ldots\}$에 대한 전이확률행렬을 생각한다. 이때, 상태 Δ는 $\boldsymbol{X}$의 흡수상태다. $X_{n-1} = s$인 상태에서 n번째 타석에 선 타자의 타격결과가 $H_n = h$일 때 r득점하고 주자상태가 $X_n = s'$가 될 확률

$$P^{(r)}(h; s, s') = P(R_n = r, X_n = s' \,|\, H_n = h, X_{n-1} = s)$$

를 성분으로 갖는 25×25행렬 $P^{(r)}(h) = (P^{(r)}(h; s, s'))$는 다음과 같은 형태가 된다.

$$P^{(r)}(h) = \begin{pmatrix} A_0^{(r)}(h) & B_0^{(r)}(h) & C_0^{(r)}(h) & D_0^{(r)} \\ O & A_1^{(r)}(h) & B_1^{(r)}(h) & D_1^{(r)} \\ O & O & A_2^{(r)}(h) & D_2^{(r)} \\ O & O & O & 1 \end{pmatrix}.$$

단, $A_k^{(r)}(h)$, $B_k^{(r)}(h)$, $C_k^{(r)}(h)$는 크기가 8×8인 블록행렬이고 O은 영행렬이다. $P_H^{(r)}(h)$의 첫 번째 블록에 있는 행렬 $A_0^{(r)}(h)$, $B_0^{(r)}(h)$, $C_0^{(r)}(h)$는 각각 아웃된 타자의 수가 0인 상황에서 타석에 선 타자의 타격 결과가 h라는 가정하에 r득점을 하고 아웃된 타자수가 $0, 1, 2$만큼 증가하면서 주자 상태의 변화를 나타낸다. 마찬가지로 두 번째와 세 번째 블록에 있는 행렬들은 아웃된 타자 수가 각각 1과 2인 상황에서 타격결과에 따른 득점과 아웃된 타자수, 주자상태의 변화를 나타내며 $D_k^{(r)}$는 각 상황에서 r득점 하고 3아웃이 될 확률을 나타낸다.

또한 $X_{n-1} = s$일 때, n번째 타석에 선 타자의 타격확률을 $P_n^H(s; h) = P(H_n = h \,|\, X_{n-1} = s)$이라 하면 전확률공식에 의하여 다음을 얻는다.

$$\begin{aligned} P_n^{(r)}(s, s') &= P(R_n = r, X_n = s' \,|\, X_{n-1} = s) \\ &= \sum_{h=1}^{6} P^{(r)}(h; s, s') P_n^H(s; h), \quad s, s' \in \mathcal{S}_X,\ r = 0, 1, 2, 3, 4. \end{aligned} \tag{2.1}$$

2.2. 자료로부터 모형의 모수 결정

본 연구에서는 장기간에 걸친 자료 확보의 어려움으로 인하여 야구동호인 사이트인 아이스텟(www. istat.co.kr)의 2007년부터 2012년까지 6년간의 한국프로야구자료를 이용하였다. 먼저 자료로부터 Moon 등 (2003)과 같이 $P^{(r)}(h)$을 구하였다. 또한 타자의 타격확률 $P_n^H(s;h)$을 구하기 위하여 각 (아웃, 주자) 상황에서 타격의 발생횟수를 바탕으로 타격 결과가 아웃 상태에 의존하는지, 주자상태에 의존하는지에 대하여 동질성 검정을 하였다. 그 결과 2루타와 3루타, 홈런에 대해서는 유의확률이 각각 0.130, 0.378, 0.232이고 1루타와 볼넷 또는 사구, 아웃의 경우에는 유의확률이 모두 10^{-5}보다 작은것으로 나타났다. 또한 1루타, 2루타 등 여기서 고려한 6가지 타격결과 각각에 대하여 8가지 주자상태에는 의존하는지 동질성 검정을 실시한 결과 모든 경우에 유의확률이 10^{-5}보다 작은것으로 나타났다. 따라서 타자의 평균타격확률은 2루타와 3루타, 홈런은 아웃된 타자 수에 의존하지 않으나 주자상태에의 의존하고 기타 타격에 대한 확률은 아웃 수나 주자상태에 의존하도록 정하였다. 모든 타자는 평균타격확률을 따라서 타격하는 것으로 가정하였다. 즉, 식 (2.1)의 $P_n^{(r)}(s,s')$은 $P_n^H(s;h) = P^H(s;h)$, $n = 1,2,\ldots$과 같이 타자에 의존하지 않는 것으로 가정하였다.

2.3. 기대득점

(아웃, 주자)가 $s = (k,b)$인 상태에서 시작하여 n명의 타자가 타석을 마쳤을 때, r득점하고 주자 상태(아웃카운트 포함)가 $s' = (k,b)$에 있을 확률을 $U_n(r;s,s')$이라 하고 $U_n(r;s) = (U_n(r;s,s'),\ s' \in \mathcal{S}_X)$, $n = 0,1,2,\ldots$이라 하자. 단, $U_0(0;s,s) = 1$이고, $r \neq 0$ 또는 $s' \neq s$에 대해서는 $U_0(r;s,s') = 0$이다. n명의 타자가 타석을 마쳤을 때, r득점할 확률은 $n-1$명의 타자가 타석을 마쳤을 때까지 $r-j$득점하고 n번째 타자가 타석을 마친 다음 j $(0 \le j \le \min(r,4))$득점하면 되므로 전확률공식에 의하여 다음이 성립함을 알 수 있다.

$$U_n(r;s) = \sum_{j=0}^{\min(r,4)} \sum_{s' \in \mathcal{S}_X} U_{n-1}(r-j;s,s') P_n^{(j)}(s,s'), \quad r = 0,1,\ldots.$$

따라서 (아웃, 주자)상태가 $s = (k,b)$인 상태에서 시작하여 그 이닝이 끝날 때까지 r득점할 확률 $R(r|s)$와 기대득점 $R_E(s)$는 각각 다음과 같다.

$$R(r|s) = \lim_{n\to\infty} U_n(r;s,\Delta), \quad r = 0,1,2,\ldots, \tag{2.2}$$

$$R_E(s) = \sum_{r=0}^{\infty} rR(r|s), \quad s \in \mathcal{S}_X. \tag{2.3}$$

마르코프연쇄(MC)와 경험분포(EMP)를 이용하여 구한 기대득점 $R_E(k,b)$와 편차(MC-EMP)를 Table 2.1에 나타내었다. Table 2.1은 마르코프연쇄를 이용한 방법이 몇몇 경우(2아웃)를 제외하고는 실제 데이타로부터 구한 기댓값보다 크고 무사상태에서의 편차가 다른 경우보다 상대적으로 크게 나타난다는 것을 보여준다.

2.4. 득점가치

(아웃, 주자)가 $s = (k,b)$인 상태에서 타격결과가 $H_n = h$일 때 r득점하고 상태가 $s' = (k',b')$으로 바뀌었다면 타격에 의하여 얻는 득점의 기댓값은 $r + R_E(s')$점이다. 타격전 상태의 득점기댓값과 타격후의 득점기댓값의 차를 그 타격의 득점가치(RV)라고 한다. $s = (k,b)$인 상태에서 타격 h의 득점가치는

Table 2.1. Run expectancy (RE)

아웃	0			1			2		
주자	MC	EMP	편차	MC	EMP	편차	MC	EMP	편차
0	0.549	0.484	0.065	0.267	0.266	0.001	0.095	0.099	−0.005
1	1.042	0.804	0.238	0.556	0.514	0.042	0.215	0.220	−0.005
2	1.258	1.112	0.145	0.700	0.680	0.020	0.316	0.323	−0.007
3	1.536	1.261	0.275	1.042	0.928	0.114	0.360	0.350	0.010
12	1.756	1.397	0.359	0.972	0.895	0.077	0.439	0.439	0.000
13	2.053	1.587	0.466	1.342	1.125	0.217	0.494	0.480	0.014
23	2.225	1.785	0.440	1.490	1.299	0.191	0.581	0.527	0.055
123	2.734	2.105	0.629	1.662	1.517	0.144	0.729	0.747	−0.018

MC = Markov chain method, EMP = empirical results.

Table 2.2. Run values (RV) by a hit

아웃	주자	1루타	2루타	3루타	홈런	걷기*	아웃
0	0	0.484	0.705	0.987	1.000	0.494	−0.281
	1	0.756	1.155	1.494	1.507	0.714	−0.447
	2	0.723	1.000	1.278	1.291	0.499	−0.403
	3	0.506	0.722	1.000	1.013	0.517	−0.459
	12	1.048	1.478	1.780	1.793	0.978	−0.567
	13	0.762	1.179	1.483	1.496	0.681	−0.639
	23	0.818	1.033	1.312	1.324	0.509	−0.564
	123	1.043	1.498	1.802	1.815	1.000	−0.831
2	0	0.122	0.222	0.265	1.000	0.120	−0.095
	1	0.217	0.663	1.145	1.880	0.225	−0.215
	2	0.687	1.001	1.044	1.779	0.123	−0.316
	3	0.856	0.956	1.000	1.735	0.134	−0.360
	12	0.760	1.710	1.921	2.655	0.290	−0.439
	13	0.946	1.663	1.866	2.601	0.235	−0.494
	23	1.511	2.119	1.779	2.513	0.148	−0.581
	123	1.132	2.194	2.631	3.366	1.000	−0.729

*: 볼넷 또는 사구

다음과 같이 구한다.

$$R_V(h;s) = \sum_{r=0}^{4} \sum_{s' \in \mathcal{S}_X} P^{(r)}(h;s,s')(R_E(s')+r) - R_E(s), \quad h = 1,2,\ldots,6,\ s \in \mathcal{S}_X.$$

아웃된 타자 수가 0과 2인 경우에 각 주자상태에 대하여 타격의 득점가치를 Table 2.2에 나타내었다. 여기서 아웃인 경우 음수는 기대득점이 아웃으로 인하여 감소했음을 나타낸다.

3. 기대승리확률

3.1. 한 이닝 공격을 마친 팀의 기대승리확률

상대팀과의 득점차가 j점인 상태로 i번째 이닝의 공격을 마친 상태에서 경기를 계속한다고 할 때, 방금 공격을 마친 팀 $a \in \mathcal{S}_A = \{초, 말\}$가 경기에서 이길 확률을 $p_i^a(j)$이라하자. 단 '초'와 '말'은 각각 원정

Table 3.1. Comparison of Markov chain method (MC) and Empirical results (EMP) for WE with no runners on base

이닝	아웃	득점차	−2	−1	0	1	−2
3회말	0	MC	0.309	0.411	0.551	0.691	0.776
		EMP	0.298	0.440	0.542	0.700	0.814
		편차	0.011	−0.030	0.008	−0.010	−0.038
	1	MC	0.277	0.374	0.518	0.670	0.763
		EMP	0.278	0.409	0.520	0.695	0.836
		편차	−0.001	−0.035	−0.002	−0.025	−0.073
	2	MC	0.258	0.350	0.497	0.657	0.754
		EMP	0.271	0.389	0.490	0.659	0.815
		오차	−0.014	−0.038	0.006	−0.002	−0.060
7회말	0	MC	0.210	0.327	0.564	0.815	0.904
		EMP	0.194	0.306	0.577	0.824	0.910
		편차	0.017	0.022	−0.013	−0.009	−0.006
	1	MC	0.167	0.266	0.512	0.795	0.891
		EMP	0.178	0.265	0.550	0.795	0.894
		편차	−0.012	0.000	−0.039	0.000	−0.003
	2	MC	0.141	0.225	0.475	0.781	0.883
		EMP	0.158	0.230	0.485	0.766	0.904
		편차	−0.017	−0.004	−0.009	0.015	−0.022

팀(매회 초 공격)과 홈팀(매회 말 공격)을 나타낸다. 직관적으로 $p_i^a(j)$는 다음을 만족하여야 함을 알 수 있다.

(1) $j_1 \le j_2$이면 $p_i^a(j_1) \le p_i^a(j_2)$이다. 즉, 점수차가 클 수록 승리할 확률이 커진다.

(2) $j > 0$이면 $p_1^a(j) \le p_2^a(j) \le \cdots$이고 $j < 0$이면 $p_1^a(j) \ge p_2^a(j) \ge \cdots$이다. 즉, 점수가 앞서있을 때는 이닝이 진행될수록 승리할 확률이 커지고, 점수가 뒤져있을때는 이닝이 진행될수록 승리할 확률이 작아진다.

한국프로야구 자료에서 홈팀이 3이닝을 1점 앞선 사태로 마친 횟수는 459회이며 그 중 298회를 승리한 것으로 나타났다. 이 때 $p_3^{\text{말}}(1) = 298/459 = 0.649$로 정하였다. 이와 같이 자료로부터 구한 결과는 도수가 아주 적은 경우를 제외하고는 $p_i^a(j)$가 만족해야할 위 조건 (1)을 만족하였다. 조건 (2)를 만족하지 않는 경우는 도수가 큰 값을 기준값으로 하여 선형보간하여 조건 (2)를 만족하도록 보정하였다.

3.2. 기대승리확률

공격하고 있는 팀 $a \in \mathcal{S}_A$가 그 경기에서 이길 사건을 W^a라하자. i번째 이닝에서 득점차가 j이고 (아웃, 주자)가 $s = (k, b)$인 상태에서 경기를 계속 진행한다고 할 때 공격하는 팀이 경기에서 이길 확률을 그 상태에서의 기대승리확률(WE)이라 하고 기호로 $W_i^a(j, k, b)$로 나타내기로 한다. 기대승리확률 $W_i^a(j, k, b)$는 다음과 같이 구한다.

$$\begin{aligned} W_i^a(j, s) &= P(W^a | Z_n = (j, s)) \\ &= \sum_{m=0}^{\infty} p_i^a(j+m) R(m|s), \quad (j, s) \in \mathcal{S}. \end{aligned} \tag{3.1}$$

직관적으로 $W_i^a(j, k, b)$는 다음을 만족하여야 함을 알 수 있다.

Table 3.2. Win expectancy (WE) in the 1 out state

이닝	주자	−4	−3	−2	−1	0	1	2	3	4
3회초	0	0.078	0.134	0.216	0.313	0.449	0.613	0.707	0.825	0.896
	1	0.100	0.162	0.250	0.353	0.486	0.641	0.732	0.840	0.904
	2	0.108	0.174	0.264	0.372	0.510	0.654	0.749	0.850	0.908
	3	0.127	0.202	0.297	0.418	0.567	0.686	0.790	0.874	0.918
	12	0.133	0.205	0.300	0.407	0.538	0.677	0.764	0.859	0.915
	13	0.155	0.236	0.336	0.457	0.597	0.711	0.807	0.884	0.926
	23	0.169	0.252	0.358	0.481	0.606	0.728	0.813	0.886	0.929
	123	0.194	0.278	0.378	0.491	0.616	0.731	0.812	0.888	0.931
3회말	0	0.123	0.164	0.277	0.374	0.518	0.670	0.763	0.807	0.866
	1	0.145	0.196	0.311	0.413	0.551	0.691	0.777	0.823	0.880
	2	0.151	0.213	0.325	0.433	0.573	0.704	0.784	0.831	0.888
	3	0.165	0.251	0.358	0.483	0.625	0.736	0.798	0.851	0.908
	12	0.179	0.245	0.360	0.465	0.594	0.719	0.798	0.844	0.897
	13	0.196	0.287	0.396	0.517	0.649	0.752	0.814	0.866	0.918
	23	0.217	0.301	0.419	0.538	0.658	0.756	0.823	0.873	0.925
	123	0.240	0.324	0.435	0.542	0.661	0.764	0.830	0.876	0.923
7회초	0	0.026	0.067	0.119	0.207	0.426	0.682	0.813	0.939	0.972
	1	0.044	0.093	0.162	0.269	0.480	0.714	0.835	0.945	0.975
	2	0.050	0.101	0.175	0.300	0.516	0.732	0.853	0.950	0.976
	3	0.064	0.119	0.205	0.374	0.604	0.777	0.897	0.961	0.979
	12	0.074	0.140	0.228	0.349	0.551	0.753	0.862	0.953	0.978
	13	0.091	0.161	0.263	0.429	0.642	0.801	0.907	0.965	0.982
	23	0.102	0.179	0.301	0.465	0.652	0.816	0.905	0.965	0.983
	123	0.139	0.224	0.329	0.471	0.653	0.810	0.902	0.965	0.983
7회말	0	0.046	0.077	0.167	0.266	0.512	0.795	0.891	0.958	0.984
	1	0.066	0.111	0.214	0.332	0.565	0.816	0.904	0.963	0.986
	2	0.071	0.124	0.229	0.366	0.605	0.830	0.914	0.967	0.988
	3	0.081	0.155	0.262	0.450	0.703	0.863	0.937	0.976	0.991
	12	0.101	0.168	0.285	0.415	0.631	0.843	0.920	0.969	0.989
	13	0.114	0.203	0.324	0.505	0.732	0.877	0.943	0.979	0.992
	23	0.134	0.220	0.366	0.543	0.734	0.884	0.944	0.979	0.992
	123	0.172	0.266	0.390	0.538	0.729	0.881	0.942	0.978	0.992

W1. $j_1 \leq j_2$이면 $W_i^a(j_1, k, b) \leq W_i^a(j_2, k, b)$이다. 즉, 같은 이닝 같은 아웃카운트, 주자 상황에서는 점수차가 클 수록 승리할 확률이 커진다.

W2. $k_1 \leq k_2$이면 $W_i^a(j, k_1, b) \geq W_i^a(j, k_2, b)$이다. 즉, 같은 이닝에서 아웃된 타자 수가 많을 수록 승리할 확률이 작아진다.

W3. 점수가 앞서있을 때는 이닝이 진행될수록 승리할 확률이 커지고, 점수가 뒤져있을때는 이닝이 진행될수록 승리할 확률이 작아진다. 즉, $j > 0$이면 $W_1^a(j, k, b) \leq W_2^a(j, k, b) \leq \cdots$ 이고 $j < 0$이면 $W_1^a(j, k, b) \geq W_2^a(j, k, b) \geq \cdots$ 이다.

실제 자료에서 발생횟수가 충분히 크지 않은 경우는 조건 W1–W3을 만족하지 않는 경우가 자주 나타났으나 식 (3.1)과 같이 구한 기대승리확률은 조건 W1–W3를 모두 만족하였다. 실제 자료에서 발생횟수가 충분히 큰 (6년간 150회 이상) 주자 없는 상황에 대하여 마르코프연쇄 방법과 경험분포를 이용하여

Table 3.3. Win values (WV) by a hit at 5th inning of home team

타격	아웃	주자	−3	−2	−1	0	1	2	3
1루타	0	0	0.060	0.072	0.069	0.065	0.037	0.032	0.014
		1	0.096	0.111	0.101	0.093	0.053	0.044	0.021
		123	0.115	0.115	0.119	0.079	0.061	0.036	0.019
	2	0	0.014	0.019	0.019	0.018	0.011	0.009	0.004
		1	0.025	0.034	0.031	0.034	0.017	0.016	0.007
		123	0.127	0.168	0.208	0.185	0.119	0.098	0.036
홈런	0	0	0.102	0.154	0.143	0.179	0.082	0.092	0.036
		1	0.195	0.224	0.251	0.196	0.136	0.097	0.043
		123	0.295	0.249	0.215	0.146	0.088	0.071	0.037
	2	0	0.087	0.157	0.143	0.210	0.085	0.111	0.040
		1	0.229	0.281	0.333	0.277	0.186	0.142	0.058
		123	0.516	0.497	0.450	0.356	0.205	0.175	0.088
아웃	0	0	−0.034	−0.042	−0.040	−0.040	−0.022	−0.020	−0.009
		1	−0.057	−0.066	−0.063	−0.055	−0.033	−0.026	−0.013
		123	−0.114	−0.110	−0.094	−0.068	−0.045	−0.030	−0.017
	2	0	−0.010	−0.015	−0.014	−0.016	−0.008	−0.008	−0.003
		1	−0.025	−0.033	−0.033	−0.034	−0.019	−0.017	−0.007
		123	−0.090	−0.112	−0.113	−0.105	−0.061	−0.052	−0.023

구한 WE를 Table 3.1에서 비교하였다. Table 3.1로부터 마르코프연쇄를 이용한 방법과 실제 자료로부터 구한 WE가 매우 유사함을 할 수 있다. Table 3.2는 3회와 7회 1사 상태에서 원정팀과 홈팀의 기대승리확률을 나타낸 것이다.

3.3. 타격의 승리가치

5이닝초, 1점 뒤지고 1사 주자 12루에 있는 상황에서 타석에 선 타자가 안타를 쳐서 1득점하고 주자상태가 13루가 되었다면 득점차는 0이 되고 이때 WE는 $W^{초}(5,0,1,13) = 0.611$이 된다. 타자가 타격하기 전 상태의 WE는 $W^{초}(5,-1,1,12) = 0.316$이므로 이 안타에 의한 기대승리확률의 변화량은 $0.611 - 0.316 = 0.295$이다. 타격전 상태의 기대승리확률과 타격후의 기대승리확률의 차를 그 타격의 승리가치라고 한다. i번째 이닝, 득점차가 j이고 (아웃, 주가)가 $s = (k, b)$인 상태에서 타격 h의 승리가치(WV)는 다음과 같이 구한다.

$$W_V^a(h; i, j, s) = \sum_{r=0}^{4} \sum_{s' \in \mathcal{S}_X} P^{(r)}(h; s, s') W_i^a(j + r, s') - W_i^a(j, s), \quad h = 1, 2, \ldots, 6.$$

Table 3.3은 5회말 점수차가 −3점부터 3점까지 일 때 홈팀의 WV일부를 나타낸 것이다.

4. 결론

본 논문에서는 마르코프연쇄를 이용하여 한국프로야구에서의 기대득점, 기대승리확률, 타격의 득점가치와 승리가치를 구하는 방법을 제시하였다. 이 논문에서 제안한 방법에 의한 결과와 실제 데이터로 부터 구한 것과 비교하였다. 그 결과 발생횟수가 충분히 커서 통계적으로 유의미한 경우에는 본 논문에서 제시한 방법이 매우 효과적임을 알 수 있었다.

본 논문에서 제시한 방법은 각 팀별 타자 혹은 투수에 대한 자료가 통계적으로 의미가 있을 만큼 충분하다면 타자의 타격확률과 투수진 등을 고려하여 각 팀별 특성에 맞는 RE, WE 등을 구하는데 적용할 수 있을 것이다. 또한 RE, WE를 바탕으로한 선수평가 모형을 개발한다면 한국프로야구에서 경기 상황을 반영한 타자의 능력이 득점이나 경기의 승패에 미친 기여도를 효과적으로 평가할 수 있는 새로운 형태의 평가 지표로 사용될 수 있을 것이다. 또한 이를 이용하여 선수들의 기량을 효과적으로 평가하여 타순 결정이나 선수의 연봉 협상을 위한 객관적인 평가지표로 사용할 수 있으리라 기대한다. 마지막으로 특정 상황에서 팀이 승리할 확률 등을 기반으로 승패를 예측하는 모델 개발에도 사용할 수 있으리라 기대한다.

References

Bukiet, B., Harold, E. R., and Palacios, J. L. (1997). A Markov chain approach to baseball, *Operations Research*, **45**, 14–23.

Hirotsu, N. and Wright, M. (2003). A Markov chain approach to optimal pinch hitting strategies in a designated hitter rule baseball game, *Journal of the Operations Research Society of Japan*, **46**, 353–371.

Hirotsu, N. and Wright, M. (2005). Modelling a baseball game to optimise pitcher substitution strategies incorporating handedness of players, *IMA Journal of Management Mathematics*, **16**, 179–194.

Jeong, J. S. (2014). Efficient Estimation Model of Hitter using Big Data Analysis in Korean Baseball League, Master Thesis, Changwon National University (in Korean).

Moon, H. W., Woo, Y. T., and Shin, Y. W. (2013). Analysis of the Korean baseball league using a Markov chain model, *The Korean Journal of Applied Statistics*, **26**, 649–659.

Sokol, J. S. (2003). A robust heuristic for batting order optimization under uncertainty, *Journal of Heuristics*, **9**, 353–370.

Tango, T., Lichtman, M., and Dolphin, A. (2006). *The Book: Playing the Percentages in Baseball*, TMA Press.

한국 프로야구 경기에서 기대득점과 기대승리확률의 계산

문형우[a] · 우용태[b] · 신양우[c,1]

[a]창원대학교 산업기술연구원, [b]창원대학교 컴퓨터공학과, [c]창원대학교 통계학과

(2015년 11월 4일 접수, 2015년 11월 29일 수정, 2016년 1월 25일 채택)

요 약

감독이 작전을 구사하는 상황이나 타자의 타격 가치를 평가하는데 유용하게 사용될 수 있는 지표로서 미국프로야구에 대해서는 기대득점과 기대승리확률, 타격의 득점가치, 타격의 승리가치 등이 제시되었다. 기대득점은 각각의 아웃카운트와 주자 상황에서 그 이닝이 끝날 때까지 얻는 점수의 기댓값이다. 기대승리확률은 이닝, 점수차, 아웃카운트, 주자상태가 주어진 상태에서 경기를 계속 한다고 할 때, 공격하고 있는 팀이 승리할 확률이다. 타격의 득점가치는 타격전 상황의 기대득점과 타격결과에 의하여 변화된 상황의 기대득점 사이의 차이를 말한다. 타격의 승리가치는 타격 전후 상황의 기대승리확률간의 차이로서 타격 결과가 승리에 미치는 영향을 나타낸다. 한국프로야구에서는 장기간 축적된 자료의 부족으로 총 발생횟수에 대한 특정 상황의 상대돗수를 이용하여 구한 이들 지표가 통계적인 의미를 갖지 못하는 경우가 종종 나타난다. 이와 같은 문제점을 극복하기 위하여 본 논문에서는 마르코프연쇄를 이용하여 한국프로야구에서 기대득점, 기대승리확률, 득점가치와 승리가치를 구하는 방법을 제시한다.

주요용어: 기대득점, 기대승리확률, 기대득점가치, 기대승리가치, 마르코프연쇄.

이 논문은 2014년도 하이브레인넷 산학융합연구과제와 2015∼2016년도 창원대학교 자율연구과제 연구비 지원으로 수행된 연구결과임.

[1]교신저자: (51140) 경남 창원시 의창구 창원대학로 20, 창원대학교 통계학과. E-mail: ywshin@changwon.ac.kr