



## MARKOV CHAIN을 이용한 한국프로야구 분석 및 최적 타순 계산을 위한 모델연구

---

저자  
(Authors) 황일회, 조현주, 김대영, 이민석

출처  
(Source) [대한산업공학회 추계학술대회 논문집](#), 2011.11, 821-828 (8 pages)

발행처  
(Publisher) [대한산업공학회](#)

URL <http://www.dbpia.co.kr/Article/NODE01960378>

APA Style 황일회, 조현주, 김대영, 이민석 (2011). MARKOV CHAIN을 이용한 한국프로야구 분석 및 최적 타순 계산을 위한 모델연구. 대한산업공학회 추계학술대회 논문집, 821-828.

이용정보  
(Accessed) 한양대학교(서울)  
166.104.4.194  
2016/03/25 20:07 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.

이 자료를 원저작자와의 협의 없이 무단게재 할 경우, 저작권법 및 관련법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

The copyright of all works provided by DBpia belongs to the original author(s). Nurimedia is not responsible for contents of each work. Nor does it guarantee the contents.

You might take civil and criminal liabilities according to copyright and other relevant laws if you publish the contents without consultation with the original author(s).

# MARKOV CHAIN을 이용한 한국프로야구 분석 및 최적 타순 계산을 위한 모델연구

**KAIST**

산업및시스템공학과

황일회(20050715)

조현주(20082045)

김대영(20070093)

이민석(20090633)

Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

2011년 대한산업공학회 추계학술대회

## 1.1 연구 배경

스포츠산업의 발전 → 산업공학의 적용 기회

- Penn State가 100년 전 **제조 시스템의 효율적 설계 및 운영**이라는 필요를 인지하여 **산업공학**이라는 학문 분야를 만들어냄
- 이어서 서비스 산업의 비율이 경제에서 차지하는 비중이 커져감에 따라 서비스 엔지니어링 등장. **서비스 분야의 효율적 설계 및 운영**으로 **산업공학의 응용분야를 확장**하겠다는 것에서 출발한 개념



**스포츠산업에 산업공학 적용!**

Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

# 1.1 연구 배경

## Sports Analytics 분야의 성장

최근 다수의 OR의 대가들이 Sports Analytics 분야의 연구를 시작



MIT Sloan Professor  
**Dimitris Bertsimas**

Statistical model that predict the number of game that Boston Red Sox baseball team would win



MIT Professor  
**Stephen C. Graves**

How to Catch a Tiger:  
Understanding Putting  
Performance on the PGA TOUR,

## MIT Sloan Sports Analytics Conference



관련 학회가 최근 5년 사이에 9배 규모로 성장할 정도로 Sports Analytics에 대한 관심이 증가하고 있음

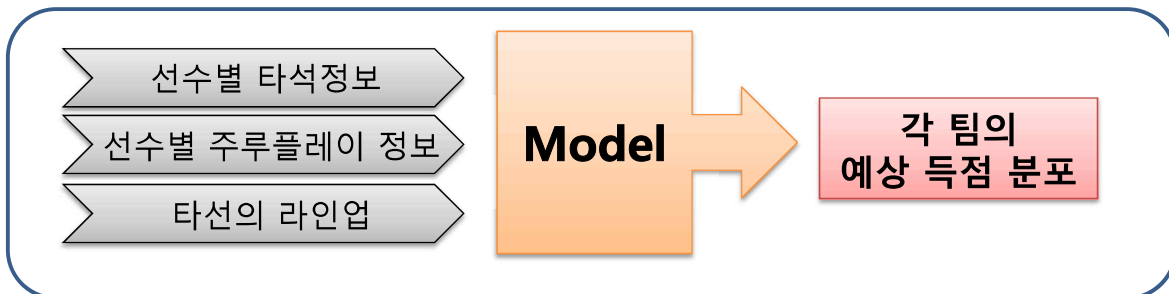
Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

# 1.2 본 연구의 의의

한국프로야구에 대한 높은 관심에 발맞춰, 프로야구에 산업공학 적용

-종전의 진부한 방법에서 벗어나, 과학적이고 수학적 방법을 이용해 한국프로야구를 체계적으로 분석

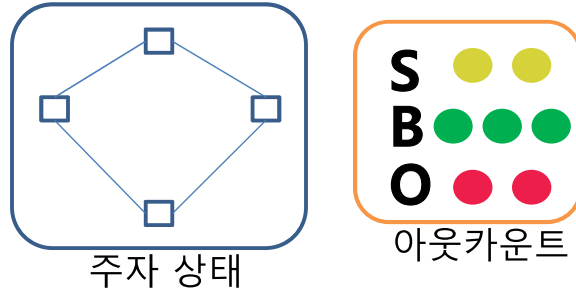
-이를 통해 한국프로야구가 더 나은 데이터 분석으로 한 단계 도약할 수 있도록 그 방법론을 제시



- 1 제시한 모델이 한국프로야구에 매우 적합하다는 것을 실제 경기 결과와 비교, 분석함으로써 보임
- 2 이 모델을 이용하여 각 팀별로 최고의 기대득점을 갖는 최적 타순을 계산한다.

Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

## 2.1.1 야구경기에서의 25가지 STATE



$(S_0, S_1, S_2, S_3)$

아웃카운트  
0~2

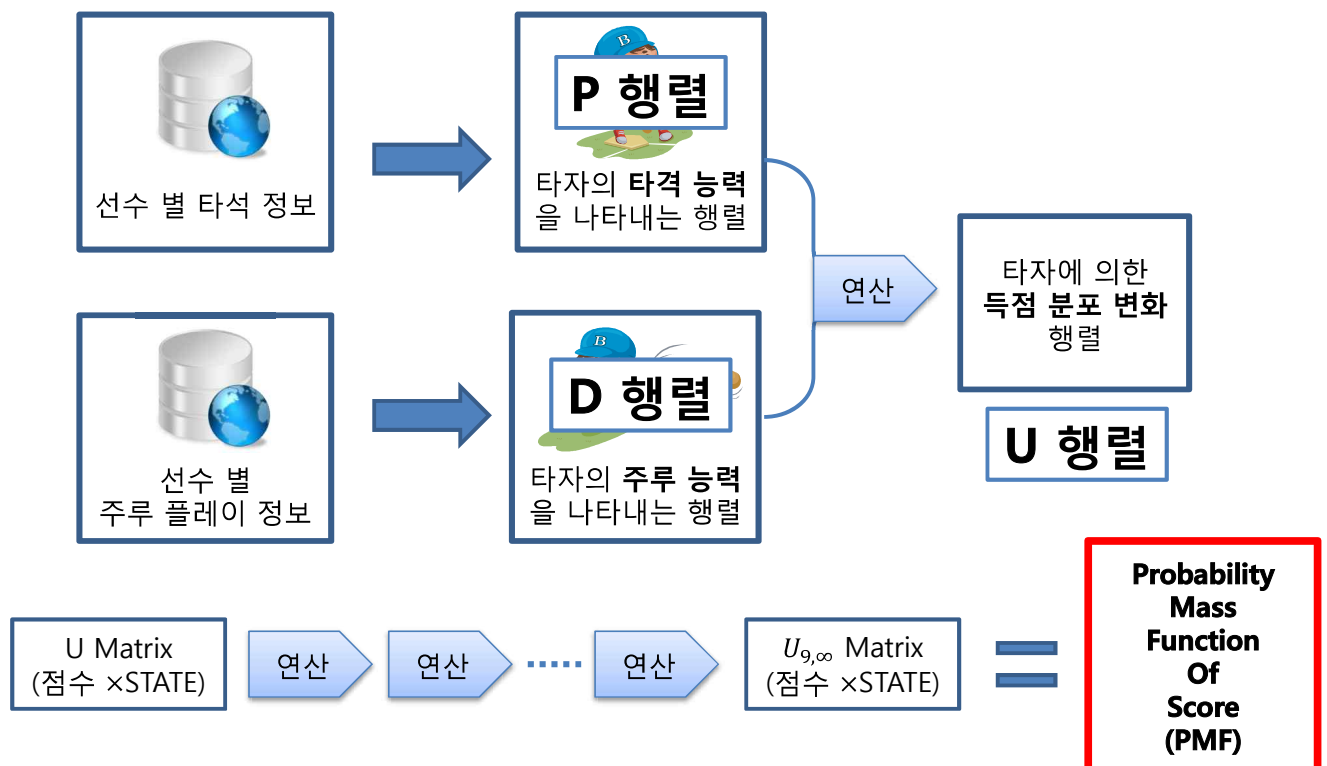
주자상태  
주자 없으면 0, 있으면 1

Ex) 무사 1, 2루 경우에는 (0,1,1,0)  
2사 1, 3루 경우에는 (2,1,0,1)

이닝의 종료를 나타내는 3아웃 상태 (3,0,0,0)를 추가하여 야구경기의 매 상황을 총 25가지의 state로 나타내기로 한다.

Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

## 2.1.2 MODELING PROCESS



Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

## 2.1.3 P MATRIX 와 U MATRIX

P 행렬 : 한 타자가 State Transition 을 일으킬 확률들의 전이 행렬

U 행렬 : Row 에 점수가, Column 에 State 정의되며, 득점 분포를 나타내는 행렬

따라서,  $U_n = n$ 명의 타자가 transition을 일으키고 난 후의 U 행렬

→ U행렬과 P행렬을 연산하면 state transition이 일어난 후의 U행렬을 구할 수 있다.

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,25} \\ p_{2,1} & & & & p_{2,25} \\ \vdots & & \ddots & & \vdots \\ p_{25,1} & p_{25,2} & p_{25,3} & \cdots & p_{25,25} \end{bmatrix} \quad U = \begin{bmatrix} p_{0,1} & p_{0,2} & p_{0,3} & \cdots & p_{0,24} & p_{0,25} \\ p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,24} & p_{1,25} \\ p_{2,1} & & & & & p_{2,25} \\ \vdots & & & \ddots & & \vdots \\ p_{20,1} & p_{20,2} & p_{20,3} & \cdots & p_{20,24} & p_{20,25} \end{bmatrix}$$

$p_{i,j}$ :  $i$  state에서  $j$  state로 transition이 일어날 확률

$p_{n,i}$ : 점수가  $n$ 점이고 state  $i$ 에 있을 확률

➤ U 행렬을 update 하는 방법

$$U_{n+1} = U_n * P_{n+1}$$

$$U_{n+1}(\text{row } j) = U_n(\text{row } j)P_0 + U_n(\text{row } j-1)P_1 + U_n(\text{row } j-2)P_2 + U_n(\text{row } j-3)P_3 + U_n(\text{row } j-4)P_4$$

여기서  $P_0, P_1, P_2, P_3, P_4$  란 P 행렬의 분해 행렬로써 각각의 점수를 낼 전이 행렬을 분리해 낸 것이다. EX)  $P_0 \rightarrow$  P 행렬에서 0점을 득점할 State Transition 만 분리

Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

## 2.1.3 U MATRIX 연산을 통한 PMF 계산

$U_{9,\infty}$  = 9번째 이닝 종료후의 U행렬

여기서 25번 state 는 3아웃 state 이므로 연산을 계속해 나가면, 모든 확률은 3아웃 상태로 absorbing 된 다.

$$U_{i,\infty} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & \cdots & 24 & 25 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & p_0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & p_1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & p_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & p_{20} \end{bmatrix}$$

$p_0, p_1, p_2, \dots, p_{20}$  : Absorption state

: 9번째 이닝 종료 후의 점수 분포(pmf)  
즉, 한 팀 이 한 경기에서 득점할 수의 PMF 이다. (모델의 최종 OUTPUT)

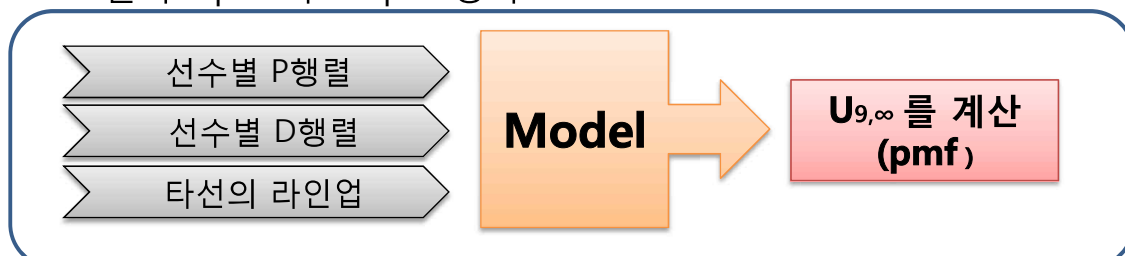
도루 상태 전이를 나타내는 D matrix

D행렬은 25\*25형태를 갖게 되며 각각의 행은 도루 이전 상태, 각각의 열은 도루 이후 상태  
→ 도루 시도율과 도루 성공율을 통해 D행렬을 만들 수 있다

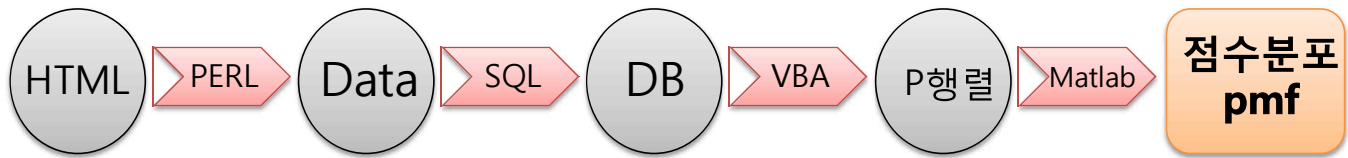
$$U_{i,\infty} = U_{i,0} * P_1 D_1 * P_2 D_2 * \cdots * P_9 D_9 * P_1 D_1 * P_2 D_2 * \cdots$$

여기서 \* 연산은 앞 슬라이드 마지막의 연산을 의미한다.

➤ 모델의 input 과 output 정리



## 2.2 IMPLEMENTATION 과정



수집된 데이터를 자동으로 데이터베이스화(저장 및 업데이트) 하기 위해 사용된 언어는 **PERL**

데이터베이스 도구로는 **Microsoft SQL Server 2008**과 **Microsoft Access**를 사용

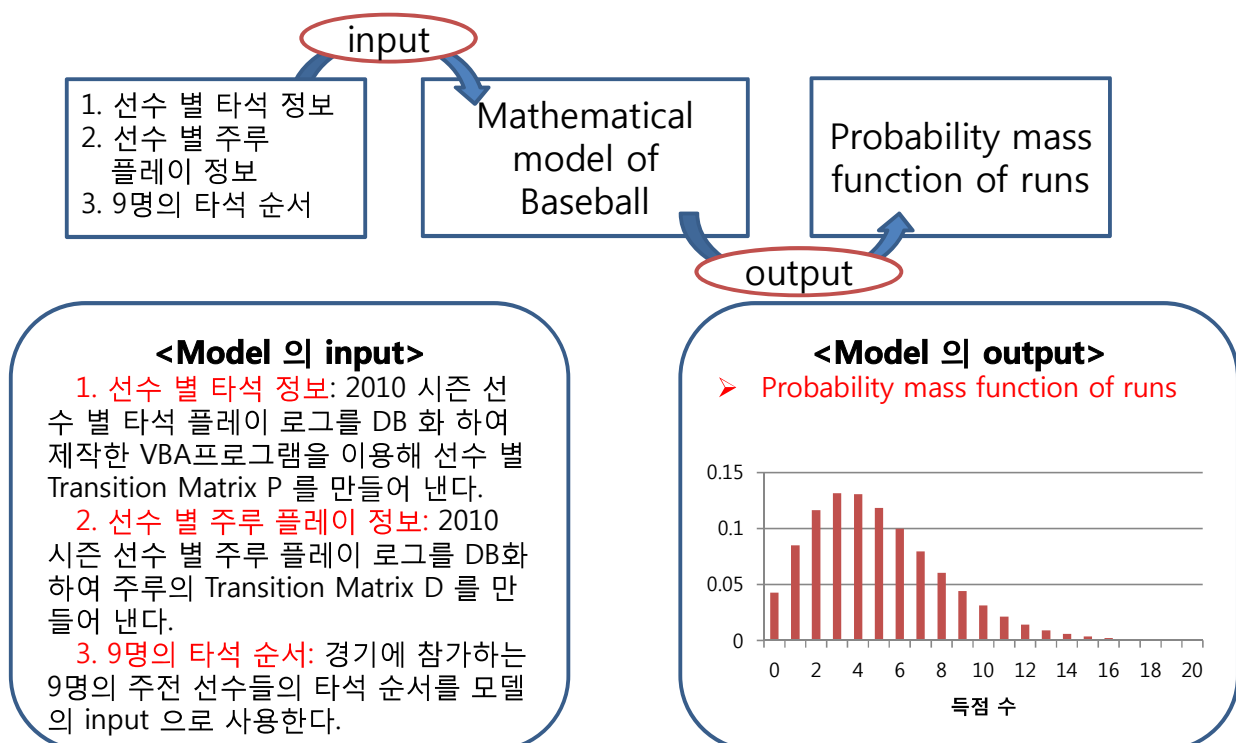
**VBA(Visual Basic for Applications)**를 사용하여 모든 선수별 P행렬을 생성

제시한 이론을 구현하고 모델링 및 분석하기 위해, 주요 도구로는 **Matlab**을 사용

Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

## 2.3 모델의 타당성 분석

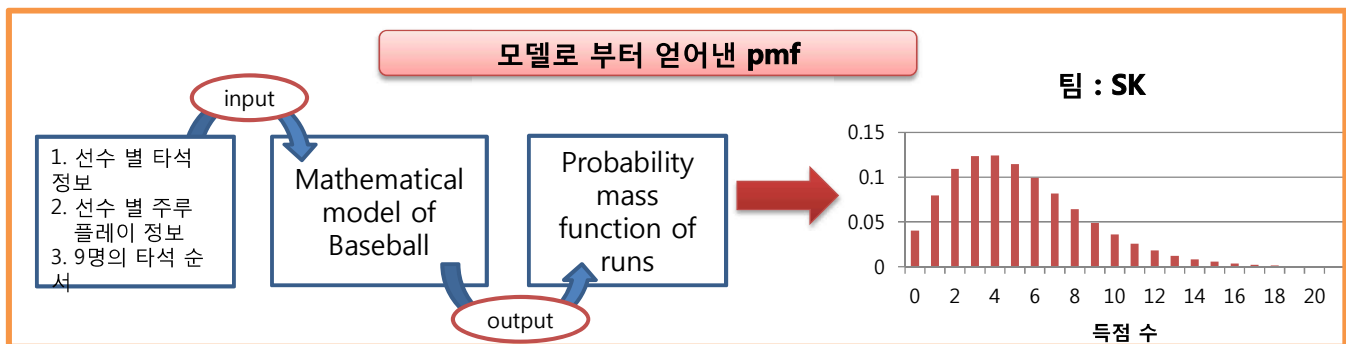
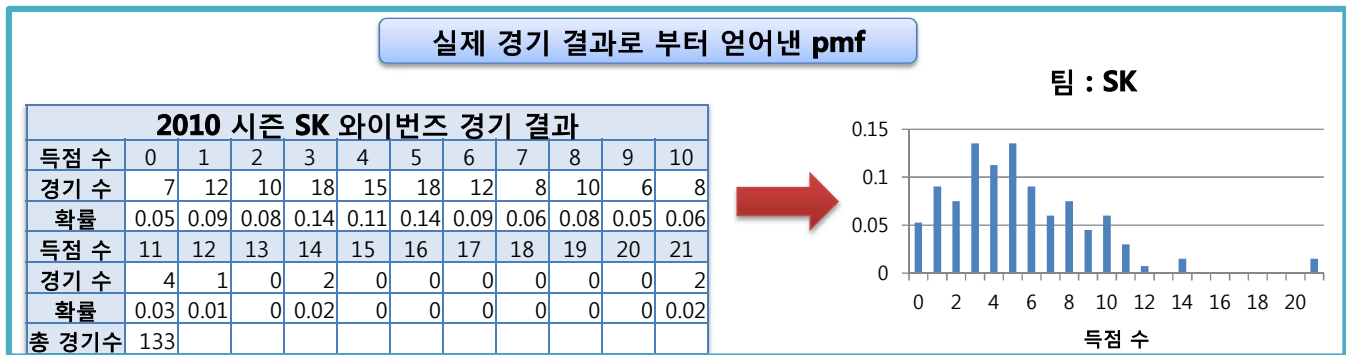
### ➤ Mathematical model of Baseball Game



Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

## 2.3 모델의 타당성 분석

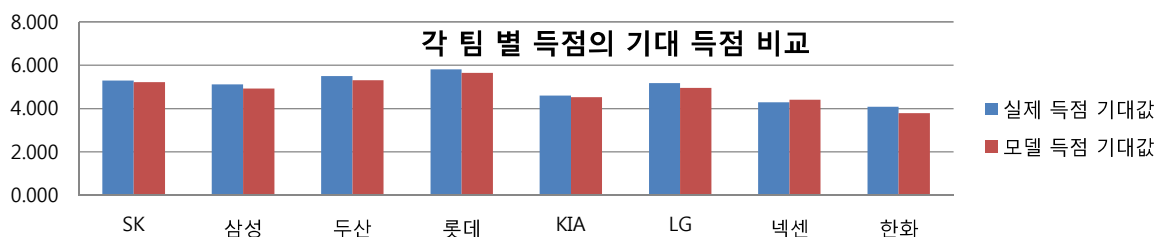
- ▶ 앞 슬라이드에서 모델로부터 pmf 를 얻어내었으므로 다음으로 2010년 시즌 실제 경기 결과를 바탕으로 probability mass function 을 생성하여 모델로 부터 얻어낸 probability mass function 과 비교한다



Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

## 2.3 모델의 타당성 분석

- ▶ 두 개의 probability mass function 으로 부터 득점의 기대 득점 비교
  - 먼저 모델로 구한 각 팀 별 Probability mass function of runs 로 부터 expected runs 를 구한다.
  - 그 후 2010년 시즌 실제 경기 결과로 부터 pmf를 얻어내고, 평균득점을 구하여 모델과 비교해본다.



	SK	삼성	두산	롯데	KIA	LG	넥센	한화
실제경기 결과로부터의 득점 기대 값	5.293	5.120	5.496	5.812	4.594	5.173	4.286	4.083
모델로부터의 득점 기대 값	5.218	4.926	5.306	5.646	4.519	4.963	4.395	3.792
오차	1%	4%	3%	3%	2%	4%	3%	7%
실제경기 결과의 기대 값 순위	3	5	2	1	6	4	7	8
모델로부터의 기대 값 순위	2	5	3	1	6	4	7	8

- 한 팀을 제외한 대부분의 팀이 5% 미만의 오차를 보인다.
- 또한, 기대 값 순위도 상당한 양의 상관 관계를 보임을 알 수 있다.

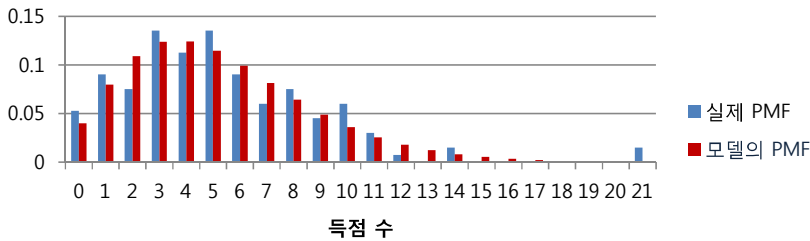
Copyright (c) 2011 KAIST SDM Lab. All rights reserved.



## 2.3 모델의 타당성 분석

### ➤ 팀 SK 의 probability mass function 비교

Probability mass function 의 비교: 실제 vs 모델 (팀: SK)



-파란 색 그래프는 실제 경기 결과로 부터 얻어낸 PMF 이고  
-빨간 색 그래프는 모델로 부터 얻어낸 PMF 이다.  
- 이 두 PMF 가 Trend 상의 일치성을 보임  
알 수 있다.

## 2.4 응용 - 최적 타순 구하기

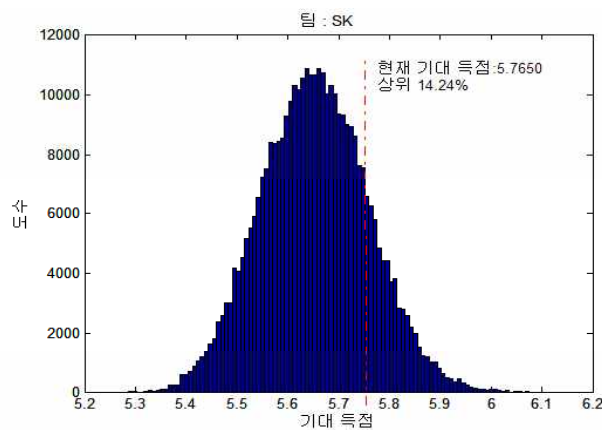
- 모델은 하나의 타순을 입력 받아서 하나의 Probability mass function 을 만들어 낸다. 또한 그 PMF 를 통해 expected runs 즉, 기대득점을 만들어 낼 수 있다.
- 따라서 한 팀의 주전 선수 9명의 타순을 계속 바꿔가며 모델을 통한 연산을 진행 하면 각각의 타순 별로 다른 expected runs 를 얻어 낼 수 있다.
  - 한 팀의 주전을 9명으로 가정하였을 때 총 9! 개의 타순이 존재한다.
  - 이를 각각 모델에 input 하면 총 9! 개의 expected runs 가 출력되고 이들 값 중에 가장 높은 expected runs 를 제공하는 타순이 최적화된 타순이 된다.

Copyright (c) 2011 KAIST SDM Lab. All rights reserved.

## 2.4 응용 - 최적 타순 구하기

### ➤ Result : 팀 SK 와이번즈

2010 시즌 결과	
타순	이름
1	정근우
2	박재상
3	김재현
4	박정권
5	최정
6	김강민
7	나주환
8	박경완
9	조동화
기대득점	5.7650



최고 기대득점 타순	1	4	3	6	9	7	2	5	8
기대 득점	6.0754								
최저 기대득점 타순	1	8	2	5	7	9	3	6	4
기대 득점	5.2859								

- MATLAB 을 이용하여 한 팀의 주전선수 9명의 타순을 바꿔가며 9!번의 모델 연산을 통해 나온 9! 개의 기대득점으로 그린 histogram.
- 2010년 시즌 SK 의 평균 기대 득점은 5.7650 점으로 전체 기대득점의 분포 중 상위 14.24% 를 차지 한다.
- 위 결과를 바탕으로 최적 타순을 구성한다면 더 좋은 시즌 결과를 기대 할 수 있다.

Copyright (c) 2011 KAIST SDM Lab. All rights reserved.



### 3. 결론

#### ➤ 이번 연구에서 기존의 연구와 차별화 되는 점

1. 기존의 모델을 바탕으로 연구를 진행하였지만, 이전 모델이 도루를 포함하지 않았던 것에 비해, 도루 모델을 구현하여 포함해서 모델을 발전 시켰다는데 의미가 있다.
2. DB 기술을 바탕으로 실제 모든 transition data 를 수집하여서, transition matrix를 만드는데 사용하였다.
  - 기존의 연구는 실제의 transition data 를 사용하지 않았으며, 수많은 가정을 통해 transition matrix 를 만들었다. (기존의 모델은 병살타, 희생타, 도루 등을 고려하지 못하였다.)
  - 이번 연구에서는 MS sql server 를 이용한 대규모 Data 처리를 통해, 가정을 부수고 좀더 실제와 부합하는 transition matrix 를 만들어 내었다.
3. 회귀 분석이 주를 이루던 국내 야구 연구와는 다른 방법으로 Markov chain 을 통해 pmf 를 만들어 내었고 과 실제 data 를 사용해 유용한 결과를 만들어 내었다.
  - 단순한 원인과 결과의 분석이 아닌, 직접 probability mass function 을 만들어 내어 많은 응용이 가능한 model 을 국내에 소개 하였다.

#### ➤ Reference

- [1] A Markov Chain Approach To Baseball, Bruce Bukiet and Elliotte Rusty Harold, 1994
- [2] DEA OERA를 이용한 프로야구 선수들에 대한 성과 측정, 이덕주, 양원모, 2004
- [3] 스탯티즈, statiz.co.kr
- [4] 한국야구위원회(KBO), www.koreabaseball.com
- [5] 한국필사용자모임 www.perlor.kr
- [6] Database Processing, David M. Kroenke
- [7] MATLAB: A Practical Introduction to Programming and Problem Solving, Stormy Attaway
- [8] Baseball Hacks, Joseph Adler