数理解析研究所講究録
1194巻 2001年 87-96



# IMPROVED OPTIMAL BATTING ORDER WITH SEVERAL EFFECTS FOR BASEBALL

武井貴裕 (Takahiro Takei), 瀬古進 (Susumu Seko)
and 穴太克則 (Katsunori Ano) 南山大学 (Nanzan University)

**Abstract**

In this paper, optimal batting order model with the effects of stealing, double play and batting average with runners in scoring position is studied.

**Keywords**: Markov Chain, Baseball, Optimization, Batting Order

## 1 Introduction

Many indices evaluating player's batting ability are considered. Such major indices are Homerun (HR), Run Batted In (RBI) and Batting Average (BA), etc. These indices, however, do not compare the batter who has high batting average and low slugging average with the batter who has low batting average and high slugging average. Which player is better? To answer this question, two models have been studied. One is the Offensive Earned-Run Average (OERA) index proposed by Cover and Keilers (1977), and the other is the Modified Offensive Earned-Run Average (MOERA) index with steal effect proposed by Ano (1999). These models calculate the expected number of runs that a player would score if he batted in all nine position in the lineup. One of merits of the MOERA index quantitatively describes steal effect by SE (Steal Effect) := MOERA − OERA. Although the steal is important, there is no index measuring the steal effect of each players. In general, a single hit followed by a successful steal sometimes has the same offensive effect as a double, in addition, an infield grounder in the state with runner on first base sometimes has the same offensive effect as taken two outs. In section 2, we propose the DOERA index with both effects of steal and double play which yields DPE (Double Play Effect) := DOERA − MOERA, and the IOERA (Improved OERA) index with the effects of steal, double play and batting average with runners in scoring position.

These DOERA and IOERA estimate individual player's abilities. Moreover we consider the offensive performance of the team and study the improved model originated Optimal Batting Order (OBO) proposed by Bukiet, Harold and Palacios (1997). This OBO model finds the optimal batting order among all of the 9! possible orders which maximizes the expected runs of a game. The OBO, the same as OERA and MOERA, is the model based on Markov Chain with an absorbing state. Embedding the effects of steal, double play and batting average with runners in scoring position, we make the OBO index more realistic and call it IOBO (Improved OBO). The models we consider here are summarized as follows.

1. IOERA is the model with effects of steal, double play and batting average with runners in scoring position. Note that IOERA includes both MOERA and DOERA. These models yield:
   (a) Effect of steal for the batter, namely SE (Steal Effect) := MOERA − OERA.
   (b) Effect of double play for the batter, namely DPE (Double Play Effect) := DOERA − MOERA.
   (c) Effect of batting average with runners in scoring position for batter, namely BASPE (Batting Average with runners in Scoring Position Effect) := COERA − DOERA.
2. IOBO is the optimal batting order with effects of steal, double play and batting average with runners in scoring position. MOBO is the order with only steal effect, DOBO is the order with both effects of steal and double play. These models yield:
   (a) Effect of steal for the team, namely TSE (Team Steal Effect) := The expected run determined by MOBO − The expected run by OBO.
   (b) Effect of double play for the team, namely TDPE (Team Double Play Effect) := The expected run by DOBO − The expected run by MOBO.
   (c) Effect of batting average with runner in scoring position for the team, namely TBASPE (Team Batting Average with runners in Scoring Position Effect) := the expected run by COBO − The expected run by DOBO.

Some examples are presented for Nippon Professional Baseball (NPB) player's and teams.

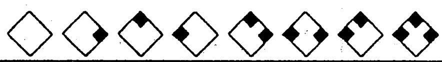
# 2 Improved Offensive Earned-Run Average

## 2.1 The State Space of Baseball, Probabilities of Batting and Stealing and Runner Advancing Rule

The IOERA index evaluating a player's ability figure out the expected number of runs that a batter would score if he batted in all nine position in the lineup. We have to make clear the out count and the state of runner and define the offensive probabilities for each players by their baseball statistics; AB, H, 2B, 3B, HR, BB, SB, CS, IG (AB: At Bats, H: Hits (Singles), 2B: Double, 3B: Triple, HR: Homerun, BB: Base on Balls, SB: Stolen Bases, CS: Caught Stealing, IG: Infield Grounder).

### 2.1.1 The State Space of Baseball

The state space of baseball consists of 25 states. Twenty-four of these states corresponds to zero or one runners on each of three bases and zero, one, or two outs. The twenty-fifth state corresponds to the end of the inning when the third out occurs. We define the twenty-fifth state as 0, and no outs, no-runner as 1, $\cdots$, two-out, three runners as 24 as follows.



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| no outs | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| one out | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| two outs | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| three outs | 25 | | | | | | | |

So the state space $S$ is given by $S = \{0, 1, \cdots, 24\}$ . The state 0 is regarded as an absorption state in a Markov chain. Let $S^\alpha$ be the substate space with runners in scoring position. Let $S^\beta$ be the substate space with no-runner or runner on only first base. The state space can be divided into $S^\alpha$ and $S^\beta$ such that $S = \{S^\alpha, S^\beta\}$ where $S^\alpha = \{3, \cdots, 8, 11, \cdots, 16, 18, \cdots, 24\}$ and $S^\beta = \{0, 1, 2, 9, 10, 17, 18\}$.

### 2.1.2 Probabilities of Batting and Stealing

Probabilities of batting and stealing with runners in scoring position are as follows. $p_0^\alpha = Pr$(easy fly or grounder (out) except for an infield grounder), $p_{IG}^\alpha = Pr$(infield grounder), $p_{B_1}^\alpha = Pr$(base on balls and caught stealing), $p_{B_2}^\alpha = Pr$(base on balls and no steal), $p_{B_3}^\alpha = Pr$(base on balls and steal second base), $p_1^\alpha = Pr$(single and caught stealing), $p_2^\alpha = Pr$(single and no steal),

$p_3^\alpha = Pr$(single and steal second base), $p_4^\alpha = Pr$(double and caught stealing), $p_5^\alpha = Pr$(double and no steal), $p_6^\alpha = Pr$(double and steal third), $p_7^\alpha = Pr$(triple and caught stealing), $p_8^\alpha = Pr$(triple and no steal), $p_9^\alpha = Pr$(triple and steal home), $p_{10}^\alpha = Pr$(homerun), $p_B^\alpha = p_{B_1}^\alpha + p_{B_2}^\alpha + p_{B_3}^\alpha = Pr$(base on balls).

Similarly, probabilities of batting and stealing with no runners in scoring position , $p_0^\beta, p_{IG}^\beta, \cdots, p_9^\beta$ and $p_{10}^\beta$ can be defined on the substate space $S^\beta$.

### 2.1.3 Runner Advancing Rule

We assume the following advancing rule.

1. Sacrifices are not counted at all.
2. Errors are counted as outs.
3. Runners do not advance on outs.
4. All single and doubles are assumed to be long. That is, a single advances a baserunner two bases, and a double scores a runner from first base.
5. A player can steal only one base. That is, a player does not steal continuously both two base and three base, or both three base and home base.
6. The batter makes double play if he hits an infield grounder following states.
    1. no-outs, runner on first base(state 2) $\rightarrow$ two-outs, no-runners(state 17)
    2. no-outs, runner on first and second base(state 5) $\rightarrow$ two-outs, runner on first base(state 18)
    3. no-outs, runner on first and third base(state 6) $\rightarrow$ two-outs, runner on third base(state 20)
    4. no-outs, bases loaded(state 8) $\rightarrow$ two-outs, runner on second and third base(state 23)
    5. one-out, runner on first base(state 10) $\rightarrow$ three-outs (state 0)
    6. one-out, runner on first and second base(state 13) $\rightarrow$ three-outs(state 0)
    7. one-out, runner on second and third base(state 14) $\rightarrow$ three-outs(state 0)
    8. one-out, bases loaded(state 16) $\rightarrow$ three-outs(state 0)

## 2.2 Transition Matrix

The $25 \times 25$ block matrices, $P = (P_{ij}) = p(j|i), i, j = 0, 1, \cdots, 24$, for each players is derived by one step transition probability, $p(j|i), i, j = 0, 1, 2, \cdots, 24$, of the player and is described as follows.

$$P = \begin{bmatrix} 1 & 0 \\ T & Q \end{bmatrix}$$

According to the runner advancing rule, $T(8\times 8)$ and Q $(24\times 24)$can be determined. For example, an element in Q, $p(4|1)$, from the state $s=1$ (no-outs, no-runners) to the state $s=4$ (no-outs, runner on third base) is obtained by $p(4|1)=Pr$(double and steal third)$+Pr$(triple and no steal) $=p_6^\beta+p_8^\beta$. The other transition probabilities can be obtained in the same way.

## 2.3 IOERA Index

Let $R(j,i)$ denote the number of runs scored by a transition from the state $i$ to the state $j$. Let $E(i)$ denote the expected number of runs scored in an inning beginning in state $i$. Then $E(i)$ satisfies the following equation: for $i=1,\cdots,24$,

$$E(i)=\sum_{j=1}^{24}p(j|i)\{R(j,i)+E(j)\}. \tag{1}$$

The expected number of run scored in an inning is given by $E(1)$, corresponding to beginning the inning no-outs and no-men on base. Let $R(i)=\sum_{j=1}^{24}p(j|i)R(j,i)$ as the expected number of runs in state $i$, and let $R,E$ denote respectively

$$R=\begin{bmatrix} R(1)\\ \vdots \\ R(24)\end{bmatrix},\quad E=\begin{bmatrix} E(1)\\ \vdots \\ E(24)\end{bmatrix}.$$

Then equation (1) can be rewritten as $E=R+QE$. Therefore we have

$$E \;=\; (I-Q)^{-1}R. \tag{2}$$

Thus the expected number of runs per game is simply given by IOERA$=9E(1)$.

According to the runner advancing rule, the expected number of runs in state $i$, $R(i)$, can be obtained by the runner advancing rule.

## 2.4 Examples: Nippon Professional Baseball

Table 1 shows Top 5 players of IOERA for Central Leagues in 1999 season.

# 3 Improved Optimal Batting Order Model

In this section, we explain the IOBO model on the basis of the OBO model proposed by Bukiet, Harold and Palacios.

Table 1: Top 5 players of IOERA for Central Leagues in 1999 season

| Player | IOERA | DOERA | BASPE |
|---|---|---|---|
| Petagine | 9.8581 | 8.9510 | 0.9071 |
| Rose | 7.4050 | 7.8075 | -0.4025 |
| Matsui | 6.7474 | 6.5461 | 0.2013 |
| Ogata | 6.5943 | 6.3143 | 0.2800 |
| Etoh | 6.3641 | 6.4380 | -0.0739 |

## 3.1 The State Space, Probabilities of Batting and Stealing and Runner Advancing Rule

The state space, probabilities of batting and stealing and runner advancing rule are the same as IOERA index, provided that absorption state (the state of three-outs) is a state $s = 25$ for convenience sake.

## 3.2 Transition Matrix

The $25 \times 25$ block matrices, $P = (P_{ij}) = p(j|i), i, j = 1, 2, \cdots, 25$, for each players is derived by one step transition probability, $p(j|i), i, j = 1, 2, \cdots, 25$, of the player and is described as follows.

$$P = \begin{bmatrix} A & B & H_1 & 0 \\ 0 & A & B & H_2 \\ 0 & 0 & A & F \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where submatrices of P can be determined by the advancing rule. For example, an element in A, $p(3|4)$, from the state $s = 4$ (no-outs, runner on third base) to the state $s = 3$ (no-outs, runner on second base) is $p(3|4) = Pr$(single and steal second) $+Pr$(double and no steal) $= p_3^\alpha + p_5^\alpha$.

## 3.3 Formulation of The Runs Scored In an Inning

Let $1 \times 25$ row vectors $u_0$ be the

$$\begin{array}{cccccc} & 1 & 2 & \cdots & 25 & \\ u_0 = & [\ 1 & 0 & \cdots & 0\ ] & \end{array}$$

It describes the state of the beginning of the inning. Let $u_n$ be the state that $n$th player finished batting. For example, if $u_n$ is

$$\begin{array}{cccccc} & 1 & 2 & \cdots & 25 & \\ u_n = & [\ 0 & 1 & \cdots & 0\ ], & \end{array}$$

it means that the state is no-outs, runner on first base when $n$th batter finished batting. Suppose $P_{n+1}$ is the transition matrix of batting and stealing for $(n+1)$th batter, $u_{n+1} = u_n P_{n+1}$ describes the state that $(n+1)$th batter finished batting. In what follows, let $U_0$ be the matrix denoting score(column) and the state(row).

$$U_0 = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & \cdots & 25 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ 20 \end{array} & \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{bmatrix} \end{array}$$

We do not give consideration to the probability that the team scores more than twenty in an inning, since the maximum runs scored in Major League is eighteen (Chicago, 1883). Let $U_n$ be the matrix described the score and state that $n$th batter finished batting. $U_n$ is formulated as $U_{n+1}(\mathrm{row}j) = U_n(\mathrm{row}j)P0 + U_n(\mathrm{row}j-1)P1 + U_n(\mathrm{row}j-2)P2 + U_n(\mathrm{row}j-3)P3 + U_n(\mathrm{row}j-4)P4$. (3)

Here, $P0, P1, P2, P3$ and $P4$ denote the transition matrix that one batter scores zero runs, one run, two runs, three runs and four runs. i.e., it is decomposition of $P$ such that $P = P0 + P1 + P2 + P3 + P4$.

## 3.4 Algorithm for Calculation of $U_n$

We show how to calculate $U_n$ by six steps.

1. (Set $P^i$) Let $P^i$ be the transition matrix corresponding to $i$th batter. Decompose $P^i$ into transition matrices $P0^i, P1^i, \cdots$ and $P4^i$ such that $P^i$ $= P0^i + P1^i + P2^i + P3^i + P4^i$.

2. (Set $U_0$) Let $U_0$ be the matrix describing 0th batter finished batting.

$$U_0 = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & \cdots & 25 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ 20 \end{array} & \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{bmatrix} \end{array}$$

3. (Calculate $U_1$) Row $j$ of $U_1$ can be calculated by using (3) and we have

$$U_1 = \begin{bmatrix} U_1(\mathrm{row}1) \\ \vdots \\ U_1(\mathrm{row}21) \end{bmatrix}$$

Needless to say, $U_0(\text{row } 0) = U_0(\text{row } -1) = U_0(\text{row } -2) = U_0(\text{row } -3) = U_0(\text{row } -4) = 0$.

4. (Calculate $U_2$) In the same way as $U_1$, row $j$ of $U_2$ can be determined by the equation (3). Same as the step 3, the values of row $1, \cdots$, and row 21 of $U_2$ can be calculated and we have

$$U_2 = \begin{bmatrix} U_2(\text{row } 1) \\ \vdots \\ U_2(\text{row } 21) \end{bmatrix}$$

5. (Calculate $U_3, U_4, \cdots$) $U_3, U_4, \cdots$ can be calculated in the same way. Computation in an inning terminates when the total value of column 25, which corresponds to the probability of three outs, goes over a certain probability (see next section), and then the next inning start with $(j+1)$th batter.

6. (Calculate the Expected Run) Let the column 25 of $U_j$ be the

$$R(25) = \begin{array}{c} 0 \\ 1 \\ \vdots \\ 20 \end{array} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{20} \end{bmatrix},$$

then the expected run $r_i$ in the $i$th inning can be given by $r_i = 0x_0 + 1x_1 + 2x_2 + \cdots + 20x_{20}$. Therefore the expected run $R$ through one game can be given by $R = r_1 + r_2 + \cdots + r_9$.

## 3.5 The Condition of the End in an Inning

Let consider the condition of the end in an inning. The property of IOBO(OBO) is theoretically given by the following equality, provided a player batted in all nine position in the lineup.

The expected run for IOBO(OBO) = IOERA(OERA) index

By using this property, we calculated the expected run of OBO and the value of OERA using several probabilities of three-outs for Hideki Matsui in 1999 season and we choose 0.99999 as the condition. Because the probability 0.99999 seems to have a high degree of confidence, and the calculation takes very long CPU time when the probability is over 0.999999.

## 3.6 Examples: Nippon Professional Baseball

The evaluation of 9! (362800) possible orders takes very enormous time. So we unavoidably set nine conditions which down the number of orders to 7! orders. The conditions is as follows:(1) let eighth batter be the catcher (2) let ninth batter be the pitcher.

### 3.6.1 Dragons in 1999 season

We tested 7! batting order for OBO and IOBO for Dragons in 1999 season. This near-optimal orders based on the previous condition takes about 60 hours of CPU time to find the best order for one team. The program was written by Mathematica version 3.0.

Next table gives the best and worst order determined by OBO, MOBO and IOBO for Dragons in 1999 season.

Table 2: The best and worst order evaluated by IOBO for 1999 Dragons

| Order | IOBO best | IOBO worst |
|---|---|---|
| 1 | Lee | Inoue |
| 2 | Sekikawa | Tatsunami |
| 3 | Gomez | Fukudome |
| 4 | Yamasaki | Yamasaki |
| 5 | Fukudome | Gomez |
| 6 | Inoue | Lee |
| 7 | Tatsunami | Sekikawa |
| 8 | Nakamura | Nakamura |
| 9 | Yamamoto | Yamamoto |
| Expected Run | 5.84873 | 3.17344 |

Runs per game of Dragons was 4.4297 in 1999 season.

# 4 Remarks

1. Although we have to make the evaluation of 9!(362800) possible orders, we narrowed condition down to 7! orders for the reason of CPU time. Therefore there are possibility that there exists optimal order with higher expected run than our order in the rest of 362213 orders.

2. We assume that if the batter hits an infield grounder in a certain state, he makes double play. If we can include the exact number of double play in a season into the model, more realistic index will be obtained.

**Acknowledgements**

We really thank Dr. Naoto Miyoshi for his comment on the batting average with runners in scoring positions, which leads our model more realistic. We also thank Ms. Riyo Endo and Mr. Shinnosuke Tsuboi for their help of Mathematica programming.

# References


[1] K. Ano, Modified Offensive Earned-Run Average with Steal Effect for Baseball, to appear in *Applied Mathematics and Computation*, 1999.

[2] R. Bellman, Dynamic Programming and Markovian Decision Processes with Application to Baseball, in Optimal Strategies in Sports, S. P. Ladany and R. E. Macholb eds., Elsevier-North Holland, New York, 1977.

[3] 1998 Baseball Recordbook, Baseball Magazine Sha, Tokyo, 1997.

[4] 1998 Baseball Recordbook, Baseball Magazine Sha, Tokyo, 1998.

[5] B. Bukiet, E. R. Harold, and J. L. Palacios, A Markov Chain Approach to Baseball, *Operations Research*, **45**, No.1, 1997, 14-23.

[6] T. M. Cover and C. W. Keilers, An Offensive Earned-Run Average for Baseball, *Operations Research*, **25**, 729-740, 1977.

[7] http://sports.yahoo.co.jp/baseball/

[8] http://sports.yahoo.co.jp/baseball/

[9] H. M. Tailor and S. Karlin, An Introduction to Stochastic Modeling, 2nd Ed, Academic press, San Diego,1994.