野球における打者評価の計算手法について

Excel によるシミュレーション

齊　藤　善　弘

概　要

野球の試合を数理モデル化し、シミュレーションすることにより、試合における戦術の検討が試みられている。本稿では、打者評価指標である OERA を求める Excel ワークシートを作成し、その道具としての利用について述べる。さらに教育面での活用法にも触れる。

1．はじめに

スポーツの試合を数理モデル化し、試合のシミュレーションや試合における戦術等に対する比較・検討の研究が行われている[6]。野球についても、走・攻・守の面で多様な研究がなされている[1]。ここでは、攻撃力の評価、すなわち打者の評価に注目する。打者に対する評価法に D'Esopo-Lefkowitz のスコアリング・インデックス（scoring index）がある[5]。これは D'Esopo-Lefkowitz モデルとも呼ばれ、野球の状態推移をマルコフ連鎖（Markov chains）としてとらえることにより、1 イニングの攻撃における得点の分布を求めることができる。スコアリング・インデックスはその得点の期待値である。また、Cover-Keilers[4]は、スコアリング・インデックスを発展させ、OERA（Offensive Earned-Run Average）を提案した。OERA は、ある打者が 1 回から 9 回まで繰り返し打席に立ったとき、1 試合当たりの得点の期待値を計算するもので、打撃能力の総合的な指標として有用であることが知られている[9]。OERA については、文献[8]に詳しい紹介がある。また、D'Esopo-Lefkowitz モデルと異なる進塁規則を適用するため、OERA モデルと呼ぶことにする。本稿では、OERA を Excel によるシミュレーションで求める手法について述べる。OERA を求めるとき、数値計算法を適用して逆行列を求める必要がある。数値計算プログラムは Fortran または C で記述するのが一般的であるが、Microsoft Excel を用いた。理由は、Excel がパソコンの標準的なソフトウェアになっていること、また、操作がしやすく、出力されたデータのグラフ化や加工が容易なことである。作成したワークシートは打者評価の道具として使え、その利用について述べる。さらに教育面での活用法にも触れる。

2．OERA モデル

本節では、T. M. Cover と C. W. Keilers が提案した OERA モデルについて解説しよう。まず、表１に示すように、イニング内の状態を１から25までの数字で定義する。

表１　アウト数とランナーの状態

		ランナーの状態							
		なし	一塁	二塁	三塁	一二塁	一三塁	二三塁	満塁
アウト数	0	1	2	3	4	5	6	7	8
	1	9	10	11	12	13	14	15	16
	2	17	18	19	20	21	22	23	24
	3	25							

OERA モデルでは、打撃結果と進塁の規則をつぎのように定めている。

a）犠打はすべて計算されない。
b）エラーはアウトとして計算される。
c）アウトによってランナーは進塁しない。
d）単打は一塁ランナーを三塁に進塁させ、二塁ランナーと三塁ランナーを生還させる。
e）二塁打と三塁打は一塁ランナー、二塁ランナー、三塁ランナーすべてを生還させる。
f）併殺はない。

打撃結果を０（アウト）、Ｂ（四死球）、１（単打）、２（二塁打）、３（三塁打）、４（本塁打）の６種類とする。ある一人の打者に対して、これらの事象の起こる確率をそれぞれ P_0、P_B、P_1、P_2、P_3、P_4 と表し、次式で計算する。

$$P_0=\frac{\mathbf{OUT}}{\mathbf{AB}+\mathbf{BB}},\ P_B=\frac{\mathbf{BB}}{\mathbf{AB}+\mathbf{BB}},\ P_1=\frac{\mathbf{1B}}{\mathbf{AB}+\mathbf{BB}},$$
$$P_2=\frac{\mathbf{2B}}{\mathbf{AB}+\mathbf{BB}},\ P_3=\frac{\mathbf{3B}}{\mathbf{AB}+\mathbf{BB}},\ P_4=\frac{\mathbf{HR}}{\mathbf{AB}+\mathbf{BB}} \tag{1}$$

ここで、**AB**：打数、**BB**：四死球数、**OUT**：アウト数、**1B**：単打数、**2B**：二塁打数、**3B**：三塁打数、**HR**：本塁打数である。

打撃結果前の状態 i の確率を s_i、その後の状態 i の確率を s'_i で表すと、上で定めた規則に従い、打撃の結果により状態確率ベクトル $\boldsymbol{s}=\left[s_1, s_2, \cdots, s_{25}\right]$ から新しい状態確率ベクトル $\boldsymbol{s}'=\left[s'_1, s'_2, \cdots, s'_{25}\right]$ に移る。つまり、野球ゲームをマルコフ連鎖とみることがで

き、25次の推移行列 $\boldsymbol{P}$ を用いて、

$$\boldsymbol{s}'=\boldsymbol{sP}$$

と数学的な定式化ができる。推移行列 $\boldsymbol{P}$ は、打撃の結果に対する確率 P_0、P_B、P_1、P_2、P_3、P_4 で定まり、

$$\boldsymbol{P}=\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & 0 & 0 \\ 0 & \boldsymbol{A} & \boldsymbol{B} & 0 \\ 0 & 0 & \boldsymbol{A} & \boldsymbol{f} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

で与えられる。ここで 8 次正方行列 $\boldsymbol{A}$ と $\boldsymbol{B}$ は

$$\boldsymbol{A}=\begin{bmatrix} P_4 & P_1+P_B & P_2 & P_3 & 0 & 0 & 0 & 0 \\ P_4 & 0 & P_2 & P_3 & P_B & P_1 & 0 & 0 \\ P_4 & P_1 & P_2 & P_3 & P_B & 0 & 0 & 0 \\ P_4 & P_1 & P_2 & P_3 & 0 & P_B & 0 & 0 \\ P_4 & 0 & P_2 & P_3 & 0 & P_1 & 0 & P_B \\ P_4 & 0 & P_2 & P_3 & 0 & P_1 & 0 & P_B \\ P_4 & P_1 & P_2 & P_3 & 0 & 0 & 0 & P_B \\ P_4 & 0 & P_2 & P_3 & 0 & P_1 & 0 & P_B \end{bmatrix},\quad \boldsymbol{B}=P_0\boldsymbol{I} \qquad (2)$$

である（$\boldsymbol{I}$ は単位行列）。また、$\boldsymbol{f}$ は 8 次のベクトル $\boldsymbol{f}=\left[P_0, P_0, \cdots, P_0\right]^T$。

吸収マルコフ連鎖（absorbing Markov chains）の理論により、つぎの24次正方行列

$$\boldsymbol{Q}=\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & 0 \\ 0 & \boldsymbol{A} & \boldsymbol{B} \\ 0 & 0 & \boldsymbol{A} \end{bmatrix} \qquad (3)$$

を調べればよく[2]、基本行列（fundamental matrix）と呼ばれる $(\boldsymbol{I}-\boldsymbol{Q})^{-1}$ を利用する。

$$\boldsymbol{I}+\boldsymbol{Q}+\boldsymbol{Q}^2+\cdots\cdots=(\boldsymbol{I}-\boldsymbol{Q})^{-1}$$

より、基本行列の（i, j）成分は、状態 i から状態 j へ移るとき、訪れる回数の期待値を表すことがわかる。よって、ある打者がノーアウトランナーなし（状態１）から始まって、各状態を訪ねる回数の期待値がわかれば、１イニングの得点の期待値がわかる。つまり、$(\boldsymbol{I}-\boldsymbol{Q})^{-1}$ の第１行目にその情報があることになる。状態 j における得点の期待値は、進

塁の規則により定まり、24次のベクトル

$$\boldsymbol{r}=\begin{bmatrix}\boldsymbol{r}_1\\ \boldsymbol{r}_1\\ \boldsymbol{r}_1\end{bmatrix} \tag{4}$$

の第 j 成分となる。ここでベクトル $\boldsymbol{r}_1$ は 8 次のベクトルで、

$$\boldsymbol{r}_1=\begin{bmatrix} P_4 \\ 2P_4+P_3+P_2 \\ 2P_4+P_3+P_2+P_1 \\ 2P_4+P_3+P_2+P_1 \\ 3P_4+2P_3+2P_2+P_1 \\ 3P_4+2P_3+2P_2+P_1 \\ 3P_4+2P_3+2P_2+P_1 \\ 4P_4+3P_3+3P_2+2P_1+P_B \end{bmatrix} \tag{5}$$

あるイニングにおける状態 $\boldsymbol{s}$ からの得点の期待値は、

$$\boldsymbol{e}=(\boldsymbol{I}-\boldsymbol{Q})^{-1}\boldsymbol{r} \tag{6}$$

となるから、状態 1 から始まる 1 イニングの得点の期待値はベクトル $\boldsymbol{e}$ の第 1 成分 $\boldsymbol{e}(1)$ となる。よって、ある 1 打者の 1 試合あたりの得点の期待値を表す OERA の値は

$$\mathrm{OERA}=9\boldsymbol{e}(1) \tag{7}$$

となる。

3．OERA のシミュレーション

OERA モデルの 1 イニングの得点の分布を求めるには、基本行列 $(\boldsymbol{I}-\boldsymbol{Q})^{-1}$、すなわち 24次正方行列 $\boldsymbol{I}-\boldsymbol{Q}$ の逆行列を計算する必要がある。OERA を求める手順はつぎのようになる。

1）P_0、P_B、P_1、P_2、P_3、P_4 の計算（式(1)）

2）行列 $\boldsymbol{I}-\boldsymbol{Q}$ の計算（式(2)、(3)）

3） $\boldsymbol{I}-\boldsymbol{Q}$ の逆行列の計算

4） $\boldsymbol{e}(1)$（$(\boldsymbol{I}-\boldsymbol{Q})^{-1}\boldsymbol{r}$ の第 1 成分）の計算（式(4)、(5)、(6)）

5）OERA の計算（式(7)）

Excel の入力フォームと出力結果を図 1 に示す。

図 1　Excel の入力フォームとシミュレーション結果

C3から C9のセルに打数、四死球数、アウト数、単打数、二塁打数、三塁打数、本塁打数を入力すれば、確率が C12から C18のセルに出力される。図 1 の例では、打数=105、四

死球数=18、アウト数=69、単打数=28、二塁打数=1、三塁打数=1、本塁打数=6とした[4]。確率の値は、$P_0=0.561$、$P_B=0.146$、$P_1=0.228$、$P_2=0.008$、$P_3=0.008$、$P_4=0.049$とわかる。これらの確率から$\boldsymbol{I}-\boldsymbol{Q}$を求め、逆行列$(\boldsymbol{I}-\boldsymbol{Q})^{-1}$は、つぎの線型方程式を解くことで得られる。

$$(\boldsymbol{I}-\boldsymbol{Q})\boldsymbol{X}=\boldsymbol{I}$$

数値解法にはガウスの消去法を選択し、計算アルゴリズムは、文献[7]を参考に VBA（Visual Basic for Applications）で記述した。実行は計算ボタンをクリックすることでなされ、C25と C26のセルに$\boldsymbol{e}(1)$と OERA$=9\boldsymbol{e}(1)$の結果が各々出力される。ここで$\boldsymbol{e}(1)$の計算には MMULT 関数を使用した。

4．おわりに

Excel のシミュレーションにより OERA を求めるワークシートを作成した。このワークシートは、打者の打撃結果（打数、四死球数、アウト数、単打数、二塁打数、三塁打数、本塁打数）を所定のセルに入力し、計算ボタンをクリックするだけで OERA を簡単に求めることができる。よって、打撃能力評価の道具として利用できる。たとえば、日本野球機構公式サイト（http://www.npb.or.jp/）から、プロ野球選手のデータを入力して OERA を求め、打撃能力の評価・比較ができよう。また、このワークシートには教育的な配慮を施している。二つ例をあげると、

ア）不要と思われる途中のデータ（行列$\boldsymbol{I}-\boldsymbol{Q}$の成分など）をセルに残している。
イ）逆行列を求めるのに、MINVERSE 関数を使わず、マクロ機能を使用した。

ア)については、OERA モデルの進塁規則を変えることにより、途中のデータの変化が見られる。また、計算手法の計算式も陽に見ることができる。イ)については、MINVERSE 関数の使用は、操作が難しいこと、完全なブラックボックスとなるためその計算手法が見えない、など難点がある。しかし VBA を使えば、計算手法が明らかとなり、数値計算の理解を促すことができる。また、ガウスの消去法以外の方法、たとえば、ガウス・ジョルダン法やガウス・ザイデル法などの数値解法の選択が可能である。さらに、文献[3]で見るように、計算コストを減らす工夫も施すことができる。

本稿では、Excel を使用したが、教育上の利便性を考えると Web 上でのツールが望まれる。たとえば Perl や Java によるシミュレーション・ツールの作成があげられる。ま

た、文献[6]および[9]のように、OERA モデルや D'Esopo-Lefkowitz モデルの拡張やその応用を考えることで、盗塁や犠打などを含めた進塁規則に基づくモデル、一人の選手の評価ではなくチーム全体の評価、打者の選定や最適打順等、興味深い課題が多い。

参考文献

[1] アルバート, J.、ベネット、J.著、加藤貴昭訳、『メジャーリーグの数理科学　上・下』、シュプリンガー・フェアラーク東京、2004

[2] Brémaud, P., 1999, *Markov Chains－Gibbs Fields, Monte Carlo simulation, and queues*, Springer

[3] Bukiet, B., Harold, E. and J. L. Palacios, 1997, "A Markov chain approach to baseball", *Operations Research*, Vol. 45, 14－23

[4] Cover, T. M. and C. W. Kelers, 1977, "An offensive earned-run average for baseball", *Operations Research*, Vol. 25, 729－740

[5] D'Esopo, D.A. and B. Lefkowitz, "The distribution of runs in the game of baseball", in "Optimal Strategies in Sports" 1977, Ladany S.P. and Machol, R.E. eds, North-Holland, 55－62

[6] 廣津信義、宮地力、「野球チームのラインナップ選定のための数理的一手法－日本代表チームの選定を例として」、オペレーションズ・リサーチ、49、2004、380－389

[7] 河村哲也、菅牧子、『エクセル数値計算入門』、インデックス出版・山海堂、2003

[8] 木下栄蔵、『野球に勝てる数学－数字から見た勝つための条件』、電気書院、1992

[9] 竹内啓、藤野和建、『スポーツの数理科学』、共立出版、1988

[10] 武井貴裕、瀬古進、穴太克則、「野球の最適打順を考えてみよう」、オペレーションズ・リサーチ、47、2002、142－147