

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。

- 犠打はすべて計算されない。
- エラーはアウトとして計算される。
- アウトによってランナーは進塁しない。
- 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。
- 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。
- 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。

$$\begin{aligned}
 P_0 &= \frac{\text{OUT}}{\text{AB} + \text{BB}}, & P_B &= \frac{\text{BB}}{\text{AB} + \text{BB}}, & P_1 &= \frac{\text{1B}}{\text{AB} + \text{BB}}, \\
 P_2 &= \frac{\text{2B}}{\text{AB} + \text{BB}}, & P_3 &= \frac{\text{3B}}{\text{AB} + \text{BB}}, & P_4 &= \frac{\text{HR}}{\text{AB} + \text{BB}}
 \end{aligned}
 \tag{1}$$

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 $\mathbf{s} = [s_1, s_2, \dots, s_{25}]$ から新しい状態確率ベクトル $\mathbf{s}' = [s'_1, s'_2, \dots, s'_{25}]$ 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 \mathbf{P} を用いて、

$$\mathbf{s}' = \mathbf{s}\mathbf{P}$$

と数学的な定式化ができる。推移行列 \mathbf{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、

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

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

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

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

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

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

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

$$\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots = (\mathbf{I} - \mathbf{Q})^{-1}$$

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、 $(\mathbf{I} - \mathbf{Q})^{-1}$ 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

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_1 \\ r_1 \end{bmatrix} \quad (4)$$

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

$$\mathbf{r}_1 = \begin{bmatrix} P_4 \\ 2P_4 + P_3 + P_2 \\ 2P_4 + P_3 + P_2 + P_1 \\ 2P_4 + P_3 + P_2 + P_1 \\ 3P_4 + 2P_3 + 2P_2 + P_1 \\ 3P_4 + 2P_3 + 2P_2 + P_1 \\ 3P_4 + 2P_3 + 2P_2 + P_1 \\ 4P_4 + 3P_3 + 3P_2 + 2P_1 + P_B \end{bmatrix} \quad (5)$$

あるイニングにおける状態 \mathbf{s} からの得点の期待値は、

$$\mathbf{e} = (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{r} \quad (6)$$

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 \mathbf{e} の第 1 成分 $e(1)$ 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

$$\text{OERA} = 9e(1) \quad (7)$$

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 $(\mathbf{I} - \mathbf{Q})^{-1}$ 、すなわち 24 次正方行列 $\mathbf{I} - \mathbf{Q}$ 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 $I-Q$ の計算 (式(2)、(3))
- 3) $I-Q$ の逆行列の計算
- 4) $e(1)$ ($(I-Q)^{-1}r$ 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。

	A	B	C	D
1				
2				
3		打数	105	
4		四死球数	18	
5		アウト数	69	
6		単打数	28	
7		二塁打数	1	
8		三塁打数	1	
9		本塁打数	6	
10				
11				
12		P0	0.560976	
13		PB	0.146341	
14		P1	0.227642	
15		P2	0.00813	
16		P3	0.00813	
17		P4	0.04878	
18				
19				
20				
21		計算		
22				
23				
24				
25		E(1)	1.082093	
26		OERA	9.738833	
27				

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 $I-Q$ を求め、逆行列 $(I-Q)^{-1}$ は、つぎの線型方程式を解くことで得られる。

$$(I-Q)X=I$$

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 $e(1)$ と $OERA=9e(1)$ の結果が各々出力される。ここで $e(1)$ 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 $I-Q$ 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]で見られるように、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. Kellers, 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

