ゲーム理論を用いたプロ野球最適化研究

M2005MM023 太田 雄大
指導教員　鈴木 敦夫

1　はじめに

スポーツの中でも野球は特に数学的な分析に適している. サッカーやアイスホッケー, バスケットボール, その他のプレイヤーが常に動いてるスポーツとは違い, プレイが区切られていて特定できる. また昨今の情報技術の進歩により, データを集めることも容易になり, 学術的な視点から野球のデータが分析され始めている. 現場において, 統計データは戦術を立てるために用いれているとともに, 選手の育成, チームの構成においても役に立っている. 野球の盛んなアメリカ合衆国では野球学会 SABR（野球のデータ分析法 "sabermetrics" の語源）があり野球に関するさまざまな統計的解析が行われている [4].

しかしゲーム理論を用いて細かいシチュエーションで最適化を計ることは今までなされていない. ピッチャーがランナー状況, アウト数, カウント数が異なる場面でどのコースにどの球種のボールを投げるのが最適であるかを求めることは OR のテーマとしても非常に興味深い. 本研究ではピッチャーとバッターそれぞれにとって有効な戦略を与えるモデルをつくり, 実際のデータを用いてその戦略を検証する.

2　データについて

本研究で用いたデータはプロ野球の中日ドラゴンズの川上憲伸投手と山本昌投手の 2003 年から 2007 年 6 月までの投球データをデータスタジアム社がリアルタイムで収集し, 電子化したものである. 内容は以下のようである.

・1 球ごとの打席結果詳細
・ピッチャの投げたボールの詳細（速度、球種、コースなど）
川上投手 (10137 球　打席数 2581)
山本投手 (7225 球　打席数 1825)

3　問題の設定

ここではゲーム理論を用いたモデルを作成し, 野球のシチュエーション別の最適な戦略を求める方法を提案する. ゲームにおけるプレイヤーをピッチャーとバッターとしてそれぞれの戦略を考え, 個々の戦略をどんな確率で行うのが最適か（混合戦略）を求める. ピッチャーの戦略はどのコースにどの球種を投げるか, バッターの戦略はスイングするかしないかとする. バッターの打席の結果を細かく分けその利得を最小 2 乗法により推定する. ピッチャーの利得は, 2 人零和ゲームなので, (−1) ×（バッターの利得）になる. ピッチャーについては期待利得が最小になるような, バッターについては期待利得が最大になるような混合戦略を線形計画法を用いて求める.

3.1　仮定と記号

本論文でおいた仮定は次のとおりである.

(1) ピッチャーがボールを投げるコースはキャッチャーの視点で 1(外角高め)、2(真中高め)、3(内角高め)、4(外角真中)、5(真中真中)、6(内角真中)、7（外角低め)、8（真中低め)、9（内角低め）とする

また, 以下では次のような記号を用いる.
T : アウトカウント数　T=0,1,2
B : ランナー状況　　B={0,1,2,3,12,13,23,123}
r : 状況が (T,B) の時そのイニング内で記録した得点
$P(r|T,B)$: 状況が (T,B) の時そのイニング内で r 得点する確率
$E(T,B)$: 状況が (T,B) の時のバッターの期待値
i : 結果の添え字集合
1 打席内での結果の回数を N_i, 結果に対するそれぞれの利得を G_i とする. ホームラン, 犠牲フライ, 安打, 四球・死球, 犠打はそれぞれの結果により状態が (T,B) になった時にもたらされる利得を $G_i(T,B)$, その結果後に期待できる得点の期待値を $g_i(T,B)$ とする.
$i=1$: ファール
$i=2$: 見送り、空振り
$i=3$: ボール
$i=4$: 凡打
$i=5$: 三振
$i=6$: 得点圏での凡打
$i=7$: 得点圏での三振
$i=8$: 併殺打
$i=9$: ホームラン
$i=10$: 犠牲フライ
$i=11$: 安打
$i=12$: 四球・死球
$i=13$: 儀打
j : あるピッチャーが対戦した打席 ID　(j=1,2,3,$\cdots$,n)
k : $G_i(T,B)$ の値を決定する回帰モデルにおけるパラメータ
c : ピッチャーが投げるボールのコース（c=1,2,3,$\cdots$,9)
b : バッターの戦略 1:スイングする　2：スイングしない　(b=1,2)
p : ピッチャーの戦略 (コース c に球種 s を投げる）
x_p : ピッチャーが戦略 p を選択する確率
y_b : バッターが戦略 b を選択する確率

c_{pb}： ピッチャーが戦略 p を選択しバッターが戦略 b を選択した場合の一球あたりの平均利得
（打点がある場合, 利得に加算）

4 利得の推定

ゲーム理論を用いてピッチャー, バッターの戦略を最適化するために一球ごとの結果の利得を推定する必要がある. 打席で期待される利得をデータから計算する実測値と結果の利得パラメータから推測する利得との残差の平方和が最小になるような利得のパラメータを求める. 推定する利得は上記に仮定した G_i である. 結果球だけで考えればファール, 空振り, 見送り, ボールなどは省けるがこれらが全体の７割を占めていることやボールカウントが先行した場合, ストライクカウントが先行した場合とでは大きく出塁率が異なるためこれらの利得も推定する.

また安打, ホームラン, 犠飛, 四・死球による結果はアウト数, ランナー状況によってその結果の後に得点できる確率に差がある. 具体的にはノーアウトからランナーが 2 塁に出るのと 2 アウトから出るのでは結果の後に期待できる得点の期待値 $g_i(T,B)$ は異なる. そこで $G(T,B)$ に $g(T,B)$ の値を反映させる.

結果により状況が (T,B) になった後の打席での期待利得は

$$g_i(T,B) = \sum_{r=1}^{\infty} rP(r|T,B)(i=9,\cdots,13) \qquad (1)$$

利得は

$$G_i(T,B) = kg_i(T,B)(i=9,\cdots,13) \qquad (2)$$

で計算する.
状況 (T,B) における打席での期待利得の実測値は

$$y(T,B) = E(T,B) = \sum_{r=1}^{\infty} rP(r|T,B) \qquad (3)$$

となり [3], パラメータにより推測される期待利得の予測値は

$$\widehat{y}_j(T,B) = \sum_{i=1}^{8} G_i N_{ij} + \sum_{i=9}^{13} G_i(T,B)N_{ij} \qquad (4)$$

となる. 残差は

$$e_j(T,B) = y_j(T,B) - \widehat{y}_j(T,B) \qquad (5)$$

となり残差平方和は

$$S = \sum_{j=1}^{n} \{e_j(T,B)\}^2 \qquad (6)$$

となりこれが最小になる利得パラメータ G_i を推定する. 制約条件を満たし残差平方和を最小にするそれぞれの利得のパラメータを川上投手の対戦データ（打席総数 2581）で Excel のソルバーで解いた結果は以下ように計算された.(G_i(T,B) の結果は表 3)

G_1=0.00669 (ファール)
G_2=0.00006 (見送り、空振り)
G_3=0.02600 (ボール)
G_4=0.24826 (凡打)
G_5=0.22780 (三振)
G_6=0.22436 (得点圏での凡打)
G_7=0.21687 (得点圏での三振)
G_8=0.32964 (併殺打)
k=1.03228

表 1, 表 2 はそれぞれ状況 (T,B) の時の期待値 $E(T,B)$, 結果によって (T,B) になった後のイニングでの得点期待値 $g_i(T,B)$ の計算結果である.
また標準偏差 s_x=0.0562, s_y=0.0400, 共分散 s_{xy}=0.0019 さらに相関係数 r_{xy}=0.8525, 決定係数 R^2=0.7268 となり最小 2 乗法による当てはめは良い.

表 1 よりノーアウト 2, 3 塁のときが最も得点が期待できるが, 2 アウトランナーなしからは最も低い. 表 3 より最も利得が大きいのは四球によりノーアウトフルベースになっとときであり, 川上投手にとって最も失点する可能性が高いことがわかる.

表 1 状況が (T,B) の時の期待値 $E(T,B)$(川上投手)

T,B	E(T,B)
0,0	0.357
1,0	0.195
2,0	0.080
0,1	0.723
1,1	0.467
2,1	0.232
0,12	1.071
1,12	0.611
2,12	0.536
0,13	1.400
1,13	0.782
2,13	0.206
0,2	1.088
1,2	0.343
2,2	0.315
0,23	1.800
1,23	1.375
2,23	0.667
0,123	0.714
1,123	1.333
2,123	0.840
0,3	1.500
1,3	0.428
2,3	0.461

表 2　結果によって (T,B) になった後の得点期待値 $g_i(T,B)$（川上投手）

T,B	安打	四・死球	犠打	犠飛	本塁打
0,0					0.423
1,0				0.000	0.080
2,0				0.000	
0,1	0.711	0.263			
1,1	0.404	0.500		0.000	
2,1	0.298	0.266		0.000	
0,12	1.200	1.166			
1,12	0.684	0.222		0.000	
2,12	0.629	0.333			
0,13	1.692				
1,13	0.888	0.500			
2,13	0.666	1.000			
0,2	1.500				
1,2	0.360		0.571		
2,2	0.642		0.416	0.000	
0,23	1.666				
1,23	3.000				
2,23	1.750				
0,F	0.333	3.500			
1,F	3.000	1.400			
2,F	0.400	1.000			
0,3	2.000				
1,3	0.250				
2,3	1.0008				

表 3　推定された利得 $G_i(T,B)$ のパラメータ（川上投手）

T,B	安打	四・死球	犠打	犠飛	本塁打
0,0					0.485
1,0					0.090
2,0					
0,1	0.734	0.271			
1,1	0.417	0.516			
2,1	0.308	0.275			
0,12	1.238	1.204			
1,12	0.706	0.229			
2,12	0.649	0.344			
0,13	1.746				
1,13	0.917	0.516			
2,13	0.688	1.032			
0,2	1.548				
1,2	0.258		0.589		
2,2	0.663		0.430		
0,23	1.720				
1,23	3.096				
2,23	1.806				
0,F	0.344	3.612			
1,F	3.096	1.445			
2,F	0.412	1.032			
0,3	2.064				
1,3	3.096				
2,3	1.032				

5　ゲーム理論を用いたモデル

ゲーム理論によりピッチャーとバッターの 2 人零和ゲームを考える．ピッチャーががどのコースどの球種をにどんな割合で投げるのが最適か、バッターはスイングする、しないのどちらが最適かを混合戦略を求める．この計算のために次のようなモデル化をした．
ピッチャーの期待利得を

$$u = \max_{b \in \{1,2\}} \sum_{p=1}^{54} c_{pb} x_p \tag{7}$$

バッターの期待利得を

$$v = \min_{p \in \{1,\cdots,54\}} \sum_{b=1}^{2} c_{pb} y_b \tag{8}$$

とする．また

$$U = \min_{x_p, p=1,\cdots,54} \{u(x_1, x_2, \cdots, x_{54})\} \tag{9}$$

$$V = \max_{y_b, b=1,2} \{v(y_1, y_2)\} \tag{10}$$

とするとミニマックス定理より $U = V$ となるような最適な混合戦略 x_p と y_b が存在する [2]
ピッチャーにとっての最適な混合戦略を求める線形計画法は次のようになる．

$$\begin{aligned} &\min \quad u \\ &\text{制約条件}: \sum_{p=1}^{54} c_{pb} x_p \leq u \ (b = 1, 2) \\ &\sum_{p=1}^{54} x_p = 1 \\ &x_p \geq 0, \ p = 1, 2, \cdots, 54 \end{aligned}$$

またバッターにとっての最適な混合戦略を求める線形計画法は次のようになる．

$$\begin{aligned} &\max \quad v \\ &\text{制約条件}: \sum_{b=1}^{2} c_{pb} y_b \geq v \ (p = 1, 2, \cdots, 54) \\ &\sum_{b=1}^{2} y_b = 1 \\ &y_j \geq 0, \ b = 1, 2 \end{aligned}$$

6　利き腕別、カウント別の結果

ピッチャーの対戦打者には大きく分けて 2 つのタイプがある. 右打ち打者と左打ち打者である．一般に右投げピッチャーは左打ち打者に弱く, 左投げピッチャーは右打ち打者に弱い．ピッチャーとバッターの利き腕が異なるのと同じでは結果が全く異なる．そこで結果は打者の利き腕別，カウント別に検証する必要ある．まずノーストライクノーボールで右打者に対しての利得行列を川上投手のデータから計算し利得表の線形計画問題を Excel のソルバーで解く (表 4) と, y_1=0.069377, y_2=0.930623, x_7=0.791869, x_8=0.208131 となりゲームの値は U=V=0.004864845 である．

よって川上投手はノーストライクノーボールの時に右打者に対してコース 2(中央高め) にスライダーを 79 %, コース 3（内角高め）にカーブを 21 %の確率で投げるのが最適だという結果になった．また右打者のバッターは 93 %振る, 7 %振らないが最適解となった. 他のカウントや山本投手についても同様に検証した．

表 4　利得行列と結果　　カウント (0,0) 右打者

振らない y2	振る y1			y
				1
0.069377	0.930623	変数		
0.015155	0.424695	0	x1	1カーブ
0.014752	0.176705	0	x2	1カット
0.01622	0.114626	0	x3	1スト
0.012362	0.491474	0	x4	1スラ
0.010408	0.350865	0	x5	2カット
0.000634	0.190804	0	x6	2スト
0.000634	0.00518	0.791869	x7	2スラ
0.020964	0.003665	0.208131	x8	3カーブ
0.019693	0.172943	0	x9	3カット
0.020522	0.248098	0	x10	3シュート
0.02344	0.16899	0	x11	3スト
0.000634	0.126469	0	x12	4カーブ
0.007564	0.107583	0	x13	4カット
0.006613	0.173257	0	x14	4スト
0.006987	0.423234	0	x15	4スラ
0.000634	1.503348	0	x16	5カーブ
0.000634	0.514887	0	x17	5スト
0.018312	0.158946	0	x18	6シュート
0.009874	0.150771	0	x19	6スト
0.020833	0.25076	0	x20	7カーブ
0.018059	0.042725	0	x21	7カット
0.015714	0.288895	0	x22	7スト
0.017936	0.146848	0	x23	7スラ
0.026046	0.135415	0	x24	7フォーク
0.01334	0.824208	0	x25	8シュート
0.01697	0.198182	0	x26	8スト
0.019693	0.230762	0	x27	8フォーク
0.011928	0.150424	0	x28	9シュート
0.013892	0.006696	0	x29	9スト

7　特定のバッター

上記の結果はバッターの利き腕別にデータから求めた解であるが, 期待利得の回帰モデルにおいて実測値と予測値が平均して大きく異なるなる回数の多い選手も多く見られた．そこでバッター個別に 1 打席あたりの実測値と予測値の差 (残差）の平均の絶対値を計算し標準偏差と比較して外れ値を多くとる選手を調べる．以下にこれに該当する一部の選手を挙げる．こういった選手には上記の最適解を実行するのに注意が必要である．

川上投手
・鳥谷　敬 (左打ち)　　　阪神タイガース
・新井　貴浩 (右打ち)　　阪神タイガース
・ラミレス (右打ち)　　　読売ジャイアンツ
・田中　浩康 (右打ち)　　東京ヤクルトスワローズ
・大島　裕行 (左打ち)　　西武ライオンズ
山本投手
・廣瀬　純 (右打ち)　　　広島東洋カープ
・脇谷　亮太 (左打ち)　　読売ジャイアンツ
・西岡　剛 (右打ち)　　　千葉ロッテマリーンズ
・金子　誠 (右打ち)　　　北海道日本ハムファイターズ

8　個人対個人の最適戦略

個々の対戦となるとデータ数が限られているのですべての選手に対し最適戦略を求めるのは難しいがデータ数の一番多かった阪神タイガースの左打ち打者の赤星憲広選手と川上選手の個々の対戦での最適化を検証した結果, 川上投手はカウント (0,0) において外角高めにストレートを 54.9 %, 内角低めにカットボールを 37 %の確率で投げるのが最適で，赤星選手はスイングするが 65 %，しないほうがいいが 35 %となった．

9　まとめ

プロ野球のピッチャー, バッターの対戦において, ピッチャーがカウント別にどのコースを狙って投げればバッターを抑えることができる可能性が高いかを検証した結果，川上投手はカウントにより異なるがストレートを外角に狙い，山本投手は内角低めの変化球という最適戦略が多く見られた．利き腕別の結果を比較すると全く異なる戦略となり，キャッチャーはバッターの利き腕別にリードを変える必要がある．またすべてのランナー状況についてはデータ不足から検証できなかったが得点圏と得点圏外ではバッターの打率は大きく異なるのでこの 2 パターンは分けて検証する必要があった．個々の対戦を見ても選手によって不得意なコースがあることがわかった. 現在のデータでは，選手個々の対戦はデータ不足から限られた選手のみ検証できるが, 今後データを蓄積し，研究を進めていくと実用が可能になるであろう．

10　おわりに

本研究では, プロ野球においてピッチャー，バッターの最適な戦略をゲーム理論の基礎的な手法を用いて求めた．バッターの利き腕別のピッチャの最適戦略は, 一応妥当のようである．さらにモデルの精度を高めるには，より多くのデータが必要である．またカウント別のみ最適戦略を検証し，個々の対戦は今回 1 パターンだけの試みであったが，今後のデータ蓄積を待って解析すればランナー状況，アウトカウントなども考慮でき，大変に興味深く，また実戦でも有効であろう．

本研究によってプロ野球のバッターとピッチャーの対戦に関心を持ち, 実際の試合での対戦を観て楽しんで頂けたら幸いである．本研究の結果を活用し，日本野球会の水準向上に役立てることができればと野球ファンかつＯＲ研究者として望むものである．

参考文献

[1] データスタジアム株式会社作成：川上投手、山本投手の投球データ.(2003 年 4 月-2007 年 6 月)

[2] 小和田正，澤木勝茂，加藤豊：OR 入門 意思決定の基礎，実教出版, 1984, 東京.
第 5 章, 決定分析とゲームの理論, pp.104-111.

[3] ShaulP.Ladany, RobertE.Machol :Optimal Strategies in Sports, North-Holland Publishing Company, 1977, U.S.A, pp.1-30.

[4] J. アルバート, J. ベネット:メジャーリーグの数理科学 (上), シュプリンガー・フェアラーク東京株式会社, 2004, 東京.