プロ野球の数理科学

鳥越　規央

メジャーリーグでは，セイバーメトリクスというプロ野球のデータ解析手法によって生み出された指標に基づいて選手を評価している．こうした動きは日本のプロ野球においても徐々に浸透している．近年はプロ野球データも1球ごとに細かく収集されているため，より精度の高い分析手法の開発も期待できる．本報告では，アウトカウント，塁状況別シチュエーションにおける得点差別の勝率確率の算出方法を解説する．これによりWPA（Win Probability Added）と呼ばれる，選手の貢献度を示す指標を用いて選手評価を行うことができる．さらには，ある打順が1試合当たりに得る平均得点の算出方法を紹介し，より多くの得点を得た打順にどのような傾向が見られるかを考察した．

キーワード：セイバーメトリクス，マルコフ連鎖，Win Probability Added，最適打順

1.　はじめに

1970年代にアメリカの野球ライターのビル・ジェームスが提唱したといわれているセイバーメトリクスとは，野球というスポーツを数値データによって客観的，科学的に分析しようとする研究分野のことである[3]．彼は統計学的に野球を分析し，バントや盗塁の効力を否定するなどの衝撃的な研究結果を発表していった．最初はアメリカでもこの理論は受け入れ難いものであった．しかしながら地道な研究を続けた結果，それらの分析に共感した野球ファンからの支持を得ることになる．球団経営でもセイバーメトリクスの理論が使われるようになったのは，80年代から90年代にかけてオークランド・アスレチックスが財政状況の厳しい中で行ったチーム運営戦略が功を奏したことを記した書籍「Moneyball」[4]の出版がきっかけとされている．

野球というスポーツを数理科学の対象として扱った最初の研究論文は1959年にOperations Researchに掲載されたLindseyによるものとされている[3]．この論文では，イニング別の得点期待値や，得点確率のモデルを構築している．その後，多種多様な職種の研究者によって多くの論文や書籍が発表されることになるのだが，日本においても野球の数理科学に関する論文が発表されることになる．1979年に本誌に掲載された鳩山由紀夫による「野球のOR」である[2]．「Moneyball」が出版される4年前のことで，まだ日本にはセイバーメトリクスという言葉が普及していない時代である．論文の中で「セイバーメトリクス」という言葉は現れないが，各アウトカウント，塁状況別の得点期待値やOERA（Offensive Earned Run Average）といったセイバーメトリクスの基礎的な指標や分析方法を紹介している．また2002年には武井，瀬古，穴太による最適打順の決定方法についての研究論文が発表されている[7]．

とりごえ　のりお
東海大学 理学部
〒259-1292 平塚市北金目4-1-1

近年はプロスポーツのデータ収集を行う企業，データスタジアム社が，日本のプロ野球の1プレイごとの細やかなデータを収集，発信していることで，日本におけるセイバーメトリクスに関する研究の幅が格段に広がっていく傾向にある．

本稿では，第2節で著者がデータスタジアム社との最初の共同研究で，セイバーメトリクスの研究に注力するきっかけとなった，日本プロ野球の試合における勝利確率（Win Probability）の推移を求めるための数理モデルを紹介する．第3節ではコンピュータシミュレーションによる最適打順の決定方法に関する考察について述べる．

2.　勝利確率の推移

2.1　イニング終了時における得点差別の勝利確率

X_kを先攻チームの第kイニングの得点，Y_kを後攻チームの第kイニングの得点を表す確率変数とする．またp_kを第kイニングにおける打者がアウトになる確率とし，関数$FT_k(x)$，$FB_k(x)$を

$FT_k(x)$：第kイニング表終了時，x点差がついたときの後攻チームの勝利確率，

表 1　後攻チームのイニング終了時における勝利確率の推定値

点差	1 回表	1 回裏	…	6 回表	6 回裏	7 回表	7 回裏	8 回表	8 回裏	9 回表
5	0.9236	0.9104	…	0.9805	0.9765	0.9901	0.9880	0.9970	0.9963	1
4	0.8829	0.8636	…	0.9613	0.9535	0.9780	0.9734	0.9921	0.9903	1
3	0.8264	0.7993	…	0.9254	0.9108	0.9527	0.9430	0.9798	0.9753	1
2	0.7519	0.7157	…	0.8620	0.8359	0.9021	0.8827	0.9507	0.9401	1
1	0.6596	0.6141	…	0.7571	0.7135	0.8075	0.7708	0.8866	0.8628	1
同点	0.5534	0.5	…	0.5751	0.5	0.5840	0.5	0.5995	0.5	0.6124
−1	0.4430	0.3859	…	0.3764	0.2865	0.3351	0.2292	0.2717	0.1372	0.1629
−2	0.3396	0.2843	…	0.2371	0.1641	0.1934	0.1173	0.1400	0.0599	0.0716
−3	0.2500	0.2007	…	0.1416	0.0892	0.1058	0.0570	0.0683	0.0247	0.0296
−4	0.1774	0.1364	…	0.0810	0.0465	0.0554	0.0266	0.0320	0.0097	0.0117
−5	0.1217	0.0896	…	0.0446	0.0235	0.0281	0.0120	0.0145	0.0037	0.0044

$FB_k(x)$：第 k イニング裏終了時，x 点差がついたときの後攻チームの勝利確率，

と定義する．また，第 k イニング時に y 点獲得する確率を示す関数を $f_k(y)$ とすると

$$f_k(x)=0.35\phi_1(x)+0.65\phi_2(x)$$

で表される[6]．ここで

$$\phi_1(0)=\phi(0;p_k)+\phi(1;p_k),$$
$$\phi_1(x)=\phi(x+1;p_k)for\quad x=1,2,3,...,$$
$$\phi_2(0)=\phi(0;p_k)+\phi(1;p_k)+\phi(2;p_k),$$
$$\phi_2(x)=\phi(x+2;p_k)for\quad y=1,2,3,...,$$
$$\phi(x;p_k)=\binom{x+2}{x}p_k^3(1-p_k)^x,$$

である．この関数を用いて，イニング終了時における得点差別の勝利確率の理論値を求める方法を紹介する[8]．まず両チームの実力は等しいとし，X_k と Y_k $(k=1, ..., 9, 10, 11, 12)$ は独立であると考える．このとき $k\geqq 9$ のときの $FB_k(x)$ は

$$FB_k(x)=\begin{cases}1 & (x\geqq 1),\\ 0.5 & (x=0),\\ 0 & (x\leqq -1),\end{cases}$$

となる．最初に 9 回表終了時の後攻チームの勝率理論値を考察する．9 回表終了時に 1 点以上リードしていることは勝利が確定しているので

$$FT_9(x)=1\quad(x\geqq 1)$$

である．また 9 回表終了時に x 点差 $(x\leqq 0)$ の場合，9 回裏に $-x+1$ 点獲得すれば勝利となる．また 9 回裏に $-x$ 点とれば延長戦に入る．延長戦での勝利確率は 0.5 と考える．これを式に表すと

$$FT_9(x)=1-P(Y_9\leqq -x)+0.5P(Y_9=-x)$$
$$=1-\sum_{i=0}^{-x}f_9(i)+0.5f_9(-x)$$

となる．次に $k\leqq 8$ のとき第 k イニング裏終了時の後攻チームの勝利確率は

$$FB_k(x)=\sum_{j=0}^{\infty}f_{k+1}(j)FT_{k+1}(x+j)$$

と表され，第 k イニング表終了時の後攻チームの勝利確率は

$$FT_k(x)=\sum_{j=0}^{\infty}f_k(j)FB_k(x+j)$$

と表される．この計算式をもとに 2010 年以前の日本プロ野球のデータから推定される，後攻チームのイニング終了時における得点差別の勝利確率を表 1 に示す．

表 2　ランナーとアウトカウントの状態

ランナー	0アウト	1アウト	2アウト
なし	1	9	17
1塁	2	10	18
2塁	3	11	19
3塁	4	12	20
1塁2塁	5	13	21
1塁3塁	6	14	22
2塁3塁	7	15	23
満塁	8	16	24

2.2　アウトカウント，塁状況別勝利確率

野球の 1 イニングにおけるアウトカウントは 0，1，2 の 3 通りで，ランナーの状況は，ランナーなし，1 塁，2 塁，3 塁，1 塁 2 塁，1 塁 3 塁，2 塁 3 塁，満塁の 8 通りある．つまり 1 イニングにおける状態は 24 通りあると考えられる．その状態における勝利確率の算出モデルを考察する．なお，D を後攻チームの点差を表す確率変数とし，状態を表す変数 S のとり得る値を表 2 に示す．

まず 9 回裏同点，2 アウト満塁の状態を想定する．ここでは盗塁や暴投，パスボールの類いは考慮しないものとする．この状況で後攻チームの打者がアウトに

ならなければサヨナラ勝ちとなり，アウトになれば延長戦に入る．このことより，この状態における勝利確率は

$$\mathrm{P}(D=0,S=24)$$
$$=\mathrm{P}(\{\text{出塁}\}|S=24)+0.5\mathrm{P}(\{\text{アウト}\}|S=24)$$

と表すことができる．

次に，2アウト2塁3塁の状態を考えると，打者がアウトで延長，ヒット，もしくは相手方の失策があれば，サヨナラ勝ちであるが，四死球の場合は2アウト満塁の状況に推移する．よって

$$\mathrm{P}(D=0,S=23)$$
$$=\mathrm{P}(\{\text{ヒット}\}\cup\{\text{失策出塁}\}|S=23)$$
$$+0.5\mathrm{P}(\{\text{アウト}\}|S=23)$$
$$+\mathrm{P}(D=0,S=24)\mathrm{P}(\{\text{四死球}\}|S=23)$$

が成り立つ．このように試合終了に近い状態から逆算して各状態での勝利確率を求めていく．

9回裏2アウト満塁で，1点差で負けている状況の場合は，アウトで敗戦が確定し，二塁打以上の長打が出れば逆転サヨナラ勝ちとなる．また四球となれば，同点で2アウト満塁の状態に推移する．さらに単打，もしくは失策出塁であれば，そのときのランナーの進塁状況において，結果が変化する．ここでランナーの進塁状況を以下の記号で示す．

1塁ランナーが本塁生還：R1→4

2塁ランナーが本塁生還：R2→4

1塁ランナーが3塁進塁：R1→3

2塁ランナーが3塁進塁：R2→3

1塁ランナーが2塁進塁：R1→2

2塁ランナーが本塁憤死：R2→o

1塁ランナーが本塁憤死：R1→o

2塁ランナーが生還すれば逆転勝ちとなり，3塁に止まれば同点2アウト満塁となる．また本塁でアウトになれば，同点で9回裏が終了し，延長戦となる．これらのことを踏まえ，一般に第kイニング裏の攻撃中の後攻チームのある状態での勝利確率を数式化する．一例として，第kイニング裏でx点差，2アウト満塁の状態における勝利確率は

$$\mathrm{P}_k(D=x,S=24)$$
$$=\mathrm{P}(\{\text{本塁打}\}|S=24)\mathrm{P}_k(D=x+4,S=17)$$
$$+\mathrm{P}(\{\text{三塁打}\}|S=24)\mathrm{P}_k(D=x+3,S=20)$$
$$+\mathrm{P}(\{\text{二塁打}\}|S=24)\{\mathrm{P}(\mathrm{R1}\to4)\mathrm{P}_k(D=x+3,S=19)$$
$$+\mathrm{P}(\mathrm{R1}\to3)\mathrm{P}_k(D=x+2,S=23)$$
$$+\mathrm{P}(\mathrm{R1}\to\mathrm{o})FB_k(x+2)\}$$
$$+\mathrm{P}(\{\text{単打}\}\cup\{\text{失策出塁}\})[\mathrm{P}(\mathrm{R2}\to4)$$
$$\{\mathrm{P}(\mathrm{R1}\to4)\mathrm{P}_k(D=x+3,S=18)$$
$$+\mathrm{P}(\mathrm{R1}\to3)\mathrm{P}_k(D=x+2,S=22)$$
$$+\mathrm{P}(\mathrm{R1}\to2)\mathrm{P}_k(D=x+2,S=21)$$
$$+\mathrm{P}(\mathrm{R1}\to\mathrm{o})FB_k(x+2)\}$$
$$+\mathrm{P}(\mathrm{R2}\to3)\mathrm{P}_k(D=x+1,S=24)$$
$$+\mathrm{P}(\mathrm{R2}\to\mathrm{o})\mathrm{FB}_k(x+1)]$$
$$+\mathrm{P}(\{\text{四死球}\})\mathrm{P}_k(D=x+1,S=24)+\mathrm{P}(\{\text{アウト}\})FB_k(x)$$

で表される．また0アウトランナーなしの状態における勝利確率は

$$\mathrm{P}_k(D=x,S=1)$$
$$=\mathrm{P}(\{\text{本塁打}\}|S=1)\mathrm{P}_k(D=x+1,S=1)$$
$$+\mathrm{P}(\{\text{三塁打}\}|S=1)\mathrm{P}_k(D=x,S=4)$$
$$+\mathrm{P}(\{\text{二塁打}\}|S=1)\mathrm{P}_k(D=x,S=3)$$
$$+\mathrm{P}(\{\text{単打}\}\cup\{\text{失策出塁}\}\cup\{\text{四死球}\}|S=1)$$
$$\mathrm{P}_k(D=x,S=2)+\mathrm{P}(\{\text{アウト}\}|S=1)\mathrm{P}_k(D=x,S=1)$$

で表される．これらの計算式によって，2010年以前の日本プロ野球のデータから推定される，後攻チームのアウトカウント，塁状況別の勝利確率が求められる．その結果は，セイバーメトリクスにおける選手評価の指標となるWPA（Win Probability Added）の根拠となる．図1は2009年8月23日に行われた東北楽天とオリックスの試合における東北楽天の勝利確率の推移を示したものである．この試合において，楽天の3番打者である鉄平選手は6回裏同点0アウト1塁2塁の状態で3ランホームランを打ち，勝利確率を0.636から0.918に増加させた．この打撃結果は0.282の貢献

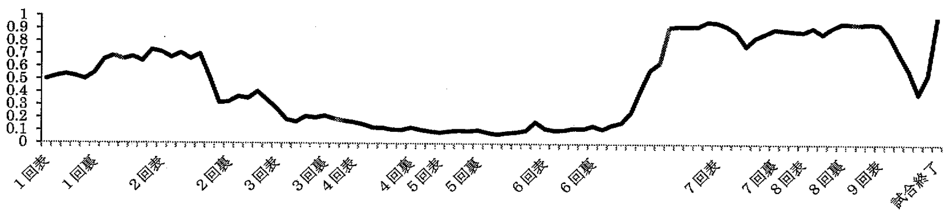

図1 2009年8月23日楽天 vs オリックスの試合における勝利確率の推移

表 3　後攻チームが1点差で負けている状況での勝利確率

イニング	0アウト1塁	1アウト2塁
1回裏	0.4765	0.4621
2回裏	0.4736	0.4587
3回裏	0.4699	0.4539
4回裏	0.4643	0.4471
5回裏	0.4559	0.4370
6回裏	0.4417	0.4204
7回裏	0.4188	0.3944
8回裏	0.3803	0.3514
9回裏	0.3211	0.2838

があったとして，この値を試合やシーズンを通じて加算していく．これがWPA算出の概要である．なおこの試合における鉄平選手のWPAは0.240で，最も貢献度の大きい打者として評価されることになる．

また状態別の勝利確率が算出されたことで，送りバントや盗塁，ヒットエンドランといった戦術の評価も行うことができる．表3は各イニングにおける1点ビハインド，0アウト1塁と1アウト2塁の状態での勝利確率を示している．

どのイニングにおいても，1アウト2塁の勝利確率が0アウト1塁より小さいことがわかる．このことから「犠牲バント」という戦術が，勝利確率を上昇させないということがわかる．ビル・ジェームスもメジャーリーグのデータからこの研究成果を発表していたが，日本のプロ野球のデータからもこのことが判明したのである[9].

3.　最適打順に関する考察

24の状態推移をマルコフ連鎖として捉え，9人の打者が9イニング攻撃して得られる得点の期待値を算出するモデルが，Bukiet, Harold and Palacios[1]によって提案されている．ここではそのモデルを利用し，簡略化したルールにおける最適打順を求める方法について考察を行う．また，モンテカルロシミュレーションによって得られた最適打順との比較についても行う．

まず，イニング中の状態は表1に準じ，3アウトの状態を25で表す．次に打撃結果に関する確率を

p_o：アウトになる確率
p_w：四球になる確率
p_{H1}：単打になる確率
p_{H2}：二塁打になる確率
p_{H3}：三塁打になる確率
p_{H4}：ホームランになる確率

とし，$p_o+p_w+p_{H1}+p_{H2}+p_{H3}+p_{H4}=1$を仮定する．つまり，犠打，犠飛，併殺打などの結果は考慮しないことにする．また進塁に関しては，塁上のランナーは打者の塁打数と等しい値だけ進塁することにする．打者が攻撃を行うことによって推移する状態を表す推移確率行列Pは以下のようになる．

$$P=\begin{pmatrix} A & B & O & \mathbf{0} \\ O & A & B & \mathbf{0} \\ O & O & A & \boldsymbol{p}_O \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 \end{pmatrix}$$

ここで

$$A=\begin{pmatrix} p_{H4} & p_W+p_{H1} & p_{H2} & p_{H3} & 0 & 0 & 0 & 0 \\ p_{H4} & 0 & 0 & p_{H3} & p_W+p_{H1} & 0 & p_{H2} & 0 \\ p_{H4} & 0 & p_{H2} & p_{H3} & p_W & p_{H1} & 0 & 0 \\ p_{H4} & p_{H1} & p_{H2} & p_{H3} & 0 & p_W & 0 & 0 \\ p_{H4} & 0 & 0 & p_{H3} & 0 & 0 & p_{H2} & p_W+p_{H1} \\ p_{H4} & 0 & 0 & p_{H3} & p_{H1} & 0 & p_{H2} & p_W \\ p_{H4} & 0 & p_{H2} & p_{H3} & 0 & p_{H1} & 0 & p_W \\ p_{H4} & 0 & 0 & p_{H3} & 0 & 0 & p_{H2} & p_W+p_{H1} \end{pmatrix}$$

$$B=diag(p_O\cdots p_O),\ \boldsymbol{p}_O=(p_O\cdots p_O)',$$

$\mathbf{0}$は零ベクトルである．またPは

$$P=P0+P1+P2+P3+P4$$

と分解できる．ここで$P0, P1, P2, P3, P4$は打者の攻撃によって得られる得点がそれぞれ0，1，2，3，4となる推移行列である．

この推移行列は選手によって異なるので，打順iの選手の推移確率行列をP_iとする．この推移行列を用いた得点期待値を算出するアルゴリズムを以下に示す．

STEP 1：9人の打者の推移確率行列$P_i(i=1, ..., 9)$を定める．

STEP 2：U_0を（1，1）成分が1で，他の成分はすべて0の31×25行列とする．

STEP 3：漸化式

$$U_{n+1}(j\text{行})=U_n(j\text{行})P0+U_n(j-1\text{行})P1 +U_n(j-2\text{行})P2+U_n(j-3\text{行})P3 +U_n(j-4\text{行})P4$$

を用いて，U_1を計算し，以下同様に$U_2, U_3, ...$と順次計算する．

STEP 4：U_jの25列の成分を$(y_0\ y_1\ y_2 \cdots y_{30})'$としたとき，$y_0+y_1+y_2+\cdots+y_{30}$があらかじめ設定した限界値を越えたときに，そのイニングを終了する．そのイニングの得点期待値は$y_1+2y_2+\cdots+30y_{30}$である．今回は限界値を0.99999とする．

STEP 5：次の打者を先頭打者として，STEP 2に

戻る．

STEP 6：9イニングが終了した時点で，それまでの各イニングの得点期待値を加算することによって1試合の得点期待値が計算される．

まずは簡略なモデルによる最適打順を考察する．9人のうち3人を強打者とした打順を考える．強打者とそれ以外の打者のパラメータを表4と表5に示す．

表4では，強打者は0.6の確率で単打を打ち，長打はないものとしている．また表5での強打者は0.6の確率でホームランを打ち，それ以外はアウトという極端な選手を想定している．このパラメータを持つ選手を並べ替え，84通りのオーダーで得点期待値の比較を行ったところ，以下のような結果を得た．

この結果より，強打者が単打のみを確実に打つ打線の場合，強打者をつなげておくことによって，得点期待値を大きくすることができる．さらにいえば，従来の打線でいうところのクリーンナップに強打者を置くことで，高い得点期待値が得られることがわかる．しかし，強打者がホームランのみを確率高く打つ場合は，1人を先頭打者とし，3人の強打者を離して配置することで，より大きな得点期待値を得ることができるという傾向が見られる．なお，比較のため，同じパラメータを用いたモンテカルロシミュレーションを10000試行，行ったところ，同様の傾向が見られた．

ここで，実際の例として，2011年の福岡ソフトバンクホークスの打撃データを用いて，その最適打順を求めてみる．表8は2011年のソフトバンクホークスの主力打者9名の打撃パラメータである．

打者9人を並べ替え，9！＝362880通りのオーダーで得点期待値の比較を行ったところ，表9の組み合わせのときに，得点期待値が大きくなることが確認された．ホームランの確率が大きいカブレラと松田が5番，6番と連続したオーダーとなっており，次いで大きい小久保，内川も2番，3番と連続している．また単打と四球の確率が内川に次いで2位の長谷川が1番打者となっているときに最適であるというのも興味深い結果である．

表4　打者パラメータ（パターン1）

	p_0	p_{H1}
強打者	0.4	0.6
その他	0.9	0.1

表5　打者パラメータ（パターン2）

	p_0	p_{H1}	p_{H4}
強打者	0.4	0	0.6
その他	0.9	0.1	0

表6　パターン1における得点期待値の大きい打順

強打者が配置される打順			得点期待値	シミュレーション
3番	4番	5番	0.568	0.545
4番	5番	6番	0.568	0.532

表7　パターン2における得点期待値の大きい打順

強打者が配置される打順			得点期待値	シミュレーション
1番	4番	7番	10.351	8.906
1番	2番	5番	10.330	8.714

4. おわりに

2011年のプロ野球で大きな改革が行われた．それは12球団が使用するボールが統一されたことである．そしてそのボールの反発係数が以前のものより小さく

表8　2011年ホークスの打撃パラメータ

	p_0	p_w	p_{H1}	p_{H2}	p_{H3}	p_{H4}
内川	.555	.050	.311	.047	.001	.027
松田	.565	.096	.233	.053	.005	.048
小久保	.593	.071	.252	.061	.000	.023
本多	.623	.077	.258	.033	.009	.000
長谷川	.585	.125	.226	.054	.000	.010
カブレラ	.625	.085	.219	.041	.000	.030
川崎	.651	.062	.251	.025	.011	.000
多村	.651	.098	.206	.035	.000	.010
細川	.751	.042	.169	.028	.005	.005

表9　得点期待値の大きい打順の組み合わせ上位3パターン

1番	2番	3番	4番	5番	6番	7番	8番	9番	得点期待値
長谷川	小久保	内川	細川	カブレラ	松田	本多	多村	川崎	4.795
長谷川	小久保	内川	細川	カブレラ	松田	本多	川崎	多村	4.792
長谷川	小久保	内川	細川	カブレラ	松田	多村	本多	多村	4.775

なったとされている．そのため，2011年の打者の成績が打率，ホームランともに減少している．2010年には両リーグ合わせて27人いた3割打者が，2011年は9人となっている．逆に投手の成績は軒並み上昇しており，2010年に防御率3点以下の投手は7人だったのに対し，2011年は26人となっている．また2011年にパ・リーグでチーム防御率が最低である千葉ロッテの3.40という値は，2010年のチーム最高防御率である日本ハムの3.52よりも良い値であり，2011年セ・リーグで最高の阪神のチーム打率.256は2010年の最低チーム打率.255と均衡している．

このような用具の変化，さらにはルール改正，球場の変化などによるデータのバイアスに関しては，細心の注意を払わなければならない．2010年以前のデータと2011年以降のデータを比較する際には，そのバイアスを考慮した評価が必要となる．

参考文献

[1] B. Bukiet, E.R. Harold and J.L. Palacios: A Markov Chain Approach to Baseball, Operations Research, 45, 1, 14-23 (1997).

[2] 鳩山由紀夫：野球のOR，オペレーションズ・リサーチ：経営の科学，24，4, 203-212 (1979).

[3] B. James: 1977 Baseball Abstract, (1997).

[4] M. Lewis: Moneyball, W.W. Norton & Company, 2003.

[5] G.R. Lindsey: Statistical Data Useful for the Operation of a Baseball Team, Operations Research, 7, 2, 197-207 (1959).

[6] G.R. Lindsey: The Progress of the Score During a Baseball Game, Journal of the American Statistical Association, 56, 295, 703-728 (1961).

[7] 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：経営の科学，47，3, 142-147 (2002).

[8] 鳥越規央：Win Probability Added in Sabermetrics, 数理解析研究所講究録1703 (2010).

[9] 鳥越規央：9回裏無死1塁でバントはするな，祥伝社，2011.