数理解析研究所講究録
第 1703 巻 2010 年 1-9

Win Probability Added in Sabermetrics

東海大学・理学部　鳥越 規央（Norio Torigoe）
School of Science,
Tokai University

１．はじめに

1970 年代のアメリカで，野球のプレイデータから選手や戦術を統計学的な手法を用いて評価する試みが B. James（1977）によって提唱された．このような研究分野はB. Davids らによって設立されたSociety for American Baseball Research（アメリカ野球学会）の頭文字に由来して Sabermetrics（セイバーメトリクス)と呼ばれている．しかしそれ以前にも野球のプレイを統計学的に分析する手法は，多くの研究者によって提案されていた．Lidsey (1960) は１イニングの得点分布をモデル化して，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メジャーリーグのデータをもとに算出している．

この研究では，アウトカウントや塁状況を基にしたアルゴリズムを構築し，2004 年から 2008 年に行われた日本プロ野球（NPB）の公式戦のデータを代入することによって，イニング終了時，さらには，アウトカウント，塁状況，ボールカウント別シチュエーションにおける勝利確率の算出について考察した．これにより WPA（Win Probability Added）と呼ばれる選手の試合における貢献度を示す指標を細かく計算することができ，選手評価の一助となる．なおこの研究はデータスタジアム（株）の協力を得て行っている．

２．１イニングの得点の確率分布

ここに 2004 年から 2008 年のプロ野球のデータから求められた１イニングに獲得する得点の相対度数分布を表２．１に表す．まずはこの分布のモデル化を Lindsey (1960)の手法に倣い行う．

表２．１．　プロ野球における１イニングの得点の相対度数

得点	0	1	2	3	4	5	>5
相対度数	.751	.138	.060	.027	.015	.006	.003

１イニングで３アウトになる前に出塁する打者の数を表す確率変数を Y とする。このとき Y の分布は負の二項分布に従うことが知られている。負の二項分布の確率関数は

$$f_k(y) = 0.35\phi_1(y) + 0.65\phi_2(y)$$
where
$$\phi_1(0) = \phi(0;p_k) + \phi(1;p_k)$$
$$\phi_1(y) = \phi(y+1;p_k)\ for\ y = 1,2,3,\cdots$$
$$\phi_2(0) = \phi(0;p_k) + \phi(1;p_k) + \phi(2;p_k)$$
$$\phi_2(y) = \phi(y+2;p_k)\ for\ y = 1,2,3,\cdots$$

と定義する．ここで

$$\phi(y,p_k) = \binom{y+3-1}{y} {p_k}^3 (1-p_k)^y$$

である．また p_k の算出方法はイニングごとのデータを用いて

$$1 - p_k = \{(安打)+(四球)+(死球)+(失策)-(併殺打)\}/(打席)$$

である．2004 年から 2008 年のデータから推定される各イニングの打者がアウトになる確率は表２．３．のようになった。

表２．３．　イニングごとの打者が凡退する確率の推定値

イニング	1	2	3	4	5	6	7	8	9
得点分布	.668	.701	.692	.689	.691	.680	.686	.689	.701

３．イニング終了時における得点差別の勝利確率の理論値

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

$FB_k(x)$：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．

まず前提として，両チームの実力は等しいとし，X_k と Y_k ($k = 1,\ldots,9,10,11,12$) 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

$$FB_k(x) = \begin{cases} 1 & (x \geq 1) \\ 0.5 & (x = 0) \\ 0 & (x \leq -1) \end{cases}$$

となる．最初に９回表終了時の後攻チームの勝利確率の理論値を考察する．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

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

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

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

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

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

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

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

と表される．この計算式をもとに，データから推定されたイニングごとの打者がアウトになる確率を代入して，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算出した．その数値を表３．１に示す．

表３．１　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	２回表	２回裏	３回表	３回裏	４回表	４回裏	５回表	５回裏
5	0.9236	0.9104	0.9314	0.9220	0.9436	0.9350	0.9562	0.9488	0.9682	0.9625
4	0.8829	0.8636	0.8918	0.8775	0.9075	0.8938	0.9245	0.9121	0.9419	0.9317
3	0.8264	0.7993	0.8349	0.8143	0.8533	0.8326	0.8743	0.8546	0.8972	0.8799
2	0.7519	0.7157	0.7576	0.7293	0.7766	0.7468	0.7992	0.7693	0.8256	0.7975
1	0.6596	0.6141	0.6594	0.6229	0.6748	0.6349	0.6941	0.6517	0.7182	0.6755
同点	0.5534	0.5	0.5439	0.5	0.5500	0.5	0.5564	0.5	0.5625	0.5
－1	0.4430	0.3859	0.4241	0.3771	0.4195	0.3651	0.4110	0.3483	0.3960	0.3245
－2	0.3396	0.2843	0.3153	0.2707	0.3040	0.2532	0.2877	0.2307	0.2648	0.2025
－3	0.2500	0.2007	0.2243	0.1857	0.2102	0.1674	0.1917	0.1454	0.1681	0.1201
－4	0.1774	0.1364	0.1534	0.1225	0.1395	0.1062	0.1225	0.0879	0.1022	0.0683
－5	0.1217	0.0896	0.1013	0.0780	0.0893	0.0650	0.0754	0.0512	0.0599	0.0375

$$\phi(y;p) = P(Y = y) = \binom{y+3-1}{y} p^3 (1-p)^y$$

で表される。ここで p は打者がアウトになる確率である。p の算出方法は

$$1 - p = \{(\text{安打}) + (\text{四球}) + (\text{死球}) + (\text{失策}) - (\text{併殺打})\} / (\text{打席})$$

である。データより p=0.687 と算出された。例えば１イニングでランナーが一人も出ない確率 $\phi(0;p)$ は y=0, p=0.687 を代入して

$$\phi(0;p) = \binom{0+3-1}{0} 0.687^3 (1-0.687)^0 = 0.325$$

となる。つまり３者凡退の確率は 0.325 であるといえる。そこで Lindsey は１イニングに x 点入る確率をモデル化する確率関数 $f(x)$ として、次のような関数を提案した。

$$f(x) = \lambda\phi_1(x) + (1-\lambda)\phi_2(x)$$

where

$$\phi_1(0) = \phi(0;p) + \phi(1;p)$$

$$\phi_1(x) = \phi(x+1;p) \ \ for\ x = 1,2,3,\cdots$$

$$\phi_2(0) = \phi(0;p) + \phi(1;p) + \phi(2;p)$$

$$\phi_2(x) = \phi(x+2;p) \ \ for\ x = 1,2,3,\cdots$$

この式の根拠として、１イニングに x（x=1,2,3,...）点とるためには x+1 人、もしくは x+2 人出塁することが必要なものとしている。その両者の比率を λ ： $1-\lambda$ に設定している。０点の場合は、３者凡退、１人出塁、２人出塁の確率から算出しているものと考えられる。ここでは少数の例外は加味されていない。次の表２．２．では λ=0.35 として計算された $f(x)$ の確率分布と実際の相対度数とを比較している。

表２． ２．　得点確率分布 $f(x)$ の値と相対度数との比較

得点　x	0	1	2	3	4	5	>5
相対度数	.751	.138	.060	.027	.015	.006	.003
$f(x)$	.753	.131	.065	.030	.013	.005	.003

表２． ２．よりこの $f(x)$ は実際の得点分布に対して当てはまりがよい。
λ の値は実際のデータより、最適な値を推定することができ、より適合するモデルを確定することができると考えられる。

そこで我々は, この式をもとに第 k イニング時に y 点獲得する確率を示す関数 $f_k(y)$ を

表３．１．　（つづき）

点差	6回表	6回裏	7回表	7回裏	8回表	8回裏	9回表	9回裏
5	0.9805	0.9765	0.9901	0.9880	0.9970	0.9963	1	
4	0.9613	0.9535	0.9780	0.9734	0.9921	0.9903	1	
3	0.9254	0.9108	0.9527	0.9430	0.9798	0.9753	1	
2	0.8620	0.8359	0.9021	0.8827	0.9507	0.9401	1	
1	0.7571	0.7135	0.8075	0.7708	0.8866	0.8628	1	
同点	0.5751	0.5	0.5840	0.5	0.5995	0.5	0.6124	0.5
−1	0.3764	0.2865	0.3351	0.2292	0.2717	0.1372	0.1629	0
−2	0.2371	0.1641	0.1934	0.1173	0.1400	0.0599	0.0716	0
−3	0.1416	0.0892	0.1058	0.0570	0.0683	0.0247	0.0296	0
−4	0.0810	0.0465	0.0554	0.0266	0.0320	0.0097	0.0117	0
−5	0.0446	0.0235	0.0281	0.0120	0.0145	0.0037	0.0044	0

４．アウトカウント，塁状況別勝利確率の推移

１イニング内におけるアウトカウントとランナーの塁状況は２４の状態があるといえる．そこでその状況を表４．１のように記号化する．

表４．１．　アウトカウントとランナーの塁状況

	０アウト	１アウト	２アウト
ランナーなし	OUT=0, R=0	OUT=1, R=0	OUT=2, R=0
ランナー１塁	OUT=0, R=1	OUT=1, R=1	OUT=2, R=1
ランナー２塁	OUT=0, R=2	OUT=1, R=2	OUT=2, R=2
ランナー３塁	OUT=0, R=3	OUT=1, R=3	OUT=2, R=3
ランナー１，２塁	OUT=0, R=12	OUT=1, R=12	OUT=2, R=12
ランナー１，３塁	OUT=0, R=13	OUT=1, R=13	OUT=2, R=13
ランナー２，３塁	OUT=0, R=23	OUT=1, R=23	OUT=2, R=23
満塁	OUT=0, R=123	OUT=1, R=123	OUT=2, R=123

また打者がヒットを打ったときの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

さらに凡打時にアウトカウントが a から１つ増え，得点の増加がなかった状況でのランナー進塁状況を以下の記号で示す．

ランナー状況が２，３塁から２，３塁のまま：OUT $a \rightarrow a+1$, R23→23

ランナー状況が２，３塁から１，３塁　　　：OUT $a \rightarrow a+1$, R23→13

ランナー状況が２，３塁から１，２塁　　　：OUT $a \rightarrow a+1$, R23→12
ランナー状況が１，３塁から２，３塁　　　：OUT $a \rightarrow a+1$, R13→23
ランナー状況が１，３塁から１，３塁のまま：OUT $a \rightarrow a+1$, R13→23
ランナー状況が１，３塁から１，２塁　　　：OUT $a \rightarrow a+1$, R13→23
ランナー状況が１，２塁から２，３塁　　　：OUT $a \rightarrow a+1$, R12→23
ランナー状況が１，２塁から１，３塁　　　：OUT $a \rightarrow a+1$, R12→23
ランナー状況が１，２塁から１，２塁のまま：OUT $a \rightarrow a+1$, R12→23
ランナー状況が b 塁から c 塁 (b=1,2,3, c=1,2,3)：OUT $a \rightarrow a+1$, Rb→c

まず第 k イニング裏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．点差を示す確率変数を D とし，D=x のときの各状態での勝利確率の理論値を求める式の一部を示す．まず第 k イニングで x 点差で２アウト満塁の状況における勝利確率は

$$
\begin{aligned}
&P_k(D=x, OUT=2, R=123)\\
&= P(\text{本塁打} \mid OUT=2, R=123)P_k(D=x+4, OUT=2, R=0)\\
&\quad +P(\text{三塁打} \mid OUT=2, R=123)P_k(D=x+3, OUT=2, R=3)\\
&\quad +P(\text{二塁打} \mid OUT=2, R=123)\{P(R1\rightarrow 4)P_k(D=x+3, OUT=2, R=2)\\
&\qquad\quad +P(R1\rightarrow 3)P_k(D=x+2, OUT=2, R=2,3)+P(R1\rightarrow o)FB_k(x+2)\}\\
&\quad +P(\text{単打，失策出塁} \mid OUT=2, R=123)[P(R2\rightarrow 4)\\
&\qquad \{P(R1\rightarrow 4)P_k(D=x+3, OUT=2, R=1)+\{P(R1\rightarrow 3)P_k(D=x+2, OUT=2, R=13)\\
&\qquad\ +P(R1\rightarrow 2)P_k(D=x+2, OUT=2, R=12)+\{P(R1\rightarrow o)FB_k(x+2)\}\\
&\qquad\ +P(R2\rightarrow 3)P_k(D=x+1, OUT=2, R=123)+P(R2\rightarrow o)FB_k(x+1)]\\
&\quad +P(\text{四死球} \mid OUT=2, R=123)P_k(D=x+1, OUT=2, R=123)\\
&\quad +P(\text{凡打} \mid OUT=2, R=123)FB_k(x)
\end{aligned}
$$

で表される．また１アウトランナー１，２塁の状況における勝利確率は

$$
\begin{aligned}
&P_k(D=x, OUT=1, R=12)\\
&= P(\text{本塁打} \mid OUT=1, R=12)P_k(D=x+3, OUT=1, R=0)\\
&\quad +P(\text{三塁打} \mid OUT=1, R=12)P_k(D=x+2, OUT=1, R=3)\\
&\quad +P(\text{二塁打} \mid OUT=1, R=12)\{P(R1\rightarrow 4)P_k(D=x+2, OUT=1, R=2)\\
&\qquad +P(R1\rightarrow 3)P_k(D=x+1, OUT=1, R=2,3)+P(R1\rightarrow o)\ P_k(D=x+1, OUT=2, R=2)\}\\
&\quad +P(\text{単打，失策出塁} \mid OUT=1, R=12)[P(R2\rightarrow 4)
\end{aligned}
$$

$\{P(R1\to 4)P_k(D=x+2,OUT=1,R=1)+\{P(R1\to 3)P_k(D=x+1,OUT=1,R=13)$

$+P(R1\to 2)P_k(D=x+1,OUT=1,R=12)+\{P(R1\to o)P_k(D=x+1,OUT=2,R=2)\}$

$+P(R2\to 3)P_k(D=x,OUT=1,R=123)+P(R2\to o)P_k(D=x,OUT=2,R=13)]$

$+P(\text{四死球} \mid OUT=1, R=12)P_k(D=x,OUT=1,R=123)$

$+P(\text{犠打} \mid OUT=1, R=12)P_k(D=x,OUT=2,R=23)$

$+P(\text{併殺打} \mid OUT=1, R=12)FB_k(x)$

$+P(\text{凡打} \mid OUT=1, R=12)\{P(OUT1\to 2, R12\to 12)P_k(D=x,OUT=2,R=12)$

$+P(OUT1\to 2, R12\to 13)P_k(D=x,OUT=2,R=13)$

$+P(OUT1\to 2, R12\to 23)P_k(D=x,OUT=2,R=23)\}$

で表される．そしてノーアウトランナーなしの状況では

$P_k(D=x, OUT=0, R=0)$

$= P(\text{本塁打} \mid OUT=0, R=0)P_k(D=x+1, OUT=0, R=0)$

$+P(\text{三塁打} \mid OUT=0, R=0)P_k(D=x,OUT=0,R=3)$

$+P(\text{二塁打} \mid OUT=0, R=0)P_k(D=x,OUT=0,R=2)$

$+P(\text{単打，失策出塁，四死球} \mid OUT=0, R=0)P_k(D=x,OUT=0,R=1)$

$+P(\text{凡打} \mid OUT=0, R=0)P_k(D=x,OUT=1,R=0)$

で表される．

５．勝利確率と WPA

表５．１，５．２において同点時，１点差負けの状態における各回裏の状況別の勝利確率を示す．この勝利確率の推移を用いることによって WPA(Win Probability Added) と呼ばれる指標を算出することができる．プレイによる勝利確率の変化量の累積によって，選手の試合における勝利貢献度を数値で表すことができる．これによって客観的に選手の試合における活躍具合を評価することができる．

参考文献

J. Albert and J. Bennett (2001) Curve Ball; Baseball, Statistics, and the Role of Chance in the Game, Springer.

James, B. (1977) The 1977 Baseball Abstract.

Lindsey, G. R. (1961) The progress of the score during a baseball game, Journal of the American Statistical Association, 56, 703-728

表５．１．同点時における勝利確率の推定値

アウト	塁状況	イニング								
		１回裏	２回裏	３回裏	４回裏	５回裏	６回裏	７回裏	８回裏	９回裏
2	123	0.5717	0.5755	0.5802	0.5861	0.5934	0.6034	0.6156	0.6315	0.6503
	23	0.5597	0.5631	0.5674	0.5728	0.5796	0.5888	0.6001	0.6145	0.6307
	13	0.5490	0.5519	0.5556	0.5605	0.5669	0.5761	0.5886	0.6065	0.6305
	12	0.5462	0.5489	0.5524	0.5569	0.5627	0.5711	0.5822	0.5980	0.6189
	3	0.5395	0.5421	0.5456	0.5502	0.5565	0.5661	0.5796	0.6001	0.6290
	2	0.5367	0.5391	0.5423	0.5466	0.5523	0.5609	0.5730	0.5912	0.6163
	1	0.5236	0.5250	0.5269	0.5293	0.5324	0.5369	0.5428	0.5510	0.5617
	なし	0.5105	0.5112	0.5121	0.5133	0.5149	0.5173	0.5207	0.5257	0.5326
1	123	0.6399	0.6469	0.6556	0.6666	0.6802	0.6988	0.7222	0.7534	0.7920
	23	0.6264	0.6334	0.6423	0.6536	0.6678	0.6876	0.7129	0.7469	0.7889
	13	0.6111	0.6174	0.6256	0.6362	0.6499	0.6698	0.6966	0.7352	0.7872
	12	0.5883	0.5930	0.5990	0.6066	0.6162	0.6296	0.6469	0.6705	0.7005
	3	0.5923	0.5983	0.6061	0.6166	0.6307	0.6519	0.6819	0.7271	0.7908
	2	0.5719	0.5763	0.5820	0.5896	0.5995	0.6140	0.6340	0.6632	0.6956
	1	0.5525	0.5555	0.5594	0.5645	0.5709	0.5801	0.5921	0.6088	0.6309
	なし	0.5267	0.5283	0.5305	0.5333	0.5370	0.5425	0.5500	0.5610	0.5756
0	123	0.6952	0.7040	0.7151	0.7288	0.7454	0.7678	0.7952	0.8310	0.8744
	23	0.6724	0.6811	0.6923	0.7064	0.7238	0.7477	0.7776	0.8171	0.8650
	13	0.6634	0.6720	0.6834	0.6979	0.7165	0.7430	0.7783	0.8285	0.8649
	12	0.6281	0.6345	0.6428	0.6531	0.6660	0.6838	0.7062	0.7363	0.7731
	3	0.6342	0.6421	0.6528	0.6669	0.6854	0.7129	0.7509	0.8070	0.8818
	2	0.6088	0.6151	0.6235	0.6345	0.6488	0.6699	0.6985	0.7404	0.7832
	1	0.5842	0.5887	0.5949	0.6027	0.6128	0.6272	0.6461	0.6727	0.7186
	なし	0.5534	0.5439	0.5500	0.5564	0.5625	0.5751	0.5840	0.5995	0.6124

表５．２．１点差負けの状態における勝利確率の推定値

アウト	塁状況	イニング								
		１回裏	２回裏	３回裏	４回裏	５回裏	６回裏	７回裏	８回裏	９回裏
2	123	0.4641	0.4605	0.4555	0.4480	0.4369	0.4182	0.3886	0.3392	0.2575
	23	0.4502	0.4460	0.4401	0.4316	0.4194	0.3994	0.3683	0.3173	0.2337
	13	0.4381	0.4329	0.4258	0.4157	0.4011	0.3773	0.3407	0.2807	0.1877
	12	0.4353	0.4300	0.4226	0.4121	0.3969	0.3722	0.3342	0.2722	0.1760
	3	0.4269	0.4212	0.4133	0.4021	0.3862	0.3605	0.3216	0.2586	0.1637
	2	0.4242	0.4182	0.4100	0.3984	0.3819	0.3554	0.3151	0.2499	0.1516
	1	0.4109	0.4040	0.3944	0.3809	0.3618	0.3310	0.2844	0.2092	0.0958
	なし	0.3968	0.3889	0.3780	0.3626	0.3410	0.3062	0.2538	0.1695	0.0435
1	123	0.5394	0.5404	0.5413	0.5418	0.5414	0.5388	0.5320	0.5171	0.4874
	23	0.5225	0.5231	0.5235	0.5236	0.5230	0.5207	0.5151	0.5034	0.4794
	13	0.5050	0.5043	0.5032	0.5011	0.4974	0.4903	0.4778	0.4550	0.4150
	12	0.4817	0.4793	0.4758	0.4704	0.4621	0.4478	0.4244	0.3844	0.3179
	3	0.4826	0.4807	0.4781	0.4742	0.4684	0.4585	0.4427	0.4161	0.3738
	2	0.4621	0.4587	0.4539	0.4471	0.4370	0.4204	0.3944	0.3514	0.2838
	1	0.4422	0.4373	0.4306	0.4209	0.4070	0.3842	0.3492	0.2918	0.2147
	なし	0.4141	0.4073	0.3980	0.3849	0.3663	0.3363	0.2908	0.2174	0.1094
0	123	0.6025	0.6066	0.6118	0.6178	0.6244	0.6321	0.6392	0.6447	0.6442
	23	0.5742	0.5775	0.5818	0.5870	0.5931	0.6006	0.6086	0.6167	0.6171
	13	0.5628	0.5654	0.5689	0.5730	0.5776	0.5830	0.5882	0.5923	0.5892
	12	0.5261	0.5261	0.5261	0.5254	0.5234	0.5184	0.5082	0.4882	0.4506
	3	0.5282	0.5290	0.5304	0.5319	0.5334	0.5347	0.5350	0.5331	0.5283
	2	0.5018	0.5009	0.4996	0.4976	0.4940	0.4874	0.4759	0.4554	0.4210
	1	0.4765	0.4736	0.4699	0.4643	0.4559	0.4417	0.4188	0.3803	0.3211
	なし	0.4430	0.4241	0.4195	0.4110	0.3960	0.3764	0.3351	0.2717	0.1629