

項目反応理論を用いた野球選手の 能力評価指標の提案

阿部 興¹・作村 建紀²・鎌倉 稔成²

(受付 2016 年 12 月 26 日；改訂 2017 年 3 月 1 日；採択 4 月 25 日)

要 旨

本研究は野球選手の、特に打撃の能力に関する新しい評価指標の提案を行う。打者の能力を評価する指標として、最も広く使用されるのは打率である。しかし、打率を一つの統計モデルと捉える場合、これは現実的でない仮定のもとに成立するものとも言える。打率はヒットを打つ確率がピッチャーの能力に関わらず常に一定であると見て計算されている。提案モデルはピッチャーによってヒットを打つ難しさが異なり、対戦するピッチャーが選手ごとに異なるという状況の下で、打撃の能力の選手間での比較を可能にする。我々が提案するモデルは、項目反応理論で用いられる 1 パラメータロジスティックモデル(ラッシュモデル)の拡張である。ラッシュモデルでは、潜在的な能力パラメータは各人一つとされているが、本論文では打者の能力に対応するマルコフパラメータを導入した。すなわち、我々は打席結果にマルコフ性を仮定して調子の波を表現する。これにより調子の波が前の打席の結果を受けて生じると解釈できる。モデルのパラメータの推定にはハミルトニアン・モンテカルロ法を用いる。パラメータ推定の安定性はシミュレーションによって評価する。提案手法の有益性は日本プロ野球の実際のデータを分析することで示す。

キーワード：階層ベイズモデル、MCMC、セイバーメトリクス、ロジスティックモデル。

1. はじめに

本研究は野球選手の、特に打撃の能力に関する新しい評価指標の提案を行う。選手の能力を公平かつ客観的に評価することは、球団を運営する上で重要な課題である。打者の能力を評価する指標として、最も広く使用されるのは打率である。次いで有名な指標として、長打率(SLG)、出塁率(OBP)、OPS などがある(Albert and Bennett, 2003)。長打率は単打に 1、二塁打に 2、三塁打に 3、本塁打に 4 の重みをつけた、ヒット数の重み付き平均である。出塁率はヒットと四球と死球の数を足したものを、打数と四球と死球と犠飛の数を足したもので割った指標である。OPS は出塁率と長打率を足し合わせた指標である。Albert and Bennett (2003)では、これらの指標をチームごとの平均的な打撃評価指標と得点との相関という観点から検討し、整理している。

より分析的な打者の能力評価指標として、Albert (2008)によるものがある。一般に打者には

¹ 中央大学大学院 理工学研究科：〒112-8551 東京都文京区春日 1-13-27

² 中央大学 理工学部：〒112-8551 東京都文京区春日 1-13-27

「調子の波」が存在すると言われている。Albert (2008)はベータ・二項モデルを用いて調子の波を表現した。このベータ・二項モデルは20打席ごとのヒットの数が二項分布に従うとし、二項分布の成功確率パラメータ p がベータ分布に従うとしたベイズモデルである。ここでのベータ分布は以下のようにパラメタライズされる。

$$(1.1) \quad \frac{1}{B(K\eta, K(1-\eta))} p^{K\eta-1} (1-p)^{K(1-\eta)-1} \quad (K > 0, 0 < \eta < 1).$$

ここで $B(\cdot)$ はベータ関数である。 η は分布の中心を表す。 K は精度パラメータで、大きいほど散らばりが小さい。すなわち、 K が小さく推定された選手ほど、調子の波が激しいと解釈できる。このモデルは成功確率 p の変化を表現し、これまで評価の難しかった調子の波を定量的に評価することを可能にした。

上述の指標はいずれもピッチャーの能力を考慮していない。ピッチャーによってヒットを打つ難しさが異なり、対戦するピッチャーが選手ごとに異なるにも関わらず、打撃の能力を選手間で比較したい場合、どのようにすればよいか。項目反応理論 (item response theory; IRT) の研究成果は、この問題に一つの解を与える。IRT は、教育における達成度評価のような、対象となる人物 (被験者) にある課題 (負荷) が与えられたときに得られる反応から能力測定を行うために考案されたモデルである (Lord, 1952)。IRT の大きな特徴の一つは、被験者に課された課題への反応から、被験者の能力と課題の難しさを同時に評価することである。つまり、被験者の能力を示す能力パラメータと課された課題のレベルを示す項目パラメータを分離して評価するため、有益な情報を提供し得る。反応を表すモデルは、課題が持つパラメータの数および得られる反応値の種類によってさまざまなものが考案されている (Hambleton et al., 1991; De Ayala, 2008; Baker, 1992)。リンク関数にはロジスティック関数やプロビット関数が用いられる。またその推定手法についても、周辺最尤推定法 (Bock and Aitkin, 1981) やマルコフ連鎖モンテカルロ法 (MCMC) を用いたベイズ推定 (Patz and Junker, 1999a, 1999b) など、さまざまな手法が確立されており、それを実装したソフトウェアやパッケージなども多い (Bilog-MG, 2005; Rizopoulos, 2006; Chalmers, 2012)。もともとは教育工学や心理学の分野で発展した理論であるが、近年ではマーケティングへの応用も報告されている (Jong et al., 2008; Raykov and Calantone, 2014)。

この手法を野球へ応用し、打席ごとのヒット・アウトという応答の確率を、打者の能力を表すパラメータと、投手の難易度 (投球の打たれやすさ) を表すパラメータに分離して推定することが可能である。IRT における能力パラメータが打者の能力パラメータに、項目パラメータが投手の難易度パラメータにそれぞれ対応する。我々が提案するモデルは、IRT で用いられるラッシュモデル (1パラメータロジスティックモデル) の拡張である。ラッシュモデルでは、潜在的な能力パラメータは各人一つとされているが、本論文では打者の能力に対応するマルコフパラメータを導入した。ラッシュモデルはその単純さから、解釈が容易であるとされる。我々のモデルもオッズ比を用いることで、選手間の比較を容易なものにできる。

我々は Albert (2008) と同様に、成功確率 p が変化することを仮定する。ただし、上述したベータ二項モデルにおける p の変化は、これまでの打席結果や、打席の状況に依存しないため、解釈がむずかしい。そこで我々は打席結果にマルコフ性を仮定して調子の波を表現する。これにより調子の波が前の打席の結果を受けて生じると解釈できる。3節では、MCMC を用いたパラメータの推定方法について述べる。4節で、パラメータ推定が可能であることを確かめる。5.1節では、2013年の日本プロ野球の実際のデータに対して分析を行う。

2. 提案モデル

$y_{i,j}$ を打者 j ($j = 1, \dots, n$) の打席 i ($i = 1, \dots, l_j$) での結果(アウトならば 0, ヒットならば 1 の値を取る)とする. ここで l_j は打者 j の合計の打席数である. $x_{i,j}$ を打者 j が打席 i で対戦した投手とする. $x_{i,j}$ は $(1, 2, \dots, m)$ のいずれかの値をとる. m はリーグ内の投手の数である. 表 1 は, 選手 j について記録されるデータを, 模式的に示したものである.

$y_{i,j}$ はパラメータ $p_{i,j}$ のベルヌーイ分布に従うと仮定し ($p_{i,j} = \Pr(y_{i,j} = 1)$), $p_{i,j}$ に以下のような回帰型の構造を仮定する.

$$(2.1) \quad \text{logit}(p_{i,j}) = \beta_{0,j} + \beta_{1,j}y_{i-1,j} + \tau(x_{i,j}).$$

ここで, $\tau(x)$ は,

$$(2.2) \quad \tau(x) = \begin{cases} b_1 & x = 1 \\ b_2 & x = 2 \\ \vdots & \vdots \\ b_m & x = m \end{cases}$$

なる関数である. 投手の難易度パラメータ b_k ($k = 1, \dots, m$) は平均が 0 になるよう基準化されているとする. 事前分布として, b_k は平均 0, 分散 σ^2 の正規分布に従うとする. $\beta_{0,j}$ は平均 μ , 分散 ξ^2 の正規分布を仮定する. $\beta_{1,j}$ には, 無情報階層事前分布として, 区間 $(-\infty, \infty)$ の一様分布を仮定する. 階層事前分布として, σ, ξ には区間 $[0, \infty)$, μ には区間 $(-\infty, \infty)$ の一様分布を仮定する.

$\beta_{1,j}$ はヒットを打つ確率に対する, 直前の打席の影響を表すパラメータである. 我々は $\beta_{1,j}$ の 90% 信用区間を求め, これが 0 を含まない選手を「調子の波がある選手」, 0 を含む選手を「調子の波がない選手」とする.

2.1 オッズ比

ロジスティックモデルはオッズ比による解釈が容易である. 本研究では 3 種類のパラメータに対してオッズ比を評価する. 一つ目は打者同士のベースラインとなる能力を比較するためのオッズ比である. 直前の打席を失敗とし, ピッチャーを特定の選手に固定した場合にヒットを打つ確率に着目すると, 打者 j の打者 h に対するオッズ比は,

$$(2.3) \quad \exp(\beta_{0,j}) / \exp(\beta_{0,h})$$

である. 二つ目は投手の能力を比較するためのオッズ比である. 投手難易度パラメータは平均が 0 になるよう基準化されていることを仮定したため, ここでは 0 を基準としたオッズ比,

表 1. 解析対象となるデータの一例.

| i | $y_{i,j}$ | $x_{i,j}$ |
|----------|-----------|-----------|
| 1 | 1 | 1 |
| 2 | 0 | 1 |
| 3 | 0 | 2 |
| \vdots | \vdots | \vdots |
| l_j | 1 | 4 |

$$(2.4) \quad \exp(b_j)$$

を考える．このオッズ比は平均的なピッチャーと比べて、どの程度ヒットという事象が起こりやすいかという指標であり、値が小さいほど能力の高いピッチャーと解釈できる．三つ目は打者の調子の波を把握するためのオッズ比である．

$$(2.5) \quad \exp(\beta_{1,j})$$

これは直前の打席結果がアウトだったときを基準として、直前の打席結果がヒットだったときはどの程度ヒットを打ちやすいかという指標である．5.1 節の事例研究ではこれらを使用して選手を評価する．

2.2 定常分布

本モデルでは、投手の難易度パラメータの平均が 0 になるよう基準化されていることを仮定した．打者 j が平均的な投手と対戦した場合のヒットの確率は、

$$P(y_{i,j} = 1) = \text{logit}^{-1}(\beta_{0,j} + \beta_{1,j}y_{i-1,j}).$$

である．一打席前でヒットを打てなかった場合にヒットを打つ確率を、

$$(2.6) \quad p_{0,j} = \frac{1}{1 + \exp(-\beta_{0,j})}$$

一打席前でヒットを打った場合にヒットを打つ確率を、

$$(2.7) \quad p_{1,j} = \frac{1}{1 + \exp(-(\beta_{0,j} + \beta_{1,j}))}$$

と表すことにする．これは遷移行列

$$(2.8) \quad \mathbf{P}_j = \begin{pmatrix} 1 - p_{0,j} & p_{0,j} \\ 1 - p_{1,j} & p_{1,j} \end{pmatrix}$$

の 2 状態マルコフ連鎖を形作る．このマルコフ連鎖は既約かつ非周期的 (Karlin, 1969) であり、以下を満たす行ベクトル π_j が定義できる．

$$(2.9) \quad \pi_j \mathbf{P}_j = \pi_j$$

$\pi_j = (\pi_{0,j}, \pi_{1,j})$ について解くと、

$$(2.10) \quad \pi_{0,j} = \frac{1 - p_{1,j}}{1 + p_{0,j} - p_{1,j}}, \quad \pi_{1,j} = \frac{p_{0,j}}{1 + p_{0,j} - p_{1,j}}$$

である． π_j は定常分布と呼ばれる． $\pi_{1,j}$ は十分な時間が経過したときの、選手 j がヒットを打つ割合と解釈でき、初期の状態には影響されない．これも指標の一つとして 5.1 節の事例研究で使用される．

3. パラメータの推定方法

式(2.1)で示したモデルの事後分布の統計量を求めるために、本研究ではハミルトニアン・モンテカルロ法(ハイブリッドモンテカルロ法ともいう)による MCMC を用いた (Bishop, 2006)．MCMC は推定の対象となる事後分布に従う乱数を、マルコフ連鎖を用いてシミュレートする方法である．近年、IRT の分野では MCMC が積極的に利用されており、MCMC による推定の性質は Fox (2010) で議論されているが、ハミルトニアン・モンテカルロ法を用いる方法に

については記載がない．MCMC 法のバリエーションとしてギブスサンプリング法やメトロポリス・ヘイスティングス法はよく知られており，IRT の分野でも盛んに利用されている．ギブスサンプリング法ではパラメータごとに全条件付き分布を導出し，変数ごとにサンプリングを行う．メトロポリス・ヘイスティングス法ではパラメータごとに適切な提案分布を選ぶ必要がある．本解析のようにパラメータが多い場合，ハミルトニアン・モンテカルロ法が便利である．

MCMC による IRT モデルのパラメータ推定法の中でも特に有名な手法として Albert (1992) により提案された，データ拡大 (data augmentation) アルゴリズムがある．0 または 1 の値を取る打席結果データ $y_{i,j}$ に対して，連続値を取る潜在変数 z を仮定する． z は平均 $\eta_{i,j} = \beta_{0,j} + \beta_{1,j}y_{i-1,j} + \tau(x_{i,j})$ ，分散 1 の正規分布に従うとし， $z_{i,j} > 0$ ならば $y_{i,j} = 1$ ， $z_{i,j} \leq 0$ ならば $y_{i,j} = 0$ となる．言い換えると， $y_{i,j} = 1$ という観測は 0 で右打ち切りされた $z_{i,j}$ の観測値であり $y_{i,j} = 0$ という観測は 0 で左打ち切り (Turnbull, 1976) された $z_{i,j}$ の観測値である．この場合尤度関数は

$$(3.1) \quad \prod_{j=1}^n \prod_{i=2}^{l_j} \Phi(0|\eta_{i,j}, 1)^{y_{i,j}} (1 - \Phi(0|\eta_{i,j}, 1))^{1-y_{i,j}}$$

と書ける．本研究ではリンク関数にロジット関数を用いる．この場合も Albert (1992) と同様の議論により，潜在変数 z を仮定すると z はロケーションパラメータが $\eta_{i,j} = \beta_{0,j} + \beta_{1,j}y_{i-1,j} + \tau(x_{i,j})$ のスケールパラメータが 1 のロジスティック分布に従うと考える． $F(z|\eta)$ をスケールパラメータ 1，ロケーションパラメータ η のロジスティック分布の分布関数とすると，尤度関数は，

$$(3.2) \quad \prod_{j=1}^N \prod_{i=2}^{l_j} F(0|\eta_{i,j})^{y_{i,j}} (1 - F(0|\eta_{i,j}))^{1-y_{i,j}}$$

と書ける．これはベルヌーイ分布にロジットリンク関数を仮定した場合の尤度と全く同一である．事後分布は，

$$(3.3) \quad \pi(\beta_0, \beta_1, b|y) = C \left(\prod_{i=1}^m \phi(b_i, 0, \sigma) \right) \left(\prod_{i=1}^m \phi(\beta_{0,i}, \mu, \xi) \right) \prod_{j=1}^n \prod_{i=2}^{l_j} F(0|\eta_{ij})^{y_{ij}} \{1 - F(0|\eta_{ij})\}^{1-y_{ij}}$$

となる．ここで ϕ は正規分布の密度関数， C は規格化定数である．Albert (1992) は各パラメータについての全条件付き分布を導出しているが，ハミルトニアン・モンテカルロ法ではその必要がなく，尤度関数と事前分布の積 (3.3) が分かればサンプリングできる．

4. シミュレーション

IRT の分野では，項目パラメータと能力パラメータが同時に推定される際，最尤推定法や条件付き最尤推定法を用いた場合，項目パラメータは一致推定量にならないことが知られている (豊田, 2005)．そのため，パラメータが安定して推定されるか検討が必要である．ここでは，シミュレーションを通じて，パラメータ推定の性質を調べる．シミュレーションでは打者を 75 人，投手を 165 人に設定し，打数は 250 で固定した．これらの設定は事例研究で用いるデータと近い数字にしている．打者能力パラメータ $\beta_{0,j}, \beta_{1,j}$ ($j = 1, \dots, 75$) と，投手難易度パラメータ b_k ($k = 1, \dots, 165$) を標準正規乱数で生成した．ピッチャー $x_{i,j}$ は各打者の各打席ごとに一樣乱数でランダムに選んだ．選択されたピッチャーに対応するパラメータ b_k から確率 $\text{logit}^{-1}(\beta_{0,j} + b_k)$ で 1，確率 $1 - \text{logit}^{-1}(\beta_{0,j} + b_k)$ で 0 を生成し，続けて成功確率

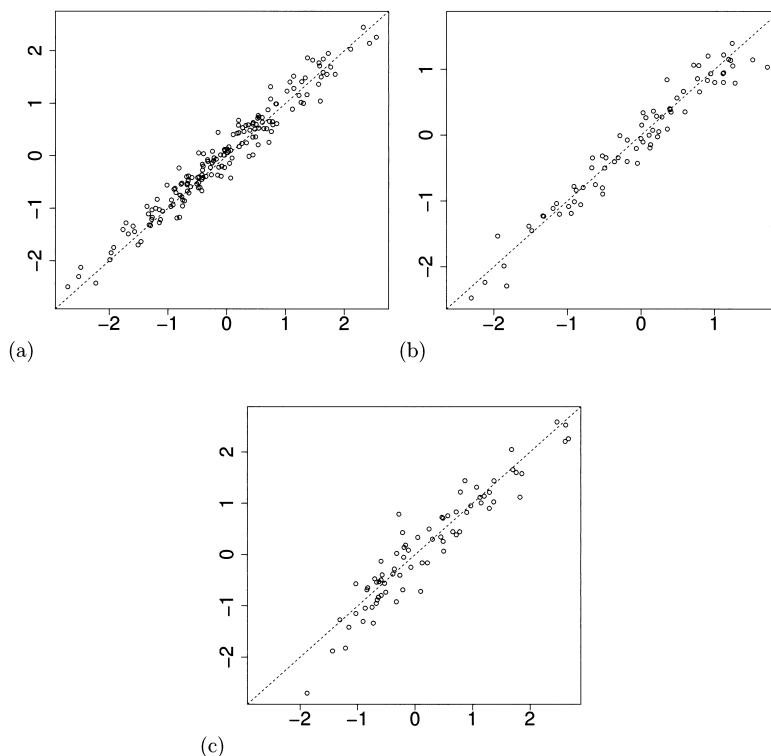


図1. 推定値と真値の比較. 横軸がシミュレーションで仮定した真値, 縦軸が推定値である.
(a) 投手難易度パラメータ b , (b) 打者能力パラメータ β_0 , (c) 打者能力パラメータ β_1 .

$\text{logit}^{-1}(\beta_{0,j} + \beta_{1,j}y_{i-1,j} + b_k)$ のベルヌーイ試行を繰り返して $y_{i,j}$ を生成した. こうして生成した $y_{i,j}$ と $x_{i,j}$ から元のパラメータを推定する. MCMC の試行回数は 2000 回とし, うち前半の 1000 をバーン・イン期間として捨てた. 連鎖の構成数は 3 とした. 事後期待値をパラメータの点推定値とした. 推定値と元のパラメータを比較したプロットを図 1(a)–1(c) に示す. これらの図は, 横軸にシミュレーションで生成した真値, 縦軸に推定値を取ってプロットした. 両方でパラメータが一致していれば, 散布図の各点が直線状に並ぶが, 点はおおむね直線状に分布している. 選手間の能力の相対的比較という観点からは, パラメータの大小の順序の関係が保たれていることが必要であり, そのことは相関係数で評価できる. 推定値と真値の相関係数は, b については 0.98, β_0 が 0.97, β_1 が 0.95 と高い値になった. 決定係数はそれぞれ 0.95, 0.94, 0.89 であった.

また, 区間推定には 90% 信用区間を使用した. 事後分布の両端から確率 $\alpha/2$ の部分を切り取って残った, 中央部の確率 $(1 - \alpha)$ に対応する区間を $100 \times (1 - \alpha)$ 信用区間と呼ぶ. ベイズモデルではパラメータが確率変数であると考えるので, 90% 信用区間は真値を含む確率が 90% となる区間である. 本解析では, MCMC サンプルの標本 5% 点と 95% 点を取り, それぞれを信用下限, 信用上限とすることで, 数値的な 90% 信用区間を得た. 表 2 より, 信用区間がその中にシミュレーションで仮定した真値を含む割合を計算すると, 90% に近い値になっている. このことが, 推定の信頼性を高めている.

投手難易度パラメータの事前分布のパラメータ σ については, 事後期待値が 1.05, 打者能

表 2. 推定された 90% 信用区間が真値を含んだ回数と割合.

| | b | β_0 | β_1 |
|----|------|-----------|-----------|
| 回数 | 151 | 70 | 68 |
| 割合 | 0.92 | 0.93 | 0.91 |

力パラメータの事前分布のパラメータについては μ が -0.11 , ξ が 0.98 であった. それぞれシミュレーションで仮定した真値 $\sigma = 1$, $\mu = 0$, $\xi = 1$ に近い値が推定された.

5. 事例研究

ここで扱うデータは, 2013 年のセ・リーグ, パ・リーグの試合の打席結果である. ただし, 四球, 死球, 敬遠, 犠打, 犠飛は打席から除いた. 年間を通じて 100 回以上打席に立った打者と, その選手に対して投球を行った投手を解析対象とした. セ・リーグとパ・リーグは分けて分析を行った. 対象となる打者はセ・リーグが 75 名, パ・リーグが 74 名, 対象となる投手はセ・リーグが 165 名, パ・リーグが 162 名である. 条件を満たす打者の打席数の平均はセ・リーグで 264.75, パ・リーグで 284.36 であった.

5.1 推定結果の概要

これまでに述べたように, 推定にはハミルトニアン・モンテカルロ法を用いた. MCMC の試行回数は 10000 回とし, うち前半の 5000 をバーン・イン期間として捨てた. 連鎖の構成数は 3 とした.

収束の判定はトレースラインの目視による確認と Gelman (1996) の収束判定指標 \hat{R} により行った. Gelman (1996) では, \hat{R} が 1.2 または 1.1 より小さければ収束したと判断してよいとされている. 本解析においてはすべてのパラメータについて, \hat{R} が 1.1 より小さいことが確認された.

投手難易度パラメータ b_k は値が小さいほど, 打つのが難しい投手である. 事後期待値の値が小さい順に上位 10 名の選手の推定結果を表 3 に示す. 90% 信用区間とオッズ比, 被安打率もあわせて表示した. 被安打率はピッチャーの能力を表す指標のひとつで, 被安打数を打数で割ったものである. また b_k の事後期待値を全投手についてヒストグラムで図示した(図 2 (a)–(b)). 図の点線は b_k の事後期待値の平均である. これらの図から b_k は 0 付近を中心に分布していることがわかる.

表 3 を見ると, 被安打率が 0.2 の金子選手が 1 位にランキングされ, 被安打率が 0.16 のウィリアムス選手が 10 位にランキングされるという逆転が見られる. ただし, ウィリアムス投手は打席数が少ないため, 信用区間の幅は広く推定されている. この結果がなぜ生まれたかを検討する. 金子選手とウィリアムス投手に関して, 打率三割以上の打者と対戦したときの被安打率と, 打率三割未満の打者と対戦したときの被安打率を集計した(表 4). ウィリアムス投手は打率の低い選手に対しての成績が好調な反面, 打率の高い選手には多くヒットを打たれている. 対照的に金子投手はウィリアムス投手と比べ, 打率の低い選手に対しての被安打率は低くないが, 打率の高い選手のヒットを少なく抑えている. そのため, 被安打率と比較したときに特異な値を取る. 本モデルは打者の能力と投手の能力を同時に評価しているため, このような場合, 能力の高い打者を多くアウトにしている投手が高く評価される. このように提案モデルは, 被安打率だけでは見落としてしまうおそれのある優秀な(能力の高い打者に強い)投手をピックアップできる.

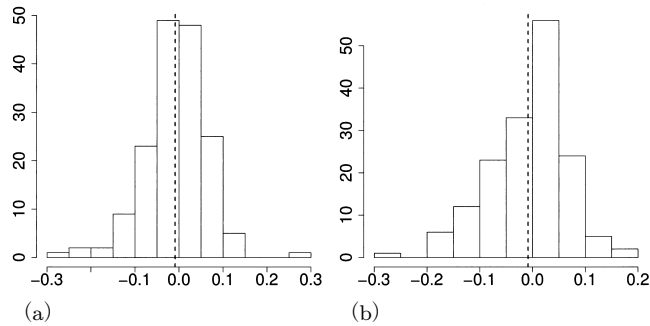


図 2. 投手難易度パラメータ b_i の事後期待値の分布 ((a)セ・リーグ, (b)パ・リーグ).

表 3. 投手難易度パラメータの推定結果.

| | | b | 90%信用下限 | 90%信用上限 | オッズ比 | 被安打率 | 打席数 |
|-------|----------|-------|---------|---------|------|------|-----|
| セ・リーグ | 前田 健太 | -0.26 | -0.44 | -0.09 | 0.77 | 0.19 | 525 |
| | マシソン | -0.24 | -0.49 | -0.03 | 0.78 | 0.17 | 182 |
| | 西村 健太郎 | -0.21 | -0.44 | -0.01 | 0.81 | 0.19 | 217 |
| | 榎田 大樹 | -0.19 | -0.39 | -0.00 | 0.83 | 0.19 | 247 |
| | 藤浪 晋太郎 | -0.17 | -0.35 | -0.00 | 0.84 | 0.21 | 412 |
| | ホールトン | -0.15 | -0.34 | 0.02 | 0.86 | 0.21 | 297 |
| | 山口 鉄也 | -0.15 | -0.37 | 0.04 | 0.86 | 0.19 | 193 |
| | 岡田 俊哉 | -0.14 | -0.33 | 0.04 | 0.87 | 0.21 | 270 |
| | 加藤 康介 | -0.12 | -0.34 | 0.07 | 0.89 | 0.20 | 133 |
| | 久保 康友 | -0.12 | -0.32 | 0.06 | 0.89 | 0.23 | 184 |
| パ・リーグ | 金子 千尋 | -0.29 | -0.45 | -0.13 | 0.75 | 0.20 | 652 |
| | ファルケンボーグ | -0.18 | -0.42 | 0.03 | 0.84 | 0.14 | 115 |
| | 田中 将大 | -0.17 | -0.32 | -0.03 | 0.84 | 0.22 | 598 |
| | 森福 允彦 | -0.17 | -0.38 | 0.02 | 0.85 | 0.20 | 172 |
| | 攝津 正 | -0.16 | -0.32 | -0.02 | 0.85 | 0.22 | 515 |
| | 嘉弥真 新也 | -0.16 | -0.37 | 0.03 | 0.85 | 0.20 | 165 |
| | 佐藤 達也 | -0.15 | -0.35 | 0.03 | 0.86 | 0.19 | 221 |
| | サファテ | -0.15 | -0.37 | 0.04 | 0.86 | 0.16 | 162 |
| | 則本 昂大 | -0.14 | -0.30 | 0.01 | 0.87 | 0.23 | 488 |
| | ウィリアムス | -0.12 | -0.34 | 0.07 | 0.88 | 0.16 | 126 |

表 4. 打率で層別した被安打率.

| | 打率 0.3 未満の打者 | 打率 0.3 以上の打者 |
|--------|--------------|--------------|
| ウィリアムス | 0.16 | 0.35 |
| 金子 千尋 | 0.19 | 0.26 |

打者能力パラメータ $\beta_{0,j}$ は、直前の打席結果がアウトだったときのヒットを打つ能力を表す。野球ではヒットよりもアウトのほうが起こりやすいため、 $\beta_{0,j}$ は選手 j のベースラインとなる能力と解釈できる。解釈しやすくなるよう、逆ロジット関数で変換した $\beta_{0,j}$ の事後期待値に加えてその 90% 信用区間、オッズ比、2.2 節で検討した π_{1j} の $\beta_{0,j}, \beta_{1,j}$ からのプラグイン推定値を算出した。オッズ比は $\beta_{0,j}$ の値を降順に並べて 10 番目の選手を基準としたオッズ比である。表 5 がセ・リーグの選手、表 6 がパ・リーグの選手についての結果である。これらは

表 5. 打者能力パラメータの推定結果(セ・リーグ).

| | $\text{logit}^{-1}(\beta_0)$ | 90%信用下限 | 90%信用上限 | π_1 | β_1 | 打率 | オッズ比 |
|---------|------------------------------|---------|---------|---------|-----------|------|------|
| ルナ | 0.38 | 0.32 | 0.45 | 0.36 | -0.25 | 0.35 | 1.30 |
| 松元 ユウイチ | 0.36 | 0.28 | 0.45 | 0.33 | -0.46 | 0.32 | 1.19 |
| 木村 昇吾 | 0.35 | 0.27 | 0.45 | 0.34 | -0.25 | 0.32 | 1.16 |
| バレンティン | 0.35 | 0.30 | 0.41 | 0.34 | -0.15 | 0.33 | 1.16 |
| 村田 修一 | 0.34 | 0.29 | 0.39 | 0.34 | -0.04 | 0.33 | 1.09 |
| 廣瀬 純 | 0.34 | 0.27 | 0.40 | 0.31 | -0.45 | 0.30 | 1.07 |
| 阿部 慎之助 | 0.33 | 0.28 | 0.39 | 0.30 | -0.56 | 0.29 | 1.06 |
| ブランコ | 0.33 | 0.28 | 0.37 | 0.34 | 0.19 | 0.33 | 1.02 |
| 藤井 淳志 | 0.32 | 0.26 | 0.39 | 0.32 | -0.06 | 0.31 | 1.01 |
| 梶谷 隆幸 | 0.32 | 0.26 | 0.39 | 0.37 | 0.58 | 0.36 | 1.00 |

表 6. 打者能力パラメータの推定結果(パ・リーグ).

| | $\text{logit}^{-1}(\beta_0)$ | 90%信用下限 | 90%信用上限 | π_1 | β_1 | 打率 | オッズ比 |
|--------|------------------------------|---------|---------|---------|-----------|------|------|
| 岡島 豪郎 | 0.35 | 0.28 | 0.41 | 0.33 | -0.20 | 0.32 | 1.12 |
| 浅村 栄斗 | 0.34 | 0.29 | 0.38 | 0.33 | -0.05 | 0.33 | 1.08 |
| 中村 晃 | 0.34 | 0.29 | 0.39 | 0.31 | -0.35 | 0.31 | 1.07 |
| 長谷川 勇也 | 0.33 | 0.29 | 0.38 | 0.33 | -0.05 | 0.33 | 1.06 |
| 今江 敏晃 | 0.33 | 0.28 | 0.38 | 0.33 | -0.02 | 0.32 | 1.04 |
| 平野 恵一 | 0.33 | 0.27 | 0.39 | 0.31 | -0.26 | 0.31 | 1.03 |
| 鶴岡 慎也 | 0.32 | 0.26 | 0.39 | 0.31 | -0.26 | 0.30 | 1.01 |
| 柳田 悠岐 | 0.32 | 0.27 | 0.38 | 0.31 | -0.16 | 0.31 | 1.01 |
| 角中 勝也 | 0.32 | 0.28 | 0.37 | 0.30 | -0.35 | 0.29 | 1.00 |
| 藤田 一也 | 0.32 | 0.28 | 0.37 | 0.30 | -0.35 | 0.29 | 1.00 |

いわば対戦した投手の効果を補正した打率である。

また打者能力パラメータ $\beta_{1,j}$ は値が正であれば、直前の打席でヒットを打つと続けてヒットを打つ確率が上がり、負であれば直前の打席でヒットを打てなかった場合にヒットを打つ確率が上がると解釈できる。MCMC サンプルから $\beta_{1,j}$ の 90% 信用区間を求め、これが 0 を含まないとき、調子の波が存在する選手とする。そのような選手を表 7-8 に示した。

表 5 の $\text{logit}^{-1}(\beta_0)$ を見ると、阿部選手は打率と比較して β_0 が大きく推定されており、梶谷選手に関しては打率と比較して β_0 が小さく推定されている。それに対応して、阿部選手は β_1 が小さい値を取り、梶谷選手は β_1 が大きい値を取っている。 β_0 と β_1 を総合した指標である π_1 は、打率に近い値を取っている。表 7 より、阿部選手、梶谷選手は「調子に波がある」プレイヤーである。

5.2 カイ二乗検定を用いた考察

ここでは、古典的なカイ二乗検定を用いて調子の波について分析を行い、提案モデルの方法が既存の結果と矛盾しないことを示す。調子の波を定量的に評価するもう一つの方法は、ある打席での成功・失敗と、その直前の打席での成功・失敗との関係を調べることである。このアプローチは Albert (2008) が用いたものである。いま 2×2 分割表の k 行 l 列目のセルを c_{kl} とする ($k = 1, 2, l = 1, 2$)。選手 j についての分割表を、 $c_{11} = \sum_i I\{y_{i-1,j} = 0, y_{i,j} = 0\}$, $c_{12} = \sum_i I\{y_{i-1,j} = 0, y_{i,j} = 1\}$, $c_{21} = \sum_i I\{y_{i-1,j} = 1, y_{i,j} = 0\}$, $c_{22} = \sum_i I\{y_{i-1,j} = 1, y_{i,j} = 1\}$ と

表 7. 打者能力パラメータ β_1 の推定結果(セ・リーグ).

| | β_1 | 90%信用下限 | 90%信用上限 | オッズ比 | β_0 | 打率 | π_1 |
|--------|-----------|---------|---------|------|-----------|------|---------|
| 福留 孝介 | -1.12 | -2.14 | -0.22 | 0.33 | 0.23 | 0.20 | 0.20 |
| 堂林 翔太 | -0.76 | -1.43 | -0.13 | 0.47 | 0.25 | 0.22 | 0.22 |
| 中村 悠平 | -0.74 | -1.53 | -0.00 | 0.48 | 0.26 | 0.23 | 0.24 |
| クラーク | -0.56 | -1.13 | -0.03 | 0.57 | 0.27 | 0.24 | 0.24 |
| 阿部 慎之助 | -0.56 | -1.02 | -0.09 | 0.57 | 0.33 | 0.29 | 0.30 |
| 鳥谷 敬 | -0.48 | -0.89 | -0.07 | 0.62 | 0.31 | 0.28 | 0.28 |
| 山田 哲人 | 0.46 | 0.00 | 0.93 | 1.59 | 0.26 | 0.28 | 0.29 |
| 梶谷 隆幸 | 0.58 | 0.10 | 1.06 | 1.79 | 0.32 | 0.36 | 0.37 |

表 8. 打者能力パラメータ β_1 の推定結果(パ・リーグ).

| | β_1 | 90%信用下限 | 90%信用上限 | オッズ比 | β_0 | 打率 | π_1 |
|--------|-----------|---------|---------|------|-----------|------|---------|
| 大崎 雄太朗 | -1.30 | -2.53 | -0.24 | 0.27 | 0.27 | 0.22 | 0.23 |
| 鬼崎 裕司 | -0.75 | -1.36 | -0.17 | 0.47 | 0.31 | 0.26 | 0.27 |
| バルディリス | -0.46 | -0.89 | -0.05 | 0.63 | 0.32 | 0.29 | 0.29 |
| 松田 宣浩 | -0.45 | -0.87 | -0.04 | 0.64 | 0.29 | 0.26 | 0.27 |
| ヘルマン | 0.42 | 0.06 | 0.78 | 1.52 | 0.30 | 0.32 | 0.33 |
| 大引 啓次 | 0.52 | 0.06 | 0.97 | 1.68 | 0.24 | 0.26 | 0.27 |
| 西川 遥輝 | 0.56 | 0.02 | 1.09 | 1.75 | 0.24 | 0.26 | 0.27 |

表 9. 阿部慎之助選手についての分割表.

| | 0 | 1 |
|---|-----|----|
| 0 | 158 | 76 |
| 1 | 76 | 22 |

して作成する. I は指標関数であり, 等式を満たすとき 1, 他の場合 0 の値を取る. 具体例として阿部慎之助選手についての分割表を, 表 9 に示した.

この表に対して独立性についてのカイ二乗検定を行った. 検定の結果, 帰無仮説が棄却された選手は, 直前の打席での成功・失敗が, 次の打席結果に影響を与えると解釈できる. 10%水準で有意になった選手を「調子の波がある」選手とする. 「調子の波がある」選手を表 10 にまとめた. 理解を容易にするために, クラメールの連関係数も併記した. カイ二乗検定では有意とされない選手についても, 表 7-8 に上げた「調子の波がある」選手については, 並べて記載した. 有意となった選手には, p 値の右に傍点を振った.

連関係数が正の値をとるときは, 直前の打席結果がヒットだった場合, 次の打席でのヒットが出やすくなると解釈でき, 負の値をとるとき, 直前の打席結果がアウトだった場合, 次の打席でのヒットが出やすくなると解釈できる. 表 7-8 と表 10 を比較すると, カイ二乗検定で「調子の波がある」と判断された選手は, 提案手法で「調子の波がある」と判断された選手にすべて含まれている. このことは提案手法が従来手法と矛盾しないことを示している. 逆に提案手法で「調子の波がある」と判断された選手が, カイ二乗検定で「調子の波がある」と判断されるとは限らない. これは提案手法がヒット・アウトの系列の情報や相手投手の情報を利用して, より感度の高い検出を行っているためである. また仮説検定では, 打席結果の情報を, 「有意」か「有意でない」という二値の解釈に落としている. 提案手法は各パラメータから調子の波の強さ

表 10. カイ二乗検定の結果 10% 水準で有意になった打者一覧(表 7-8 に上げた選手もあわせて記載).

| | | p 値 | 連関係数 | 打席数 |
|-------|--------|--------|-------|-----|
| セ・リーグ | 梶谷 隆幸 | 0.05 . | 0.14 | 219 |
| | 鳥谷 敬 | 0.09 . | -0.09 | 444 |
| | 阿部 慎之助 | 0.09 . | -0.10 | 333 |
| | 堂林 翔太 | 0.10 . | -0.11 | 295 |
| | 山田 哲人 | 0.10 | 0.10 | 295 |
| | 福留 孝介 | 0.11 | -0.12 | 212 |
| | クラーク | 0.16 | -0.08 | 346 |
| | 中村 悠平 | 0.22 | -0.10 | 201 |
| パ・リーグ | ヘルマン | 0.06 . | 0.10 | 434 |
| | 大引 啓次 | 0.07 . | 0.11 | 327 |
| | 鬼崎 裕司 | 0.08 . | -0.12 | 258 |
| | 西川 遥輝 | 0.09 . | 0.12 | 232 |
| | バルディリス | 0.10 | -0.09 | 420 |
| | 松田 宣浩 | 0.11 | -0.08 | 484 |
| | 大崎 雄太朗 | 0.14 | -0.15 | 133 |

や、その信用区間といった、より多くの情報を与えることができる。

5.3 ヒットの予測

提案モデルでは打者 j とその直前の打席結果、および投手 k が決まれば、ヒットを打つ確率を予測できる。直前の打席結果がアウトだった場合、打者 j と投手 k の対戦でヒットが出る確率は、

$$(5.1) \quad p_{0,j,k} = \text{logit}^{-1}(\beta_{0,j} + b_k).$$

直前の打席結果がヒットだった場合、

$$(5.2) \quad p_{1,j,k} = \text{logit}^{-1}(\beta_{0,j} + \beta_{1,j} + b_k)$$

と表す。直前の打席結果がアウトだった場合の、打者 j と投手 k の対戦回数を $n_{0,j,k}$ 、そのヒットの回数を $h_{0,j,k}$ と表す。直前の打席結果がヒットだった場合の、打者 j と投手 k の対戦回数を $n_{1,j,k}$ 、そのヒットの回数を $h_{1,j,k}$ と表す。実際のヒットの割合 $h_{0,j,k}/n_{0,j,k}$ 、 $h_{1,j,k}/n_{1,j,k}$ と、提案モデルによる確率の推定値 $p_{0,j,k}$ 、 $p_{1,j,k}$ を比較する。以降、単に実際の割合といった場合は $h_{0,j,k}/n_{0,j,k}$ または $h_{1,j,k}/n_{1,j,k}$ 、推定確率といった場合は $p_{0,j,k}$ または $p_{1,j,k}$ を指すものとする。推定確率と実際の割合の差を、図 3(a)および図 3(b)に示す。 $n_{0,j,k}$ または $n_{1,j,k}$ が大きくなるにつれ、実際の割合と予測確率の差は小さくなる傾向が見られた。

$n_{0,j,k}$ または $n_{1,j,k}$ が一定以上大きいにも関わらず、予測確率と実際の割合が大きく食い違う場合、その打者と投手の組み合わせに固有の特殊な事情がある可能性を考える。 $n_{0,j,k}$ または $n_{1,j,k}$ が 10 を超える組み合わせについて、実際の割合と予測確率の差が 0.2 を超えたものを表 11、12 にまとめた。

阿部慎之助選手はベースラインとなる能力パラメータ β_0 が大きく推定された(表 5)優れた選手だが、スタンリッジ投手とは 13 回対戦して 1 回しかヒットを打っていない。いわばスタンリッジ投手は阿部慎之助選手の天敵といえるだろう。本モデルは打者の能力と投手の能力を測る基準となるため、このような特異な関係性を見出すことに利用できる。

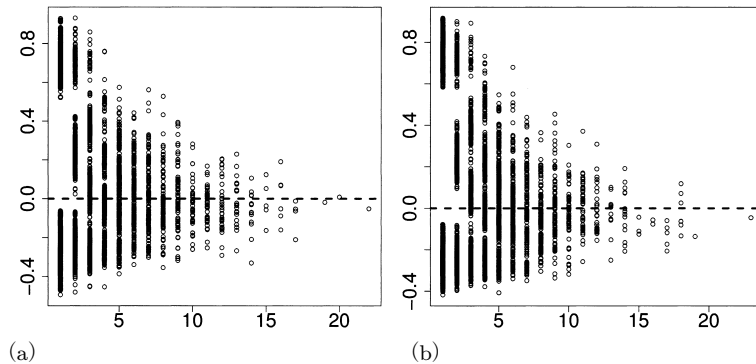


図 3. 推定確率と実際の割合の差. 横軸に n , 縦軸に差を取った((a)セ・リーグ, (b)パ・リーグ).

表 11. 予測確率と実際の割合の比較(セ・リーグ).

| 打者 | 投手 | 一打席前 | 実際の割合 | 予測確率 | n |
|--------|---------|------|-------|------|-----|
| 鳥谷 敬 | 大竹 寛 | 0 | 0.00 | 0.33 | 14 |
| 新井 良太 | 藤井 秀悟 | 0 | 0.00 | 0.29 | 12 |
| 阿部 慎之助 | スタンリッジ | 0 | 0.08 | 0.34 | 13 |
| ルイス | 菅野 智之 | 0 | 0.00 | 0.25 | 11 |
| 鳥谷 敬 | 三浦 大輔 | 0 | 0.09 | 0.34 | 11 |
| 鳥谷 敬 | 八木 亮祐 | 0 | 0.08 | 0.34 | 12 |
| 長野 久義 | 大野 雄大 | 0 | 0.54 | 0.31 | 13 |
| 梵 英心 | メッセンジャー | 0 | 0.09 | 0.31 | 11 |
| バレンティン | メッセンジャー | 0 | 0.12 | 0.33 | 17 |
| 菊池 涼介 | メッセンジャー | 0 | 0.00 | 0.21 | 14 |
| マートン | 澤村 拓一 | 0 | 0.50 | 0.30 | 12 |
| 新井 貴浩 | 三浦 大輔 | 0 | 0.50 | 0.30 | 12 |

表 12. 予測確率と実際の割合の比較(パ・リーグ).

| 打者 | 投手 | 一打席前 | 実際の割合 | 予測確率 | n |
|--------|-------|------|-------|------|-----|
| 大引 啓次 | 田中 将大 | 0 | 0.06 | 0.26 | 17 |
| 松井 稼頭央 | 十亀 剣 | 0 | 0.07 | 0.28 | 14 |
| 内川 聖一 | 西 勇輝 | 0 | 0.08 | 0.33 | 13 |
| 本多 雄一 | 牧田 和久 | 0 | 0.50 | 0.19 | 12 |
| 中村 晃 | 牧田 和久 | 0 | 0.64 | 0.26 | 11 |
| 大谷 翔平 | 田中 将大 | 0 | 0.00 | 0.26 | 11 |

5.4 パラメータ推定の安定性

4 節ではシミュレーションデータからパラメータ推定の安定性を議論したが, ここでは実データを用いて改めて検討する. 一般に項目反応理論では項目の困難度パラメータを正確に推定するためには, 多くの被験者が必要であるとされている. 被験者の数を少なくともった場合でも, 被験者の数を多くした場合と同様の推定結果が得られるならば, モデルとそのパラメータはより信頼できるものとなる. そこでパラメータ推定の対象となる打者が約半数となるように

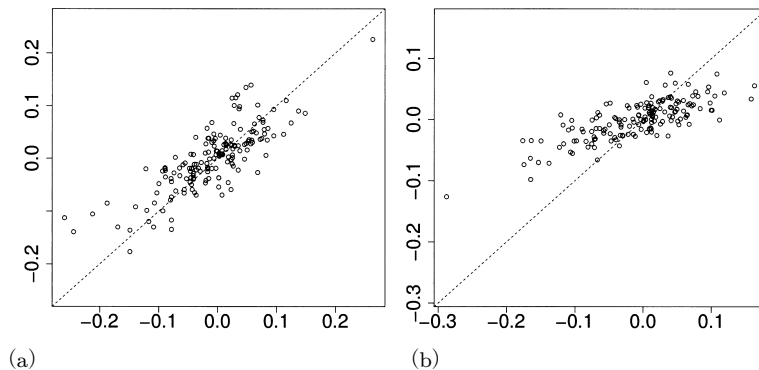


図 4. 投手難易度パラメータの推定結果の比較((a)セ・リーグ, (b)パ・リーグ).

ランダムサンプリングで打者を選び直した新たなデータセットに対してモデルのフィッティングを行い、5.1 節で推定した投手難易度パラメータと、新たなデータセットから推定した投手難易度パラメータを比較した。セ・リーグについての比較が図 4(a)、パ・リーグについての比較が図 4(b)である。これらの図は縦軸にランダムサンプリングしたデータセットから推定した投手難易度パラメータ、横軸に 5.1 節で推定した投手難易度パラメータをとりプロットした。点はおおむね直線状に分布しており、相関係数はセ・リーグで 0.81、パ・リーグで 0.77 と高い値になった。決定係数はそれぞれ 0.77、0.48 であった。このことはサンプルサイズが小さくても結果に矛盾が生じないことを意味しており、安定したパラメータの推定が行えた見込みが高いといえる。ただし、図 4(b)においては、線形回帰した場合に傾きが 1 より小さくなるような系統的な差が生まれている。ベイズモデルではサンプルサイズが小さい場合、事前分布の影響で推定値がシュリンクするためである。順序の関係は保たれているため、選手間の相対的比較をする上では大きな問題とならないが、提案モデルを利用する際は、この性質を認識する必要がある。

6. まとめ

本研究では打席結果を 0 または 1 にコード化したデータから、投手の能力と打者の能力を推定するモデルを提案した。ハミルトニアン・モンテカルロ法を用いたパラメータ推定の安定性はシミュレーションと事例研究で確かめた。事例研究では、従来用いられる被安打率では低く評価されるおそれのある優れた選手を見つけることができた。また 0 または 1 のプロセスにマルコフ性を仮定することで、打者の「調子の波」を捉えることができた。

謝 辞

本研究の貸与データは情報・システム研究機構の新領域融合研究プロジェクト『社会コミュニケーション』データ中心科学リサーチコモンズ事業『人間・社会データ』の支援を受けたものです。また、本研究の一部は科学科研究費(若手研究(B)No.15K21379)の助成を受けたものです。心より感謝申し上げます。

参 考 文 献

- Albert, J. (1992). Bayesian estimation of normal ogive item response curve using Gibbs sampling, *Journal of Educational Statistics*, **17**(3), 251–269.
- Albert, J. (2008). Streaky hitting in baseball, *Journal of Quantitative Analysis in Sports*, **4**(1), DOI: 10.2202/1559–0410.1085.
- Albert, J. and Bennett, J. (2003). *Curve Ball*, Springer-Verlag, New York. (加藤貴昭, 後藤寿彦 訳 (2004). 『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.)
- Baker, F. B. (1992). *Item Response Theory: Parameter Estimation Techniques*, Marcel Dekker, New York.
- Bilog-MG (2005). Scientific Software International, <http://www.ssicentral.com/irt/index.html>.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, Cambridge. (元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇 訳 (2008). 『パターン認識と機械学習 下(ベイズ理論による統計的予測)』, 丸善出版. 東京.)
- Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm, *Psychometrika*, **46**(4), 443–459.
- Chalmers. R. (2012). Mirt: A multidimensional item response theory package for the r environment, *Journal of Statistical Software*, **48**(1), 1–29.
- De Ayala, R. J. (2008). *The Theory and Practice of Item Response Theory*, The Guilford Press, New York.
- Fox, J. P. (2010). *Bayesian Item Response Modeling Theory and Applications*, Springer-Verlag, New York.
- Gelman, A. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC Interdisciplinary Statistics, London.
- Hambleton R. K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*, Vol. 2, Sage Publications, New York.
- Jong, M. G. D., Steenkamp, J. B. E. M., Fox, J. P. and Baumgartner, Hans (2008). Using item response theory to measure extreme response style in marketing research: A global investigation, *Journal of Marketing Research*, **45**(1), 104–115.
- Karlin, S. (1969). *A First Course in Stochastic Processes*, Academic Press, Cambridge. (佐藤健一, 佐藤由身子 訳 (1974). 『数理解析とその周辺 3 確率過程講義』, 産業図書, 東京.)
- Lord, F. M. (1952). A theory of test scores, *Psychometric Monographs*, No. 7, Psychometric Corporation, Richmond.
- Patz, R. J. and Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response models, *Journal of Educational and Behavioral Statistics*, **24**(2), 146–178.
- Patz, R. J. and Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses, *Journal of Educational and Behavioral Statistics*, **24**(4), 342–366.
- Raykov, T. and Calantone, R. J. (2014). The utility of item response modeling in marketing research, *Journal of the Academy of Marketing Science*, **42**(4), 337–360.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response analysis, *Journal of Statistical Software*, **17**(1), 1–25.
- 豊田秀樹 (2005). 『項目反応理論[理論編]』, 朝倉書店, 東京.
- Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society, Series B (Methodological)*, **38**, 290–295.

Measurements of Baseball Players' Batting Abilities

Ko Abe¹, Takenori Sakumura² and Toshinari Kamakura²¹Graduate School of Science and Engineering, Chuo University²Department of Industrial and Systems Engineering, Chuo University

Statistics of player performance is an important part of baseball. Many stats have been proposed to measure a batter's performance, including batting average, on-base percentage, and slugging percentage. In the field of baseball analytics, the "streakiness" of batter's ability is often discussed using a binary sequence of hitting outcomes for a player during a season.

Unlike previous studies, which use data from the batter, we take a different approach. To analyze a batter's performance, we simultaneously model the pitcher and batter's ability. To model a batter's streakiness, we employ an extension of a one-parameter logistic item response model. Item response theory (IRT) estimates both the subject's ability and item difficulty. In this study, the ability parameter and item difficulty parameter correspond to the batter's ability and pitcher's ability, respectively. Although simplicity is thought to make the one-parameter logistic model easy to interpret, our model incorporates numerous parameters. However, using the odds ratio allows athletes to be compared.

We express streakiness by the interactions of previous at bats and imposing the Markov property on batting data. Specifically, we use MCMC in the Hamiltonian Monte Carlo method (also called the hybrid Monte Carlo method). The computation of Gibbs sampling is complex and time consuming, but the Hamiltonian Monte Carlo method is easily computed once the prior distribution and the likelihood function are defined. Our simulation study shows that the true and estimated values agree well. Additionally, the calculated proportion of times that the credible interval contains the true value is close to the nominal value.

To demonstrate the usefulness of our proposed method, we applied it to analyze actual data from Japanese professional baseball. Two-way tables can measure the dependence of the previous success and the current success by the Pearson chi-square statistic and the corresponding p-value of the test of independence. The results provide more information and are consistent with the results of chi-square test. Because comparing streakiness in the hypothesis test is difficult, we ranked streaky players from the credible intervals and the posterior means. IRT requires many subjects to estimate item difficulty parameters. Although we estimated the parameters using fewer batters, the results from our method are similar to those from IRT.