

状況を考慮した MLB打者の能力推定

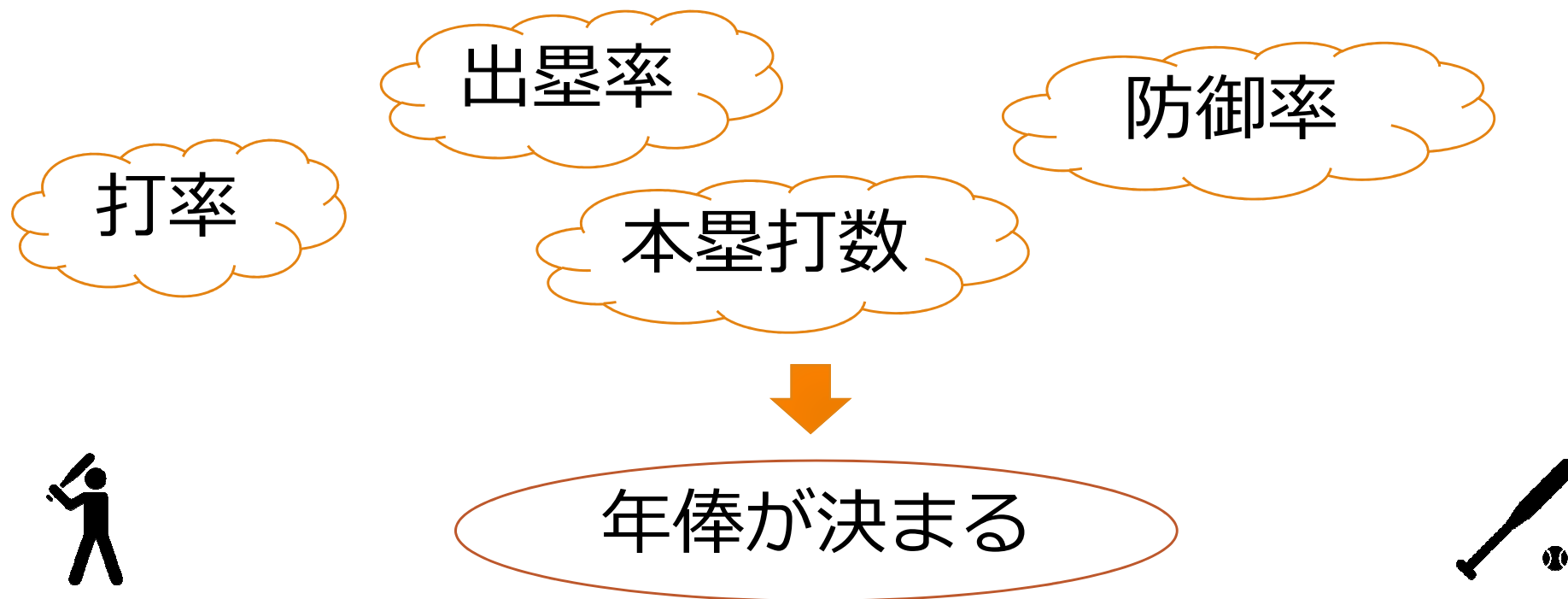
尾崎研究室 4年

古野啓太



はじめに

いろいろな能力を評価



はじめに



打率

出塁率



適切な評価指標なのか？

ここが大きな問題

はじめに

難しいピッチャーと数多く対戦してるバッターは可哀想



投手の能力や状況を考慮した数値が欲しい！

はじめに

評価の方法について分析



ベースとなる理論

項目反応理論

例（テストの場合）

テストの平均点

	試験X	試験Y
A大学	50点	
B大学		80点

B大学の学生が優秀かと言われたらそうとは限らない

テストの難易度が分からないから

例（テストの場合）

テストの平均点

	試験X	試験Y
A大学	50点	
B大学		80点

試験Xの方が難しかったら？

試験Yの方が簡単だったら？

例（テストの場合）

テストの平均点

	試験X	試験Y
A大学	50点	
B大学		80点

問題、受験者の両方の能力を考慮しないと適切な評価をすることは不可能

例（テストの場合）

これらを考慮することができる理論が

項目反応理論

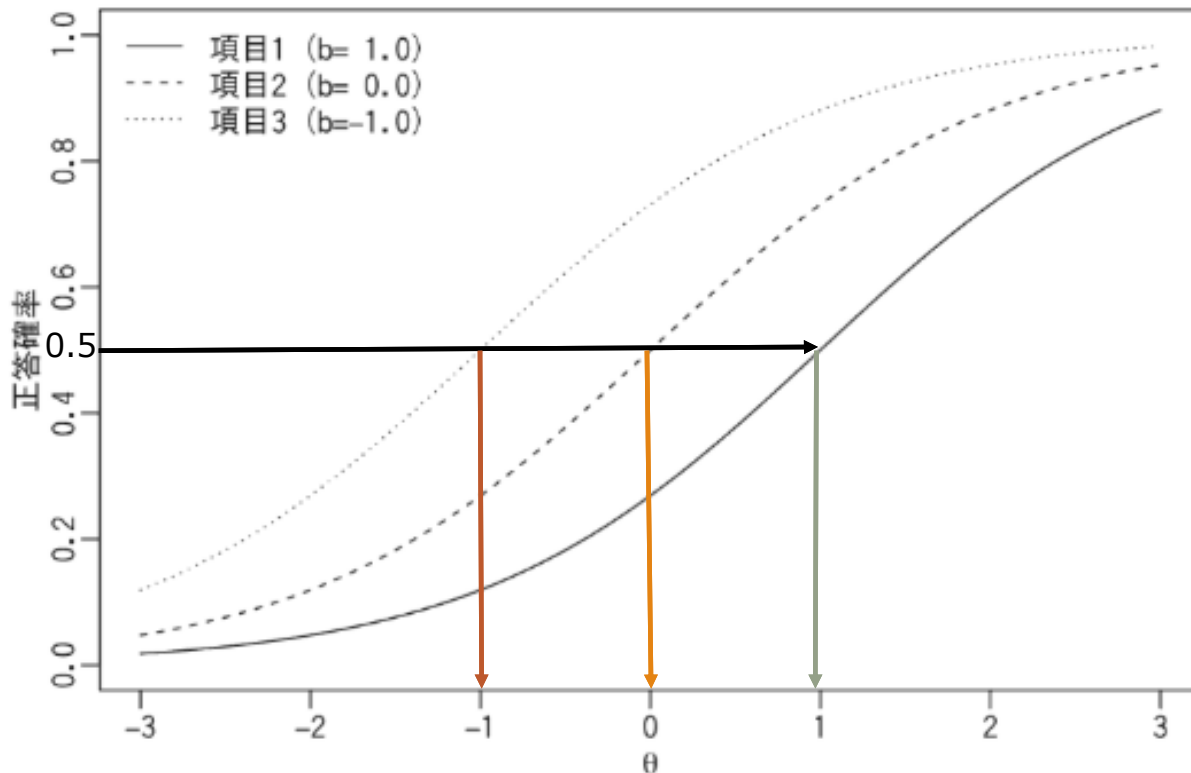
項目反応理論

ある試験の解答（正答は1、誤答は0、－は選択せず）

	問題1	問題2	問題3	問題4	問題5
Aさん	1	0	1	1	1
Bさん	0	1	1	－	1
Cさん	1	0	1	0	－

各問題の難易度、各受験者の能力を推定

項目反応理論



能力 θ の受験者が
問題 j に正答する確率

↓

$$P_j(\theta) = \frac{1}{1 + \exp(-Da(\theta - b_j))}$$

θ 受験者の能力

j 項目

b_j 問題の難易度

a 項目識別力 (今回は定数)

D 尺度因子 (今回は定数)

項目反応理論

ある試験の解答（正答は1、誤答は0、－は選択せず）

	問題1	問題2	問題3	問題4	問題5
受験者A	1	0	1	1	1
受験者B	0	1	1	－	1
受験者C	1	0	1	0	－

各問題の難易度、各受験者の能力を推定

項目反応理論

ある打席の結果（出塁は1、凡退は0、－は対戦無）

	投手1	投手2	投手3	投手4	投手5
打者A	1	0	1	1	1
打者B	0	1	1	－	1
打者C	1	0	1	0	－

各投手の難易度、各打者の能力を推定

参考文献①

- ・ 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, 作村 建紀, 鎌倉 稔成

統計数理 (2017), 第65巻 第2号 235-249

投手の能力を考慮した推定を行っている

ランナー状況などは考慮していない

参考文献②

- ・ 項目反応理論を用いたプロ野球選手の評価について
時光順平, 鳥越規央
数理解析研究所講究録 第 1804 巻 2012 年 21-29

ランナー状況を考慮している

各投手の難易度、各打者の能力を推定

参考文献②

	ランナー1塁	ランナー2塁
投手A	1	0	
投手B	0	1	
投手C	1	0	

出塁は1、凡退は0

データ

2018年のメジャーリーグの打席データを使用
イニング、打者選手ID、カウントの中身（空振りストライク、見逃しストライク、ボールなど）、打席結果など

2018年のMLB全ての試合のデータが
入っている

打者984人 投手778人

本研究

①状況を考慮しない場合

②得点圏、非得点圏

③高難易度、低難易度



①状況を考慮しない場合

ランナー状況などを考慮しない

出塁率



ランキング化

能力値



ランキング化

②-1 得点圏

ランナーが2塁か3塁にいる時

出塁率



ランキング化

能力値



ランキング化

②-2 非得点圏

ランナーなしが1塁にいる時

出塁率



ランキング化

能力値



ランキング化

分析

推定は全打者を対象に行った



ランキングなどの分析は規定打席数(502打席以上)の打者に限定


3打数3安打で出塁率10割の選手が存在してしまうため

分析（状況を考慮しない場合の例）

選手ID	出塁率	能力値	出塁率順位	能力値順位
yelic001	0.421	2.274	5	2
freef001	0.404	2.044	10	4
carpm002	0.388	1.732	16	10
bettm001	0.448	1.779	2	9
nimmb001	0.411	1.608	7	14
ramij003	0.405	1.404	9	22

③ 高難易度、低難易度

	投手1	...
打者A		
...		



	<u>状況1</u>	...
打者A		
...		

各状況の難しさが推定できる

分析

144通りのシチュエーションに分ける

アウトカウント(3)×ランナー(8)
×試合の前半か後半か(2)
×勝ってるか同点か負けてるか(3)

本研究

難易度の平均



標準偏差

難易度の平均 + 標準偏差 \leq 難易度の数値

難しい状況

難易度の平均 - 標準偏差 \geq 難易度の数値

簡単な状況

分析

高難易度の例

0アウト3塁 試合の後半 同点
0アウトランナー無 試合の前半 負けている
0アウトランナー無 試合の前半 同点

分析

低難易度の例

0アウト満塁 試合の前半 同点
2アウト2,3塁 試合の後半 同点
2アウト満塁 試合の前半 同点

本研究

難しい状況

	投手A	...
打者A		
...		

簡単な状況

	投手A	...
打者A		
...		

同じように当てはめて推定

分析

全状況



ランキング化

難しい状況



ランキング化

簡単な状況



ランキング化

打者を対象としたランキング

分析

各打者の順位の例

選手ID	全状況の順位	高難易度順位	低難易度順位
mazan001	69	7	73
vottj001	5	6	46
rojam002	76	122	3
olsom001	63	116	13

まとめ

- ・ 状況を高難易度、低難易度に分けての分析は意味のある研究

- ・ 高難易度、低難易度に分けた結果、高難易度に強い打者、低難易度に強い打者を見つけ出すことができたことが、新しい発見

今後の課題

- ・ 今回は2018年のデータを用いたが、他の年でも分析を試みる
- ・ 出墨率以外の能力も対象にして推定