

Harville の公式と Gini 係数を用いた競馬の着順予想

佐藤 裕也 華山 宣胤

尚美学園大学大学院 芸術情報研究科 情報表現専攻

1. はじめに

競馬においてある出走馬の着順は、馬の体調や斤量などに関する競走馬の条件や、馬場の状態、気象、レースに関する条件に依存すると考えられる。そこで、馬券購入者がこれらの条件を正確かつ十分に検討していると考えた場合、出走馬の得票率は、各馬の1着になる確率（馬券購入者が考える主観確率の平均）を表したものと考えられる。しかし、図表1が示す2013年5・6月に中央競馬（JRA）で開催された重賞レース（ハンデ戦を除く）における1番人気の得票率と実際に1着になった馬の人気順からわかるように、20レース中1番人気の馬が1着となるレース数の得票率に基づく期待値がおよそ6回であるのに対し、実際に1番人気の馬が1着となったレースは8レースであった。

図表1. 2013年5・6月に中央競馬（JRA）で開催された重賞レース（ハンデ戦を除く）における1番人気の得票率と実際に1着になった馬の人気順

レース名	一番人気の得票率	一着馬の人気順	レース名	一番人気の得票率	一着馬の人気順
京都新聞杯	0.59	1	鳴尾記念	0.23	6
新潟大賞典	0.27	6	安田記念	0.20	1
NHKマイルC	0.27	10	エフソnC	0.18	4
京王スプリングC	0.17	1	マーメイドS	0.28	8
ヴィクトリアマイル	0.25	1	函館スプリントS	0.38	6
平安S	0.56	1	ユニコーンS	0.34	3
オークス	0.24	9	東京ジャンプS	0.32	4
京都ハイジャンプ	0.40	1	宝塚記念	0.33	2
日本ダービー	0.29	1	ラジオNIKKEI賞	0.21	8
目黒記念	0.22	4	CBC賞	0.22	1
得票率に基づき一番人気が一着になるレース数の期待値		一番人気馬が一着だったレース数			
5.97		8			

このように人気の高い出走馬が実際の結果に見合うほどの得票を必ずしも集めないという傾向は「本命バイアス」と呼ばれている。⁽¹⁾そして、このような傾向は、出走馬の人気に偏りがあるほど強く表れると言われている。⁽²⁾一方、得票数の偏りだけでなく、馬券購入者のリスクを表す得票率の分散も本命バイアスと関係を持つと考えられる。⁽²⁾

そこで、本稿では、本命バイアスの傾向が顕著に見られる1番人気馬の得票率に注目し、1番人気馬が勝つレースと1番人気以外が勝つレースの判別分析を行い、1番人気馬が勝つレースの特徴の把握を行う。判別分析を行うための、説明変数としては1番人気の馬の得票率、人気の偏りを表す得票率のGini係数、馬券購入者のリスクを表す得票率の分散を用いる。

2. データ

使用したデータは、JRAのホームページ上にまとめられている2013年のJRA主催の重賞レースの成績表である。⁽³⁾この成績表には、レースの結果の他に馬の名前、年齢、勝馬の得票数等が記載されている。なお、ハンデキャップ戦のデータは除外した。

3. 手法

本研究での分析手順は以下の通りである。

- (1) データから得票率の推計
- (2) Gini係数を算出
- (3) 分散を算出

(4) 1番人気の馬の得票率とGini係数と標準偏差を用いて判別分析を行う

なお、各手順では以下のようなことを行う。

(1) <得票率の推計>

取得したデータには得票率が含まれていない。そこで、下記の計算式で各レースごとに各馬の得票率の推計を行う。

得票率 = (最終オッズ × 勝馬への投票数 ÷ 全投票数) ÷ 最終オッズ

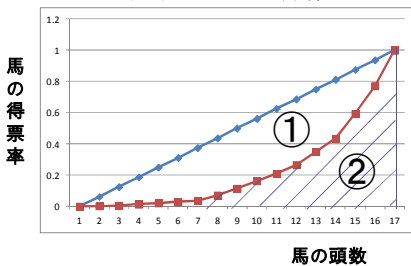
(2) <Gini係数の算出>

(1)で推計した得票率を用いて、各レースごとにGini係数を算出する。Gini係数の算出方法は、レースに出走する馬ごとに得票率の低い馬から図表1の赤線のように得票率を累積する。

A study of prediction of finishing order of Horse Racing using Gini coefficient and Harville's approximation
Yuya Sato Shobi University
Nobutane Hanayama Shobi University

その後、①と②の面積を計算し、下記の式で計算を行うことで Gini 係数を算出することができる。しかしながら多くの競馬レースにおいて 1 着になる馬は、5 番人気以上の馬である。そこで本稿では、6 番人気以下の馬が 1 着になることはイレギュラーとして考え、5 番人気以上の馬のみの得票率を用いて Gini 係数を算出する。

図表 2.Gini 係数のイメージ



$$\text{ジニ係数} = \frac{\text{①の面積}}{\text{①の面積} + \text{②の面積}}$$

(3) <分散の算出>

(1)で推計した得票率を用いて、各レースごとに分散を算出する。ただし、(2)と同じように 5 番人気以上の馬のみの得票率を用いて分散を算出する。

(4) <判別分析>

本研究の目的である一番人気が勝つレースの特徴を把握するために判別分析を行う。目的変数には実際の各レースにおいて 1 番人気の馬が 1 着のレースか 1 番人気以外の馬が 1 着のレースかを設定した。説明変数には 1 番人気の馬の得票率と(2)で算出された Gini 係数と(3)で算出された分散を用いた。

4. 分析結果の検討

図表 3.分析の結果

判別式	判別係数	標準判別係数	t値	p値	判定
log5番人気Gini	0.507	0.238	0.50	0.479	[]
5番人気分散	39.864	0.765	3.89	0.051	[]
log1番人気得票率	-0.659	-0.305	0.80	0.374	[]
定数項	-0.866				

前項で判別分析を行った結果が、上記の図表の通りである。

図表 3 から得票率の偏りを表す Gini 係数と馬券購入者のリスクを表す分散の標準判別係数がそれぞれ 0.238 と 0.765 という正の値として推定された。このことから、競馬レースにおいて得票率の分散と得票率の Gini 係数が高いほど 1 番人気で 1 着になる可能性が高いと言える。一方、1 番人気の得票率の標準判別係数が -0.305 と負の値として推定された。このことから、1 番人気の得票率が高いほど 1 番人気で 1 着になる可能性

が低くなると言える。これらのことから 1 番人気で 1 着になるレースの特徴は、得票率の偏りと分散が大きく、1 番人気の馬の得票率が低いレースと言える。

次に、図表 4 に示す分析精度を見てみよう。図表から、判別の中率が 66.7%と十分な精度とは言えない。

図表 4. 分析精度

判別の中率 (%)	66.7%
誤判別の確率 (%)	36.1%
マハラノビス平方距離	0.503846
相関比	0.105

しかし、この分析結果にしたがって、1 番人気で 1 着になると判別されたレースについてだけ馬券を購入し、1 番人気以外が 1 着となると判別されたレースについては馬券を購入しないという賭け方を考えてみよう。すると、図表 5 に示す判別クロス表から分かるように、1 番人気で 1 着になると判別されて実際に 1 番人気で 1 着になったレース数は 1 番人気で 1 着になると判別されたが実際は 1 番人気以外が 1 着になったレースの数を上回っているため、年間を通しては黒字となるのが期待できる。さらに、単純に控除率を 25%と考えた場合、毎回 100 円分の馬券を購入すると、年間の黒字額はおよそ 990 円であると算出できる。

図表 5.判別クロス表 (n 表)

	1	0	全体
0以上	17	15	32
0未満	20	53	73
全体	37	68	105

5. まとめ

本稿では、競馬における 1 番人気の馬に注目し、一番人気で勝つレースと 1 番人気以外で勝つレースの判別分析を行った。その結果、1 番人気の馬の勝利確率は、人気の偏りを表す得票率の Gini 係数と馬券購入者のリスクを表す得票率の分散が大きいほど高く、1 番人気の得票率が高いほど低くなるのが分かった。

参考文献

- (1) 小幡 積、太宰 北斗、競馬とプロスペクト理論、行動経済学会、2012
- (2) Kaftal Max、"The Wise Guy's Juice: A Reduced Form Skewness Analysis within Australian Fixed-Odds Horse Racing." Prevalence、2012
- (3) JRA ホームページ年毎全成績表
<http://www.jra.go.jp/datafile/seiseki/report/2013.html>