

情報学者が競馬予想に踏み出す ときに知っておくべきこと

寺沢憲吾 | 公立はこだて未来大学

AIブームを迎えて

AI（人工知能）や機械学習が注目される中で、これを競馬の予想に使ってみようと思ったり、あるいは指導する学生にそういうテーマで研究してみたいと希望されたりすることも少なくないのではないと思う。本稿では、必ずしも我が国の競馬事情に詳しくない情報系の研究者が、情報処理技術を駆使した競馬予想を用いて馬券で利益を出す試みに踏み出す際に知っておくべき基本事項を解説するとともに、利益を出す手法の開発の可能性について私見を述べたい。

日本の競馬のしくみ

我が国で行われている競馬は、農林水産省所管の特殊法人である日本中央競馬会（JRA）が主催する「中央競馬」と、地方公共団体が主催する「地方競馬」があるが、本稿では主に中央競馬について解説する。中央競馬で売られている馬券（正式には「勝馬投票券」という）には単勝、複勝、枠連、馬連、馬単、ワイド、3連複、3連単、WIN5の9種類の券種（賭式）がある。本稿では議論を簡潔にするため主に単勝式を例として解説するが、基本的な考え方はほかの賭式にも応用可能である。

オッズの計算法

単勝式の馬券とは、レース（競走）においてどの馬が勝つか（＝1着となるか）を当てるものであり、

予想が的中した場合、購入金額×オッズ（倍率）の払戻が行われる。オッズの設定方法には2種類あり、主にイギリスなどで見られる「ブックメーカー方式」は賭けを主催する者が独自にオッズを設定して客に提示するものであり、レースの結果によっては賭けの主権者が損失を被るリスクを負う。一方、日本で採用されている「パリティ方式」は、馬券の総売上から一定率の税金や経費を控除した残金を的中者で按分するという方式で、馬ごとの売上額に応じてオッズが変動する。最初に税金と経費を控除するため賭けの主権者はリスクを負わず、安定した経営ができる利点がある反面、客（買う側）としては、買った時点でオッズが確定していないという不利益があると言われているが、現在ではこの不利益は当然のものとしてファンに受け入れられている。

的中者で按分する払戻総額は上述のとおり総売上額から一定率を控除したものであり、この率のことを控除率と呼んでいる。中央競馬の単勝式馬券における控除率は20%と定められているため、オッズは次式で計算される。

$$\text{オッズ} = (\text{総売上額}) / (\text{その馬の売上額}) \times 0.80 \quad \dots(1)$$

仮に総売上額に占めるその馬の売上額の割合が50%ならばオッズは1.6倍であり、1%ならばオッズは80倍である。このように中央競馬のオッズは主催者の裁量で上げ下げしているものではなく、売上額に基づいて機械的に定められるものである。

オッズは集合知で決まる

馬券を購入する競馬ファンは「この馬のオッズが2.5倍もつくなら買う」あるいは「この馬のオッズが2.3倍しかつかないなら買わない」といったように、オッズを見ながら購入行動を決定している。式(1)によれば、ある出走馬について、その馬の馬券が買われればオッズは下がり、それ以外の馬の馬券が買われればオッズは上がる。多数の競馬ファンがそれぞれの判断基準で購入行動を決定することにより、最終的に収束したオッズは、その馬の勝利確率を集合知¹⁾により求めたものが反映されたものと考えることができる。

表-1は、2014年6月から2018年10月までに中央競馬で行われた全競走における、オッズ別に集計した出走馬数と勝馬数およびその割合を示したものである。オッズが1.1倍から2.0倍へ増加するにつれ勝率がほぼ単調減少となっており、また回収率(オッズ×勝率)も大きくなばつきがないことが見て取れる。オッズの調査範囲を100倍まで拡大したものが図-1であるが、こちらを見ても勝率はオッズに対して単調減少であり、回収率はおおむね一定であることが分かる。これは集合知により形成されたオッズが驚くほどに正確であることを示していると言える。

■表-1 オッズと回収率および勝率の関係

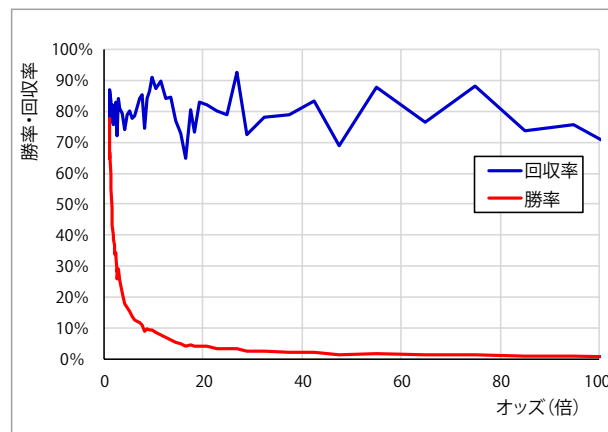
オッズ	出走	勝利	勝率	回収率
1.1	54	42	77.78%	85.56%
1.2	94	61	64.89%	77.87%
1.3	169	113	66.86%	86.92%
1.4	287	171	59.58%	83.41%
1.5	381	209	54.86%	82.28%
1.6	440	214	48.64%	77.82%
1.7	600	290	48.33%	82.17%
1.8	626	269	42.97%	77.35%
1.9	654	261	39.91%	75.83%
2.0	606	234	38.61%	77.23%

優れた競馬予想アルゴリズムとは

上記のような前提を知らずにAIで競馬予想をしようと考えた場合に陥りやすい誤りは、「最も勝率が高い馬を求めること」を目的と設定してしまうことである。教師あり機械学習を単純に適用するとついこれをやってしまいがちであるが、これはオッズを無視して的中率を高めることのみを目的に設定したことに相当するので、この方法では利益(回収率)は上がらない。

冒頭で述べたように「競馬予想で利益を出す」ことを目指すのであれば、「最も勝率が高い馬を求める」のではなく、「全出走馬の勝率を正確に求め、それにその馬のオッズを掛け、期待値を求める」ことが必要となる。その上で、期待値が1.0を上回れば買う、という作戦をとれば、(求めた「勝率」が正確ならば)目指していた「利益を出す」ことが可能となる。つまりAIを用いた競馬予想で目的とすべきことは、分類問題のように「勝つ馬」「勝たない馬」を分類するのではなく、それぞれの馬について勝率を正確に算出することなのである。

しかし競馬の競走は繰り返し試行ではない(まったく同じ競走が複数回行われることはない)ため、算出した「確率」が正確であったか否かを検証する直接的な方法はない。検証できることはせいぜい「決



■図-1 オッズと回収率および勝率の関係

まったアルゴリズムに基づいて購入行動を続けた場合の回収率」程度である。

集合知との乖離を見つける

「勝つ確率」を高い精度で見積もりたい。どうすればよいだろうか。1つの簡単で（かつ役に立たない）案は、オッズを使うことだ。前述の通り、集合知により形成されたオッズはきわめて精度の高い勝利確率の予測指標となっている。具体的には、オッズの逆数（の0.8倍）を出力すれば、それは高い精度で勝利確率を見積もったことになる。しかし、このように見積もった勝利確率では「勝利確率×オッズ」は一定となり、役に立たない。

重要なことは、自ら求めた「勝つ確率」と、集合知により形成される（オッズという形で表された）確率との乖離を発見することである。つまり「集合知はこの馬の勝つ確率を10%と見積もっているが、私は15%と判断する」ような場合に馬券を購入するという作戦を立てるということである。集合知の見積りより自らの見積りが正しければ、長期的には利益が出る。

しかし、すでに述べた通り、集合知により形成されたオッズは驚くほど正確である。その集合知よりも正しく確率を見積もるとは何を意味するのだろうか。集合知によるオッズが正確であるということの意味は、たとえば「単勝2.0倍の馬を全部まとめれば、勝率40%である」ということである。しかし、その「単勝2.0倍の馬」の中には、「真の勝利確率は40%より上」の馬と「真の勝利確率は40%より下」の馬が混ざっていると仮説を立て、それらを正確に識別することができれば、目標が達成されたことになる。

以上により、目指すべきことは、自らの手法で勝利確率を正確に算出し、かつそれがオッズと乖離していることであると分かった。しかし実際にこれを達成することは容易ではない。雑に考案した手法で求めた勝利確率はオッズと乖離しているだろうが、

それは恐らくその勝利確率が正確でないことの表れにすぎないだろう。

株価予測との違い

蛇足ではあるが、ここで競馬予想と株価予測との違いについても述べておこう。

株価とは本質的にはその会社が将来にわたって生み出す利潤から逆算したその会社の現在価値ということになるから、その価値を自分の評価で算定し、それが集合知による評価（＝株価）より上回っていれば購入すべし、というのが原則である。これは「生み出す利潤の見積り＝勝つ確率の見積り」と考えれば競馬の予想と同じ構造である。

馬券の購入と株式の購入の大きな違いは、株式を購入する場合、その株式を転売する権利を同時に買っているということである。現在株価が自分の算定した本来価値よりも高値であったとしても、それよりもさらに高い価格で売却できる（買ってくれる人が存在する）ことに確証が持てるのならば、株式の場合は購入してよい。一方馬券の場合は馬券を額面より高く買ってくれる人はいないので（あなたから額面より高く買うくらいならJRAが額面通りで売ってくれる）、そのような転売目的の購入はあり得ない。その点が大きな違いである。

集合知を上回るには

目指すべきは勝利確率を集合知よりも正しく見積もることであると分かったが、すでに述べた通り集合知は非常に正確であり、手強い。この手強い集合知を上回るにはどのような策があり得るだろうか。ここではいくつかの案を紹介する。

オッズの歪みを見つける

これまで述べてきた通り、オッズはファンの集合知によって形成される。オッズは馬券発売開始直後には歪みがあったとしても、徐々に修正されて適正

なオッズに収束する。この修正プロセスの当事者である多くのファンは、オッズを目で見て、頭で考えて、購入行動を決めている。これは人力作業であるから、修正すべき対象が多すぎると修正が追いつかなくなる。先に例にとった単勝式（1着の馬を当てる）馬券は、（中央競馬の競走の最大出走頭数は18頭であるから）最大でも18個の数値を修正すればいいので人力作業でも十分間に合うが、3連単（1着と2着と3着をすべて当てる）馬券は、組合せ数が最大で $18 \times 17 \times 16 = 4,896$ 通りとなり、集合知による修正が追いつかず、歪んだオッズが適正なオッズに収束しないまま馬券発売終了時刻を迎えることがあり得る。

もしこの歪みを検知する手法を開発することができれば、期待値が1.0を超える組合せだけを選択的に購入することができるかもしれない。すなわち、単勝オッズが示す勝利確率が正確であると仮定して3連複や3連単などの組合せ数の多い券種のオッズを合成し、合成オッズと実オッズの乖離を見つけるのである。実際、この手法で利益を上げている人々は存在すると噂されている²⁾。

他者が使っていないデータを使う

集合知を上回るためには、集合知を形成する集団があまり注目していない要素にあえて注目することも一案であろう。したがって、オッズを形成するファンが主に何に注目しているかを知ることでもまた大切である。

競馬を予想する要素には、馬や競走に関する過去のデータ、専門紙などを通じて得られる関係者のコメント、調教に関するデータ、実際に馬を見て得られる当日の馬の調子の判断、などが挙げられる。一方で、JRAの2017年の総売上2兆7,689億円のうち、開催している競馬場における売上は864億円（3.1%）にすぎない³⁾という事実は、多くのファンは現場で実際に馬を見て予想するというよりは、事前に得られるデータと当日テレビやインターネット

で配信される映像に基づいてレースの予想をして競馬を楽しんでいることを示している。さらに、たとえば同年の有馬記念においては27億円の単勝売上のうち16億円以上はパドック（下見所）に馬が現れる前に売れていることから、当日の馬体を映像ですら見ることなく、もっぱらデータと事前情報に基づいて馬券を購入しているファンが半分以上であることが分かる。つまり、昨今では馬体を見るファンよりもデータを見るファンが多いのである。

集合知を形成する集団にデータ派が多いのなら、それを上回るためにむしろ馬体に注目するという策が考えられる。パドック映像を解析して、その馬の当日の体調を判断し、予想に反映させるのである。馬体診断（馬の目利き）に関する書籍は数多く出版されている一方で、現状ではこれを重視しているファンもその具体的方法はもっぱら人間による目視診断であり、AI等を用いた画像解析でこれを行う試みは私の知る限り先例がない。JRA-VANが提供している「データマイニング予測」⁴⁾や（株）ドワンゴが主催する競馬予想アルゴリズム競技会「電脳賞（春）」⁵⁾においても、取り扱うデータは客観的な数値データに限られているようである。競走馬の当日の体調を映像からAIで判断する分野は未開拓なので、ここにチャンスがあるのではないかと思う。

人間の犯しやすい誤りを知る

集合知を上回るための作戦として、集合知はどのようなときに誤るのかを考えるアプローチもあり得る。馬のことを考えるのではなく、オッズを形成する人間側のことを考えるのである。

競馬予想に限らず、人間、あるいは人間の集団がどのような誤りを犯しやすいかにはいくつかのパターンがあるのではなかろうか。たとえば、過去に得た成功体験に基づく法則を、それがout-of-dateであるにもかかわらず適用したがる、メディアからの「この馬は強い／弱い」という情報に流される、

格言や法則にはそれが適用される条件があるはずなのに例外ケースにもそのまま適用してしまう、などといったことは、いかにも人間が犯しそうな過ちである。このような過ちのパターンを定式化し、それを検知することができれば、集合知による（誤った）確率と真の確率との乖離を発見できるかもしれない。

注意点

競馬予想で利益を上げる策の案について述べてきたが、ここで1つ重要なことがある。それは、仮に儲かる手法が構築できたとしても、それを決して公開してはいけないということである。なぜなら公開されたが最後、その情報は集合知に組み入れられてオッズが修正され、期待していた「集合知と真の確率との乖離」は消失してしまうからである。

競馬予想を楽しむこととは

競馬予想で利益を出すためには、勝利確率の見積りについて、真値と集合知の乖離を見つけることが重要であることについて述べた。また、それを達成するには、馬についてよく知ることで真値を追求していくだけでなく、人間についてよく知ることで集合知の犯しやすい誤りを探す策もあり得ることについても述べた。なお、本稿では説明を簡潔にするため、いくつかの例外（JRA プレミアムなど）は省略して説明していることをご了承いただきたい。

実のところ、いち競馬ファンの心情としては、「勝利確率の真値」が算出され、それが広く共有されたら、競馬予想という娯楽は終わりだと考えている。どの馬を買っても期待値が一緒というのであれば、これはサイコロを振るのと同じで予想の楽しみがな

い。「競馬とは思いつきと思ひ込みと思ひ違ひを楽しむ遊び」——これはある競馬評論家の言であるが、データによる裏付けがなくとも思い思いに仮説を立てて予想を披露し、それに説得力があるとかないとか言って根拠のない評論をし、数時間後に判明する結果との差異を楽しむことこそ、まさに競馬という娯楽の醍醐味である。

期待値がマイナスな遊びに自分はなぜ参入するのかと考えるとき、そもそも世の中のおよそあらゆるレジャーは期待値がマイナスであることに気がつく。ほとんどの先進国で存在している競馬という文化は、英国などでは貴族に支えられているかもしれないが、貴族制度のない我が国においてはこれを支えているのはファンである。ファンが馬券を購入するお金が、巡り巡って、馬産地の牧場や、育成、調教、馬具装丁などにかかわる多くの関係者の生活を支えている。つまりは馬事文化に貢献している。文化への投資は、散財である。菊池寛の以下の言葉で本稿の締めくくりとしたい。「馬券買ひは道楽也。散財也。真に金を儲けんとせば正道の家業を励むに如かず。」⁶⁾

参考文献

- 1) ジェームズ・スロウィッキー（著）、小高尚子（訳）：「みんなの意見」は案外正しい、角川書店（2006）。
- 2) たとえば http://blog.livedoor.jp/suda_takao/archives/66132379.html など
- 3) 平成 29 事業年度事業報告書、日本中央競馬会
- 4) <http://jra-van.jp/fun/dm/mining.html>
- 5) <http://ch.nicovideo.jp/dennoushou/>
- 6) 菊池 寛：我が馬券哲学、青空文庫。

（2018年11月14日受付）

■寺沢憲吾（正会員） kterasaw@fun.ac.jp

1998年東京大学工学部土木工学科卒業。2006年公立はこだて未来大学大学院博士（後期）課程修了、博士（システム情報科学）。2012年より公立はこだて未来大学准教授。現在の研究テーマは主に古典籍などの難読難解な文書画像の認識と理解。競馬は趣味の1つである。