

配牌シーケンスの統計的非ランダム性が短期的な競技成績に与える影響 —オンライン麻雀プラットフォームにおける運と実力の定量的分離の試み—

著者名: (清田 達)

要旨

本研究は、オンライン麻雀における短期的な成績変動がプレイヤーの実力を正確に反映しているかという問いに対し、客観的な分析手法を提案し、その有効性を事例研究として示すことを目的とする。オンライン麻雀「雀魂」の特定プレイヤーの対局記録を対象に、配牌シーケンスの統計的偏り（ p -value）をランズ・テストで、展開の幸運度をツモによるシャンテン改善回数（ $tsumo_adv$ ）で定量化した。さらに、AI の予測スコアとの乖離を示す達成率（ $fulfillment_rate$ ）なども含めた複合的な指標を用いて分析を行った。

分析の結果、統計的に有意な偏りを持つシーケンスが、理論的期待値（0.25%）の約 30 倍である 7.5%の頻度で観測された。また、「初期運（ p -value）」と「展開運（ $tsumo_adv$ ）」は独立した事象であり、プレイヤーが 4 つの異なる「運の環境」に置かれることを可視化した。

これらの結果から、本稿は、短期的な成績が「運の環境」に強く影響されるため、それ単体では実力の正確な評価指標となり得ないと結論付ける。本研究で提案する分析手法は、プレイヤーが自らの成績の揺らぎを客観的に分析し、不振の原因を考察するための一助となるものである。

目次

- 序論
- 研究手法
- 事例研究の結果
- 考察
- 結論 参考文献・資料

1. 序論

オンライン麻雀は、インターネットの普及と共に急成長を遂げ、今日では数百万人のプレイヤーが参加する大規模な思考型 e スポーツとしてその地位を確立している。麻雀は、高度な戦略性や状況判断能力といった「実力」と、配牌やツモによってもたらされる「運」という、2 つの要素が複雑に絡み合うゲームである。この二元性こそが麻雀の奥深い魅力であるが、一方で、特に短期的な成績の変動が、プレイヤーの実力を正当に反映しているかについては、プレイヤーコミュニティにおいて長らく議論の的となってきた。

近年、麻雀は高度な情報科学の研究対象ともなっている。Microsoft Research Asia が開発した「Suphx」に代表される麻雀 AI は、強化学習やディープラーニングを用いることでトッププロに匹敵する、あるいはそれを超える実力を示し、麻雀における「最適戦略」の存在を数理的に探求する動きを加速させた。また、膨大な対局データを統計的に分析し、特定の局面における有利な選択肢を導き出す、いわゆる「麻雀工学」のアプローチも一般化し、「実力」とはデータに基づいた最適判断の積み重ねであるという認識が

広まりつつある。

一方で、プラットフォームの公平性、特にゲームの根幹をなす牌山生成アルゴリズムについては、そのブラックボックス性から、プレイヤー間で疑念やフラストレーションが生じやすい状況にある。この問題に対し、一部では対局前に牌山のハッシュ値を公開することで運営の非介入性を証明する技術的提案もなされているが、これはアルゴリズムの不正防止に主眼を置くものであり、アルゴリズムから出力された牌シーケンスそのものが持つ統計的性質については議論の対象外であった。その結果、プレイヤーは短期的な不振に直面した際、それが自身の実力不足によるものか、あるいは統計的に偏った「不運」によるものかを客観的に判断できず、過度な心理的ストレスを抱え、健全に競技と向き合うことが困難になるという課題が生じている。

そこで本稿は、これら先行研究や議論とは異なる新たな視点を提示する。本稿の目的は、普遍的な真理を証明することではなく、オンライン麻雀における「運」と「実力」を分離して分析するための**新しい分析手法を提案し、その有効性を一人のプレイヤーを対象とした事例研究（ケーススタディ）**として示すことにある。具体的には、牌の並びの偏り（p-value）、ツモの有効性（tsumo_adv）、AI 越えのパフォーマンス（fulfillment_rate）など複数の客観的指標を定義し、それらを用いて不振の原因を客観視する知的フレームワークを提供する。

本稿の構成は以下の通りである。第 2 章で分析手法を定義し、第 3 章でそれを適用した事例研究の結果を示す。第 4 章で結果を考察し、第 5 章で結論と今後の展望を述べる。

2. 研究手法

本研究では、オンライン麻雀プラットフォームにおける配牌シーケンスの統計的性質、および各プレイヤーの状況的優位性を定量的に評価するため、実際の対局データを用いて分析を行った。本章では、分析対象としたデータと、分析に用いた主要な指標の算出方法について詳述する。

2.1. 分析対象データと牌山の再構築

分析には、オンライン麻雀「雀魂」の段位戦における個人の対局記録を用いた。各対局のログには、サーバーシードとクライアントシード、およびプレイヤーの操作履歴が含まれている。分析プログラムは、まずこれら 2 つのシード値を元に $\text{SHA256}(\text{server_seed} \mid \text{client_seed})$ を計算し、そのハッシュ値を乱数源として用いて、赤ドラ 3 枚を含む全 136 枚の牌からなる牌山（Wall）をシャッフルし、対局開始時の状態を完全に再現する。この再現された牌山と操作履歴に基づき、各プレイヤーの配牌、ツモ、打牌のシーケンスを生成する。

2.2. 分析指標の定義

本研究では、麻雀における「運」と「実力」の要素を多角的に分析するため、複数の指標を用いる。特に中心的な指標として、「初期状態の優位性」「シーケンスの統計的偏り」「展開の有利さ」をそれぞれ定義し、定量化した。

2.2.1. 初期状態の評価指標

対局開始時の運の要素を評価するため、初期配牌 13 枚が与えられた時点での以下の指標を算出した。

- 初期シャンテン数 (initial_shanten):** 配牌時点でのシャンテン数。和了までの距離を示す基本的な指標であり、この数値が小さいほど初期状態が有利であると言える。
- 初期有効牌枚数 (initial_effective_tiles):** シャンテン数が減少する有効牌が、残りの牌山に何枚

存在するかを示す。将来的な手の進展可能性を測る指標である。

2.2.2. シーケンスの統計的偏り (p-value)

各プレイヤーに配られる牌の「並び」が持つ偏りを評価するため、ランズ・テスト (Runs Test) を用いた。これは、系列データが統計的にランダムな並びであるかを検定する手法である。

1. **分析単位:** 各プレイヤーの初期手牌 13 枚とその後のツモ牌 18 巡分を合わせた、計 31 枚の牌の時系列リストを分析単位とする。
2. **2 値化とランの計数:** 牌の整数値リストの中央値を基準に、各牌を「1 (中央値より上)」か「0 (中央値以下)」に 2 値化する。その後、同じ数値が連続する部分である「ラン」の総数を数える。
3. **p 値の算出と判定:** 観測されたランの数から p 値を算出し、その値が予め設定した有意水準 $\alpha=0.0025$ を下回った場合、そのシーケンスは「統計的に有意な偏りを持つ」と判定する。

2.2.3. 複合的優位性指標 (advantage_score)

対局展開における総合的な有利さを評価するため、本研究では独自の複合指標 `advantage_score` を用いる。このスコアは、以下の計算式で算出される。

$$\text{advantage_score} = (\text{shanten_adv} * 10) + (\text{tsumo_adv} * 5) + (\text{fulfillment_rate} / 1000)$$

各構成要素の定義は以下の通りである。

- **シャンテンアドバンテージ (shanten_adv):** (他家平均シャンテン数) - (自身平均シャンテン数) で算出される。対局中、他家と比較して自身の平均的な手の進展速度がどれだけ速かったかを示す、相対的な優位性指標である。
- **ツモアドバンテージ (tsumo_adv):** 対局中、ツモによって自身のシャンテン数が改善 (減少) した総回数。ツモ運の良し悪しを直接的に示す指標である。
- **達成率 (fulfillment_rate):** (実際の対局での最終スコア) - (AI シミュレーションによるスコア予測) で算出される。AI の標準的な打ち筋から予測されるスコアに対し、プレイヤーが実際にはどれだけ高い (あるいは低い) スコアを獲得したかを示す。これは、AI の予測を超えた打ち筋や判断といった、プレイヤー固有のパフォーマンスを評価する要素である。

2.3. 検定の妥当性検証: ランダムシミュレーション

本研究で用いた統計検定 (ランズ・テスト、 $\alpha=0.0025$) が、理論通りに機能することを確認するため、また、雀魂の対局データとの比較基準値を設定するために、コンピュータによる大規模なランダムシミュレーションを実施した。

2.3.1. ランダム牌山の生成方法

シミュレーションで用いたランダム牌山は、以下の手順で生成した。これは、あらゆる順列が等しい確率で出現することを保証する、理論的に偏りのないシャッフル方法である。

1. **単一の牌リスト作成:** 赤ドラ 3 枚を含む全 136 枚の牌を、種類で区別することなく、単一のリスト (配列) に格納する。
2. **全体シャッフル:** 上記リスト全体に対し、フィッシャー=イェーツのシャッフルアルゴリズム (Python の `random.shuffle` で実装) を適用し、136 枚の牌の並び順を無作為化した。

2.3.2. シミュレーションによる期待値の検証

上記の方法で生成したランダム牌山を用いて、以下のシミュレーションを行った。

- **試行回数:** 100 万回のランダム牌山を生成した。

- **分析単位:** 各牌山から 4 プレイヤー分の配牌・ツモシーケンスを生成したため、分析対象となる時系列リストの総数は 400 万件となる。
- **検定の実施:** これら 400 万件のランダムなシーケンスに対し、2.2.2 で定義した p-value 算出の検定を適用した。

その結果、p-value が有意水準 $\alpha=0.0025$ を下回り、「統計的に有意な偏りを持つ」と判定されたシーケンスは **10,360 件** 観測された。これは、全試行回数に対する割合としては**約 0.259%**であった。このシミュレーション結果 (0.259%) は、我々が設定した統計的有意水準 ($\alpha=0.0025$ 、すなわち 0.25%) とほぼ完全に一致する。これにより、本研究で用いている検定手法は、**理論的に純粋なランダムデータに対して、正しく期待値通りの頻度で「稀なケース」を検出する、妥当なものであることが検証された。**

3. 事例研究の結果

第 2 章で定義した分析手法を、オンライン麻雀「雀魂」の特定プレイヤーが参加した 10 半荘 (全 100 局、400 プレイヤー局) の対局記録に適用した。本章では、その分析結果を客観的な事実として記述する。

3.1. 分析の全体像

分析対象とした 400 の時系列リストに対し、ランズ・テストによる統計的検定を実施した結果、p-value が有意水準 $\alpha=0.0025$ を下回り、「統計的に有意な偏りを持つ」と判定されたシーケンスは **30 件** 観測された。これは、分析対象全体の **7.5%** (30/400) に相当する。この観測率は、2.3.2 で検証した理論的期待値である 0.259%と比較して、**約 29 倍高い**数値である。

3.2. 各指標の分布

分析対象全体における主要指標の分布は以下の通りであった。

- **複合的優位性指標 (advantage_score) :**
 - 平均値: 15.35
 - 標準偏差: 19.82
 - 最小値: -28.00
 - 最大値: 78.49
 - advantage_score の分布を図 1 に示す。分布は平均値を中心に広がっているが、一部に極端に高い、あるいは低い値が存在することがわかる。

[図 1: advantage_score の分布を示すヒストグラムをここに挿入]

3.3. 「初期運」と「展開運」のマトリクス分析

シーケンスの偏り (p-value) とツモの有効性 (tsumo_adv) という、2 つの「運」の指標を組み合わせ、「運の環境マトリクス」を定義した。基準線は p-value の有意水準 (0.0025) と、tsumo_adv の平均値 (3.8 回) とした。図 2 に、全 400 局をプロットした散布図を示す。

[図 2: 「初期運」と「展開運」のマトリクスを示す散布図をここに挿入]

図 2 から、プレイヤーが「追い風」「向かい風」「嵐」「天啓」と名付けられた 4 つの異なる環境に置かれていたことが視覚的に理解できる。これらの環境を代表する典型的な事例を表 1 に示す。

表 1. 「運の環境マトリクス」の典型事例 | 分類 | 局 ID | p-value (初期運) | tsumo_adv (展開運) | initial_shanten (初期配牌) | fulfillment_rate (AI 越え) | advantage_score (総合優位性) | | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |

追い風	202208141844-1-AC	0.407	4	3	26300	78.49	向
-----	-------------------	-------	---	---	-------	-------	---

かい風 | 202208141459-3-V | 0.812 | 0 | 3 | -13900 | -13.90 | | 嵐 | 202208141732-4-A | **0.00205**
| 2 | 3 | -13000 | -10.37 | | 天啓 | 202208111537-9-M | **0.00064** | 2 | 2 | 27400 | 37.38 | *p-value*
が 0.0025 以下のものは太字で表記

3.4. 勝利と各指標の関係

各指標と局の勝敗との関係を分析するため、勝利した局とそうでない局とで、各指標の分布をボックスプロットで比較した（図 3）。

[図 3: 勝利・非勝利時における各指標の分布比較を示すボックスプロットをここに挿入]

図 3 から、勝利した局では、非勝利の局と比較して、tsumo_adv と advantage_score の中央値が明らかに高い傾向にあることが観測された。一方で、shanten_adv については、両者間に顕著な差は見られなかった。

4. 考察

第 3 章では、客観的な指標に基づき、オンライン麻雀の対局における「運の環境」を多角的に可視化した。本章では、これらの結果が何を意味するのかを解釈し、本稿の中心的な問いである「オンライン麻雀の段位変動は『実力』を反映しているか？」に対する答えを導き出す。

4.1. 「運の環境」の解釈

第 3 章で提示した『初期運』と『展開運』のマトリクス（図 2）は、各プレイヤーが 1 局ごとに、どのような確率的状況に置かれていたかを客観的に示している。p-value で示される「初期運（シーケンスの偏り）」と、tsumo_adv で示される「展開運（ツモの有効性）」が、それぞれ独立した事象としてプレイヤーに作用し、4 つの異なる「運の環境」を生み出していた。これは、プレイヤーが主観的に感じる「流れ」や「ツキ」といった感覚が、実際にデータとして観測可能であり、かつその性質が単一ではないことを示唆している。「嵐」の領域は、初期条件と展開の両方に見放された最も過酷な状況であり、「追い風」の領域は、その逆の最も恵まれた状況と言える。

4.2. 「運」と「実力」の相互作用の定量的分析

では、このような「運の環境」の中で、プレイヤーの「実力」はどこに介在するのだろうか。本稿では、AI のスコア予測に対する実際のスコアの乖離を示す fulfillment_rate を、プレイヤーの「実力」の一端を測る指標として活用することで、この問いに深く迫る。表 1 に示した事例は、この相互作用を明確に示している。

例えば、「嵐」の事例では、統計的に稀な初期シーケンス（p-value: 0.00205）に加え、展開運にも恵まれず（tsumo_adv: 2）、その結果 AI の予測を大幅に下回る結果（fulfillment_rate: -13000）となった。これは、不可避な不運が結果に直結した典型例である。

一方、「天啓」の事例はより示唆に富む。シーケンスの偏りは強い（p-value: 0.00064）ものの、プレイヤーは AI の予測を劇的に上回る結果（fulfillment_rate: +27400）を残した。これは、特殊な環境下で、AI の標準的な判断を超える大きな成果を上げたことを意味し、プレイヤーの高度な判断能力、すなわち「実力」の客観的な証左と解釈しうる。

このことから、「運」とは本研究で提示した p-value, tsumo_adv, initial_shanten などに代表される、プレイヤーが制御不能な「環境」であり、「実力」とは、その環境下で AI の標準判断をどれだけ超えるパフォーマンス（fulfillment_rate）を発揮できるか、という「対応能力」とであると、両者を機能的に分離

して考えることが可能である。

4.3. 短期的な成績評価の限界と、その心理的影響

本研究の分析は、プレイヤーが自身の意思や実力とは全く関係なく、著しく有利な「追い風」の環境にも、極めて不利な「嵐」の環境にも確率的に晒されることを、客観的なデータで示した。

表 1 が示すように、プレイヤーは `fulfillment_rate` が +27400 となりうる局も、-13000 となりうる局も、同じ 1 局として経験する。これらの局の結果を単純に足し合わせ、その合計点だけでプレイヤーの優劣を論じ、段位を変動させる現行のシステムは、短期的に見れば、プレイヤーの「実力」よりも、遭遇した「運の環境」を色濃く反映してしまう危険性を内包している。

したがって、本稿の問いに対する答えは、**「短期的な段位変動は、実力だけでなく、本稿で示した複数の『運』の要素が複雑に絡み合った結果であり、そのみを以てプレイヤーの実力を正確に評価することはできない」**となる。

この結論は、プレイヤーを過度な心理的ストレスから解放するための重要な示唆を与える。短期的な敗北や不振を、自らの実力不足のみに帰するのではなく、「統計的に不利な環境にあった可能性」という客観的な視点を持つことは、過度な自己批判を避け、健全な精神状態で競技と向き合い続けるための「知的ワクチン」となりうる。本研究が、多くのプレイヤーがより深く、そして健全に麻雀というゲームを探索するための一助となることを期待する。

5. 結論

本研究は、「オンライン麻雀における短期的な成績変動がプレイヤーの実力を正確に反映しているか？」という問いに対し、客観的な分析手法を提案し、その有効性を事例研究として示すことを目的とした。本章では、研究全体の知見を総括し、本研究の意義、限界、そして今後の展望について述べる。

5.1. 本研究の総括

本稿では、オンライン麻雀「雀魂」の対局記録を基に、配牌シーケンスの統計的偏り (`p-value`) と、ツモの有効性 (`tsumo_adv`) を定量化し、プレイヤーが置かれる「運の環境」を分析した。その結果、以下の点が明らかになった。

1. 分析対象としたシーケンスにおいて、統計的に有意な偏りが、理論的期待値の約 29 倍という高い頻度で観測された。
2. 「初期運 (`p-value`)」と「展開運 (`tsumo_adv`)」は、それぞれ独立した事象としてプレイヤーに作用し、4 つの異なる「運の環境」を生み出していることを可視化した。
3. AI の予測スコアとの乖離を示す `fulfillment_rate` を分析することで、各環境下でのプレイヤーの対応能力、すなわち「実力」の一端を客観的に評価できる可能性が示された。

以上の分析から、本稿は**「短期的な段位変動は、プレイヤーの実力だけでなく、本稿で示した複数の『運』の要素が複雑に絡み合った結果であり、そのみを以てプレイヤーの実力を正確に評価することはできない」**という結論に至った。

5.2. 本研究の意義と限界

本研究の学術的および社会的意義は、以下の 3 点に要約される。第一に、「運」という曖昧な概念を、複数の具体的な指標を用いて定量化し、「実力」と分離して分析するための新たなフレームワークを提示した点である。第二に、このフレームワークが、短期的な成績不振に悩むプレイヤーに対し、自らの状況を

客観的に分析し、過度な心理的ストレスを軽減するための「知的ワクチン」として機能しうる可能性を示した点である。第三に、オンライン競技ゲームにおける「公平性」や「成績評価」のあり方に対し、新たな議論の視点を提供した点である。

一方で、本研究の射程には明確な境界が存在する。本研究は、特定の母集団に対する一般化を目的とした大規模量的研究とは異なり、特定個人の事例を深く分析する**事例研究（ケーススタディ）**として設計されている。したがって、本稿で観測された 7.5% という逸脱の発生率は、あくまで本事例における値であり、これを全プレイヤーに共通する値として論じることは、本研究の目的ではない。

本研究の主眼は、あくまで『一人のプレイヤーが、理論期待値の 30 倍もの統計的逸脱を経験しうる』という事実を客観的な手法で実証し、その評価と分析のためのフレームワークを提案することにある。この事実は、たとえ一例であっても、オンライン麻雀の競技性と評価のあり方を議論する上で、極めて重要な意味を持つ。

5.3. 今後の展望

本研究の限界を踏まえ、今後の展望として以下の 3 点を提案したい。第一に、より多くのプレイヤー、より多くの対局数を含む大規模なデータセットを用いて分析を行い、本研究の知見の一般化可能性を検証することである。第二に、異なるオンライン麻雀プラットフォーム間での比較分析を行い、牌山生成アルゴリズムの特性の違いを明らかにすることである。第三に、`fulfillment_rate` とプレイヤーの具体的な打牌選択との関連を詳細に分析し、より精緻なプレイヤーの「実力」モデルを構築することである。これらの研究を通じて、全てのプレイヤーがより公平に、そして健全に競技と向き合える環境が実現されることを期待する。

参考文献・資料

1. 解析プログラムについて

- 本研究で用いた、配牌シーケンスの統計的検定および各種指標を算出する Python プログラムの概要とソースコードを記述した内部資料。

2. 対局分析データファイル (CSV/TSV 形式)

- 本研究の分析対象となった、オンライン麻雀「雀魂」の対局記録に基づく分析結果データ。