

麻雀における Rating システムに関する改善案

——忘れられた重大な欠陥を持つ『Rating システム』

平成 21 年 7 月 13 日
と つ げ き 東 北

——あなたは、野球選手のバッティング技術を比較するとき、「打率」を見るか。
それとも、「そのバッターが参加した試合の勝率」を見るか——

ネット麻雀における R (レーティング) 変動式のうち、順位等に関する項 (試合数による補正及び他家 R との調整分を除く) は、一般的な麻雀のルール等を十分に考慮したうえで決定されるべきである。

たとえば、25,000 点持ち 30,000 点返し、5-10 のウマをつけた場合、X を各プレイヤーの最終持ち点 (オカ、ウマを考慮したもの) として、以下のとおり与えられるべきである。

$$R \text{ 変動 (順位相当分)} = (X - 25000) \div 2000 \times 0.373$$

1. 既存の Rating (R) システムの致命的な——忘れられた——欠陥
2. 麻雀に特化した R システムの構築のために
3. 実際に R 変動式を変更するための考え方
4. R 変動式に対する補正
5. R 変動式の変更によるメリット
6. さらなる改善へ向けての課題

変更履歴

09/07/13 初版

※とある避けがたい事情により、本資料は1日で急遽作成したため、計算ミス等により (涙)、後に数値の細部が修正される可能性があります……ご了承を。

1. 既存の Rating システムの致命的な——忘れられた——欠陥

近年、各種ネット麻雀においては、チェス等で伝統的に用いられてきた Rating システム（以下、R システム）による実力評価が採用されている。記憶頼りになりがちな「麻雀の技術」を定量化する試みは極めて有意義であり、好ましい傾向である。

R システムは、ある種の 2 人対戦ゲームにおいて、R というスカラー値で「強さ」を定量化し、R の差によって各々の勝率を計算することを目的として作られた仕組みである。たとえば、プレイヤー A とプレイヤー B の対戦成績が「60 勝 40 敗」、プレイヤー A とプレイヤー C の対戦成績が「70 勝 30 敗」だったとする。各々個別の勝率はわかったが、ではプレイヤー B とプレイヤー C が対戦すると、勝率はどの程度になるか、という問題を、B と C を直接対決させることなく R 値で判断しようというのが、R システムの着想である（実際に R システムが理想的にかつ完全に機能するには、種々の強い仮定が必要となるが、多くのプレイヤーのおおまかな強さを把握するためには十分に機能するようである）。

チェス等、成績に対して実力が強い影響を与えるゲームにおいてさえ、必ずしも「強い側」が 100 戦 100 勝とはいかない。結局、1 試合あたりの「結果の良さ」は、実力の影響と運の影響（その時の体調や、たまたまの失敗等）を合算した「良さ」で表されることとなるだろう。そこで、結果のうちの「運の要素」の部分は、「運」であるがゆえに、ガウス分布すると仮定して設計されたものが R システムなのである。

ネット麻雀『東風荘』を筆頭として、各種ネット麻雀が、4 人対戦ゲームである麻雀に R システムを適用したのは、非常に画期的・意欲的な試みであり、賞賛されるべき功績である。そうした野心的な試みが、麻雀研究を一気に加速させたのだった。

ただ、R システムや麻雀の各種性質にかかる研究が進んだ現在となつては、少しだけ——麻雀を研究する者から見ると看過できない——誤算があつたことを指摘できる。

チェスや将棋のように、「強さ」が定量化困難なゲームでは、「試合結果」をもとに R を変動させる以外の方法がなかったが、麻雀においては「点棒のやりとり」によって、各個別の局での「強さ」が定量化可能なはずである。

ところが、麻雀の R はチェス等に倣って、順位によって決定されるというシステムが一般化してしまった経緯から、より詳細に計測できるはずの「強さ」に関する情報が欠落することとなってしまった。つまり、多くのネット麻雀では「順位戦」ベースでの「強さ」だけが R に反映される形で定着してしまつたのである（2009 年 7 月現在でもそうである）。通常の麻雀のルールでは、オカやウマを含めて「得点をよりたくさん持っていること」が強さの度合いを意味するのであり、6 万点のトップと 3 万点のトップが常に「順位に応じて」同一評価を受けるのは、一般的な麻雀の感覚からすると不合理だ。

これはちょうど、野球での打者のバッティング技術を正確に評価するために、「打率が何割か」を見るかわりに、「その打者が出場した試合で、チームが勝ったか否か」で判断してしまうようなものである。無意味な量子化（点数の差を、わずか 4 種類に押し込めてしまうこと）に伴う、あからさまな情報欠損が生じているのである。

2. 麻雀に特化した Rating システムの構築のために

そこで、R システムを麻雀に特化させ、適正化させることを考えよう。

ここでいう適正化とは、いうまでもなく、「実態としてのリアル麻雀での評価」に近い形にし、実力を的確に表すものにするのである。たとえば、ハコラスは、ギリギリの 4 位よりも低い評価を受けるべきだし、近い点数であっても 2 位は 3 位よりも高い評価を受けるべきだ。麻雀での一般的なオカ・ウマ等の影響も加えよう。加えて、R の分布が既存の R システムの分布と矛盾したり、著しく乖離したりしないよう留意して設計することが必要条件となる。

理想的には、次の全性質を併せ持つ R システムを構築することを考える。

- (1) 実力 R が正規分布するとすれば、みかけ R もまた最終的に正規分布すること
- (2) 実力 R は、十分な試合数における偶然の R 上下変動分を加味して、みかけ R の分布よりも少し小さな分散を持つこと（当該偶然の変動分の分散は計算により求めること）
- (3) 実力 R の差に比例する形で、1 試合あたりの「実力による得点」分が決定されること
- (4) 1 試合あたりの最終的な試合結果（得点）の分散は、「実力による得点分散 + 1 試合あたりのゲーム性に依存した偶然による得点分散」と一致すること
- (5) 実力 R と他プレイヤーの実力 R 平均の差に対して、試合ごとに線形な補正が加わること（強いプレイヤーに勝つほど、高い R を得ることができること）
- (6) 上記条件を満たしたうえで、長期間、多数のプレイヤーが打った結果、初期のみかけ R 分布と最終的なみかけ R 分布の統計的性質が一致し、系統的誤差が発生しないこと
- (7) 既存の R の分布と、新しい R システムでの R の分布が大きく乖離しないこと

まずは、そもそもプレイヤーの実力が、現状の R システムにおいて、どのような分布をしているかについて調べ、1 試合あたりの「偶然による得点」をどう見積もるべきかを決定する必要がある。

手持ちの、「東風荘」のプレイヤー全体のみかけ R 分布に関するデータを利用した。やや古いデータだが、みかけ R は標準偏差約 215 の正規分布におおよそ従う。実力 R は、みかけ R よりも若干分布の幅が小さくなり、おおよそ 207.88 程度であることがわかった。これは、WM 氏の研究「東風荘における実力 R の分散（筆者註：十分な試合数を打った場合のみかけ R が、実力から乖離する分布に関する解析的研究）」で示された分散に基づいた。

実際にシミュレーションを行った結果、1 試合あたりの「運による得点」の分散として取るべき値が、R にして約 1000^2 程度と求まった。実力 R と、長期間打った後のみかけ R が一致することを大前提に、分散の差等についても考慮して総合的に考慮した結果、このパラメータにおいて、上記(1)(3)(4)(5)(6)が満たせたと判断した（下図 1）。

この結論は WM 氏の結論「990²」とわずかに異なる（(2)の点）が、みかけ R の分散と実力 R の分散の誤差平方の差は充分小さく、当該誤差よりも、実際のみかけ R 分布と長期間打った後のみかけ R 分布の系統的一致性に着目して、当方での結論を優先する形とした。

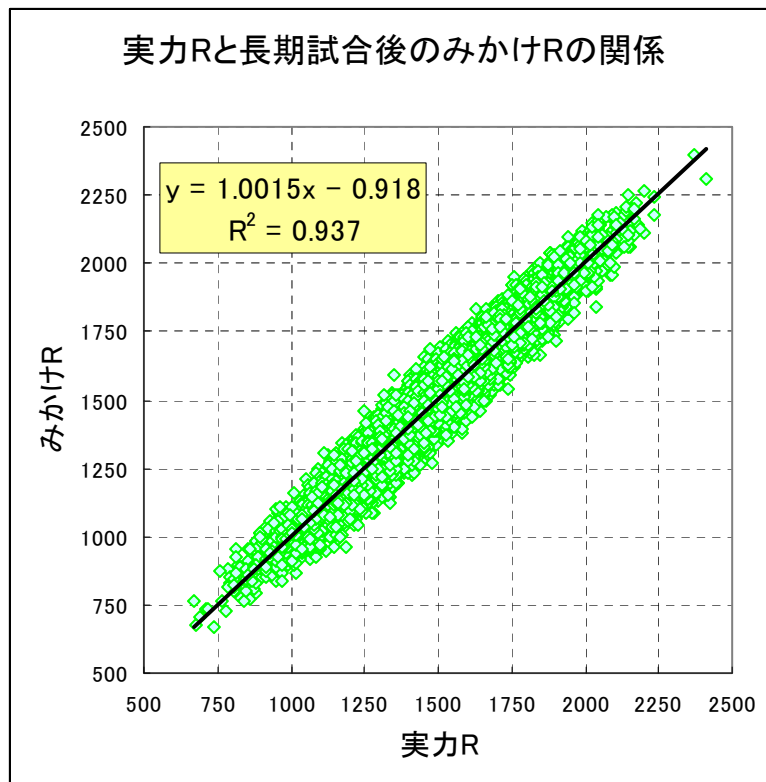


図1：実力Rと長期試合後のみかけRの関係（30000名 1000試合）

なお、実力 R = みかけ R と仮定した場合であっても、1試合あたりの運による得点分散自体には大きな違いは出ない（同様の方法で、約 1030^2 程度と試算できる）。よほど特殊な R 変動式やルールを用いていない限りは、東風戦1試合あたりの運による実力への影響として、 R にして $1000^2 \sim 1030^2$ 程度を採用すれば問題ないことが確認できる（これが仮に 1100^2 程度となったとしても、恐らく以降の結論に深刻な影響は生じないだろう）。

3. 実際に R 変動式を変更するための考え方

現状、ネット麻雀『東風荘』（あるいは『天鳳』）の R システムにおいて、400 試合以上プレイしたプレイヤーが、他プレイヤーと同じみかけ R 値を持っていた場合、1位～4位の R 変動値はそれぞれ $[6, 2, -2, -6]$ となる。

一方、東風荘における1位～4位の平均的な最終持ち点（オカ・ウマ考慮せず。初期持ち点は 27,000 点）は下図2のとおりである（※東風戦超上級ランキング卓における、のべ 185,195 試合の結果から得られた統計に基づく。なお、リーチ棒等が出たまま試合が終了した場合等に生ずる点数不足分は、単純に、全員に均等に加算する形で補正した）。

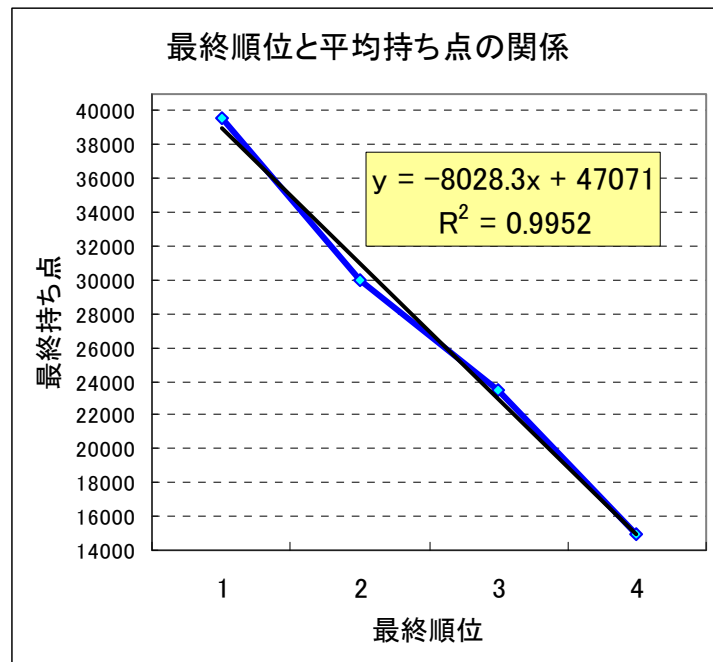


図 2：第一東風荘超ランにおける最終順位と、平均的な持ち点との関係

図 2 からわかるとおり、実際には順位は得点とほぼ線形な関係を持っており、1 位の順位変化ごとに、持ち点がおおよそ 8,000 点変化する。東風荘がルール上、劇的に「順位戦」であるにもかかわらず、である。

このことから、単純には、トップのオカ及び順位のウマも考慮した上で、開始時の持ち点から 2,000 点ごとに R が 1 上昇する仕組みとすることで、既存の R システムと同等の性質を持ちつつ、連続的な（量子化されていない）実力の評価が可能になることがわかる。

ただし、このような変更を安易に与えると、R の変動式に「分散を大きくする変更」が生じ、各プレイヤーが長期間打ち続けた場合の、実力 R とみかけ R が、現状よりもばらつくであろう。たとえば、これまで「実力 R2000」で、R1900~R2100 前後を上下していたプレイヤーが、突如 R1700~R2300 の間にばらつくように変化したとすれば、プレイヤーにとっては R システム全般への理解の妨げとなるし、偶然性の要素が強く感じられることによるモチベーションの低下をもたらしかねない。実際、単純に 1 位~4 位に対して、オカとウマ (5,000-10,000) を考慮し、[+12000,+4000,-4000,-8000]の点数を与えるようにしたうえでシミュレーションを行ったところ、みかけ R の標準偏差は 550 程度 (2 倍以上) となり、現状と比較して過大になった (R2500 のプレイヤーが存在するような値となる)。

この種の影響を抑え、現状の R システムとかけ離れない程度の長期的分散とする必要がある。そのためには、2,000 点ごとに R 1 ではなく、もう少し大きな点数に対して R が 1 変動するよう補正せねばならない。

具体的数値について、次章で検証する。

4. R 変動式に対する補正

理論的にもっとも適切なのは、「他家 R との差による R 調整」が、ちょうど期待される増加 R と一致するような値となる補正值を探し出すことだ。他家に対する R の高さに比例して、原点よりも高い点数を取得できることを前提として（可能ならばそのように R が分布するように）R システムを設計してきたわけだから、当然、R の相対的高さに応じた「R 差によるペナルティ」が課せられるべきである。果たしてそれは可能だろうか。

もともと、R の差に対して線形な調整が加わっていたところに、線形でない要素を持つ（1位が非常に優遇され、2位が少し優遇され、3位と4位は相対的に不遇な扱いを受ける）配分を与えるため、これは困難に見える。

しかし、結論から言えば、この問題は近似的に解消できる希望がある。

麻雀の性質上、相当に実力差がある場合であっても、1試合あたり平均+8,000点（Rにして、補正前の値で+4）取れば「最強」と呼んで良い実力であり、これは一般的な場においても、選抜が加わった場においても変わらない事実である。どのようなレベルの層が集まっても、結局のところ、各プレイヤーの持ち点はある種の点数範囲内に収まり、それ以上の差はつかない。実際、東風荘超上級卓において最強水準の人物でも、長期的に見れば、試合終了時の持ち点は平均 28,000 点（開始時より+1,000 点）程度である。個々のプレイヤーに対して、得点を求めるインセンティブが働けばもう少し差がつくだろうが、R 変動自体は「トップの獲得」「順位の獲得」そして「順位にかかわりのない高いアガリ」に大きく依存することとなろう。逆説的になるが、順位を目指す麻雀とは、結局のところ（大部分は）得点を目指すゲームとなる——誤解ないように注釈をつけるとすれば、そのインセンティブの変化と、得点差が大きくなる場合（入門者と上級者に生ずるような圧倒的な点差）に反応する、適切な R 変動が重要なのだが。

順位に無関係な得点の多寡であっても、局所的に見れば各プレイヤーにとって重要な R 変動の要因となる。それは十分にスリリングな体験である。順位のみを目指す麻雀では、このようなインセンティブは発生せず、成績に対して系統的な影響を与えない。得点の多寡がインセンティブとなる空間では、十分に大きな実力差がある場合（たとえば全てのユーザが同一空間で対戦する場）等には特に、平均的な持ち点の影響は強く働くであろう。

上述した条件を考慮し、順位 1 位の差に対して点数にして 8,000 点分の差をもたらし、なおかつトップ等を優遇した場合の R の分布が、既存のシステムにおける R 分布と乖離しないよう R 変動式に一定の「比例定数」をかけ、補正することを考える。「比例定数」は当然、R 変動式のうちの、「他家 R との調整項」以外の全ての項にかかる。

実際に、上記前提に従った R 変動に可変の「比例定数」をかけたうえで、再度シミュレーションを行った。オカの影響は+15、ウマは 5,000-10,000 としてある。既存の R 分布と、このルールにおける R 分布がおおよそ一致する「比例定数」を、妥当な R 変動での「比例定数」として取ることとする。

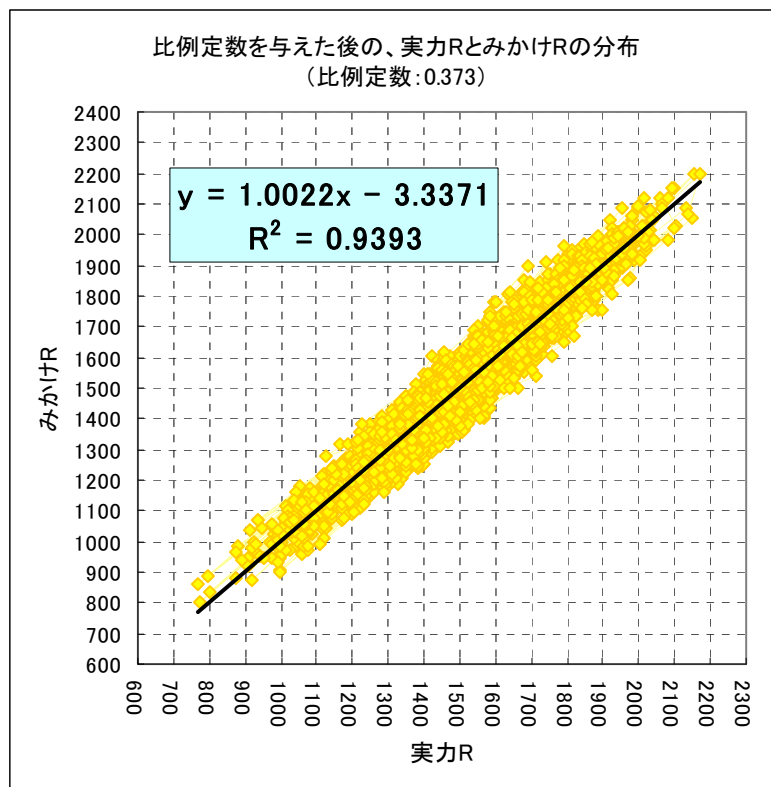


図3：「比例定数」により変動を抑えた後の、実力RとみかけRとの分布 (3000名 5000試合)

図に示したように、「比例定数」を0.373程度にすることで、

- ・ 現状のR分布との誤差が小さなR分布を実現できる
- ・ 1試合あたりのR変動の分布は、現状と大差ない程度に抑えられる
- ・ 現状のR変動システムとまったく同じ式で「他家Rとの差の調整」が保持できる
- ・ 持ち点の多寡が実力に反映される「一般的なルール」に近づけることができる

ことが期待できる。

なお、0.373という係数は、各プレイヤーが等しい実力を持った対戦を繰り返した場合の、新しい評価Rの標準偏差(24.03)と、従来の評価Rの標準偏差(8.944)との比である0.372にほぼ等しい。また、決定係数は図1の決定係数とほぼ一致しており、「偶然によるみかけRの分散」以外の誤差は特段生じていないと考えられる。みかけRの標準偏差は216.3であり、個々のプレイヤーにとってのRの「変動」も既存システムと乖離しない。

個々のプレイヤーにとって変化することは、「より高い得点を目指し、オカやウマも考慮したプレイを心がける」ようにすることだけであり、それは「一般のルールに近い麻雀に近づく」「よりいっそう正確な(納得できる)実力評価が可能となる」ことを意味する。各プレイヤーは、一般的なルールに合致した打ち方に変えるという対処を行うことで、現状と同じR(の分布)を維持することが可能であると期待できる。

5. R変動式の変更によるメリット

不遇にも「順位戦」に特化することになってしまった歴史的経緯を持つ R システムにおいて、実際の麻雀と同様、「得点の多寡」が R 変動を引き起こす変更は、(特にダントツやダンラスにおいて) プレイヤーのモチベーションを劇的に変化させる可能性を持つ。いわゆる「消化試合」も少なくなる。筆者とて、東1で親倍に放銃してしまえば、酒でも飲みながら適当に打つ——可能なら早く飛んで次の試合に行きたいと、ふと思わずにはいられない——経験があるし、逆に東2あたりで親倍をアガれば、惰性で3900程度に振り込んで局を流すこともある。しかしそれは「一般的な」麻雀のあり方とはやはり乖離している。

点数の多寡によって R 変動が行われるとすれば、そういった局面からでも「マンガン2回アガれば R がいくら上昇する」といったようにモチベーションを維持することができる。1位であっても4位であっても、また2位や3位であっても、各局において最善の態度で麻雀に取り組むことができるのである。そしてその技術の巧拙は、麻雀という偶然性の高いゲームの中で、ごく長期的に見れば、「順位」だけでは表せない、系統的な差を生み出すはずなのだ。これは特に、実力差がある者同士が打つ場合に顕著に現れるだろう。

各ネット麻雀等の事情によるが、「赤・チップ」等の影響を直接的に R に反映させることも可能となり、また、東南戦など長期間に渡る試合(実力差が本来発揮されやすいはずの試合)において、「順位」に還元され得ない適切な評価が可能になる点も重要である。

何よりも、「一般的な麻雀におけるインセンティブ」と、ネットにおけるそれを近づけることは、とりもなおさず「ネットのルールは順位戦だから実戦と違う」「ルールが特殊で、ラスだけ回避すればいい」といった、一種の「偏見」をなくすことにも貢献するだろう。それは、ネット外からネット内へと、麻雀プレイヤーを流入させるためにも有益である。

6. さらなる改善へ向けての課題

実際には、プレイヤーのインセンティブの変化により、予測困難な R 分布がもたらされることも想定せねばならない。場合によっては、しばらく様子を見た後、実測結果に基づいて R 変動の式(「比例定数」等)を「適切に」変更する必要があるかもしれない(とはいえ、それは運営として問題のある行為ではない——むしろ、一定以上の R を持ったプレイヤーが全体の上位 X%程度存在するよう調整している、と明確に示される方が、ユーザに対して親切である)。また、より理論を洗練しつつ、実測値と比較することにより、「比例定数」よりも緻密な(麻雀における正確な実力を把握可能な) R システムに改善する余地は充分にあると考えられる。

さらなる改善を考えた場合は、ルールの差、R マッチングによる影響、一定以上の R のプレイヤーのみが参加できる卓が存在する場合や、実際の上級者が新しい R 変動システムに適応した際の理論的な結果と実測の試合結果との誤差等に関する研究が必要である。そのうちいくつかは比較的容易に検証可能なものだが、実測値との比較が必要となる場合、

どうしても現実的な時間を要するという問題が残る。

運営の方針にもよるが、麻雀のゲーム性を考えれば、打ち手のタイプに依存して複数の R システムを導入するといったことも将来的に必要なようになってくる可能性もある。

しかしながら、少なくとも同じ順位を争った際に、より高い得点を保有していることが正当に評価され、なおかつこれまでの R システムによる評価値とさほど乖離しない分布を持つ R システムを——やや荒い形ではあるが——提案することができたと考える。

「ヤクマンをアガったトップと 30,000 点のトップとが等価に、あるいは 5 回放銃して飛んだラスとツモられただけのラスとが等価にしか評価されないゲームは、どこかしら『一般的な麻雀』とは異なる何かかもしれない。『一般的な麻雀』という概念が架空のものでなければ、だが」

- ※ 本件についての概要は、2009 年 7 月 12 日に『東風荘』に、また同年 7 月 13 日に『天鳳』に対して提案した。
- ※ 実際に R 変動式を変更し、長期的に運用するにあたっては、適時見直しを行う等による、信頼性の担保が必要になると想定される。

麻雀研究家 とつげき東北

【主著】『科学する麻雀』（講談社現代新書）