

トーナメントシステムの評価に関する一考察

鈴木知道

東京理科大学 理工学部 経営工学科

概要

スポーツやゲームなど様々な競技において、プレーヤーの順位づけを行うにあたって様々なトーナメントが実施されている。競技は、マラソンやポーリングのように競技の結果に対してタイムや採点など個々のプレーヤー（あるいはチーム）ごと数値化が可能な競争や採点競技と、サッカーやレスリングのように1対1の直接対決の結果が得られる対人競技に大別される。本研究は後者を対象とする。トーナメントシステムはトーナメントの規模や期間に依存するが、主要な目的はプレーヤーの競技の結果（成績、出来）に応じて適切な順位付けを行うことである。多くのトーナメントシステムが提案されているが、それらを的確に評価するのは難しいのが現状である。本研究では、統計的な視点からトーナメントシステムをとらえ、評価するにあたって考慮すべき観点に関する考察を行った。

Evaluating Tournament Systems

Tomomichi Suzuki

Department of Industrial Administration, Tokyo University of Science

Abstract

Various tournament systems are run in order to rank the players who participated in a competition of sports or games. There are competitions that all teams or players compete for time or score such as marathon or ski jumping. There are also competitions that teams or players face each other in a match such as soccer or wrestling. In this paper, the latter is considered. The tournament system to be used depends on the size and the period of the competition. In any case, one major objective the tournament is to appropriately rank the players according to the result of the competition. Today, many tournament systems are proposed, yet we do not have the methodology to evaluate their effectiveness. In this paper, various aspects are discussed from statistical point of view on how to evaluate the tournament systems.

1. はじめに

スポーツやゲームなど様々な競技において、プレーヤーの順位づけを行うにあたって様々なトーナメントが実施されている。競技は、マラソンやスキージャンプのように競技の結果に対してタイムや採点など個々のプレーヤー（あるいはチーム）ごと数値化が可能な競争や採点競技と、サッカーやレスリングのように1対1の直接対決の結果が得られる対人競技に大別される。トーナメントシステムはトーナメントの規模や期間に依存するが、主要な目的はプレーヤーの競技の結果（成績、出来）に応じて適切な順位付けを行うことである。

競争や採点競技は、タイムや採点が出そろえばその後の順位づけはその結果に応じて行うことが可能である。しかしながら、対人競技の場合はプレイヤーの順位をつけるのが一般的に困難である。最大の理由として、すべてあるいは数多くの対戦の組合せを実現させることが難しいことが挙げられる。

対人競技の順位付けに関しては、総当たり形式、ノックアウト形式などの多くのトーナメントシステムが用いられている。いろいろな工夫が行われている一方で、これらのシステムが、どのくらい良いシステムになっているかを評価することが難しいのが現状である。本研究では統計的な側面から、トーナメントシステムを評価するにあたって考慮すべき観点に関する考察を行った。

2. トーナメントシステム

対人競技におけるトーナメントシステムに関して以下に述べる。

2.1 ラウンドロビン形式（総当たり）

もっとも一般的な対戦形式の一つである。すべてのプレイヤー（競技者、選手）が自分以外の他のプレイヤーすべてと同じ回数だけ対戦する形式である。サッカーの各国のリーグ戦やワールドカップの予選、野球のペナントレース、ラグビーの6カ国対抗など、多くのトーナメントがこの形式を採用している。各対戦の試合数が1の場合はシングルラウンドロビン、2の場合はダブルラウンドロビンと呼ばれる。後者は、サッカーのホームとアウェー、チェスの先手と後手などのように実力が同じと考えられるプレイヤー同士が対戦しても勝率が50%からある程度からずれる場合などによく用いられる。

ラウンドロビンの長所は、すべての対戦組み合わせが実施されるので、プレイヤー全員に対して順位を決定できることである。また、全体としての試合数が増えるので、比較的正確に順位を決定することができる。短所としては、試合数が増えてしまうがために、実施に時間がかかってしまうことが挙げられる。また、興行的には明確な決勝戦が存在しないという点もある。

2.2 ノックアウト形式（勝ち上がり、トーナメント、ウィンブルドン方式）

ラウンドロビンと並んでよく用いられる大会形式である。各対戦で勝ったプレイヤーが次のラウンドに進む形式である。プレイヤーは1回負けると、その大会での競技は終了となる。テニスの4大大会、甲子園での高校野球など、この形式も非常に多く採用されている。

最後に決勝戦があることは興行的には非常に有利である。また試合数が少なく済むというのも長所といえよう。しかしながら、プレイヤーの順位を決定するという観点からは、実に有効性に乏しい形式といえる。上位の若干名を除けば、順位をつけること自体が難しいことがほとんどである。また、対戦組み合わせ（どのプレイヤーと対戦するか）の影響が大きい。これらの短所を補う方法として、シードの設定や、ダブルエリミネーション（敗者復活を可能とする）などの方法が提案されている。[1][2]

また、ボーリングの決勝トーナメントでよく用いられるパラマストーナメント（ステップラダー）や、ソフトボールの決勝トーナメントでみられるページシステムは、ノックアウト形式の変形であると考えられる。

2.3 スイス式

スイスで行われたチェスの大会で採用されたことから名がついたトーナメント方式である。すべての参加プレイヤーが同じ回数だけ試合を行う。各ラウンドにおいてプレイヤーはトーナメントのそれまでの成績が同じあるいは近い相手と対戦する。また、原則としては同じプレイヤーと

複数回対戦することはない。主に頭脳スポーツの分野で用いられている。

スイス式の長所は、すべてのプレーヤーに順位をつけることが可能であり、その精度はラウンドロビンほどではないにしろ、比較的正確につくことである。プレーヤー側からの視点では、1回負けたら終わり、ではないことも長所であるといえる。一方、最大の短所は、その運営の難しさにあるといえる。成績が似たもの同士が対戦するという事は、トーナメントの結果に応じて、次の対戦組み合わせ（ペアリング）を決定しなければいけないということである。ペアリング決定のルールは基本的には明確になってはいるもののバリエーションも多く[3][4]、知識や経験を必要とする。また、さまざまな工夫も提案[5][6][7][8][9]されている。近年は、ペアリングのための市販のソフトウェア[10][11]も販売されているが、ソフトウェアを用いる場合でも基本的な知識は必要であるし、対戦ラウンドごとにペアリングを決定しなければいけないことには変わりはない。

2.4 トーナメントシステムの階層構造

前節までで述べたように様々なトーナメント形式が実施されている。実際のトーナメントでは、これらのトーナメント形式を複数用いてトーナメントを運用していることが多い。サッカーの世界カップ本戦では、ラウンドロビンの予選とノックアウトの決勝を併用している。実はこの形式を用いている競技は多い。また、人数やグループを変えてラウンドロビンを連続して実施している場合もある。このように各トーナメントシステムの組み合わせまで考慮すれば、トーナメント形式は実に広がりを持つことが分かる。

3. 一対比較法

統計手法の一つである一対比較法に関して以下に述べる。

3.1 シェッフエ(Scheffe)の方法

JIS Z 8144 (=ISO 5492) によれば「数個の試料が存在するとき、それらを2個ずつ組にして評価者に呈示し、比較判断によって評価する試験方法」と定義される[12]。このISOを作成しているTC34は農産食品のための技術専門委員会であり、人間が測定器となる官能評価、感性品質において、食品における味の評価を行うに際し、2つのサンプルを（食べて）比較してどちらがより美味しいかを判断した結果をもとに、全体のサンプルの美味しさを評価するのに用いることが可能である。シェッフエの方法は、一対比較法としてよく用いられる手法であり、2個の試料の善し悪しの判定に程度をつける（AはBよりたいへん良い、少し良い、など）ことも可能である。シェッフエの方法では、以下の式で表される統計モデルを仮定している。[13]

$$x_{ijk} = (\alpha_i - \alpha_j) + \gamma_{ij} + \delta_{ij} + \varepsilon_{ijk} \quad (1)$$

ここで x は評点を、 α は各試料の効果（嗜好の度合い）を、 γ は組合せ効果の効果を、 δ は順序の効果、 ε は誤差をそれぞれ表す。添え字の i, j はそれぞれ1番目と2番目に呈示した試料のサンプル番号を、 k はその組合せでの繰り返しを表す。なお、誤差は互いに独立で平均がゼロ、分散が一定の正規分布にしたがうと仮定している。全ての試料の組合せについて複数回の繰り返しを行い、知りたい各試料の効果を中心に各種効果を推定する方法である。

3.2 サーストン(Thurstone)の方法

一対比較法の基礎となる方法である。サーストンが1927年に心理学の分野で発表したもの[14]である。シェッフエの方法もこの方法が基礎となっている。呈示される2つの試料（サーストン

は「刺激」としている)の判断をモデル化した。サーストン¹⁾は5つの場合分けを行い、それぞれの場合について、刺激の評価を行う方法を提案した。そのうちの第5番目のケース(実は一番シンプルなケースである)は以下の式と同等である。

$$x_{ijk} = (\alpha_i - \alpha_j) + \varepsilon_{ijk} \quad (2)$$

(2)式での記号の説明は(1)式に準ずる。すなわち、2つの試料を比較した場合、その判定は、2つの試料の真の効果の差だけに依存するということになる。この差についての推定を行う手法である。

3.3 ブラッドレー・テリー(Bradley and Terry)の方法

シェッフエの方法と並んでよく用いられる一対比較法の手法である。BTモデルとも書かれる。ブラッドレー・テリーの方法[15][16]では、以下のモデルが用いられる。

$$p_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \quad (3)$$

(3)式において、 p_{ij} はサンプル*i*の方がサンプル*j*より優れていると判断される確率であり、 π は各試料の効果を表している。一対比較の結果より、各試料の効果²⁾を推定することが目的となる。効果の推定には通常、最尤推定法が用いられる。

4. 順位付けの手法としてのトーナメントの位置づけ

4.1 一対比較法におけるトーナメントシステムの位置づけ

対人競技におけるトーナメントシステムを、一対比較法の枠組みで説明する。対人競技におけるひとつひとつの試合は、一対比較法において2つの試料を比較することに相当する。トーナメントシステムは、試合を行うプレイヤーの対戦組合せを決める方法であり、最終的に順位を決定する。すなわち、一対比較法において、どのように試料の組合せを決めていくかという実験計画法であるといえる。知りたい効果(ここでは各プレイヤーの真の効果)を、一連の一対比較(実験)を通して解析するという意味で、「実験計画法」と用語を用いる。

3章で述べた一対比較法の手法のうち、シェッフエの方法とサーストンの方法は、すべての試料の組合せについての情報が必要となる。しかしながらこの条件をみたしているトーナメントシステムは、ラウンドロビンだけである。その他のトーナメント形式の場合、既存の手法をそのまま用いることはできない。

ラウンドロビンのようにすべての条件(組合せ)での実験が行われている計画は完備型実験計画と呼ばれる。ラウンドロビン以外のトーナメント形式では、一部の条件(組合せ)に対してのみ実験が行われている。このような実験は、不完備型実験計画と呼ばれる。いずれの場合であっても、一対比較法における実験計画であることには変わりはない。

結論として、「トーナメントシステムは、一対比較法における実験計画である」と位置づけることができる。特に、スイス式は組合せの決定がそれまでの結果に依存するという性質をもっており、「逐次的な一対比較法」であるといえる。グリックマン[9]もこの観点から論文をまとめている。

4.2 どのような順位をつけたいのか

「適切にプレイヤーの順位をつける」にも様々な目的があると考えられる。一番強い優勝者だけを決めればよい場合もあるだろうし、「メダル」対象者の3名を適切に決めたい場合もあるであろう。また、その後行われるであろう決勝トーナメントへの参加者を決めるための予選会であるならば、定められた人数の予選通過者をなるべく公平に選びたい場合もあるであろう。また、参加者全員の順位付けを行いたい場合もある。

既存の文献においても、トーナメントシステムの評価を行うことは容易ではなく、様々な評価基準が用いられている。その中でも、あらかじめ仮定したシード順になるか、あるいは何らかの数値化（レーティングを設定）をしておいてその数値と比較する、という方法が主流である。しかしながら、この考えにはひとつ問題点がある。それは、強いプレイヤーが必ずしもいつも勝つわけではないからである。ばらつきを考慮しなくては行けない。

たとえばシード順が1位のプレイヤーを考えよう。参加者はいつも同じとして、何回もトーナメントを行ったら、このプレイヤーが優勝する回数が一番多いであろう。シード順が1位のプレイヤーが優勝する確率が他のプレイヤーよりも低かったら、それは良いトーナメントシステムとはいえない。しかし、シード順が1位のプレイヤーがほとんどの場合優勝してしまう場合はどうか。すると今度は2位以下の選手に不利なシステムになっている可能性がある。

極論をすれば、もし「強い人（シード順が良い人、レーティングが高い人）が勝つのが理想」であるならば、試合をしなければよいのである。あるノックアウト形式に関する論文^[1]において、組合せ表でシード順で **18672345** と並ぶのが一番良いとの結論を出したものがあつたが、これはいただけない。これはシード順1位にとっては良い組合せであるが、他のプレイヤーに対してはまったく良い組合せとはいえない。すなわち不公平なトーナメントなのである。

参加者が決定した時点で、シード順1位の選手がどのくらいの確率で優勝するのが妥当か、というものがあるはずだ。もちろんこの妥当であると判断される確率を求めること自体、非常に難しいことには間違いない。しかしながら、トーナメントシステムとしての評価は、その妥当な確率に対して、高すぎないか、低すぎないか、の両方が評価されるべきである。

4.3 提案する評価方法

本研究においては、トーナメントシステムを評価するにあたって、プレイヤーの当該トーナメントでの成績に応じて、適切に順位をつけられる能力を対象とする。すなわち、興行的な観点ではなく、純粋にプレイヤーの成績を評価することを対象とする。

前節で述べた評価したい順位づけのうち、どの場合かというのは限定するつもりはない。むしろ、上記のどの場合にも対応できるような、「適切にプレイヤーの順位をつけ方」の評価の枠組みを提案したいのである。具体的には以下を提案したい。

「トーナメントにおける順位付けは、当該トーナメントでの『出来(performance)』に応じて行われるべきである。」

『出来』も様々な定義が考えられる。大会が終わったあとから見れば、「急所」であつた試合があり、そこで勝ったプレイヤーを出来が良かったと定義することも可能ではあるはずだ。しかし、この場合はどこが急所であつたかを判断するしくみを考えなくてはならない。本当は、「ここの一番」の強さはあるかもしれない、しかしながら本研究では、当該トーナメントでの強さの平均を「出来」と定義する。いささか平凡ではあるかもしれないが、プレイヤーが発揮する1試合ごとの「出来」の平均という自然な尺度にはなっている。

5. 『出来』の評価方法について

前章で、「トーナメントにおける順位付けは、当該トーナメントでの『出来(performance)』に応じて行われるべきである」と書いたが、そのためには何が必要であろうか。大きく2つの項目があると考えられる。一つは、プレーヤーの出来の数値化、勝敗の決定方法であり、もう一つはプレーヤーの母集団の設定である。

5.1 出来の数値化、勝敗の決定方法

第一に必要なのが、トーナメントでのプレーヤーの出来の数値化である。これに関しては、3章で述べたサーストンのモデルが最も扱いやすい。サンプルの真の効果に対応するのが、プレーヤーの真の強さである。しかし実際には、試合ごとに発揮できる出来にばらつきがでる。すなわちモデル上の誤差の実現値である。この真の効果と誤差を足し合わせたものが、その試合での出来であるといえる。トーナメントの出来は、各試合の出来を平均すれば得られる。

サーストンのモデルを用いてチェス界に応用したものが、イロ(Elo)によるイロ・レイティング[17] (以後、単にレイティングと呼ぶ) である。その基本は、プレーヤーの強さの真の値が、そのプレーヤーのレイティングであり、試合ごとの出来の分布は、平均がゼロ、標準偏差が200の正規分布にしたがう、というものである。この仮定のもとでは、試合ごとの出来を数値化するのは容易である。また、プレーヤー間の勝率は正規分布を用いて容易に表すことが可能である。しかしながら、正規分布の分布関数の値を求めるには正規分布表を引く必要があるため、正規分布に近い分布を当てはめて計算を楽にしたものが実際には用いられていることが多い。

ブラッドレー・テリーのモデルは、そして Jackson モデル[2]は、プレーヤー間の勝率を与えるという観点からは申し分ないあるいは十分であるが、試合ごとの「出来」を評価するには向いていない。また、プレーヤー間の勝率として、必ず強者勝ち、あるいはいつも勝率 1/2 としてトーナメントシステムを評価する方法もあるが、この場合、何が評価できているのかを吟味する必要がある。

現状では、レイティングをもとに出来を計算するのが最も現実的であると考えられる。

5.2 プレーヤーの母集団の設定

レイティングを用いてプレーヤーの強さおよび出来を表現できさえすれば、トーナメントシステムの評価が可能かという点、まだ不十分である。トーナメントに参加しているプレーヤーの強さの分布(プレーヤーの母集団)を考慮しなくてはならない。もちろん、すべての場合を検討し尽くすのは不可能であるが、典型的に現れるいくつかのパターンについては検討するのは可能であると思われる。

ある競技のすべてのプレーヤーの強さの分布に正規分布が仮定できるとするならば、まず正規分布は候補として挙げられる。また、実際にあるトーナメントに参加するプレーヤーはその一部であることを考慮すれば、以下の4つの場合が考えられる。[18]

- a) すべての強さのプレーヤーが均等に参加：この場合は、結局正規分布となる。
- b) 上位者、最上位者が参加：競技では強い者に興味が向かうので、実はこのパターンは多いのではないかと推察される。この場合は右に歪んだ分布になる。単純な分布として三角分布が考えられる。
- c) 下位者を中心に参加：b)の逆の分布となる。
- d) 中位者を中心に参加、幅が狭い範囲内のプレーヤーの参加：この場合はプレーヤーごとの強さの差が小さい。したがって一様分布が向いている。

以上の4つの場合、いずれもそれぞれの分布のばらつき(分散)の大きさをどのくらいに設定

するのかという問題は残っている。また、以上の母集団が2つ以上混在する母集団を考慮してもよいかもしれない。

まったく別の観点から、実際に行われたトーナメントに参加したプレーヤーのレイティングをそのまま用いるという方法[19]もある。この場合は、プレーヤーの強さの分布のどの特徴が最終的にトーナメントシステムの評価に影響を与えたかを明らかにすることが、分布を仮定した場合に比べてより困難になるとは予想される。

6. まとめ

本研究では、トーナメントシステムを、順位付けを行うための実験計画であるとみなし、よいトーナメントシステムを評価するための枠組みを提案した。どんな競技においても、誰が強いという議論は尽きない。今回の提案が少しでも強さの客観的な評価につながれば、と願っている。

参考文献

- [1] McGarry T., Schutz R.W. (1997), Efficacy of traditional sport tournament structures, *Journal of the Operational Research Society*, 48, 65-74.
- [2] Marchand E. (2002), On the Comparison between Standard and Random Knockout Tournament, *The Statistician*, 51, pp.169-178.
- [3] FIDE Handbook, <http://www.fide.com/official/handbook.asp>
- [4] 日本チェス協会, <http://www.jca-chess.com/>
- [5] 瀧澤武信, 柿木義一(2002), 世界コンピュータ将棋選手権における対戦組み合わせシステムの有効性, 第7回ゲーム・プログラミングワークショップ, pp. 93-100.
- [6] 橋本剛, 長嶋淳, 飯田弘之(2002), シミュレーションによる大会方式検証の提案—世界コンピュータ将棋選手権を題材にして—, 第7回ゲーム・プログラミングワークショップ, pp.101-108.
- [7] 香山健太郎(2003), シミュレーションに基づくスイス式トーナメント組合せ方法の評価, 第8回ゲーム・プログラミングワークショップ.
- [8] 河合孝尚, 川島伸太郎, 飯田弘之(2005), トーナメントスコアに基づく順位決定の新たな指標, 情報処理学会ゲーム情報学研究発表報告, pp.27-34.
- [9] Glickman, M. E. (2005), Adaptive paired comparison design, *Journal of statistical planning and inference*, 127, 279-293.
- [10] Swiss Perfect Home Page, <http://www.swissperfect.com/>
- [11] Swiss-Manager Home Page, <http://swiss-manager.at/>, <http://schach.wienerzeitung.at/>
- [12] JIS Z 8144:2004, 官能評価分析—用語, 日本規格協会
- [13] Scheffe, H. (1952), An Analysis of Variance for Paired Comparisons, *Journal of American Statistical Association*, 147, 381-400.
- [14] Thurstone, L.L. (1927), A Law of Comparative Judgment, *Psychological Review*, 34, 278-286.
- [15] Bradley, R.A. and Terry, M.E. (1952), Rank Analysis of Incomplete Block Designs, *Biometrika*, 39, 324-345.
- [16] 藤野和建, 竹内啓(1988), スポーツの数理科学, 共立出版.
- [17] Elo, A. E., *The Rating of Chessplayers, Past and Present* (second edition), ARCO, 1978.
- [18] 野崎雄太, 皆川浩之, 鈴木知道 (2006), 競技における効果的な順位決定法に関する研究, 品質管理学会第36回年次大会発表会要旨集
- [19] Sakoda Y., Suzuki T. (2005), A Study on Sports Tournament Systems, *Proc. The 19th Asia Quality Symposium*.