Bradley-Terry モデルの改良

松 田 眞 一
E-Mail: matsu@nanzan-u.ac.jp

Bradley-Terry モデルは対戦型のスポーツにおいてチームの強さを推定するモデルである。このモデルを用いると直接対戦のないチーム間についてもその勝敗の確率を予測できるため大変有用であるが, 引き分けの存在するスポーツには適用できないという欠点があった。ここでは, 引き分けを考慮したモデルに改良する方法を提案する。

1　はじめに

スポーツに勝敗はつきものであり, どちらが強いのかという問題は最も大きな関心事である。そのため, 複数のチームが互いに対戦する場合について一方が他方に勝つ確率をモデル化する試みは古くから行われている。その中で Bradley-Terry モデル（以下では BT モデルと略す）は強さを 1 次元的に評価できて, この問題に対する簡便な解答を与えるものである。（竹内・藤野 (1985) 参照）

しかし, このモデルでは引き分けを考慮していないため, 引き分けが存在するスポーツの場合は結果から全く除外するか 0.5 勝 0.5 敗として計算に組み込むしかない。それでは, 引き分けがどのように起こっているのか判断する方法がない。

本論文では, BT モデルを改良し, 引き分けの起こる確率も組み込んだ 1 つのモデルを提案する。

2　改良型 BT モデル

本論文では, 勝敗を考える対象をチームと呼び, 全部で m チームあるとする。（モデル自体は個人競技のスポーツにも当然適用できる。）第 i 番目のチームが第 j 番目のチームに対して勝つ確率, 負ける確率, 引き分ける確率をそれぞれ p_{ij}, q_{ij}, r_{ij} とすると

$$p_{ij} + q_{ij} + r_{ij} = 1, p_{ij} = q_{ji}, r_{ij} = r_{ji}$$

が成り立つ。いま, BT モデルと同様に各チームの強さ π_i というパラメータを導入し,

$$p_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \cdot (1 - r_{ij})$$

というモデルが成立しているとする。一方, 引き分けについても何らかのモデルを導入しないと定式化できない。ここでは次のように 2 つのパラメータ α, β に依存するモデル化を考えよう。

$$r_{ij} = \alpha - \beta \left(\frac{\pi_i - \pi_j}{\pi_i + \pi_j} \right)^2$$

このモデルは 2 つのチームの強さの近さに依存して引き分けの確率が決まると考えるもので, 近ければ近いほど引き分けの確率が上がるのかどうかはパラメータ β の符号から判断

できるようになっている。なお, 引き分けの確率が強さとは無関係に定まると考えるモデルを考えることも可能ではあるが, 勝敗と引き分けを別のものとして考えることになり全体の定式化としては簡単で面白みに欠けるものとなる。

3 パラメータの推定

実際に第 i 番目のチームと第 j 番目のチームの対戦が n_{ij} 回行われた場合に第 i 番目のチームが勝つ回数の確率変数を X_{ij}, 負ける回数の確率変数を Y_{ij} , 引き分けとなる回数の確率変数を Z_{ij} とおくとそれらは多項分布に従っていると考えられ, 次のように確率分布が定まる。

$$\Pr\{X_{ij}=x_{ij}, Y_{ij}=y_{ij}, Z_{ij}=z_{ij}; 1\le i<j\le m\} = \prod\prod_{i<j} \frac{n_{ij}!}{x_{ij}!y_{ij}!z_{ij}!} p_{ij}^{x_{ij}} q_{ij}^{y_{ij}} r_{ij}^{z_{ij}}$$

この確率を L とおくと, それは改良型 BT モデルに対して次のように書き直せる。

$$\begin{aligned} L &= \prod\prod_{i<j} \frac{n_{ij}!}{x_{ij}!x_{ji}!z_{ij}!} \left(\frac{\pi_i}{\pi_i+\pi_j}\right)^{x_{ij}} \left(\frac{\pi_j}{\pi_i+\pi_j}\right)^{x_{ji}} \\ &\quad \cdot \left(1+\beta\left(\frac{\pi_i-\pi_j}{\pi_i+\pi_j}\right)^2-\alpha\right)^{x_{ij}+x_{ji}} \left(\alpha-\beta\left(\frac{\pi_i-\pi_j}{\pi_i+\pi_j}\right)^2\right)^{z_{ij}} \end{aligned}$$

これに対してまず両辺の対数を取ると

$$\begin{aligned} \ell = \log L &= const + \sum_{i=1}^{m} T_i \log\pi_i - \sum\sum_{i<j}(n_{ij}-z_{ij})\log(\pi_i+\pi_j) \\ &\quad + \sum\sum_{i<j}(n_{ij}-z_{ij})\log\left(1+\beta\left(\frac{\pi_i-\pi_j}{\pi_i+\pi_j}\right)^2-\alpha\right) \\ &\quad + \sum\sum_{i<j} z_{ij}\log\left(\alpha-\beta\left(\frac{\pi_i-\pi_j}{\pi_i+\pi_j}\right)^2\right) \end{aligned}$$

となる。ただし, T_i は第 i 番目のチームの総勝ち数を表す。すなわち,

$$T_i = \sum_{j\ne i} x_{ij}$$

である。この ℓ に対して $\sum_{i=1}^{m}\pi_i = k$ という制約の下でラグランジュの未定乗数法を用いると次のような最尤方程式が導かれる。

$$\begin{cases} \dfrac{\partial}{\partial\pi_i}(\ell-\lambda(\displaystyle\sum_{j=1}^{m}\pi_j-k))=0 & (i=1,\dots,m) \\ \dfrac{\partial}{\partial\alpha}(\ell-\lambda(\displaystyle\sum_{i=1}^{m}\pi_i-k))=0 & \\ \dfrac{\partial}{\partial\beta}(\ell-\lambda(\displaystyle\sum_{i=1}^{m}\pi_i-k))=0 & \\ \dfrac{\partial}{\partial\lambda}(\ell-\lambda(\displaystyle\sum_{i=1}^{m}\pi_i-k))=0 & \end{cases}$$

これを計算すると次のようになる。

$$
\begin{cases}
\dfrac{T_i}{\pi_i} - \displaystyle\sum_{j\neq i} \dfrac{n_{ij}-z_{ij}}{\pi_i+\pi_j} + \sum_{j\neq i}(n_{ij}-z_{ij})\dfrac{4\beta\pi_j}{(1-\alpha)(\pi_i+\pi_j)^2+\beta(\pi_i-\pi_j)^2}\dfrac{\pi_i-\pi_j}{\pi_i+\pi_j} \\
\qquad + \displaystyle\sum_{j\neq i} z_{ij}\dfrac{-4\beta\pi_j}{\alpha(\pi_i+\pi_j)^2-\beta(\pi_i-\pi_j)^2}\dfrac{\pi_i-\pi_j}{\pi_i+\pi_j} - \lambda = 0 \\
\displaystyle\sum\sum_{i<j}(n_{ij}-z_{ij})\dfrac{-(\pi_j+\pi_j)^2}{(1-\alpha)(\pi_i+\pi_j)^2+\beta(\pi_i-\pi_j)^2} \\
\qquad + \displaystyle\sum\sum_{i<j} z_{ij}\dfrac{(\pi_j+\pi_j)^2}{\alpha(\pi_i+\pi_j)^2-\beta(\pi_i-\pi_j)^2} = 0 \\
\displaystyle\sum\sum_{i<j}(n_{ij}-z_{ij})\dfrac{(\pi_j-\pi_j)^2}{(1-\alpha)(\pi_i+\pi_j)^2+\beta(\pi_i-\pi_j)^2} \\
\qquad + \displaystyle\sum\sum_{i<j} z_{ij}\dfrac{-(\pi_j-\pi_j)^2}{\alpha(\pi_i+\pi_j)^2-\beta(\pi_i-\pi_j)^2} = 0 \\
\displaystyle\sum_{i=1}^{m}\pi_i - k = 0
\end{cases}
$$

ここで第 1 式の両辺に π_i を掛けて i について和を取ると

$$
\sum_{i=1}^{m} T_i - \sum_{i=1}^{m}\sum_{j\neq i}\frac{\pi_i}{\pi_i+\pi_j}(n_{ij}-z_{ij}) + \sum_{i=1}^{m}\sum_{j\neq i}\frac{4\beta\pi_j(n_{ij}-z_{ij})}{(1-\alpha)(\pi_i+\pi_j)^2+\beta(\pi_i-\pi_j)^2}\frac{\pi_i-\pi_j}{\pi_i+\pi_j}
$$
$$
+ \sum_{i=1}^{m}\sum_{j\neq i}\frac{-4\beta\pi_j z_{ij}}{\alpha(\pi_i+\pi_j)^2-\beta(\pi_i-\pi_j)^2}\frac{\pi_i-\pi_j}{\pi_i+\pi_j} - \lambda\sum_{i=1}^{m}\pi_i = 0
$$

となり, 第 4 式を用いて

$$
\sum_{i=1}^{m} T_i - \sum\sum_{i<j}(n_{ij}-z_{ij}) - \lambda k = 0
$$

となる。$\sum_{i=1}^{m} T_i = \sum\sum_{i<j}(n_{ij}-z_{ij})$ であるから $\lambda = 0$ が導かれる。

したがって, 次のような関係式が導かれる。

$$
\begin{cases}
\pi_i = \dfrac{T_i}{\displaystyle\sum_{j\neq i}\dfrac{n_{ij}-z_{ij}}{\pi_i+\pi_j} - \sum_{j\neq i}\dfrac{4\beta\pi_j(\pi_i-\pi_j)}{\pi_i+\pi_j}\dfrac{(\alpha n_{ij}-z_{ij})P_{ij}-\beta Q_{ij}}{\{(1-\alpha)P_{ij}+\beta Q_{ij}\}\{\alpha P_{ij}-\beta Q_{ij}\}}} \\
\alpha = \dfrac{\displaystyle\sum\sum_{i<j} z_{ij} + \beta\sum\sum_{i<j}\dfrac{(n_{ij}-z_{ij})Q_{ij}}{(1-\alpha)P_{ij}+\beta Q_{ij}}}{\displaystyle\sum\sum_{i<j} n_{ij}} \\
\beta = \dfrac{\displaystyle\sum\sum_{i<j}(\alpha n_{ij}-z_{ij})}{\displaystyle\sum\sum_{i<j}\dfrac{z_{ij}P_{ij}}{\alpha P_{ij}-\beta Q_{ij}}}
\end{cases}
$$

ただし,

$$P_{ij} = (\pi_i + \pi_j)^2, \quad Q_{ij} = (\pi_i - \pi_j)^2$$

とおく。

この関係式を基に繰り返し計算で推定値を求めるとよい。

4 計算手順

実際に推定値を求めるためにはプログラムを作成する必要がある。以下にその手順を示す。

手順 1 初期値を決定する。

$$\pi_i = 50; \quad i = 1, \ldots, m, \qquad k = \sum \pi_i = 50m$$

$$\alpha = 0.1, \quad \beta = -0.05$$

手順 2 以下の関係式を用いて強さの推定値 π_i' を求める。

$$\pi_i' = \frac{T_i}{\displaystyle\sum_{j\neq i} \frac{n_{ij} - z_{ij}}{\pi_i + \pi_j} - \sum_{j \neq i} \frac{4\beta\pi_j(\pi_i - \pi_j)}{\pi_i + \pi_j} \frac{(\alpha n_{ij} - z_{ij})P_{ij} - \beta Q_{ij}}{\{(1-\alpha)P_{ij} + \beta Q_{ij}\}\{\alpha P_{ij} - \beta Q_{ij}\}}}$$

ただし,

$$P_{ij} = (\pi_i + \pi_j)^2, \quad Q_{ij} = (\pi_i - \pi_j)^2$$

である。

手順 3 π_i' が $\sum \pi_i' = k$ を満たすように基準化する。すなわち、$k\pi_i' / \sum \pi_i'$ を新たな π_i' とする。

手順 4 π_i' が次の式を満たす場合, それらを新たな π_i として手順 2 に戻って手順 2〜4 を繰り返す。

$$\sqrt{\sum \{(\pi_i - \pi_i')/50\}^2} > 10^{-8}$$

手順 5 以下の関係式より α', β' を求める。

$$\alpha' = \frac{\displaystyle\sum\sum_{i<j} z_{ij} + \beta \sum\sum_{i<j} \frac{(n_{ij} - z_{ij})Q_{ij}}{(1-\alpha)P_{ij} + \beta Q_{ij}}}{\displaystyle\sum\sum_{i<j} n_{ij}}$$

$$\beta' = \frac{\displaystyle\sum\sum_{i<j} (\alpha n_{ij} - z_{ij})}{\displaystyle\sum\sum_{i<j} \frac{z_{ij} P_{ij}}{\alpha P_{ij} - \beta Q_{ij}}}$$

ただし,

$$P_{ij} = (\pi_i + \pi_j)^2, \quad Q_{ij} = (\pi_i - \pi_j)^2$$

である。

手順 6 α' が自然な状況になるように以下の修正を施す。
もし $\alpha' < 0$ ならば $\alpha' = 0.00001$ とする。
もし $\alpha' > 1$ ならば $\alpha' = 0.99999$ とする。

手順 7 α', β' が次の式を満たす場合, 新たな $\alpha = (\alpha' + 9\alpha)/10$, $\beta = (\beta' + 9\beta)/10$ として手順 2 に戻って手順 2〜7 を繰り返す。

$$\sqrt{(\alpha - \alpha')^2 + (\beta - \beta')^2 + \sum\{(\pi_i - \pi_i'')/50\}^2} > 10^{-6}$$

ただし, π_i'' は前回の π_i の値である。

手順 8 得られた π_i, α, β を推定値とする。

実際のプログラムは Java を用いて作成した。

手順の中の初期値と終了条件の定数の設定は適用する問題によってはもっと適切なものがあるかもしれない。上の初期値設定は次の数値例を多少意識して設定してある。モーメント法で推定するなどデータに依存した初期値設定を考えるともっとよいであろう。また, 収束のさせる方法も上記のように π と α, β を分けた方がよいと思われるのだが, もっとよい方法があるかもしれない。特に α, β の改善式は上記のようにあまり大きく変動しない工夫をしなければ発散する場合がでてきた。収束の早さは初期値設定に依存するが、収束における誤差の変動を見てみると相当よい初期値を与えない限り誤差が減少する方向だけとはならないようである。数値例に示したものでも一旦減少したものの再び上昇しまた減少するというような収束の仕方を示していた。

5 数値例

サッカーの J1 における 2001 年度のデータに関しての分析結果を示し, 従来の BT モデルのとの差異を考察する。

表は第 1, 第 2 ステージの全 30 節のデータから得られた強さの推定値と引き分け係数である。通常の BT モデルに対して引き分けを 0.5 勝 0.5 敗として解いたものと引き分けを除いて解いたものおよび提案する方法の結果を示した。

提案する方法での引き分け係数の β の値は正であり, チームの力の差が大きくなると引き分けにくいことを表している。特に α と β の値がほとんど同じなので力の差が激しいとほとんど引き分けないということになる。この結果はサッカーの一般的なものかというとそうではない。同様の解析を J2（2001 年度全 44 節）に関しても行ってみたが, その場合は $\alpha = 0.0001$, $\beta = -0.239$ となった。すなわち, 力が近い場合はほとんど引き分けず, 力の差が激しいほど引き分けるということになる。このことは J2 の下位チームが引き分けを狙いにいっていることを示唆している。また, α が 0 であることは実力の近いチーム同士では引き分けにくいことを表している。その理由としては, J2 内の実力の差が激しく同じぐらいの強さのチームに勝てるかどうかが順位の上昇に大きな影響があることと, J1 への昇格がかかるチームは無理をしてでも勝ち点 3 を目指していくことが考えられる。後者の理由では先に述べた実力差があるチームで引き分けの確率が高いことと矛盾するように感じるか

表 1: J1(2001 年度）の強さの推定結果

年間順位	チーム名	引分を 0.5 勝 0.5 敗	引分を除く	提案する方法
1	磐田	263.9	290.3	357.1
2	鹿島	67.5	67.1	62.1
3	市原	54.2	53.6	46.8
4	清水	62.7	59.1	59.8
5	名古屋	58.3	60.2	49.1
6	柏	38.3	37.4	30.3
7	G 大阪	35.8	34.2	28.2
8	FC 東京	38.3	36.1	31.5
9	広島	27.2	25.3	20.5
10	浦和	27.2	26.3	19.8
11	札幌	25.4	22.6	19.7
12	神戸	25.4	21.9	19.5
13	横浜 M	22.1	18.3	17.7
14	東京 V	20.6	18.5	15.0
15	福岡	17.8	16.0	12.4
16	C 大阪	15.3	13.0	10.7
	α	—	—	0.113
	β	—	—	0.105

もしれないが、J1 とは違って J2 の上位のチームの決定力では守りに入られると延長戦の中で V ゴールを挙げられないのであろう。

引き分けを 0.5 勝 0.5 敗とした場合の結果と提案した方法の結果を比べると全体的に見て上位のチームほど変化が小さく下位チームほど大きいことが分かる。これは引き分けを一律に 0.5 勝 0.5 敗としたことにより上位チームには損に下位チームには得に働いたためであろう。この影響は磐田に一番顕著に現れている。2001 年度の磐田の強さを象徴する結果といえる。

引き分けを除いた場合の結果と提案した方法の結果を比較すると清水などのように余り変化がないチームと名古屋のように弱くなったチームが見られる。その違いは引き分け係数から考えると力の離れたチームに痛い引き分けを喫したかどうかである。引き分けを除いた場合は当然のことであるが, 引き分けを 0.5 勝 0.5 敗とした場合でもこのような強さに対する影響をうまく取り込めていない。これは引き分けの確率が強さの近いもの同士と強さの離れたチーム同士では異なっていることによるもので, それが取り込めることが提案した方法の長所である。すなわち, 提案した方法では引き分けの状況を確認できるだけでなく, 延長戦での決定力を含むようなよりよい強さの推定ができているということである。

6　おわりに

本論文では, 強さを推定する BT モデルの改良を提案したが, 引き分けを取り込む試みは成果を上げたといえる。しかし, 今回のモデル式では強さの近さ遠さに応じて引き分け確率が変化するというもので引き分けしやすいチームや引き分けしにくいチームといった個別の分析はできていない。より精密な分析にはデータが豊富に必要となるためサッカーの場合は難しい問題であるが, 通常の BT モデルの解析のカモ・苦手の分析のように何らかの方策を考える余地はあるであろう。

参考文献

竹内啓・藤野和建 (1985): “スポーツの数理科学— もっと楽しむための数字の読み方 —”, 共立出版.