

研究ノート

サッカーの得点と統計的研究 (その3 : toto の確率の考察)

鷺 崎 早 雄

(1)

サッカーは世界的に人気があるスポーツであるが、その理由にはルールが比較的簡単なこと、世界中どこに行っても統一がとれていることなどがある。また、ヨーロッパや南米を中心にサッカーくじが盛んである。日本でもJリーグ発足時からサッカーくじ（正式にはスポーツ振興くじ）の導入が検討され、平成10年5月に「スポーツ振興投票の実施等に関する法律」が公布、諸準備の後に平成12年11月に第1回が開催された。サッカーくじのルールもサッカーと同様に簡単なものであり、多くの人たちが簡単に賭けを楽しんでもらうことを念頭においている。

(2)

サッカーくじがあたる確率を高めるために、試合結果を予想する統計的手法が研究されている。Moroney (1956) や Reap et al. (1971) は、試合ごとのゴール数の分布をマクロ的に分析し、それがポアソン分布や超幾何分布に従っていることを示した。しかし、マクロな分布がわかっても、試合における両チームの対戦結果を予想することはできない。Maher (1982) は、ホームチーム、アウェイチームそれぞれの得点は独立のポアソン分布に従っていることを示した。本研究においても、研究ノート1 (文献 [6]) でイギリスのプレミアリーグのデータがポアソン分布に従うことを示した先行研究を検討し、研究ノート2 (文献 [7]) で実際のJリーグデータを用いて同じ結果を得た。しかし、Maherの方法は試合結果を多数集めて分布のパラメータを

推定しているため、事後 (ex post) 予測である。この方法では事前 (ex ante) 予測をすることはできない。Dixson and Coles (1997) は、両チームのスコアや試合結果を事前予測することができるモデルを提案した。このモデルでは、両チームのスコアは独立にポアソン分布に従うが、両チームが0点か1点台にあるときには独立ではなく交互作用が働くようなモデルによって事前予測ができるように調整する。このように、ある条件では、独立なポアソン分布に従うというマクロ的な仮定を一部変更して調整したり、時間的に最近のパフォーマンス結果を取り入れたりして事前予測能力を得る方法が提案されている。

(3)

日本のスポーツ振興くじには2種類ある。そのうち toto は開催週ごとに指定された13試合のホームチームの結果を、(勝ち)、(負け)、(その他)で予想する。totoGOAL3は開催週ごとに指定された3試合のホームチーム、アウェイチームそれぞれの得点を、(0点)、(1点)、(2点)、(3点以上)で予測する。toto、totoGOAL3とも、くじの種類にはSINGLEおよびMULTIがある。SINGLEは1つの試合について1つだけ予想するが、MULTIの場合には、1つの試合に対して迷ったら複数の結果を予想することができる。MULTIのうちDOUBLEでは1つの試合に対して2つの予想をすることができる。たとえば、A対Bが試合するとき、ホームの勝ちと引き分け、ホームの負けと引き分け、ホームの勝ちとホームの負け、合計3種類のいずれかを予想することができる。図1の

図1 くじの概念

	ホーム	アウェイ	SINGLE			MULTI					
						DOUBLE			TRIPLE		
			1	0	2	1	0	2	1	0	2
1	広島	清水			○			○			○
2	柏	鹿島	○			○	○		○	○	○
3	磐田	FC東京		○			○			○	
4	神戸	横浜			○			○			○
5	札幌	G大阪			○	○	○		○	○	○
6	名古屋	新潟	○			○	○		○		

DOUBLEの例では6試合のうち3試合をDOUBLEにしてある。またTRIPLEでは1つの試合に対して3つの予想をすることができる。したがって、ある試合をTRIPLEとすれば、結果がどうなってもその試合は正しい予想になる。図1のTRIPLEの例では6試合のうち2つの試合をTRIPLEにしてある。totoの購入方法にはもう1つランダム購入という方式がある。ある試合において予想ができないときに、その試合結果をランダムとし、勝ち、負け、その他のいずれかをコンピュータに決めてもらう方式である。この方式の極端な場合として13試合全部をランダムにすることもできる。こうなるとロータリーと変わらなくなるが、サッカーにまったく門外漢であってもくじを買うことができる。totoの販売額のうち50%が1等、2等、3等に払い戻される。1等は13試合すべて予想が正しかった場合、2等は1つだけ正しくなかった場合、3等は2つ正しくなかった場合である。くじを買う人が1枚のチケットで何かがあたる確率は1等、2等、3等の確率をすべて足した場合である。残りの50%から販売管理費用を除いた金額が収益となるが、その2/3はスポーツ教育やスポーツ振興事業に配布され、1/3は国庫に納められる。日本ばかりではなく、ヨーロッパの他の国でも、おおむね50%前後が政府関係に配分されている。

(4)

ここでtotoの当選確率を検討してみる。簡単のために事前確率を予想できる材料はないものとし、過去のデータとして得られたマクロな事後確率だけで考える。図2は2004年度のJリーグ各チームのホームゲームの戦績である。全体の試合数は240試合、そのうちホームチームが勝った試合が102試合、引き分けが64試合、ホームチームが負けた試合が74試合である。240試合の事後確率は、ホームチームの勝ちが0.425、引き分けが0.267、ホームチームの負けが0.308である。この確率についてはStefani(文献[5])がヨーロッパ各国のデータ7000試合について計算しているが、それによるとホームチームの勝ちが0.490、引き分けが0.280、ホームチームの負けが0.230である。2004年のJリーグデータと比較して、ヨーロッパのデータではホームチームの負けよりも引き分けの確率が大きいなど微妙な違いはあるが、よく似ているといえよう。

いま、ホームチームが勝つ確率を p_1 、引き分ける確率を p_0 、ホームチームが負ける確率を p_2 と置く。また、totoの13試合のうち、 n_1 をホームチームが勝つと予想する試合数、 n_0 を引き分けと予想する試合数、 n_2 をホームチームが負けると予想する試合数とする。ただし、 $n_1 + n_0 + n_2 = 13$ である。

2004年のJリーグのデータによってそれぞれの値を求めると、 $p_1 = 0.425$ 、 $p_0 = 0.267$ 、

図2 2004年度Jリーグの結果

	1st ステージ			2nd ステージ			年 間		
	1	0	2	1	0	2	1	0	2
横 浜	6	2	0	4	1	2	10	3	2
磐 田	6	1	1	1	3	3	7	4	4
浦 和	5	2	0	6	1	1	11	3	1
G 大 阪	4	1	2	5	0	3	9	1	5
鹿 島	4	3	0	4	2	2	8	5	2
F C 東 京	5	2	1	2	3	2	7	5	3
市 原	3	5	0	5	1	1	8	6	1
名 古 屋	3	3	2	3	1	3	6	4	5
東 京 V	3	2	2	3	2	3	6	4	5
大 分	4	1	2	2	2	4	6	3	6
清 水	3	1	3	2	0	6	5	1	9
神 戸	3	3	2	2	2	3	5	5	5
広 島	2	4	2	3	4	0	5	8	2
新 潟	0	2	5	4	2	2	4	4	7
柏	1	1	5	0	5	3	1	6	8
C 大 阪	1	1	6	3	1	3	4	2	9
合 計							102	64	74

$p_2=0.308$ となる。また、13試合のうち、 $n_1=6$ 、 $n_0=3$ 、 $n_2=4$ と置く。このとき、13試合のうち正しく予想する試合数 E_0 、すべて予想が正しい確率 P_0 、13試合のうち1つだけ正しくない予想がある確率、13試合のうち2つ正しくない予想がある確率は次のようになる。まず、13試合のうち確率的に正しく予想する試合数 E_0 は期待値の考え方で求められる。この場合、 E_0 は下の式で計算されるとおり4.583である。何も事前確率がない場合、過去の事後確率だけで予想すると13試合のうち、正しく予想する試合数の期待値は4.5試合でしかない。

正しい選択をする試合数の期待値

$$= n_1 p_1 + n_0 p_0 + n_2 p_2 = E_0$$

1つもエラーがなく全13試合を正しく予想する確率は、勝ちを買った数 n_1 、引き分けを

買った数 n_0 、負けを買った数 n_2 で、それぞれすべて正しい事象が発生する確率を計算することになる。式は以下のようになる。これまで与えた数字を前提にすると、 P (no error) は $1.00942E-06$ である。 P の逆数をとると、990672と計算されるから、この事象は約99万回に1回しか起こらない確率である。

$$p(\text{no errors}) = (p_1)^{n_1} (p_0)^{n_0} (p_2)^{n_2} = p_0$$

1つだけエラーがあり13試合中12試合を正しく予想する確率は、勝ちを予想したにもかかわらず1試合だけ間違えた場合、引き分けを予想したにもかかわらず1試合だけ間違えた場合、負けを予測したにもかかわらず1試合だけ間違えた場合の確率を足したものである。Stefani (文献 [5]) の表現に従えば、式は以下のようになる。これまで与えた数字を

前提にすると、P (one error) は2.55792E-05である。Pの逆数をとると、39094となるから、この事象は約3万9千回に1回しか起こらない事象である。

$$p(\text{one error}) = p_0 G \dots \text{where}$$

$$G = n_1 R_1 + n_0 R_0 + n_2 R_2$$

$$R_1 = (1 - p_1) / p_1, \quad R_0 = (1 - p_0) / p_0,$$

$$R_2 = (1 - p_2) / p_2$$

2つエラーがあり13試合中11試合を正しく予想する確率は、勝ちの予想を2つ間違えた場合、引き分けの予想を2つ間違えた場合、負けの予想を2つ間違えた場合、勝ちの予想と引き分けの予想をそれぞれ1つ間違えた場合、引き分けの予想と負けの予想をそれぞれ1つ間違えた場合、引き分けの予想と負けの予想をそれぞれ1つ間違えた場合、合計6つの場合の確率の和である。Stefani (文献 [5]) の表現に従えば、式は以下ようになる。これまで与えた数字を前提にすると、P (two error) は、0.00029695である。Pの逆数をとると3368となるから、この事象は約3300回に1回しか起こらない事象である。

$$p(\text{two errors}) = p_0 H \dots \text{where}$$

$$H = n_1(n_1 - 1) / 2 R_1^2$$

$$+ (n_0(n_0 - 1) / 2) R_0^2$$

$$+ (n_2(n_2 - 1) / 2) R_2^2$$

$$+ n_1 n_0 R_1 R_0 + n_1 n_2 R_1 R_2$$

$$+ n_0 n_2 R_0 R_2$$

以上の計算結果をまとめたものを表1に示す。

(5)

全13試合を正しく予想する確率Pから全13試合の等確率 (equivalent accuracy) を求めると、0.346であるから、過去のデータを使

用してランダムに予想した場合、各試合を正しく予想する確率は平均的に3割4分6厘でしかない。この結果は、Stefani (文献 [5]) のヨーロッパのデータを使用してもほとんど変わらない。むしろ面白いことに、過去のデータも使用しないで、常に13試合すべてホームチームが勝つとしたときの確率は、過去のデータを利用してランダムに購入するよりも等確率がよいという計算結果が得られる。Jリーグでこの方式を採用した場合、等確率は0.425であり、1等、2等、3等のうちどれかがあたる確率は0.0021、すなわち475週に1回 (約10年に1回) になる (表1参照)。この確率はDOUBLEやTRIPLEを併用することによって、上げることができる。Stefani (文献 [6]) の結果によれば、毎週3試合にTRIPLEを使用すると1等の確率は0.000798となり、毎週買い続けて30年に1回の確率で1等が当たる可能性が出てくる。しかし、DOUBLEやTRIPLEを購入すればコストは当然かかるから、30年に1回の確率で1等があたるとしても、経済的にペイするとは限らない。

(6)

以上のことから、マクロな事後確率だけでくじを購入しても、ほとんど当選を期待できるレベルにはならない。くじを買って、少しわくわくしながら試合結果を見る楽しみを得る程度である。確率を上げるためには、マクロな予想ではなく、各試合ごとの試合結果を、対戦するチームごとのパフォーマンスを勘案して統計的に予測する方法を検討する必要がある。その方法には2つある、1つは試合の対戦チームごとの得点数を予想するアプローチである。もう1つは、得点数ではなく、チームの勝ちにつながる属性データを使用してダイレクトに結果を予測する方法である。この

表1 確率計算結果

	期待値E ₀	確 率			発生までの試行回数*		
		P(no error)	P(one error)	P(two error)	P(no error)	P(one error)	P(two error)
2004年度マクロデータを用いて予測した場合	4.583	1.00942E-06	2.55792E-05	2.96951E-04	990672	39094	3368
100%アウェイ勝利と予測した場合	5.525	1.47590E-05	2.59584E-04	2.10721E-03	67755	3852	475

* toto の場合は週の単位

2つの方法は、すでにヨーロッパの統計学や計量経済学の専門家によって研究が先行している。本研究もその方向へ今後向かうつもりである。

参考文献

- [1] Dixon M.J. and Coles S.C. : Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, 1997
- [2] Maher M.J. : Modelling association football scores. *Statistica Neerlandica* 36, 1982
- [3] Moroney M.J. : *Facts from figures*, 3rd edn. Penguin : London, 1956
- [4] Reep C., Pllard R. and Benjamin B. : Skill and chance in ball games, *Journal of the Royal Statistical Society Series A* 131, 1971
- [5] Stefani, R.T. : Observed Betting Tendencies and Suggested Betting Strategies for European Football Pools. *The Statistician* 32, 1983
- [6] 鷺崎早雄『サッカーの得点と統計的研究 (その1、文献研究)』、環境と経営 第10巻 第2号
- [7] 鷺崎早雄『サッカーの得点と統計的研究 (その2、J1リーグデータによる計算実験)』、環境と経営 第11巻第1号