

# 研究ノート

## 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)

鷲崎 早雄

### はじめに

先行研究によれば、サッカーの得点分布はポアソン分布あるいは負の2項分布を用いるとうまく近似できることが報告されている。このことは前回の研究ノートに詳細を記した<sup>1)</sup>。本稿の目的は、これら先行研究の統計的方法を、日本のJリーグのデータを使用して計算実験してみようということである。計算実験にあたって、ポアソン分布と負の2項分布では、ポアソン分布が数値処理面での取り扱いに優れていること、クローチャーが行ったイギリスのプレミアリーグの結果はポアソン分布を使用していることから、本稿ではポアソン分布で実験を行うこととした。

### 1 使用データ

2004年J1リーグ前期、後期のデータを使用する。2004年J1リーグのデータは、日本サッカー協会公式ホームページに掲載されているので、その中から「Jリーグデータ2004」ダウンロードして利用した。

2004年の全240試合について、ホームチーム名、アウェイチーム名、ホームチーム得点、アウェイチーム得点の各項目をデータベースに入力した。そのデータから、ホームゲーム全試合の平均得点、アウェイゲーム全試合の平均得点を求めた。また、チームごとのホームゲーム平均得点、アウェイゲーム平均得点も同時に求めておく。

### 2 チーム別成績

表1はファーストステージ、セカンドス

テージごとに発表されているチーム別成績を合計し、勝点の順位で並べ替えたものである<sup>2)</sup>。2004年の年間結果は、第1位浦和、第2位横浜FM、第3位G大阪である。

全240試合のホームチーム平均得点は1.679、アウェイチームの平均得点は1.421であった。2002/2003のプレミアリーグと比較すると、プレミアリーグではホームチーム平均得点が1.500、アウェイチーム平均得点が1.132であった<sup>3)</sup>。このことから、ホームの得点のアウェイよりも大きいという経験則は、英国の場合と同様に日本のデータでも見ることができる<sup>4)</sup>。

特に、上位5位の鹿島まではホームゲーム平均得点が高いという傾向が顕著である。次の市原はどちらかと言えば、ホームとアウェイが拮抗している。その次の名古屋は、わずかながらアウェイの平均得点が高い。その他にはC大阪、柏もアウェイの方が大きい。チームごとのホーム平均得点とアウェイ平均得点について、上位を占めるチームはホームゲームでしっかり得点するチームであること、ホームとアウェイの得点が拮抗しているか、アウェイの方が良く得点をするようなチームは、どちらかと言えば中位から下位に来るチームであること、などの特徴があることに気がつく。プレミアリーグのデータでは、アウェイの方が大きかったチームは1チームだけであった。また、J1のデータと同

<sup>2)</sup> 磐田、鹿島、市原の勝点は48点で同点である。この3チームの並びに特に意味はない。

<sup>3)</sup> 鷲崎[1]

<sup>4)</sup> プレミアリーグのホームとアウェイの差は、J1のそれよりも大きい。この差はたまたまなのか、構造的なものなのかは別途検討が必要である。

<sup>1)</sup> 鷲崎[1]参照

表1 J1リーグチーム別、ホーム・アウェイ別成績

	ホーム	アウェイ	ホームの 相手得点	アウェイの 相手得点	ホームの 勝利率	アウェイの 勝利率	勝点前期	勝点後期	勝点合計
浦和	2.467	2.200	1.200	1.400	1.121	1.571	25	37	62
横浜FM	1.733	1.400	0.933	1.067	1.238	1.313	36	23	59
G大阪	2.733	1.867	1.667	1.533	1.464	1.217	24	27	51
磐田	2.133	1.467	1.600	1.333	1.455	1.100	34	14	48
鹿島	1.533	1.200	0.867	1.200	1.278	1.000	24	24	48
市原	1.867	1.800	1.200	1.800	1.037	1.000	20	28	48
名古屋	1.600	1.667	1.467	1.467	0.960	1.136	20	24	44
F東京	1.467	1.200	1.200	1.533	1.222	0.783	23	18	41
東京V	1.467	1.400	1.000	2.000	1.048	0.700	19	20	39
新潟	1.467	1.600	1.867	1.933	0.917	0.828	14	23	37
神戸	1.867	1.333	1.667	2.000	1.400	0.667	15	21	36
広島	1.467	1.067	1.067	1.667	1.375	0.640	15	16	31
大分	1.533	0.800	1.800	1.867	1.917	0.429	17	13	30
清水	1.467	1.067	1.667	1.933	1.375	0.552	16	13	29
C大阪	1.333	1.467	1.800	2.467	0.909	0.595	10	16	26
柏	0.733	1.200	1.733	1.667	0.611	0.720	12	13	25

様にそのチームは下位のチームである<sup>5)</sup>。

チームの力を表す指標として、ゲームにおける自得点と相手得点の比を考える。ホームゲームにおいては、ホームである自得点と相手のアウェイチームの得点の比である。アウェイゲームにおいては、アウェイである自得点と相手のホームチームの得点の比である。この比は、その試合の自分と相手の得点の相対的比率であるから、この比率が大きいということはある種の「勝つ力」を表すものと考えられる。そこで、この指標をここでは「平均勝利率」と仮称することにする。ホームでは、名古屋、新潟、C大阪、柏を除き、平均勝利率が1より大きい。しかしながら、アウェイでは上位8チームの「平均勝利率」が1より大きく、それ以下のチームの「平均勝利率」は1より小さいという具合にきれいに並んでいる。勝点の説明変数として後にこ

<sup>5)</sup> プレミアリーグとJ1リーグと同様な結果を得ているので、アウェイがホームよりも得点が多いようなチームは、勝点であまり良い成績ではないという事柄は構造的なものであると予想される。さらに多くのリーグのデータを使用して分析してみる必要がある。

の「平均勝利率」をとりあげる。

### 3 得点分布

全240試合のホームチームの得点分布、アウェイチームの得点分布は表2(a)および図1(a)のとおりであった。表2は、全240試合のホームチームの得点分布、アウェイチームの得点分布を、0点、1点、2点、3点、「4点とそれ以上」の5グループに分類して頻度を記したものである。最頻値はホーム、アウェイいずれの場合も1点である。分布の形は1点から点数の大きいほうへ裾が流れるポアソン分布の特徴を示していることがわかる。

ポアソン分布の確率分布式<sup>6)</sup>において、平均値( $\lambda$ )にホームゲームの平均得点を用いて期待値を計算すると、全240試合のうち、ホームの場合に得点が0点であった試合の期待値は44.8回と計算される。この数字は、実際の2004年度における43回にきわめて近い。得点が1点の試合の期待値は、75.2試合で、実際は79試合であるから、これもきわめて期待値

<sup>6)</sup>  $\Pr(x, y) = (e^{-\lambda_1} \lambda_1^x / x!) \times (e^{-\lambda_2} \lambda_2^y / y!)$

と実現値が近いと言ってよいであろう。さらに得点が2点の場合は期待値63.1に対して実現値64、得点が3点の場合は期待値35.3に対して実現値34、得点が4点以上の場合は期待値21.6に対して20であり、いずれの得点分類においても、統計的な期待値と実現値はきわめて近い値となった。アウェイの場合はホームほどには期待値と実現値が接近しているとは言えず、乖離が大きくなっている。もっとも大きく乖離している箇所は、2点のケースであり、実現値が期待値をかなり上回っている。

ここでJ1のホームとアウェイの得点分布を、プレミアリーグの2002/2003のケースと比較してみよう。ただし、プレミアリーグは全部で380試合であるので、これを240試合ペースに変換してある。表2(b)がプレミアリーグのケースである。また、J1リーグ、プレミアリーグの分布のグラフを図1(a)、図1(b)に示してある。総じて、J1リーグにおいても、プレミアリーグと同様ポアソン分布によくフィットしていると言えるであろう。平均得点の値そのものは年ごとに変動するから、分

表2(a) J1リーグ得点分布

得点	ホーム		アウェイ	
	observed	expected	observed	expected
0	43	44.8	51	58.0
1	79	75.2	78	82.4
2	64	63.1	70	58.5
3	34	35.3	24	27.7
4 or more	20	21.6	17	13.5
total	240	240.0	240	240.0

(\*)4 or more の平均回数は、4.7回にしてある。

表2(b) プレミアリーグ得点分布

得点	ホーム		アウェイ	
	observed	expected	observed	expected
0	51	53.6	80	77.4
1	76	80.3	88	87.6
2	72	60.2	47	49.6
3	27	30.1	18	18.7
4 or more	13	15.6	8	6.7
total	240	240	240	240

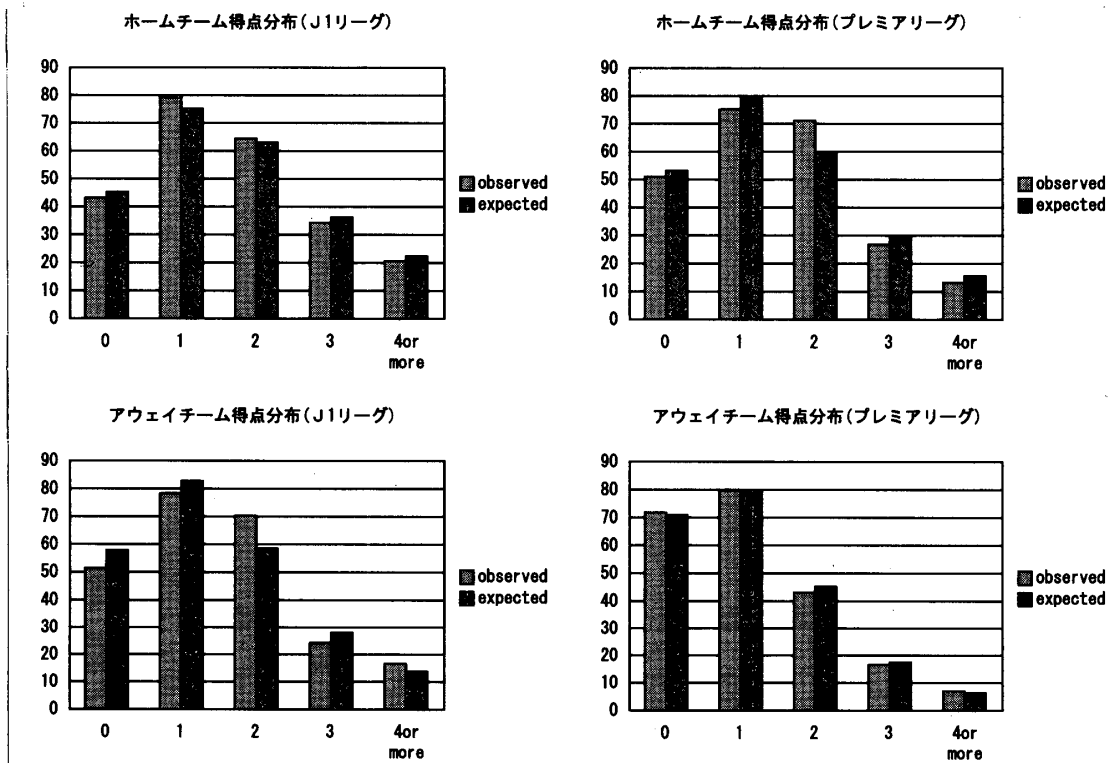


図1(a) J1リーグ度数分布(期待値vs実績)

図1(b) プレミアリーグ度数分布(期待値vs実績)

布の形について1年分のデータだけでJ1とプレミアリーグの特徴の差を示すことは難しい。いま、図1のグラフからとりあえずこの2つの分布を比較すると、J1リーグの方が総じてホームゲームにおいても、アウェイゲームにおいても得点が大きくなる傾向にあるようだ。逆に、プレミアリーグの方は、小得点で競った試合が多いという傾向にあるようだ。

#### 4 組み合わせ得点分布

対戦ごとのホームチームの得点とアウェイチームの得点を「対」にした組み合わせの分布は、両者の得点が独立であることを前提として、2次元のポアソン分布に従うと考えられる。クローチャーによれば、プレミアリーグのデータはこの仮定をよく満足するものである<sup>7)</sup>。Jリーグの場合にはどうであろうか？

表3(a)は、2004年J1リーグの全240試合について、実際の得点の組み合わせを調べて2次元のヒストグラムにしたものである。表を見ると、最頻は2-1で32ゲーム、その次は1-2で29ゲーム、3番目が1-1で25ゲームである。また、0-0のゲームは15ゲームあり、同点のゲーム全体は63ゲームあった<sup>8)</sup>。

表3(b)は、実際のデータに対して2次元のポアソン分布を仮定した場合に、各得点の期待値を計算したものである。この計算結果から幾つかのことが言える。1つは、表3(a)の実際のデータと比較して、2次元のポアソン分布がよくフィットしているという事実である。プレミアリーグにおける仮説と同様、日本のJ1リーグにおいてもそのことが確かめられた。これはほぼサッカーという競技の構造であるといつて良いであろう。同点ゲームの期待値は、57.3であるが、これは実際の同点ゲームよりも6ゲーム程度多い<sup>9)</sup>。プレミアリーグの場合は、期待値に対して実際の同点ゲームは10%程度少なかった。この限りに

おいては、J1リーグとプレミアリーグは、同点に関して逆のパフォーマンスになっている<sup>10)</sup>。

表3(c)は、組み合わせの期待値の前提となるポアソン分布の確率である。J1リーグの場合もっとも確率が大きいのは1-1のケースで10.7%である。ついで、2-1のケースで9%である。ここで、計算では最頻値が1-1であるのに、実際のゲームで2-1になるのは、同点になりそうなゲームでは、双方が勝利を狙って、いろいろな作戦を交え、普通以上に努力をするからではないかと考えられる。

以上のように、組み合わせの場合の得点予測は、確率的にもっとも多い組み合わせでも10%程度であるから、予測を当てることはなかなか難しい。しかし、ポアソン分布を用いて計算をすれば、その程度の確率では予測ができるのだといえよう。

プレミアリーグとの比較は大変面白い結果となった。ポアソン分布のフィット性という観点からは、プレミアリーグもJ1リーグもほぼ同様の結果である。イタリアや南米など、その他のリーグと比較しても、多分同様の結果が得られるであろう。このことはすなわち、サッカーという競技の特質によるものであろう。ただし、同じサッカーでも、競技力が未熟である場合、競技レベルがあまりに異なるチームが混在するリーグの場合などでは、ポアソン分布の仮定を満たさないことが予想される。競技力がハイレベルであり、リーグに参加するチームの競技力が拮抗している場合という条件が必要かもしれない。

同点ゲーム数の期待値と実際のデータのように、J1リーグとプレミアリーグにおいて、結果が逆になるものがある。J1リーグが、何らかのバイアスで期待値以上に同点試合が多くなっているようにも思えるが、このような構造的な結論を導くためにはまだデータが不足しているので、さらに多数の年次をJ1リーグとプレミアリーグで分析してみる必要がある。

<sup>7)</sup> 鷲崎[1]

<sup>8)</sup> 4点以上の分類には同点以外も含むので同点試合の数からは除いた。

<sup>9)</sup> ここでも4点以上の分類は除いて分析している。

<sup>10)</sup> この特徴が構造的なものかどうかは、さらに年次を増やして検討する必要がある。

表3(a) 対戦スコアの組み合わせ分布 (実績)

	0	1	2	3	4 or more	Total
0	15	10	12	3	3	43
1	14	25	29	8	3	79
2	9	32	18	4	1	64
3	11	9	8	5	1	34
4 or more	2	2	3	4	9	20
Total	51	78	70	24	17	240

表3(b) 対戦スコア組み合わせ分布 (期待値)

	0	1	2	3	4 or more	Total
0	10.8	15.4	10.9	5.2	2.5	44.8
1	18.2	25.8	18.3	8.7	4.2	75.2
2	15.2	21.7	15.4	7.3	3.6	63.1
3	8.5	12.1	8.6	4.1	2.0	35.3
4 or more	5.2	7.4	5.3	2.5	1.2	21.6
Total	58.0	82.4	58.5	27.7	13.5	240.1

表3(c) 対戦スコア組み合わせの確率

	0	1	2	3	4 or more	Total
0	0.0450	0.0640	0.0455	0.0215	0.0105	0.1866
1	0.0756	0.1075	0.0764	0.0362	0.0176	0.3133
2	0.0635	0.0902	0.0641	0.0304	0.0148	0.2630
3	0.0355	0.0505	0.0359	0.0170	0.0083	0.1472
4 or more	0.0218	0.0309	0.0220	0.0104	0.0051	0.0902
Total	0.2415	0.3432	0.2438	0.1155	0.0563	1.0003

## 5 勝点の回帰分析

勝点は、勝った場合が3点、引き分けた場合が1点、負けた場合は0点であり、この合計点の大きさがリーグの順位が決定する。同点の場合はそのほかの技術要素を加味して順位を決めるが、今はそのことは考えずに同点は同点のままとしておく。クローチャーは勝点を被説明変数とし、ホームにおける平均勝点とアウェイにおける平均勝点を説明変数として重回帰分析を行い、良好な結果を得た。

表4は2004年のJ1リーグデータに対してクローチャーのモデルを用いて重回帰分析を

した結果である。結果を線形式で表現すると下記のようなになる。

$$\text{勝点} = -0.412 + 8.716 * \text{ホーム勝点} + 34.538 * \text{アウェイ勝点}$$

ここで注目されることは、ホームでの勝点係数が8.72、アウェイでの勝点係数が34.54で、アウェイでの勝点係数のほうが、勝点に対する寄与度が大きいということである。アウェイの寄与度が大きいということは、プレミアリーグのデータからも同様な結果となっている。すなわち、クローチャーの結果によれば、ホームでの勝点係数は11.4であるが、アウェイでの勝点係数は17.8であり、

ホームと比較して約50%以上も大きな数字であった。J1リーグの結果はR<sup>2</sup>およびそれぞれのt値は表4に示すとおりであり、統計的な有意水準として十分な値を得ている。

勝点と勝利力の関係は、J1においてもプレミアリーグにおいても、ホームでの勝利力より以上にアウェイでの勝利力が大きいことがわかった。この結果は、日本のJリーグでもイギリスのプレミアリーグでも、ほぼ同じ傾向で変わらないから、何らかのサッカー競技特有の構造的なものがあるのかもしれない。さらにいろいろなデータで分析してみた。

ホームの得点と、アウェイの得点はこのように平均で見るとなかなか複雑である。ホームの平均得点のアウェイの平均得点を上回る傾向にあるのは、サッカー競技の構造的な性質と見られる。アウェイでは強く得点力もあるが、ホームでは弱いというチームは全体の順位では下の方になっている。リーグ戦できちんと上位に入ってくるためにはホームでしっかりと得点をする事ができるチームである。ホームは弱いアウェイは強いというチームは、上位には入ってこない確率が高い。さらに、本研究の分析からは、リーグ戦で勝点を稼いで上位に上がるためにはもう1つ条件があることがわかった。それは、アウェイでの勝利力が必要であるということだ。アウェイでは、ホームほどには得点ができず、

平均得点は少ない。しかし、アウェイにおいてきちんと相手よりも得点があげられることが勝点を左右しているのである。

## 6 サントリーチャンピオンシップの結果とポアソン分布

ここまでの分析はJ1リーグ全体の平均データを用いたものであった。このノートの最後に、個々のチームごとのデータを用いて、ある特定の対戦についてポアソン分布で勝敗の確率を予測してみることにする。データとして2004年のサントリーチャンピオンシップにおける浦和と横浜の対戦を使用し、その期待値を計算して実際の結果と比較してみる。第1戦は横浜がホームで浦和がアウェイであるから、横浜のホーム得点率1.733、浦和のアウェイ得点率2.2を使って2次元のポアソン分布から期待確率を計算する。結果は表5(a)である。

計算した期待値から、この試合で浦和が勝つ確率は51.8%と予測できる。同じく横浜が勝つ確率は27.8%、同点となる確率は21%と予測できる。浦和が横浜よりも倍程度の確率で有利であると予測されるが、実際の結果は期待値とは反対に横浜が勝つという結果になった。第2試合は、浦和が勝つ確率が61%、横浜が勝つ確率19%、同点は21%の確率であった。圧倒的に浦和が優勢な状況の中で実際に浦和が勝った。

第2試合では、期待値が最も大きいケースは浦和と横浜が2-1のケースである(表5(b))。2番目は3-1である。実際の結果1-0は期待値では10番目であった。期待値と結果が一致するのはなかなか大変である。このチャンピオンシップの結果では、実際の得点は期待値で計算した得点よりも少ない方にシフトしていたようである。普段のリーグ戦よりは締まった接戦になったといえるのだろう。

表4 勝点の回帰結果

回帰統計			
重相関 R	0.956525		
重決定 R <sup>2</sup>	0.914941		
補正 R <sup>2</sup>	0.901855		
標準誤差	3.536751		
観測数	16		
	係数	標準誤差	t
切片	-0.41224	4.897846	-0.08417
H勝利力	8.716133	3.042921	2.864397
A勝利力	34.53786	2.933835	11.77225

表5(a) 横浜ホーム、浦和アウェイの確率表

	0	1	2	3	4 or more	合 計
0	0.0196	0.0431	0.0474	0.0348	0.0389	0.1837
1	0.0339	0.0747	0.0821	0.0602	0.0674	0.3183
2	0.0294	0.0647	0.0712	0.0522	0.0584	0.2758
3	0.0170	0.0374	0.0411	0.0301	0.0337	0.1593
4 or more	0.0074	0.0162	0.0178	0.0131	0.0146	0.0690
合 計	0.1073	0.2360	0.2596	0.1904	0.2129	1.0062

表5(b) 横浜アウェイ、浦和ホームの確率表

	0	1	2	3	4 or more	合 計
0	0.0209	0.0293	0.0205	0.0096	0.0033	0.0836
1	0.0516	0.0723	0.0506	0.0236	0.0083	0.2063
2	0.0637	0.0891	0.0624	0.0291	0.0102	0.2545
3	0.0524	0.0733	0.0513	0.0239	0.0084	0.2093
4 or more	0.0608	0.0851	0.0595	0.0278	0.0097	0.2429
合 計	0.2493	0.3490	0.2443	0.1140	0.0399	0.9966

## まとめと今後の課題

本稿ではクローチャーの結果がJ1リーグのデータにも有効であるのかどうかを中心に計算実験の結果を整理した。その結果、クローチャーが得たプレミアリーグの結果はJ1リーグでも観察されることがわかった。今後の研究課題として、これら両リーグに共通に見られた結果が、サッカー競技が本質的に持っている構造的な事柄なのかどうか、もし構造的な事柄だといえるなら、それはどのような原因で構造的なのかを、より深く検討する必要がある。また、プレミアリーグの結果とJ1リーグの結果が逆になった事柄もある。これらは、たまたま使用したデータの年度でそうっただけなのか、それともJ1とプレミアリーグの構造的な差として逆になっているのかのどちらかである。本稿の範囲だけからはどちらとも言えないので、今後、データ数を増やして検討していく価値がありそうである。

## 参考文献

- [1] 鷲崎早雄「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、Vol. 10、No2、2004
- [2] Croucher, J.S., *Using Statistics to Predict Scores in English Premier League Soccer*, in Butenko, S. and Gil-Lafuente, J. and Pardalos, P.(eds), *Economics, Management, and Optimization in Sports*, Springer, 2004
- [3] 東京大学教養学部統計学研究室編「統計学入門」東京大学出版会、1991
- [4] 縄田和満「Excelによる統計入門」朝倉書店、1996