# 研究ノート

## サッカーの得点と統計的研究
## （その１、文献研究）

鷲 崎 早 雄

本研究ノートは、スポーツ分野に対する経済分析手法や経営科学手法の応用を研究していく過程で、サッカーの得点と統計的研究という点に興味をもち、クローチャの文献[1)]をノートしたものである[2)]。もともとサッカーの得点問題は、サッカーの技術的な問題意識から出たというわけではなく、サッカーくじの予測やサーカーくじビジネスのビジネスモデルなどの研究と結びついて発展してきている。したがって、この問題はスポーツ経営論の分野の１つの問題だと考える。今後、この問題をさらに掘り下げ、日本におけるＪリーグのデータや日本のサッカーくじのビジネスモデル分析に研究を進めていくことが当面の目標である。

### Ⅰ　サッカーとポアソン分布

スポーツデータを経済分析手法やＯＲ[3)]など経営科学手法によってモデル化しようという試みは第二次世界大戦後から始まっている。これは戦後におけるこれらの手法の発達と、実際の経営現場への適用事例の拡大とが関係しているものと考えられる。特にＯＲは、もともと戦争中における合理的な戦略・戦術の採用を目的としたものであるから、勝つ事を目的とし、そのための戦略・戦術を素早く決めなければならないスポーツとは、研究の背景として類似したところがある。

サッカーは他のスポーツと異なり、得点がなかなか入らない競技である。なおかつ野球やアメリカンフットボールのように、一度停止した場面から、何かの戦術を選択して次のプレーを行なうのではない。サッカーは全体として45分間の流れの中で、両チーム合わせて22人の選手がボールを蹴り続け、何かのきっかけで得点が入ったり入らなかったりする競技である。このような事象の起こり方を、現実の世の中のその他の事例で探してみると、例えば大きな不動産の契約成立のような事例と似ている。あるいは、交通事故の発生のような事例とも似ている。

不動産の契約成立の場合、不動産の商談自体は各所で無数に行なわれているが、高価な買い物であるから契約成立までこぎつける確率は非常に小さい。たまたま、ある物件と、適当な不動産屋と顧客との関係がマッチしたときに、契約成立というケースが起こる。交通事故の発生も同じように考えられる。車は無数に走っているが、たまたま車の外部環境とドライバー、場合によっては「車」というハード自身との関係において、交通事故が発生する。このような事象の発生の仕方は、確率モデルでは良く知られているように、試行回数がきわめて大きくて、発生確率が小さい（すなわち稀少である）場合に成立する「ポアソンの小数の法則」にしたがっている。

ポアソン分布は、「ポアソンの小数の法則」を２項分布に適用して、試行回数を無限大、発生確率を限りなくゼロに近くしたときの分布である。不動産の契約成立や交通事故発生の事象にはポアソン分布がよくフィット（適合）する。歴史上で最初にポアソン分布を現実の事例にあてはめた研究として、ボルトキーヴィッチの「プロシア陸軍において馬に

---

1) 文献[1]
2) したがって、本稿に含まれるデータセット、モデル、計算結果は原著からの引用であり、著者のオリジナルなものではない。
3) OR : Operations Research

蹴られて死んだ兵士数」がある[4]。馬に蹴られた兵士の数とサッカーの得点では、馬に蹴られるのは全くの偶然だろうし、サッカーの得点は監督・選手の真摯な戦術とプレーの結果であるから、事象を発生させる行為の質としては全く異なるものである。しかし、「得点がある」という事象だけをとらえてフィットする分布を考えるとすれば、サッカーボールを蹴るのも、馬が人を蹴るのも同じようにポアソン分布のような分布がまず思い浮かべられる。

## II　ポラードの結果[5]

ここでは、サッカーの得点に関するポラードの結果について整理しておく。サッカーの得点分布に関する研究は、Moroney (1951)、ポラード(1977)、ポラード(1986)、Norman (1998) などがある。これらの研究が示唆するところでは、サッカーは、試合の展開によって得点率が変化するという現象が見られるのが普通である。例えば、負けているチームは、試合のある時点からは防御を犠牲にして、相手に得点されるリスクを増加させてでも、自チームの得点率を高める戦術をとる。このような現象があるために、ポラードは、サッカーの得点分布は「負の2項分布(Negative Binomial Distribution)」にもっとも適合するということを示した。

表1はポラードが負の2項分布を当てはめた結果である。データは1967-68シーズンの英国第一部フットボールリーグ (English First Division Football League) の924試合の得点を用いている。負の2項分布の確率変数は、パラメータとして$k$と$p$を持つ下のような式(1)で与えられる。パラメータは、サンプル平均$m$と分散$s^2$を使用して推計される。

$$\Pr(\text{得点}=r) = {}_{k+r-1}C_{k-1}\,p^k(1-p)^r \qquad r=0,1,2,3.\cdots\cdots \qquad \cdots\cdots(1)$$
$$\text{但し、}k>0,\ 0<p<1$$

1967-68シーズンのデータでは、$m=1.51$、$s^2=1.75$であった。

表1　サッカーゴール数の実績値と計算期待値の対比

| ゴール数 | 実績値 | 期待値 |
|---|---|---|
| 0 | 225 | 226.6 |
| 1 | 293 | 296.4 |
| 2 | 224 | 213.9 |
| 3 | 114 | 112.6 |
| 4 | 41 | 48.3 |
| 5 | 15 | 17.9 |
| 6 | 9 | 5.9 |
| 7+ | 3 | 2.5 |
| 合計 | 924 | 924.1 |

(注) 計算は負の2項分布の期待値を求めたもの
(出所) Croucher[1]　p. 45

表1を見ると「負の2項分布」が極めてよくフィットしている。得られたP値は0.57であるから、統計的検定の信頼性においても高い値となっている。ポラード (1977) の研究以外にも、モデルの適合性を高めるために、いろいろな研究が行なわれている (Reep and Benjamin 1968[6]、Hill 1974[7]、Maher 1982[8]、Crouchere 1984[9])。これらの研究では、得点に影響を与えるいろいろな環境パラメータを取り上げている。特に、ホームグランド・アドバンテージの問題はよく研究されている。たとえば、ウィニング・マージンのような概念を導入してホームとアウェイのパフォーマンスの違いをモデルに取り込むことなどが研究されている。

4) 文献[2] p.115

5) Pollard, et al *Sport and the negative binomial distribution*. In S.P. Ladany and R.E. Machel (eds) *Optimal Strategies in Sports*. NewYork : North Holland, 188-195, 1977

6) Reep, C. and P. Benjamin. Skill and chance inassociation football,

7) Hill, I.D. Association football and statistical inference, *Applied Statistics*, 23 : 1974

8) Maher, M.J. Modelling association football scores. *Statistica Neerlandica*, 36, 1982

9) Croucher, J.S. The effect of changing competition points in English Football League. *Teaching Statistics*, 2, 1984

## III　クローチャの結果

ここでは、ホームゲームとアウェイゲームにおけるパフォーマンスの違いに焦点を当てたクローチャの研究について言及する。使用されたデータセットは、2002/03シーズンに英国プレミアリーグで行われた380試合のデータである。

クローチャは、この380試合の得点について、各チームごとの得点をホームの場合とアウェイの場合の2つに分け、ポアソン分布を用いて分布を推定し、適合性の検定を行なった。ポアソン分布は $\lambda$ をパラメータとして、得点を $r$ とすると、下の式(2)で表すことができる。

$$\Pr(\text{得点}) = r) = e^{-\lambda}\,\lambda^{r}/r! \qquad r=0,1,2,3,\cdots \qquad \cdots\cdots(2)$$

380試合の全得点のうち、ホームチームの得点合計は570点であり、1試合平均にすると $\lambda_1=1.500$点、またアウェイチームの得点合計は430点で、1試合平均 $\lambda_2=1.132$点であった。$\lambda_1$および $\lambda_2$は、各チームごとにも求められる。クローチャが使用した英国プレミアリーグの各チームの $\lambda_1$と $\lambda_2$は表2に示すとおりである。また、表の勝点は、勝った場合に3点、引き分けは1点、負けた場合には0点として計算されたものである。

表2　ホーム平均得点とアウェイ平均得点

| チーム | $\lambda_1$ | $\lambda_2$ | 勝点 |
|---|---|---|---|
| Manchester United | 2.21 | 1.68 | 83 |
| Arsenal | 2.47 | 2.00 | 78 |
| Newcastle | 1.89 | 1.42 | 69 |
| Chelsea | 2.16 | 1.42 | 67 |
| Liverpool | 1.58 | 1.63 | 64 |
| Blackburn | 1.26 | 1.47 | 60 |
| Everton | 1.47 | 1.05 | 59 |
| Southampton | 1.32 | 0.95 | 52 |
| Manchester City | 1.47 | 1.00 | 51 |
| Tottenham | 1.58 | 1.11 | 50 |
| Middlesbrough | 1.89 | 0.63 | 49 |
| Charlton | 1.37 | 1.00 | 49 |
| Birmingham | 1.32 | 0.84 | 48 |
| Fulham | 1.37 | 0.79 | 48 |
| Leeds | 1.32 | 1.74 | 47 |
| Aston Villa | 1.32 | 0.89 | 45 |
| Bolton | 1.42 | 0.74 | 44 |
| West Ham | 1.11 | 1.11 | 42 |
| West Bromwich | 0.89 | 0.63 | 26 |
| Sunderland | 0.58 | 0.53 | 19 |

（出所） Croucher[1] p. 47

統計的に見ると、$\lambda_1$と $\lambda_2$の間には相関があり、表2の場合相関係数は0.644（P値＝0.002）である。ホームゲーム・アドバンテージにより、大多数のチームはホームゲームの方が $\lambda$ の値が大きいが、Leeds United だけはその関係が逆転している。すなわち、アウェイの方が得点が大きい。しかし、これは全体で20チームのうち、わずか1チームだけである。次に、$\lambda_1$と勝点の間の相関も著しく強く、相関係数は0.890（P値＝0.001）である。ホームゲームにおける得点率の高いチームが、勝点も大きいという傾向にある。相関係数が0.890であるから、この傾向は著しく大きい。アウェイの場合の $\lambda_2$と勝点の間にも相関がある。相関係数は0.796（P値＜0.001）であった。もともと、$\lambda_1$と $\lambda_2$の間には相関があるから、$\lambda_2$と勝点の間にも相関があるということである。

380試合の全ての得点について、実際の得点の分布とポアソン分布から得られた得点の分布とを比較すると（表3）、ポアソン分布が非常によく適合していることが見てとれる。このように、クローチャの研究では実際の英国のプレミアリーグのデータがホームとアウェイのそれぞれにおいてポアソン分布によく適合する、という結果が導かれた。

表3　ホームチームの分布とアウェイチームの分布比較

| 得点 | ホーム（$\lambda$=1.500） | | アウェイ（$\lambda$=1.132） | |
|---|---|---|---|---|
| | 実績値 | 期待値 | 実績値 | 期待値 |
| 0 | 81 | 84.8 | 126 | 122.6 |
| 1 | 121 | 127.2 | 139 | 138.7 |
| 2 | 114 | 95.4 | 74 | 78.5 |
| 3 | 43 | 47.7 | 29 | 29.5 |
| 4以上 | 21 | 24.9 | 12 | 10.7 |
| 合計 | 380 | 380 | 380 | 380 |

（出所） Croucher[1] p. 48

試合における対戦相手ごとの得点分布は、得られた $\lambda_1$、$\lambda_2$ を用いてホームチームの分布とアウェイチームの分布の積として仮定する。本来は、2変数のポアソン分布を仮定して交互作用までを考慮するべきであろうが、計算上の複雑性が増すので、クローチャの研究ではホームの分布とアウェイの分布は独立であると仮定している。対戦相手ごとの確率分布は下の式(3)で求まる。

$$\begin{aligned} \Pr(x,y) &= (e^{-\lambda_1}\lambda_1{}^x/x!)\times(e^{-\lambda_2}\lambda_2/y!) \\ &= \lambda_1{}^x\lambda_2{}^y e^{-(\lambda+\lambda_2)}/x!y! \qquad \cdots\cdots(3) \end{aligned}$$

但し、$x$ はホームチームの得点、
$y$ はアウェイチームの得点
$\lambda_1$ はホームチームの1試合平均得点
$\lambda_2$ はアウェイチームの1試合平均得点

対戦相手ごと計算結果の合計は表4のようになった。表4はホームチームの得点を縦軸、アウェイチームの得点を横軸にとり、縦軸と横軸の交点に、そのスコアの実績試合数が書かれている。(　) 内は、式(3)の分布から計算された試合数の期待値である。たとえば、0－0の試合は実績で20試合あったが、計算された期待値は27.3試合であった。一方、ホームチームが1－0で勝つケースは、計算では41試合であるが、実際には50試合あった。0－0と1－0の組み合わせを除くと、他の組み合わせは計算と実績とはそれほど乖離していない。これは、どうやら0－0で試合が展開しているときに、ホームチームが頑張って1点入れるケースが、確率的に多いのではないかという可能性を示唆する。しかし、詳しく分析するためには、ホームとアウェイの2変数の交互作用を入れたモデルが必要になるため、今後の課題である。

このように、チームの得点分布を当てはめるだけではなく、試合の組み合わせによる得点分布を当てはめても、統計的に良くフィットすることが、クローチャの研究ではっきりしてきた。本稿に引用した結果は、各表ともにリーグの試合全体の合計に関する結果である。これを、各チームの対戦を考慮した予測にまで持っていければ、サッカーの試合において、もっとも勝つ可能性の高い組み合わせ

表4　ホームとアウェイの得点の組み合わせの実績と期待値

| ホーム＼アウェイ | 0 | 1 | 2 | 3 | 4以上 | 合計 |
|---|---|---|---|---|---|---|
| 0 | 20<br>(27.3) | 32<br>(30.9) | 15<br>(17.5) | 9<br>(6.6) | 5<br>(2.4) | 81<br>(84.8) |
| 1 | 50<br>(41.0) | 42<br>(46.2) | 19<br>(26.3) | 7<br>(9.9) | 3<br>(3.6) | 121<br>(127.2) |
| 2 | 35<br>(30.8) | 40<br>(34.8) | 26<br>(19.7) | 10<br>(7.4) | 3<br>(2.7) | 114<br>(95.4) |
| 3 | 17<br>(15.4) | 16<br>(17.4) | 9<br>(9.8) | 0<br>(3.7) | 1<br>(1.3) | 43<br>(47.6) |
| 4以上 | 4<br>(8.1) | 9<br>(9.1) | 5<br>(5.2) | 3<br>(1.9) | 0<br>(0.7) | 21<br>(25.0) |
| 合計 | 126<br>(122.6) | 139<br>(138.7) | 74<br>(78.5) | 29<br>(29.5) | 12<br>(10.7) | 380<br>(380) |

(出所) Croucher[1] p. 50

を求める問題や、リーグ終盤になって下位に居るチームが、下のリーグと入れ替えになる確率を求める問題や、サッカーくじの新しい商品を検討する問題など、いろいろな問題に統計的手法を応用することができるようになると期待される。

## 参考文献


[１] 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 Optimaization in Sports*, Springer, 2004

[２] 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