話題

統計物理の眼で見るサッカー

成塚拓真 〈中央大学理工学部 narizuka@phys.chuo-u.ac.jp〉
山崎義弘 〈早稲田大学理工学術院 yoshy@waseda.jp〉

1. はじめに

1.1 スポーツと物理

「一九六四年のはなやかなオリンピック東京大会もはや一昨年のこととなり，…」から始まるロゲルギストのエッセイ「陸上競技」．およそ50年前に発表されたこのエッセイでは，エネルギーと制御に基づく考察に対し，投擲（投げる）・跳躍（跳ぶ）・競走（走る）が題材となった．[1] スポーツはしばしば物理の話題になる．例えば，衝突時の運動量保存則を説明するためにカーリングが用いられ，ボールの回転と軌道との関係が野球や卓球で論じられるように．

近年，「社会物理学」が注目を集めている．*1 これは，社会現象を多数の人間で構成された系が示す集団的振る舞いと捉え，統計物理学の視点で特徴づけようとする試みである．そうであるならば，スポーツも社会物理学の対象とみなすことができるのではないだろうか．例えば，試合中の得点変動をブラウン曲線とみなした解析により，クリケットやサッカーでは時系列の分散に異常拡散が観られるとの報告がある．[3,4] また，バスケットボールにおける得点差の時系列について，得点差の符号が変わらない時間（つまり，一方のチームがリードを続ける時間）の分布が逆正弦則に従うとの報告もある．[5] 本稿では，対戦型集団スポーツの代表例ともいえるサッカーに焦点を当て，統計物理の視点から考えてみる．

1.2 対戦型集団スポーツの階層性

対戦型集団スポーツは，有限のフィールドで特定の時間内に，構成要素である選手が二つの部分系に分かれて競合している系である．選手はルールに従って，環境の影響や対戦相手の予期せぬ動きなどの不確定な要因を伴いながら，ある目的を実現しようと運動する．そして，世界中様々な場所で異なる構成要素によって，こうした系が多数実現していると考えれば，対戦型集団スポーツから試合やチームの詳細に依らない再現性ある共通の構造が抽出できるかどうか試したくなるのも無理はないだろう．

対戦型集団スポーツにおける試合の流れ（時間的推移）については，階層的な見方ができる．先ず，それぞれの時刻での各選手の運動に着目する場合を最も短い時間スケール（ミクロ）とする．次に，一試合全体，あるいは試合を単位とする時間スケールでの特徴に着目する場合を最も長い時間スケール（マクロ）とする．マクロの階層では，各選手の速さ分布といったミクロな情報の統計的性質や試合の勝敗が議論の的になる．また，ミクロとマクロとの中間にある時間スケール（メゾ）では，試合で攻守が切り換わるまでの一連の流れや，得点やシュートにつながるプレーなど，試合の様々な局面をダイナミクスとして切り出すことができる．

このような階層的な見方は，選手の空間的配置に対しても同様に導入することができよう．つまり，個々の選手に着目すれば「1体問題」，2選手間のパスや相互作用などに着目すれば「2体問題」，フォーメーションなどチーム全体の特徴を議論する場合は「多体問題」となる．

2. サッカーを題材に

2.1 トラッキングデータ

スポーツの試合中に選手やボールの動きを追跡（トラッキング）する技術が進歩し，いまや，詳細な位置データ（トラッキングデータ）が容易に手に入る状況となった．実際，野球においてはPITCHf/xという装置が導入され，ピッチャーの投げたボールの軌道や球速を詳細に測定することができる．[6] そして，投手によるボールの握りだけでなく，投げた後の軌道でストレートやカーブといった球種を判別する方法が考案されている．サッカーにおいても，日本では，2015年に日本プロサッカーリーグ（Jリーグ）でトラッキングシステムが導入され，0.04秒ごとに全選手（および，全審判）の位置データが取得されている．本稿で紹介する我々の解析では，2015年と2016年の2年間にJリーグで行われた11試合分のトラッキングデータ，ならびに，ボールタッチデータ（パスによりボールの受け渡しが行われたときのボールの座標，受け渡しを行った2選手）を使用した．これらのデータは，データスタジアム株式会社から貸与されたものである．*2

2.2 サッカーの階層性

ここで，サッカーの階層性について考えてみる．先ず，試合の時間的推移について．ミクロな時間スケールでの試合の状態は，各時刻における選手・ボールの位置，ボールの保持状態によって表される．ボールの保持状態は，どちらか一方のチームが保持しているか，どちらのチームも保持していないかのいずれかである．いま，保持チームが変わらない連続した時間帯を「シーン（scene）」と呼ぶことに

*1 社会物理学“physique sociale”という言葉は，ケトレー（A. Quételet）が1835年に出版した統計学の先駆的論文で使われている．[2]

*2 スポーツデータ解析コンペティション（http://estat.sci.kagoshima-u.ac.jp/sports/index.htm）に参加し，参加期間中に無償で貸与されたデータ．

©2017 日本物理学会

表1　サッカーの階層性とカテゴリー.

	選手個人［1体］	2選手の関係［2体］	チーム全体［多体］
各時刻 ［ミクロ］ **特徴量の定義**	**●トラッキングデータ** **各選手の位置** **●ボールタッチデータ** **ボールの位置・保持状態**	速度相関 相対距離	重心・慣性半径・秩序変数 フォーメーション フロントライン・最終ライン
シーン シークエンス ［メゾ］ **ダイナミクス**	選手の軌道	**追跡ダイナミクス** フォーメーション内での連動	**重心と慣性半径の相関** **フォーメーションの遷移**
試合全体 ［マクロ］ **統計的性質**	走行距離・スプリント回数 滞在領域のヒートマップ 速度分布・加速度分布	**向きの相関に対する，秩序変数，距離依存性** 接近領域	**フォーメーションの分類** パス回しネットワーク

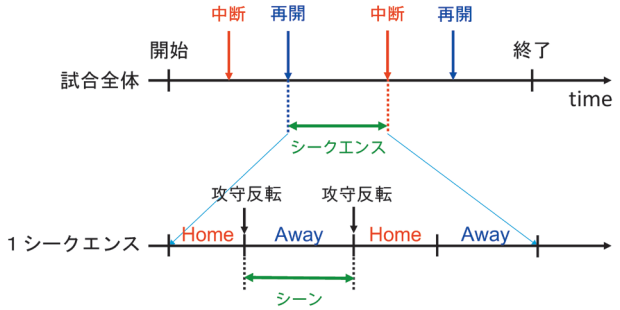

図1　サッカーにおける時間の階層性.

する．すると，試合が再開し中断するまでの時間帯は複数のシーンで構成される．この一連の時間帯を「シークエンス（sequence）」と呼ぶことにしよう．シーンやシークエンスはメゾスケールに相当し，マクロスケールである試合全体は複数のシークエンスで構成される（図1参照）．

このような階層性を考えれば，時間として各時刻・シーン（およびシークエンス）・試合全体，チームの構成要素として1体・2体・多体というそれぞれ3つの階層が存在し，統計物理の眼でサッカーを見る上で，表1のような3×3＝9つのカテゴリーを設定することができる．この表に基づいて，ミクロの階層で1体・2体・多体系としての特徴量を定義する．そして，メゾの階層ではこれら特徴量のダイナミクスに，マクロの階層ではこれらの統計的性質に着目する．

3.　サッカーの統計的性質・ダイナミクス *3

3.1　チームの重心と慣性半径の関係 [7]

サッカーでは，各チームがフォーメーションという一つのまとまりを成して，フィールド内を運動している．従って，大まかにはフォーメーションを構成する全選手の平均座標（重心）と標準偏差（慣性半径・広がり）によってチームは特徴付けられるだろう．図2(a)は，ホームチーム（赤）がアウェイチーム（青）のゴールに向かって攻撃しているシーンである．全選手の位置は▲で示され，矢印は選手それぞれの速度を表している．また，各チームの平均位置（★）の座標X_H，X_A，および，★を中心にして標準偏差σ_H，σ_Aを半径とした円が示されている．（なお，二つのゴールを結ぶ直線に平行で，ホームチームが攻撃する向きを正

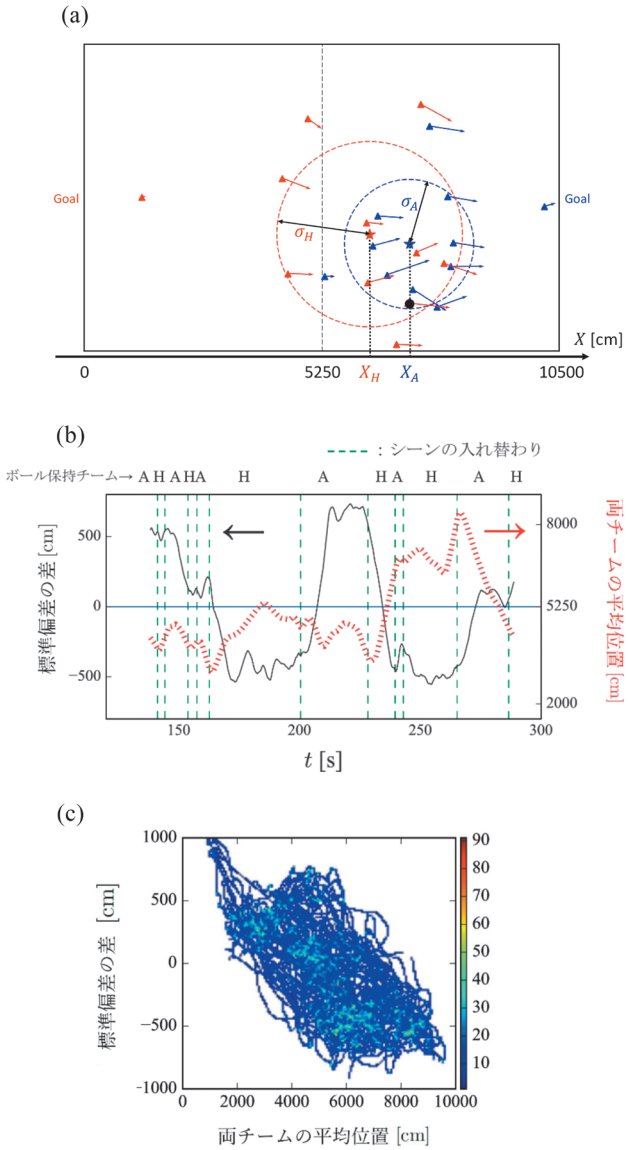

図2　各チームの重心と標準偏差の関係．(a) ホームチーム（赤）がアウェイチーム（青）のゴールに向かって攻撃しているシーン．(b) 両チームの平均位置のX座標（赤），両チームの標準偏差の差（黒）の1シークエンス分の時系列．グラフの上部にあるA・Hは，各シーンのボール保持チーム．(c) 試合全体のデータから求めた2次元頻度分布．

*3 本稿で紹介する解析結果は，ゴールキーパーの位置を考慮していない．

©2017 日本物理学会

としたX座標軸を設定した.）各時刻においてこれらの量を求め，X_{H}とX_{A}の中点（つまり，両チームの平均位置），および，標準偏差の差$\sigma_{\mathrm{A}}-\sigma_{\mathrm{H}}$について，1シークエンス分の時系列を表したものが図2(b)である．

図2(a)，(b)から，相手チームのゴールに近いフィールドで，ボールを保持して攻撃しているチームの方が標準偏差は大きくなる傾向がみられる．実際，各時刻で両チームの平均位置と標準偏差の差を求め，一試合全体にわたってこれらの2次元頻度分布を求めた結果が図2(c)で，この傾向は分布としても確認できる．このような傾向がみられる要因として，攻撃チームは得点する（つまり，ボールをゴールに入れる）可能性を増やすために，選手を分散してフィールドに配置しようとする一方，守備チームは失点を防ぐために，選手をゴールの周りに固めて配置しようとすることが考えられる．この要因はサッカーのルールを反映したもので，試合やチームの詳細に依らないといえよう．

3.2　相手選手との向き[8)]

2体問題として，選手どうしがどのくらい速度の向きをそろえているかに着目する．先ず，向きのそろい具合を表す指標として，各時刻tにおいて次の2つの量を導入しよう．(1) ある選手に対してk番目に近い相手選手との速度がなす角$\theta(k,t)$，(2) フィールド内にいる全選手の速度について，それらの単位ベクトルの平均の絶対値（いわゆる，秩序変数に相当する量）$\phi(t)$．そして，$\phi\leq\phi(t)<\phi+\Delta\phi$を満たす時刻の$\theta(k,t)$を用いて円周分散$V_\theta(k,\phi)$を求める．図3(a)は，ある試合のホームチームに対して実際に求めた$V_\theta(k,\phi)$である．この図では，ϕが大きくなるにつれて分散が小さくなる傾向がみられる．ϕが大きくなるのは全選手が向きをそろえて運動している状況なので，分散が小さくなるのは当然ともいえる．ここでは，全てのϕにおいて，最近接選手（$k=1$）のみが他の近接選手（$k\geq2$）よりも分散が小さくなっていることに注目したい．このことは，特に最近接選手と強く向きをそろえる傾向にあることを示唆している．

図3(b)は，最近接選手とのなす角度分布のϕ依存性$f_\phi(\theta)$を示している．この図では，ϕが小さくなるにつれて角度分布の裾野が広く（つまり，なす角度のばらつきが大きく）なっている．我々は，これらの分布が円周上の統計分布として知られるvon Mises分布$f_{\mathrm{vM}}(\theta)$とwrapped Cauchy分布$f_{\mathrm{wC}}(\theta)$の混合分布で良くフィットできることを確認した．実際，これらの分布の混合割合をcとして，$f_\phi(\theta)=cf_{\mathrm{vM}}(\theta)+(1-c)f_{\mathrm{wC}}(\theta)$でデータをフィットしたグラフも図3(b)に併せて描かれている．ここで，各ϕに対してフィッティングで用いたcの値が図3(c)に示されており，ϕ〜0.7で急激に変化していることが分かる．この急激な変化は，チームとしての運動モードの変化を反映したものであると推察できる．つまり$\phi>0.7$では，ほとんど全ての選手が同じ向きに運動しており，ロングパスなどをきっかけに全選手が同じ向きにボールを追いかける状況が示唆される．一

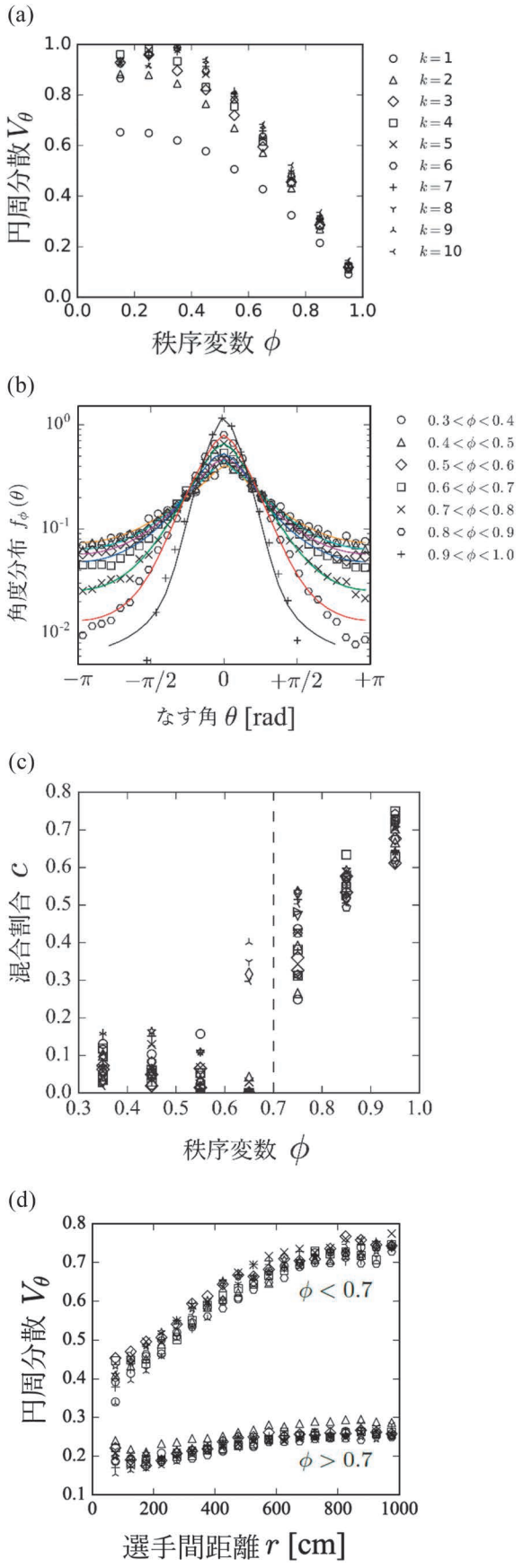

図3　両チームから一人ずつ選んだ選手の速度がなす角θについて．(a) 円周分散の秩序変数，第k近接選手依存性．最近接選手（$k=1$）との (b) 角度分布，その秩序変数依存性，(c) 混合割合cの秩序変数依存性，(d) 円周分散の選手間距離依存性．

©2017 日本物理学会

方$\phi<0.7$では，最近接選手のみが強く向きをそろえ，両チームがボールを奪い合っている状況だと考えられる．図3(c) には解析した11試合全ての結果が示されており，$\phi\sim0.7$でcが急激に変化する傾向は試合やチームの詳細に依存しない．

いま，$\phi>0.7$と$\phi<0.7$で状況を二つに分けて考えてみよう．図3(d) は，それぞれの状況における，円周分散 ($k=1$) の選手間距離依存性を示している．この図では$\phi<0.7$のとき，相手選手との間隔が5 m程度より近づくと円周分散の減少が確認できる．また我々は，相手選手が5 m程度より近づく頻度は最近接選手のみ高くなることも別の解析で確かめている．[8] 以上のことから，サッカーでは，選手に対して相手チームの最近接選手のみが頻繁に5 mより近づいて，お互いの向きをそろえていると考えられる．

さらに我々は，2選手間で行われる追跡（ランダムに向きを変えて逃げる選手を，もう一方の選手が追いかける）という行動を表現した数理モデルを構築し，図3で示した結果の再現を試みた．モデルや解析の詳細は文献8に譲るとして，このモデルには，逃げる選手がランダムに向きを変える頻度，追跡の程度，追跡を継続する時間がパラメータとして含まれている．そして，図3(b) に示された混合分布の形状や，図3(d) に見られる円周分散の減少が再現されることを数値計算によって確かめた．追跡の問題は数学の古典的な問題でもあり，[9] 最近では生物の集団系で観られる追跡に対するモデルが提案され，個体数と捕獲率の関係や揺らぎの効果などが議論されている．[10, 11]

3.3 フォーメーションとドロネーネットワーク [12, 13]

チームのフォーメーションについて，3.1では各選手の詳細な配置を考慮せず，チーム全体としてどこに存在し（重心），どのくらい広がっているか（標準偏差）という大まかな状態に着目した．ここでは，選手の配置をより詳しく取り扱うため，以下のような方法を考えよう．先ず，時刻tにおいて，チーム内の各選手を頂点としたドロネー（Delaunay）三角形分割 *4 を行い，図4(a) のように，三角形の辺をリンクとした「ドロネーネットワーク」を構築する（図中の数字は選手の背番号を表している）．このネットワークを隣接行列で表し，異なる時刻における隣接行列の違いを距離として導入する．そして，階層的クラスタリング法 *5 を用いることで，各時刻のドロネーネットワークをいくつかのクラスターに分類する．いま，この分類されたクラスターをフォーメーションと呼ぼう．図4(b) は，あるチームにこの方法を適用して得られたフォーメーションのうち，出現割合が最も高い3つの場合を示している．数字（背番号）を囲んでいる楕円は，各フォーメーション

*4 平面上に分布した点（母点）に対し，それぞれの母点との距離が最短となる領域（ボロノイ領域）を考え，隣接するボロノイ領域の母点どうしをつないでできる三角形領域による分割．

*5 要素（いまの場合，ドロネーネットワーク）どうしの類似度によって，グループ分けする手法．特に，類似度に対するしきい値により，グループ分けの程度を階層的に変化させることができる．

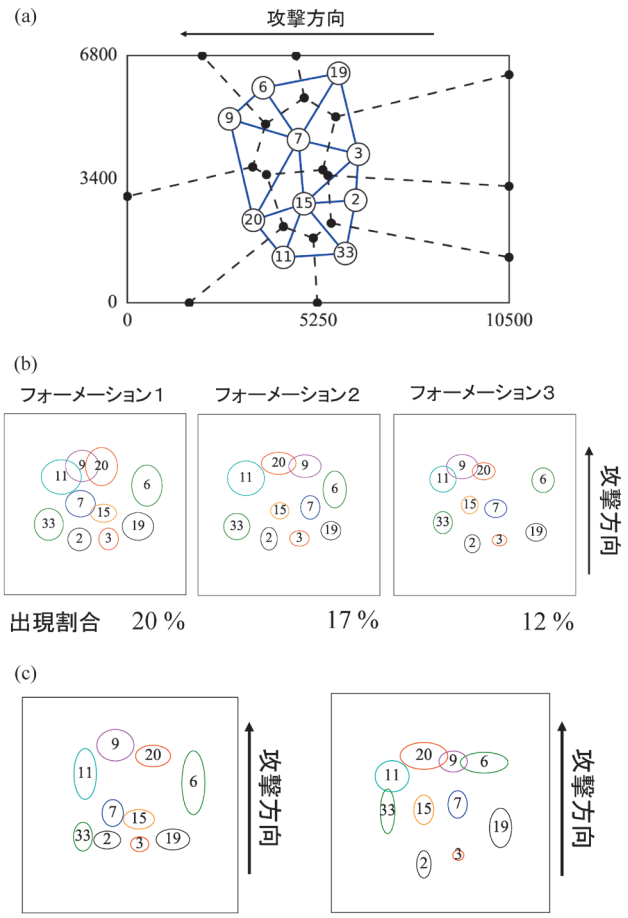

図4　ドロネーネットワークとフォーメーション．(a) ある時刻のドロネーネットワーク．点線は双対なボロノイ分割線．(b) 階層的クラスタリング法によるフォーメーションの分類．(c) 左：守備時．右：攻撃時の局面で，最頻出のフォーメーション．

に属するドロネーネットワークを用いて，重心系で各選手の位置がどの程度ばらついているかを表したものである．フォーメーション1と2では背番号7と15，および，9と20の選手がそれぞれ入れ替わっており，フォーメーション2と3では背番号9と20の選手が入れ替わっている．

ドロネーネットワーク法を用いれば，試合の局面で異なるフォーメーションを抽出することもできる．例えば，図4(c) は，あるチームの攻撃時と守備時で最も頻繁に現れたフォーメーションを表している．攻撃時にはフロントラインに選手が集まり，守備時にはバックラインに選手が集まっている様子がうかがえる．また，各時刻のドロネーネットワークはいずれかのフォーメーションに属するので，フォーメーション間の遷移確率を求めることもできる．さらに，アマとプロの差やヨーロッパ，アジアといった地域差を見出したり，過去の試合データからフォーメーションを蓄積しておけば，年ごとや相手チームごとの戦術の違いを抽出したりできるかもしれない．近い将来，次のような実況解説はあり得るだろうか…「第3シーケンス，リスタートの3.5秒後，チームPが今シーズン出現頻度5位のフォーメーションで攻撃開始．チームQは遅れて1.4秒後に頻度2位の守備フォーメーションで対抗．その1.2秒後にチー

©2017 日本物理学会

ムPは1位の攻撃フォーメーションにシフト．このシフトはやや意外（遷移確率が比較的低い）．チームQは対応が0.5秒遅れて相性の悪い（失点率の高い）守備フォーメーションに．そして，ボールを持つ選手の3 m以内に相手選手が不在，75%の確率でゴールが決まる状況となり…」．

4. さいごに

図4(a)には，ドロネーネットワークと併せて双対なボロノイ（Voronoi）分割線が点線で表されている．点線で囲まれた領域（ボロノイ領域）は各選手の守備範囲や勢力圏を表す標準的な指標「優勢領域」として良く用いられている．[14] 例えば，試合中の各チームの優勢領域の面積は，攻撃側チームの方が守備側チームよりも大きくなることが指摘されている．[15]

本稿で紹介した結果の他にも我々は，選手の位置を考慮したパスまわしのネットワークを構築し，その統計的性質を調べた．[16, 17] 実際，実データからパスまわしネットワークの次数分布を求め，試合・チームの詳細に依らない分布形状となることを見いだした．また，マルコフ連鎖モデルを構築して分布形状を再現し，さらに，分布形状を解析的に求めて複雑ネットワークにおけるフィットネスモデルとの関連を議論した．特に，ある極限で次数分布はべき分布のように振る舞うことを示した．

現時点では，不確定要因が具体的にどのようなものであり，統計に対してどのような影響を及ぼすかについて明らかになっていない．不確定要因は多様で複雑であり，明らかにするのは難しいかも知れない．しかしながら，本稿で紹介した結果は，マクロな時間スケール（試合全体での統計性）では，不確定要因を時間相関のないランダムなものと見なしてもよい（マルコフ性を仮定してよい）場合があることを示唆している．これは，シークエンス間の相関が弱い（または，ない）ことを反映しているのではないだろうか．

以上，サッカーを物理の散歩道として歩いてみた．今後，対戦型集団スポーツ特有の不変構造が見つかれば，統計学・統計物理学に根ざしたスポーツ科学が発展し，対戦についての理解が深まるかもしれない．2020年に行われる予定の東京オリンピックでは，どのようなデータが得られ，そこからどのようにすれば不変構造が抽出できるのか．統計物理の眼で見るオリンピックも面白いと思う．

本研究の貸与データは情報・システム研究機構の新領域融合研究プロジェクト『社会コミュニケーション』データ中心科学リサーチコモンズ事業『人間・社会データ』の支援を受けたものです．データを貸与して下さったデータスタジアム（株），ならびに，議論をしていただいた山本健氏・西森拓氏・松下貢氏・卯田純平氏・原嶋貴大氏に感謝の意を表します．

参考文献

1）ロゲルギスト，『第三　物理の散歩道』（岩波書店，東京，1966）．
2）豊田利幸，『物理学とは何か』（岩波書店，東京，2000）．
3）H. V. Ribeiro, S. Mukherjee, and X. H. T. Zeng, Phys. Rev. E **86**, 022102 (2012).
4）R. da Silva et al., Phys. Rev. E **88**, 022136 (2013).
5）A. Clauset, M. Kogan, and S. Redner, Phys. Rev. E **91**, 062815 (2015).
6）D. Kagan, Phys. Teacher **47**, 412 (2009).
7）成塚拓真，卯田純平，山崎義弘，統計数理研究所共同研究レポート **363**, 83 (2016).
8）T. Narizuka and Y. Yamazaki, Europhys. Lett. **116**, 68001 (2016).
9）P. Nahin., *Chases and Escapes* (Princeton Univ. Press, 2007).
10）A. Kamimura and T. Ohira, New J. Phys. **12**, 053013 (2010).
11）T. Ohira, T. Hosaka, and T. Nogawa, Artificial Life and Robotics, **20**, 257 (2015).
12）成塚拓真，卯田純平，原嶋貴大，山崎義弘，統計数理研究所共同研究レポート **380**, 79 (2017).
13）成塚拓真，山崎義弘，統計数理 **65**，出版予定（2017）—特集「スポーツ統計科学の新たな挑戦」．
14）J. Gudmundsson and M. Horton, ACM Comput. Surv. **50**, 1 (2017).
15）T. Taki and J. Hasegawa, Proc. Computer Graphics Int. 2000 (2000) p. 227.
16）T. Narizuka, K. Yamamoto, and Y. Yamazaki, Physica A **412**, 157 (2014).
17）T. Narizuka, K. Yamamoto, and Y. Yamazaki, J. Phys. Soc. Jpn. **84**, 084003 (2015).

（2017年4月30日原稿受付）

©2017 日本物理学会