

実店舗内における顧客の購買行動データの分析

Analysis of consumer's purchase behavior data in actual store

石橋健 (兵庫県立大学大学院情報科学研究科)

Ken Ishibashi (Graduate School of Information Science, University of Hyogo)

近年では、情報通信技術の発達により、実店舗内で顧客の購買に関する詳細なデータを収集することが可能になっている。例えば、スーパーマーケットではショッピングカートへ取り付けられたセンサー搭載ビーコンにより顧客の店舗内の移動経路を追跡することができる。このような実店舗内で収集したデータを用いた分析は、その分析方法が十分に確立されているとは言い難いため、さらなる研究が求められている。

本研究では、スーパーマーケットで収集した顧客動線データやアイトラッキングデータの分析に関する研究事例を用いて、データ分析の現状を紹介する。また、今後の研究の発展の可能性や分析手法における検討課題について議論する。アイトラッキングデータを用いた分析モデルとして、既存研究では Wedel ら[1]が提案したビジュアルマーケティングの理論に基づくフレームワークが検討されている。本研究でもこの理論に基づき、図1に示すフレームワークを用いた分析を検討している。まずトップダウン要因とは、人の情報処理と深くかかわる要因である。例えば、同じ商品の広告であっても見る人の購買への関心の有無によって重要性が異なる。次にボトムアップ要因とは人の視覚に刺激を与える要因である。情報処理の基盤はトップダウン要因から形成されるが、店舗内のボトムアップ要因の影響を受けて情報処理の仕方が変化する。この変化の一端が眼球運動として観測されたものであり、フレームワークではこの眼球運動が意思決定に与える効果であるダウンストリーム効果を検証することを検討している。

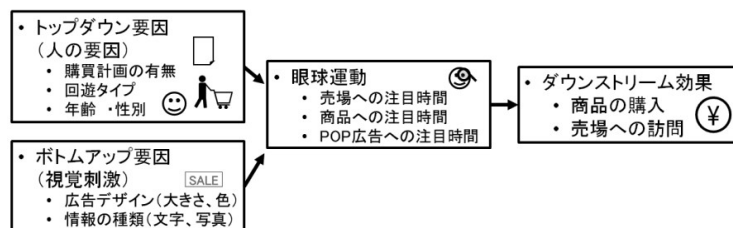


図1 分析のフレームワーク (仮)

参考文献

- [1]. Wedel, M., & Pieters, R. (2008). Eye Tracking for Visual Marketing. *Foundations and Trends in Marketing*.

サッカーのトラッキングデータを用いた 重力モデルに基づく選手能力の視覚化

土田 潤¹, 宿久 洋¹

¹ 同志社大学文化情報学部

サッカーは世界的に人気なスポーツの1つであり、多くのプロチームが存在する。プロチームの対戦において、相手チームに勝利することは大きな目的の一つである。相手チームに勝利するために、試合のプレー内容に関するデータを利活用することがあり、データ活用に関する論文も多く発表されている。例えば、Rue and Salvesen (2000) では、得失点を予測するモデルを動的一般化線形モデルを用いて構築している。これまでのデータ活用に関する論文の多くは、試合結果や試合の集計結果の分析であった。近年の情報技術の発達により、試合中の選手の位置に関するデータ（トラッキングデータ）を比較的安価かつ容易に取得できるようになった。これに伴い、トラッキングデータを用いたデータ分析が盛んにおこなわれている。例えば、神谷他 (2016) では、トラッキングデータを用いて、戦況変化の自動検出法を提案している。さらに、成塚他 (2016) では、選手間の距離や角度について、サッカー独自の特徴は何かを考察している。

トラッキングデータを用いた分析の多くは、サッカー戦術や選手の配置に関する研究が多く、選手個人に対する評価は少ない。本報告では、全体の動きを考慮しつつ、選手個人の動きを評価する指標について考える。具体的には、選手の配置が重力モデルによって表現されるとし、選手1人に対して、重力モデルの質量のパラメータを割り当てる。これを選手質量といい、選手個人の指標とする。選手質量の推定は、重力モデル式が対数線形モデル式と対応することを利用する。つまり、分割表に対する対数線形モデルの主効果を選択質量に対応させ、最尤推定を行う。選手質量は時間ごとに算出することができるため、時間ごとの選手質量の推移を視覚化することができる。時間ごとの推移を確認することで、いつ、どの選手が、全体の配置に影響を与えているかを考察することが可能となる。本報告は土田・宿久 (2017) に基づいている。

参考文献

- [1] Rue, H. and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in league, *Journal of the Royal Statistical Society: Series D*, **49**, 399–418.
- [2] 神谷哲太, 中西航, 泉裕一郎 (2016). トラッキングデータを用いたサッカーの試合における戦況変化の抽出, 「スポーツデータ解析における理論と事例に関する研究集会」, 統計数理研究所共同研究レポート, No. 363, 77–82.
- [3] 成塚拓真, 卯田純平, 山崎義弘 (2016). サッカーの対戦的特徴に現れる普遍的な統計性の探求, 「スポーツデータ解析における理論と事例に関する研究集会」, 統計数理研究所共同研究レポート, No. 363, 83–90.
- [4] 土田潤, 宿久洋 (2017). 重力モデルを用いたサッカー選手の動きの定量化, 『統計数理』, **65**, 271–286.

Social network analysis method Analysis on patent rights relations

(社会ネットワーク分析法を用いた特許権に関する社会関係の分析)

岡山大学 社会文化科学学域 特任助教 姜 佳明

Patents, in general, are a popular academic research topic, especially with economists. Early work focused on how well patents protect an innovation in terms of time to imitate or on patent output in particular industries. Patent statistics are also now commonly utilized the most reliable business performance for applied research and development activities when investigating the knowledge transfer or the technology evolution.

On the other hand, recent developments in the field of social network analysis (SNA) have resulted in software tools for visualization as well as improved analysis and interpretation of patent statistics, e.g., patent applications, patent citations or joint patent applications. Among the ways of surveying the structure of a network, centrality, structural equivalence and brokerage roles are key structural features of social networks and communicates important information about an individual's prominence or role within a given network.



Our research results focused on the firms that develop business method software, revealed that the firms positioned with a relative centrality or situated within the same structural equivalent cluster have more citations to their counterpart firms' patents. Among the

different brokerage roles, we find positive promotion to knowledge transfer when the citing and cited firms both serve the role of an itinerant as well as that of a gatekeeper/representative, while firms that act as gatekeeper/representative (alone) cite less patents from firms that do not enact this kind of a role.

Source: Jiang, Jiaming, R.K. Goel, X. Zhang, 2019, "Knowledge flows from business method software patents: influence of firms' global social networks,"

Journal of Technology Transfer, Vol.44 (4), pp.1070-1096.

生体振動データに基づく無拘束型睡眠段階推定法の検討

田島 友祐[†]
Yusuke TAJIMA[†]

[†] 滋賀大学データサイエンス・AI イノベーション研究推進センター

[†] Data Science and AI Innovation Research Promotion Center, Shiga University

睡眠は1日の三分之一を占める重要な活動であるにも関わらず、不眠症や無呼吸症候群をはじめとする睡眠障害を患っていても気が付けないことが多くある。これは、睡眠障害を知るためには脳波系を始めとする多くのセンサの装着による計測が必要であることに加え、計測されたデータを専門医師が診断する必要があるため敷居が高いことに起因する。近年は、腕時計型のウェアラブルセンサを始めとする、高精度かつ安価なデバイスが発達したことで、睡眠中の生体データを計測することが容易になっている。そこで、これらのセンサデータを用いて、睡眠の質を知ることが出来れば睡眠障害の早期発見に繋がると考えられる。

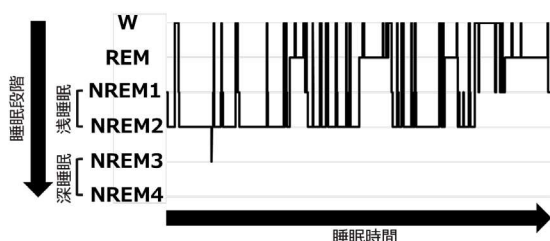


図 1. ある被験者の睡眠段階

図 1. はある被験者 1 名の睡眠段階である。睡眠段階は専門病院にて診察の際に用いられるものである。横軸は睡眠時間であり、就寝から起床までを示してある。縦軸は睡眠段階を示す。睡眠段階は眠りの深さを表す指標であり、眠りの浅いものから覚醒 (W)、レム睡眠 (REM)、ノンレム睡眠 1~4 (NREM1~4) の 6 段階ある。これから睡眠障害の有無や睡眠の質を知ることができる [1]。

本発表では、センサデータとして圧力センサによる生体振動データを用いる。圧力センサはベッドの下に敷くことで心臓の鼓動から寝返りなどの体動などの振動を計測できるセンサであり、無拘束で計測可能なセンサである。無拘束型のセンサであるため、計測者の背景に問わず、また設置も簡易であるため、計測環境も幅広いことが利点である。生体振動データには心拍や呼吸、体動などを含んだ振動データであるが、これらの値から睡眠の状態を推定することはセンサの感度

によって値が異なることになる。そこで、生体振動から体動、心拍と呼吸の 3 つの値へ周波数解析により導出し、これらの値を用いる。図 2. はある被験者の睡眠時に計測された体動、心拍と呼吸の生体情報である。

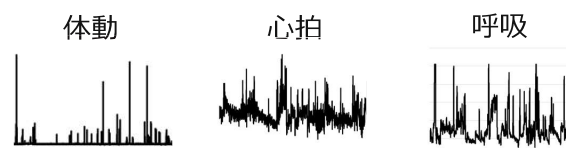


図 2. ある被験者の睡眠時に計測された生体情報

本発表で用いるデータは睡眠時の生体情報 (体動、心拍、呼吸、それぞれ 1 秒間に 1 データのサンプリングレート) と専門医師によって診断された睡眠段階 (6 段階の質的変数、30 秒間に 1 データ) である。生体情報を説明変数、睡眠段階を応答変数として推定することを検討する。具体的には、各睡眠段階の生体情報の違いを基本統計量から検討する。この時、1 秒を単位時間として捉える場合と 30 秒を単位時間として捉える 2 通りで違いがあるのかも考慮する。30 秒を単位時間として捉える場合では、30 秒間の生体情報の代表値として平均値や中央値などのどれが睡眠段階の推定に妥当かも検討する。次に、睡眠段階を 6 つのクラスに分類する問題として考え、分類手法として有名なロジスティック回帰、線形判別分析や K 最近傍法を適用し、推定精度を比較する。しかし、一般的に、就寝直後は心身の疲労を回復するために深い眠りが出やすく、朝に近づくに従って活動に向けて眠りは浅くなるなどの特徴があるように、就寝からの時間を考慮する必要があるため、時系列データとしての分析も検討する必要がある。時系列データとしても、サンプリングレートがそれぞれ存在するため、離散的な時系列データである。そこで、時系列データとして自己相関関数、偏自己相関関数、パワースペクトラムなどの統計量や定常性などから生体情報の特徴を理解するとともに推定法を検討する。

[1] 日本睡眠学会：睡眠学。初版。株式会社朝倉書店、(2009)。

グループテストにおける事後確率の近似計算法の検討

Study of Approximate Calculation Method of Posterior Probabilities in Group Test

松島 裕康[†] 田島 友祐[†] 盧 曉南^{††} 神保 雅一[‡]

[†]滋賀大学 データサイエンス・AI イノベーション研究推進センター

^{††}山梨大学 大学院総合研究部 工学域電気電子情報工学系

[‡]統計数理研究所 大学統計教員育成センター

E-mail: [†]{hiroyasu-matsushima, yusuke-tajima}@biwako.shiga-u.ac.jp,

^{††}xnlu@yamanashi.ac.jp, [‡]jimbo@ism.ac.jp

複数の検体を混ぜ合わせたグループに対して検査を行うことで、検査数を削減するグループテストと呼ばれる方法がある[1]。検査対象が n 個あるとき、それらを一つ一つ検査すると n 回の検査が必要になるが、グループテストでは、複数の検体をひとまとめにしたグループ (pool と呼ばれる) に対して検査をすることで、検査数を少なくすることができる。具体的には、図1のように、 n 個の各検体 (c_1, c_2, \dots, c_n) を m 個の各 pool (G_1, G_2, \dots, G_m) に混ぜ合わせて、pool ごとに検査が実行される。グループテストは、製品の不良品の検出や、DNA ライブラリのスクリーニング、PCR 検査などにおいて、集団の中から要請検体の検出などに応用される。グループテストでは、1つの pool に対する検査結果として陽性 (positive) か陰性 (negative) といった結果が得られるが、検査の効率を高めるために、多くの検体の中から出来るだけ少ない pool 数で目的の検体を見出すことができる実験を計画することが重要である。また、このような検査においては、実験結果に偽陽性 (false positive) や偽陰性 (false negative) などの誤りが生じることも考慮する必要があり、pool に対する検査結果から検体一つ一つの状態を推定する方法が必要となる。

本発表では、Pearl の belief propagation[2]のアルゴリズムをグループテストに適用した BNPD (Bayesian Network Positive Decoder) [3]を用いたシミュレーション実験に取り組み、実験結果からアルゴリズムの有効性について議論する。BNPD アルゴリズムの検出力について、1つのグループに混ぜ込む検体の数や用意する pool の数についてシミュレーションを行い、その検出力について比較評価する。

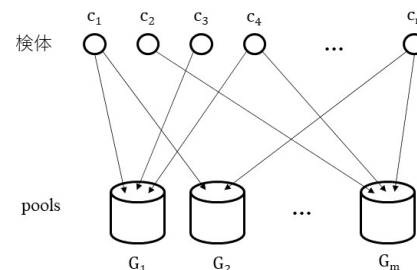


図1 Pooling design

参考文献

- [1] Aldridge, M., Johnson, O., Scarlett, J., “Group testing: An Information Theory Perspective. Found”, Trends Commu. Infor. Theory, Vol. 15, pp.196-392, 2019.
- [2] Pearl, J., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.
- [3] Ueda, H., Jimbo, M., “A Positive Detecting Code and Its Decoding Algorithm for DNA Library Screening”, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 6, No. 4, pp. 652–666, 2009.

行列正規分布の縮小推定とその適用

統計数理研究所 湯浅 良太
東京大学 久保川 達也

概要

多次元正規分布に従うデータが得られた際、標本平均を用いる事が自然である。標本平均は最尤推定量であり、一致性や不偏性など様々な良い性質を持つ。しかし、次元が3以上の時には標本平均の各成分をある1点、例えば原点の方向に縮小させることで推定精度が改善させられるという事が James and Stein (1961) により示され、Stein のパラドックスとして知られている。その後、様々な状況での縮小推定量が研究され、その1つとして Efron and Morris (1972) は複数の多次元正規分布からデータを得た際にそれぞれの平均を同時に推定する状況を考えて、これは平均行列を推定する状況となっている。この時、経験ベイズ推定量の枠組みで標本分散共分散行列の逆行列を用いた縮小推定量が提案された。この推定量は特異値を縮小するような推定量で行列としての構造を活かした推定量になっている。Yuasa and Kubokawa (2020) ではリッジ型の逆行列を用いる事で高次元の場合でも用いる事ができる推定量を提案した。ここまで挙げた推定量は分散共分散行列を既知と仮定している。Yuasa and Kubokawa (2022) では分散共分散行列が未知の場合を考え、行列平均と分散共分散行列に対する縮小推定量を提案している。

こうした縮小推定量の実際のデータへの適用については、Efron and Morris (1975) がいくつか考えており、その中には例えば野球選手のシーズン開始直後45打席の打率から、シーズン終了後の打率を知りたいという問題がある。各球団から打席数の多い選手9人の打率を考える場合には自然とデータが行列として得られる。本発表では、縮小推定について紹介するとともに、こうした行列データに対して縮小推定量の適用を行いその結果も紹介する。

参考文献

- Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika*, **59**, 335–347.
- Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.*, **70**, 311–331
- James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. 4th Berkeley Sympos. Math. Statist. Prob.*, **1**, 361–379. Univ. California Press, Berkeley
- Yuasa, R. and Kubokawa, T. (2020). Ridge-type linear shrinkage estimation of the mean matrix of a high-dimensional normal distribution. *J. Multivariate Anal.*, **178**, 1–19.
- Yuasa, R. and Kubokawa, T. (2022). Generalized Bayes estimators with closed forms for the normal mean and covariance matrices. *J. Statist. Plann. Inf.*, to appear.

統計的DEA法

趙 宇 (東京理科大学) 国友 直人 (統計数理研究所)

経営工学やオペレーションズ・リサーチなどの分野で発展しているDEA(Data Envelopment Analysis)では確率的メカニズムを十分に考慮しているとは言えない。現実の生産インプットや生産アウトプットには様々な原因から偶然性が含まれていると考えられるので、DEAに関する新しいアプローチ、統計的DEA法を提唱する。統計的DEA法は効率的フロンティアの推定に負値ノイズを含む観測データを利用する方法であり、統計的極値論を応用して数理的に正当化できる。

確率的な線形フロンティアを考え、 n 個 ($n \geq 2$)の一次元観測データ Y_1, Y_2, \dots, Y_n が $Y_i = a + bX_i + U_i$ ($i = 1, \dots, n$) にしたがうとする。ここで a, b は未知母数、 X_i は説明変数、 $U_i \leq 0$ は i.i.d. 確率変数列、分布関数を F とする。説明変数 X が存在すると、観測値 Y の最大値と想定される誤差項 U の最大値と必ずしも一致しない。母数の推定には次の方法が考えられる。簡単化のために X の値域により2分割を行い、小さい方から区間 $I_{(L)}$ と区間 $I_{(M)}$ を構成する。さらに各区間での X の平均を $X_{(L)}, X_{(M)}$ とする。各区間での Y の最大値を $Y_{(L)}, Y_{(M)}$ とし、 b を $\hat{b} = \frac{Y_{(M)} - Y_{(L)}}{X_{(M)} - X_{(L)}}$ もしくは最小二乗推定量 \hat{b}_{LS} を用いて推定する。切片係数 a は $\hat{a} = \max_{i=1, \dots, n} \{Y_i - \hat{b}X_i\}$ により推定する。すべてのデータが推定された直線の下に来る条件下で最適化するというものであり、これはより一般には「凸集合内の n 個の観測データから超平面を推定する」という統計的問題と解釈できる。データの上位値を利用することにより極値分布を利用するというアイデアは自然であるがその統計的性質は自明ではない。データ数 $n \rightarrow \infty$ のとき $\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} a \\ b \end{bmatrix}$ となる。

効率的フロンティアの非線形性を考慮すると、説明変数 X によりデータを分割して推定する方法が実際的である。ここでは分割方法の一例を示す。説明変数を順序づけて m 個の点を選び $X_{(1)} < X_{(2)} < \dots < X_{(m)}$ とする。変数 X の領域を有界として m 個の区間 I_j ($j = 1, \dots, m; m < n$) を構成し、区間 $X_{(j)} \in I_j$ とする(区間 I_j を重なりを許すが中心は単調増加系列にする)。ここで区間 I_j における未知母数 (a_j, b_j) の推定を同様に行うことにすると、「各区間内のデータ数 n_j ($j = 1, \dots, m$) を大きく、かつ区間数 m を大きくとる」ことができる。各区間内のデータ数 $n_j \rightarrow \infty$ のとき $\begin{bmatrix} \hat{a}_j \\ \hat{b}_j \end{bmatrix} \xrightarrow{p} \begin{bmatrix} a_j \\ b_j \end{bmatrix}$ ($j = 1, \dots, m$) となる。

統計的DEA法は一般化が可能である。例えば説明変数 X が複数ある場合には $\mathbf{X}_i = (X_{1i}, \dots, X_{Ki})$ に対して順序づけにより区間を構成することが考えられる。また n 個 ($n \geq 2$)の一次元観測データ Y_1, Y_2, \dots, Y_n を変数 X の順序づけた区間 $X_{(1)} < X_{(2)} < \dots < X_{(m)}$ に対して $Y_i = f(X_i, \theta) + U_i$ ($i \in I_j; j = 1, \dots, m$) という非線形モデルも考えられる。ここで θ は未知母数、 X_i は説明変数、 $U_i \leq 0$ は i.i.d. 確率変数列、分布関数を F とする。この時には経済学的な考察から f は凹関数であるので、条件 $b_1 > \dots > b_m \geq 0$ が必要となる。この関数の推定問題を区分線形関数 $f(X, \theta) = a_j + b_j X$ ($X \in I_j; j = 1, \dots, m$) の推定問題とすることが考えられる。各区間 I_j におけるデータ数 n_j を多くとれば $m \rightarrow \infty$ のとき任意の凹関数 f の推定が可能となり、新たな統計的問題として興味深い。

参考文献

- [1] Aigner, D., K. Lovell, & P. Schmidt (1977), "Formulation and Estimation of Stochastic Production Models," *Journal of Econometrics*, 6, 21-37.
- [2] Cooper, W. W., Seiford, L. M., & Tone, K. (2007), *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*, Springer.
- [3] Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling Extremal Events*, Springer.

大気レーダーの信号処理における統計学の利用

情報・システム研究機構 国立極地研究所 橋本 大志

電波を用いて大気の運動を遠隔測定する装置は大気レーダーと呼ばれ、地表付近から高度数百キロメートルに至るまでの広い高度領域を高い時間・空間分解能で計測できる唯一の測器である。その性質から気象学、特に大気力学において重要であり、地球全体の物質循環を捉えるために赤道域や両極域を中心として設置されている。

晴天時の一見透明な大気によっても電波が散乱され、適切な信号処理を施せば上空の風を測ることができる、という大気レーダーの原理は1950年台に提唱され、1970年台には図1に示す乱流風速測定の基本的な手法が確立された。大気レーダーの電波の散乱源は大気の渦によって生じる屈折率の揺らぎであり、渦がランダムな運動を行うため速度の分布はガウス型となる。一方、渦は風によって流されるので、渦の速度の分布の中心は平均的な乱流風速を与える。この推定方法も含め、大気レーダーの信号処理においては様々な統計的考え方が用いられている。また最近では、複数の受信機・周波数を用いた信号処理を行うことにより、別の方位より到来する外来不要波の抑圧や、波長・ビーム幅で決まる分解能を超えるレーダーイメージの生成など、より高度な統計的信号処理手法の適用も進んでいる。この講演では大気レーダーの原理と、その信号処理を支える統計学について紹介する。

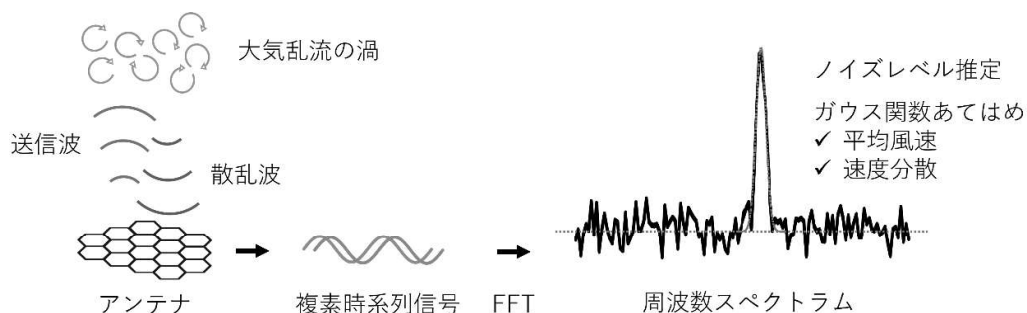


図1 大気レーダーの基本原理

「寿命」と「健康寿命」の側面から捉える線虫の可能性

中野義雄¹, 森内将貴², 福島友太朗², Suico Mary Ann², 甲斐広文², 首藤剛²

1. 東京理科大学薬学部, 2. 熊本大学薬学部

【目的】 寿命と健康寿命の間には差があることが知られており、この差はヒトだけでなく、実験動物のマウスや線虫など多くの動物で確認されている。マウスを対象にした寿命研究は2-3年必要であり、多くの検討を実施するのは困難である。そこで寿命が30日と短く、ヒトと同様に老化による筋肉萎縮が起きる実験動物の「線虫」に着目した。線虫の分子や遺伝子の機能はヒトと類似していることから、線虫を用いた寿命研究は汎用されている。しかし、線虫を用いた従来の健康寿命・寿命測定法には3つの問題点がある; 1. 侵襲性が高く、線虫の寿命に影響を与える、2. 寿命と健康寿命を同時に測定できず、対応のあるデータが得られない、3. 実験操作が煩雑で網羅的な化合物探索が困難である。本研究では、これらの問題点を克服する新規測定系の確立と、健康寿命関連因子の探索を目的に、研究を実施した。

【手法】 実験細胞の観察に用いる IncuCyte® を線虫へ活用し、自動で画像を取得する最適な飼育条件を探索した。経時的に取得した線虫の連続画像を基に、線虫の動きを捉え、「活動状態」、「不活動状態（頭と尾のみが動く状態）」、「死亡状態」の3つに区分した。各状態を基に、「寿命（生存から死亡までの期間）」、「不健康期間（不活動状態から死亡までの期間）」、「健康期間（生存から不活動状態までの期間）」と定義した。また、健康期間と不健康期間を指標に、k-means cluster で4つのクラスターに分類した。

【結果・考察】 遺伝子・飼育環境が同等の野生型線虫を12時間ごとに撮像し、寿命・不健康期間・健康期間を測定した結果、個体差があることを見出した。また、生存時から遺伝子機能を促進、もしくは抑制させた線虫（遺伝子変異線虫）と野生型線虫の寿命などをKaplan-Meier Method で解析した結果、有意に延伸もしくは短縮していることを見出した。それら遺伝子変異線虫と野生型線虫の健康期間と不健康期間に対し、k-means clustering (k=4) を実施した結果、遺伝子変異線虫ごとに特徴がることを見出した。これらの結果から、遺伝子機能は健康期間などに影響を与えることが示唆された。次に、化合物の評価も可能であるか検討するために、寿命延伸作用があるメトホルミンを線虫に投与した。その結果、メトホルミンは寿命だけでなく、健康期間も延伸させると共に、不健康期間を短縮させていることを見出した。本研究で確立した測定系・解析法はスクリーニングにも応用できることから、寿命・健康寿命に関連する薬物および遺伝子を見出すツールとなることが期待される。

がん分子標的薬のベイズ流用量探索法

東京医科歯科大学 佐藤宏征

抗がん剤の臨床開発では、開発早期に実施される第I相試験において用量（以下、推奨用量）が決定される。当該推奨用量は、その後実施される新薬のスクリーニングを目的とした第II相試験及び有効性の検証を目的とした第III相臨床試験で用いられることから、適切な推奨用量を決定することは臨床開発を成功させるために極めて重要である。また、第I相試験の対象は、通常、標準的な治療によって症状緩和や延命などの効果が得られなかった進行期のがん患者であるため、試験に組み入れられた患者に可能な限り推奨用量またはその付近の用量で治療することが望まれる。これらの要件を満たすために、第I相試験ではベイズ流統計手法などを利用した用量探索法が用いて推奨用量が決定される。

従来開発の主流であった細胞障害性抗がん剤は、毒性と有効性が概ね比例関係にあり、用量を増やすと、両方が単調に増加することが仮定される。また、その毒性の強さから、毒性のみに着目した用量探索法を用いて、安全性の観点から許容される用量の上限である最大耐用量 (Maximum Tolerated Dose; MTD) を推奨用量として決定する。一方、個別化医療の推進などを受けて、近年では開発が盛んに行われている分子標的薬は、毒性と有効性が比例関係にあるとは限らず、用量を増やしても、ある用量以上は有効性が増加しないまたは減弱する、非単調な用量-有効性関係を示す可能性がある。そのため、分子標的薬の第I相試験では、毒性に加えて、非単調な用量-有効性関係を考慮して推奨用量（生物学的最適用量; Biological Optimal Dose）を決定する必要がある。

現在までに、分子標的薬の用量-有効性関係を考慮して BOD を探索する方法が複数提案されている。例えば、Sato *et al.* (2016) は、change-point model と呼ばれるモデルを利用することにより、分子標的薬の非単調な用量-有効性関係を考慮したモデルベースのベイズ流用量探索法を提案した（以降、CP法）。シミュレーション実験において、非単調な用量-有効性関係を仮定したシナリオでは、CP法の性能は通常のプロジスティックモデルを用いる方法を上回ったが、選択するモデルの精度によって性能が変わることが明らかになった。第I相試験は患者に初めて薬が投与される段階であり、用量-有効性関係に関する情報が十分存在するとは限らないため、モデルベースの用量探索法を用いる場合、モデルの選択に不確実性に対処する必要がある。

当日の発表では、用量-有効性関係に関するモデル選択の不確実性を考慮した分子標的薬のためのベイズ流用量探索法について紹介する。

参考文献

Sato H, Hirakawa A, Hamada C. An adaptive dose-finding method using a change-point model for molecularly targeted agents in phase I trials. *Statistics in Medicine*, **35**, 4093-4109, 2016.

女性のライフコースにおける健康問題についての疫学研究

-大豆製品摂取と骨粗鬆症の関係-

群馬大学 数理データ科学教育研究センター 長井 万恵

大豆に含まれる大豆イソフラボンと女性ホルモンのエストロゲンと構造が似ており、エストロゲン様の作用を持つと言われ、植物性エストロゲンともよばれる。大豆イソフラボンによるエストロゲンの代替効果についての検証も多く行われている。本研究では、大豆摂取と骨粗鬆症の発生との関連を検討するため探索的に地域相関の検討を行った。

女性の看護職有資格者により構成された大規模コホート Japan Nurses' Health Study (JNHS) のベースライン調査参加者を対象とし、出身都道府県別(47都道府県)、年齢カテゴリ別(35歳未満、35-39歳、40-44歳、45-49歳、50-54歳、55歳以上)に、豆腐、納豆、みそ汁などの大豆製品の摂取頻度(全くとらない、週に1日程度、週に2~3日、週に4~5日、ほぼ毎日、毎日2回以上)の割合を算出し、コホート全体の年齢構成で調整した年齢調整割合についても算出を行った。『骨粗鬆症と診断されたことがある』と回答した女性の割合について、納豆を比較的摂取している地域をモデル人口とした際の標準化罹患比を算出した。

納豆の摂食習慣は、他の大豆製品の摂取習慣と比較して、地域差がみられた。骨粗鬆症の標準化罹患比では、納豆を摂取しない対象者が多い地域の方が骨粗鬆症の発生が多い可能性があり、納豆の摂取と骨粗鬆症の発症に、集団単位で関連がある可能性が考えられた。このことから、納豆の摂取が骨粗鬆症の予防の一助となりえる可能性があることを仮説として考え、前向き観察データから、個人単位での納豆の骨粗鬆症予防への影響の検討した結果についても説明する。



多目的最適化問題の幾何とその応用

The geometry of multi-objective optimization problem and its application

加葉田 雄太朗 (長崎大学情報データ科学部)
email: kabata@nagasaki-u.ac.jp

単目的最適化問題（目的関数が一つの場合の最適化問題）は、基本的には関数の最大最小問題として扱うことができ、その最適解の意味は明快である。一方で、多目的最適化問題（目的関数が複数の場合の最適化問題）では、一般には目的関数の間のトレードオフが生じるために、最適解の意味はそれほど明快ではない。例えば、図1のように2つの可微分関数 $f, g: \mathbb{R} \rightarrow \mathbb{R}$ に対しては、 f を小さくしようとすると g が大きくなってしまいうような（あるいは、 g を小さくしようとすると f が大きくなってしまいうような）トレードオフが生じる。このようなトレードオフの関係が生じる点はパレート解と呼ばれる。多目的最適化問題におけるパレート解は、1点ではなく集合として現れる。例えば、 $f, g: \mathbb{R} \rightarrow \mathbb{R}$ に関する2目的最適化問題では、パレート解集合は1次元的な集合であり、目的関数の平面上では曲線（パレートフロント）になっている。

多目的最適化問題の中で、パレート解集合を決定せよという問題は重要であるし、さらに無数の点を含むパレート解の中から最終的に選ぶべき選好解を決定するという問題も生じ、これらの問題には多くのアプローチが提案されている（例えば [2] を参照）。講演者は最近情報系や数学の研究者等とともにパレート解の幾何学的な性質について研究している（例えば [1] を参照）。本講演では特に特異点論や微分幾何学に基づいた上記の問題のアプローチについて紹介する。また、統計や情報の文脈で現れる関数の多目的最適化問題についても言及し、特異点論や微分幾何学の観点からどのような性質をパレート解が持つかを説明する。

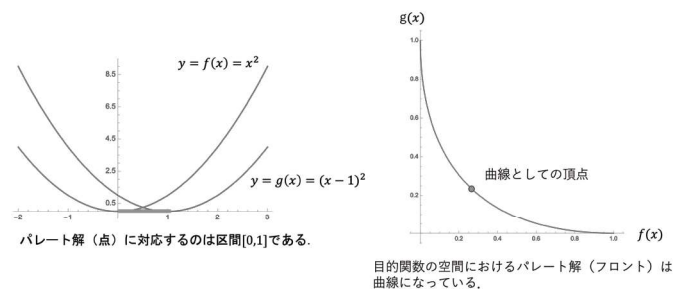


図 1: $f(x) = x^2, g(x) = (x - 1)^2$ に関する 2 目的最適化問題とパレート解

参考文献

- [1] N. Hamada, K. Hayano, S. Ichiki, Y. Kabata and H. Teramoto, Topology of Pareto sets of strongly convex problems, SIAM J. Optim., **30** (2020), 2659 – 2686.
- [2] 坂和正敏, 非線形システムの最適化, 森北出版, 1986.