

ベイズ推定と確率伝搬法による

グループテストを通じた感染者特定と有病率推定

[1] 要旨

グループテストとは、患者から採取した検体を混ぜ合わせ、混ぜ合わせた検体に対して検査を行うことで検査回数を削減する方法である。本研究では、グループテストの結果から感染者を特定する問題をベイズ推定として定式化し、確率伝搬法と呼ばれるアルゴリズムにより実装している。Expectation-Maximization(EM)法やブートストラップ法と組み合わせることで、感染者の割合という未知パラメータの推定や、信頼性評価に基づく推定性能の改善が可能であることが示されている。

[2] 本文

N 人の患者集団内に含まれる、 K 人の感染者を特定するという問題を考える。一人一人を独立に検査すれば、検査回数は N 回となる。グループテストは、患者数よりも少ない回数の検査から感染者を特定する方法であり、1943 年に米国の統計学者・経済学者である Robert Dorfman により考案された。グループテストにおいては、患者の検体を混ぜ合わせた M 個のプールを作り、このプールに対して検査を実行する。プール数は患者数より小さいため、検査数は患者数よりも小さくなる。ただし、プールに対する検査結果から感染者を特定するための推定プロセスが必要となる(図 1)。このプロセスは、圧縮されたデータの復元と解釈することができるため、情報理論的研究を通して、感染者数が十分小さい時にグループテストが有効であることが示されてきた。

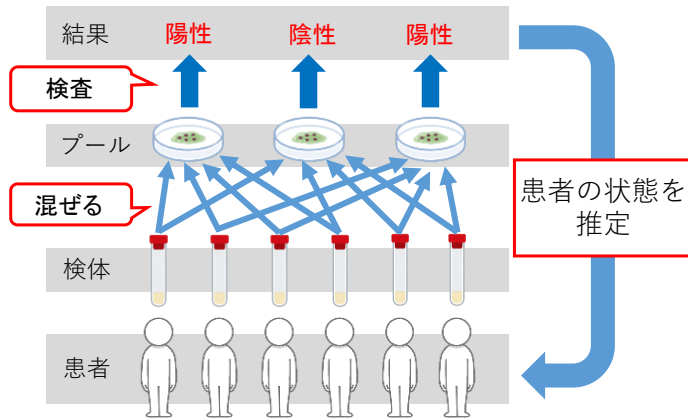
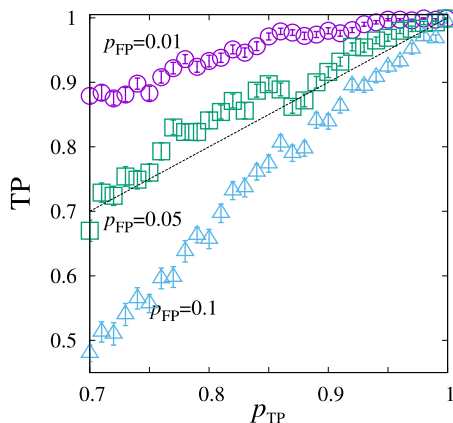


図 1. 本研究で議論するグループテストの模式図。6 人の状態を 3 回の検査から推定する例。

感染者数が十分小さいという条件下の推定は、統計学ではスパース推定として知られる。スパース推定とは、高次元変数が多いのゼロ成分を含むと仮定し、変数の次元よりも少ない数の条件式から変数の値を推定する枠組みの総称である。感染者を 1、非感染者を 0 とすれば、グループテストにおける感染者推定もスパース推定に含まれる。これまで統計物理学、特にスピングラス理論はスパース推定の発展に大きく貢献してきた。統計物理学に基づく推定アルゴリズムの動力学的記述は、数々の推定問題に定量的理解を与えている。このようなアプローチをグループテストに導入したのが本研究である。

統計数理研究所 数理・推論研究系の研究グループは、グループテストをベイズ推定により定式化し、統計物理学と関係づけられる確率伝搬法により感染者を特定する方法の研究を行った。その結果、疫学的に重要なパラメータである患者集団内の感染者の割合(有病率)を推定しつつ、一人ずつ独立に検査する場合よりも感染者を正しく特定できることを示した(図 2)。さらに、推定結果の信

頼性を評価することで精度が改善されることを示した。この成果は、日本物理学会が発行する英文誌 *Journal of the Physical Society of Japan (JPSJ)* の 2020 年 8 月号に掲載された。



[図 2] 患者数 $N=1000$, 感染者数 $K=10$ において、 $M=500$ 回のグループテストによって正しく陽性と判断された感染者の割合(TP)と、一人ずつ検査した場合に正しく陽性と判断される感染者の割合(p_{TP})の比較。非感染者を誤って陽性と判断する確率(p_{FP})を 3 種類の値に固定した例を示す。斜線は $TP=p_{TP}$ を表しており、 $TP>p_{TP}$ であれば M 回のグループテストを通して、 N 回の独立な検査よりも高い精度で患者を特定できることを意味する。(原論文の図 6 より引用)

グループテストに対する確率伝搬法は先行研究でも議論されてきたが([1]~[3]参照)、本研究は確率伝搬法により近似評価された対数尤度を用いた EM 法や、階層ベイズモデルの導入によって、感染者の特定と有病率の推定を同時に行うことに成功している。また、検査に有限の確率でエラーが生じることを踏まえ、検査結果から推定された患者の状態を確率変数と見做し、ブートストラップ法を用いた信頼性評価に基づき推定性能を向上させた。

同研究グループは現在、機械学習における効率的データサンプリング法である能動学習を用いた検査数の削減を試みており、本研究のさらなる発展が見込まれる([4]参照)。本研究は推定精度、計算時間の両面からグループテストの実用化を期待させるが、あくまでも理想的な状況を扱った理論的研究であることに注意が必要である。まず、本研究ではランダムに検体を混ぜ合わせてプールを構成しており、人手による検査では実現が困難であると考えられる。また、ベイズ推定による定式化において、検査におけるエラーの独立性を仮定しているが、この仮定は必ずしも適切でないことが示唆されている。しかしながら本研究は、今般の情勢の中で統計物理学および統計学が果たせる一つの役割を示している。本研究を契機として、提案手法を実用化させる新しい検査システムの開発や、検査の実態を踏まえた理論の改善など、理論と実用の協同的研究を通じた効率的検査方法の実現へとつながることが期待される。

[1] M. Mézard, M. Tazria and C. Toninelli, *J. Phys.: Conf. Ser.* 95 (2008).

[2] D. Sejdinovic and O. Johnson, 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2010.

[3] T. Kanamori, H. Uehara and M. Jimbo, *J. Stat. Theory Pract.* 6 (2012).

[4] A. Sakata, arXiv: 2007.13323.

原論文(7月8日公開済)

[Bayesian Inference of Infected Patients in Group Testing with Prevalence Estimation](#)

[Ayaka Sakata: J. Phys. Soc. Jpn. 89, 084001 \(2020\).](#)

<情報提供：坂田綾香（統計数理研究所 数理・推論研究系）>