

ベイズ推定に基づく適応的グループテスト

坂田 綾香 (統計数理研究所数理・推論研究系 ayaka@ism.ac.jp)

グループテストとは、患者の検体を混ぜ合わせて検査することで検査数を削減する方法である。患者集団に含まれる感染者を特定する際、一人一人検査する場合に必要な検査数は患者数と一致する。一方で、感染者の割合が十分小さいとき、グループテストを用いることで検査数を患者数より少なくすることができる。ただし、混ぜ合わせた検体に対する検査結果から患者一人一人の状態を推定する数学的手続きが必要となる。

グループテストは、数学的には論理和で定義された線形代数の問題として扱われ、また符号や情報通信の問題としても解釈できるため、情報理論に基づく数学的研究が行われてきた。もともとは医学検査の方法として提案されたグループテストであったが、問題設定を一般化すれば、集団に含まれる特別な状態にあるものを見つける手法として捉えられるため、製品の中に含まれる不良品の検出や、ゲノムライブラリのスクリーニングなどにも応用されている。

情報理論的研究では、例えば「患者数が十分大きいとき、感染者の割合が A で検査数が B 以上であれば、原理的に感染者を特定できる」といった知見が得られるが、実際に任意の検査結果から患者を推定するには何らかのアルゴリズムが必要である。そこで、信号処理の分野で提案された圧縮センシングなど他の問題との類似性を活用することで、グループテストに対する様々な決定論的アルゴリズムが開発されてきた。

しかし一般に、検査には誤りが含まれる。このような場合にはベイズ推定による感染者の特定法が効率的であり、そのアルゴリズムとして確率伝搬法が用いられている。統計物理の分野でも、グループテストにおけるベイズ推定をランダム系の問題と見做

す立場から、確率伝搬法を用いた研究が行われてきた。これらの研究から、グループテストは検査数を削減できるだけでなく、ある状況下では検査の誤りを訂正できることが示されてきた。

グループテストは、検体の混ぜ方によって非適応的グループテストと適応的グループテストに大別される。ここでは混ぜ合わせた検体をプールとよぶ。非適応的グループテストでは、プールをあらかじめ決めて固定するが、適応的グループテストでは逐次的な検査を考え、前回の検査までの結果を考慮して、次の検査を行うプールを適応的に作る。

一般に、非適応的グループテストよりも適応的グループテストの方が高い精度で患者を特定できるとされる。しかし、適応的グループテストに対する最もシンプルなアルゴリズムである binary splitting 法では、検査の誤りを修正する能力に限界があることが知られる。

我々は、ベイズ推定の枠組みで、予測分布に基づく適応的グループテスト法を考案し、確率伝搬法により実装した。提案法では、次の検査を行う候補となるプールの「不確実性」を予測分布により評価し、不確実性の高いプールを採用する、という手続きで適応的グループテストを行う。不確実性の高いプールとは、これまでの検査結果に基づく推定と整合性が取れないテスト結果を与えるプールである。このようなプールに対する検査を行うことで、推定結果を真の状態に近づけていくことができる。この手続きを確率伝搬法により実行し、非適応的グループテストよりも少ない数の検査数から、高い精度で感染者を特定できることを示した。

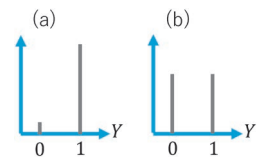
用語解説

確率伝搬法：

確率伝搬法は、ベイズ推定において事後分布の周辺化分布を近似的に評価する方法である。確率分布に含まれる変数の依存関係をグラフで表現し、メッセージとよばれる条件付き分布を再起的に更新することで、周辺化分布を評価する。その際、グラフが局所的にツリー構造をもつと仮定するため、厳密な評価には相当しないが、特に疎なグラフについては妥当な近似を与える。

予測分布：

予測分布は、すでに取得された検査結果に基づく推定が、未知の検査結果 Y にどの程度当てはまるのかの指標である。 Y が $\{0, 1\}$ の二値とすると、(a) のように $Y=1$ の値が大きく $Y=0$ の値が小さい場合は「未知の検査結果が $Y=1$ であれば、これは今までの検査結果に基づく推定と合っており、一方で $Y=0$ は検査の誤りである」と解釈できるが、(b) のように $Y=1$ と $Y=0$ の値が同程度の場合は、未知の検査結果の解釈ができず、不確実性が高いと考える。



未知の検査結果 Y が $\{0, 1\}$ の場合の予測分布の例。(a) は解釈ができる場合、(b) はできない場合の例。