2種類の陽性があるグループテストにおける周辺事後確率の推定

神保　雅一　　大学統計教員育成センター　特任教授

1．グループテストの例

水質検査

多地点の排水などに含まれる有害物質の有無を想定する際に，個々の地点の排水を個別にテストするのではなく，複数の地点の排水を混合したプールを地点の組合せを変えて作り，それらのプールの観測値から各地点の有害物質の有無を推定する．

DNA Library Screeningにおけるグループテスト Knill &Shliep(1995)

1990年代にヒトゲノムの解読のためのPCR検査にグループテストの手法が用いられた例がある．複数のクローンを含むpoolsをたくさん作り，各poolの検査結果から各クローンの陽性確率を推定した．推定にはMCMCを用いた．

COVID-19のPCR検査におけるグループテスト

中国の武漢，韓国などで，複数の検体を１つのプールにまとめてPCR検査を行い，その結果が陰性であれば，全検体を陰性とする検査が行われた．日本でも世田谷区の検査機関で「プール方式」の試みが行われようとしたが実現には至らなかった．

2．グループテストとは

n種のitems(検体)を一つ一つテストするとn回のテストが必要になるが，通常，陽性itemの比率は非常に小さいことが少なくない．

複数のitemをひとつにまとめてそのグループ(pool)に対して検査を行ない，その結果が陰性の場合には，そのpoolに含まれるすべてのitemが陰性であることを1回の検査で判定できる．また，陽性の場合には，その中のいずれかのitemが陽性であるとわかる．

しかし，false positive / false negativeの確率を加味して陽性／陰性の判定を行う必要がある．

グループテストにはadaptive/nonadaptiveの2種類がある．
１回のテストに時間がかかる場合にはnonadaptive

様々なpoolsを作りそれらのpoolsに対して検査を行い，それらの検査結果から各itemの陽性事後確率を求めれば，総item数nよりはるかに少ない数のpoolsの検査結果から各itemが陽性である事後確率を計算することができる．

各itemが陽性である周辺事後確率の計算量は$O(2^n)$で膨大
⇒ 効率の良いアルゴリズムが必要！

3．これまでの研究

○　各プールの観測値が得られたもとでitem jが陽性である周辺事後確率$P\left(X_j=1 \middle| \boldsymbol{S}=\boldsymbol{s}\right)$の近似計算アルゴリズム

MCPD：Knill&Shliep(1995), MCMC(Markov Chain Monte Carlo)を用いた事後確率計算アルゴリズム
BNPD: Uehara&Jimbo(2008), Belief Propagationを用いたアルゴリズム．MCPDより速い．
CCPD: Uehara&Jimbo(2009), pooling designに短いサイクルがある場合にBNPDに代わるアルゴリズム
LP: Malioutov&Malyutov(2012), 整数計画問題を線形計画問題に緩和して解くアルゴリズム．検査結果は{0, 1}に限定．
BP：Sakata(2020, 2021, 2022), adaptive, nonadaptive group testにおけるベイズ推定，カットオフ値の決定．
その他：Kanamori, Uehara, Jimbo(2012), BNPDの近似精度
　　　　Aldrige, Johnson, Scarlett(2019)に情報理論的解析

○　Pooling designの構成法

多数の組合せ論研究者によるd-disjunct, d-separableなどの性質を持つ組合せデザインの研究がある．ブロックデザインとの関係も深い．

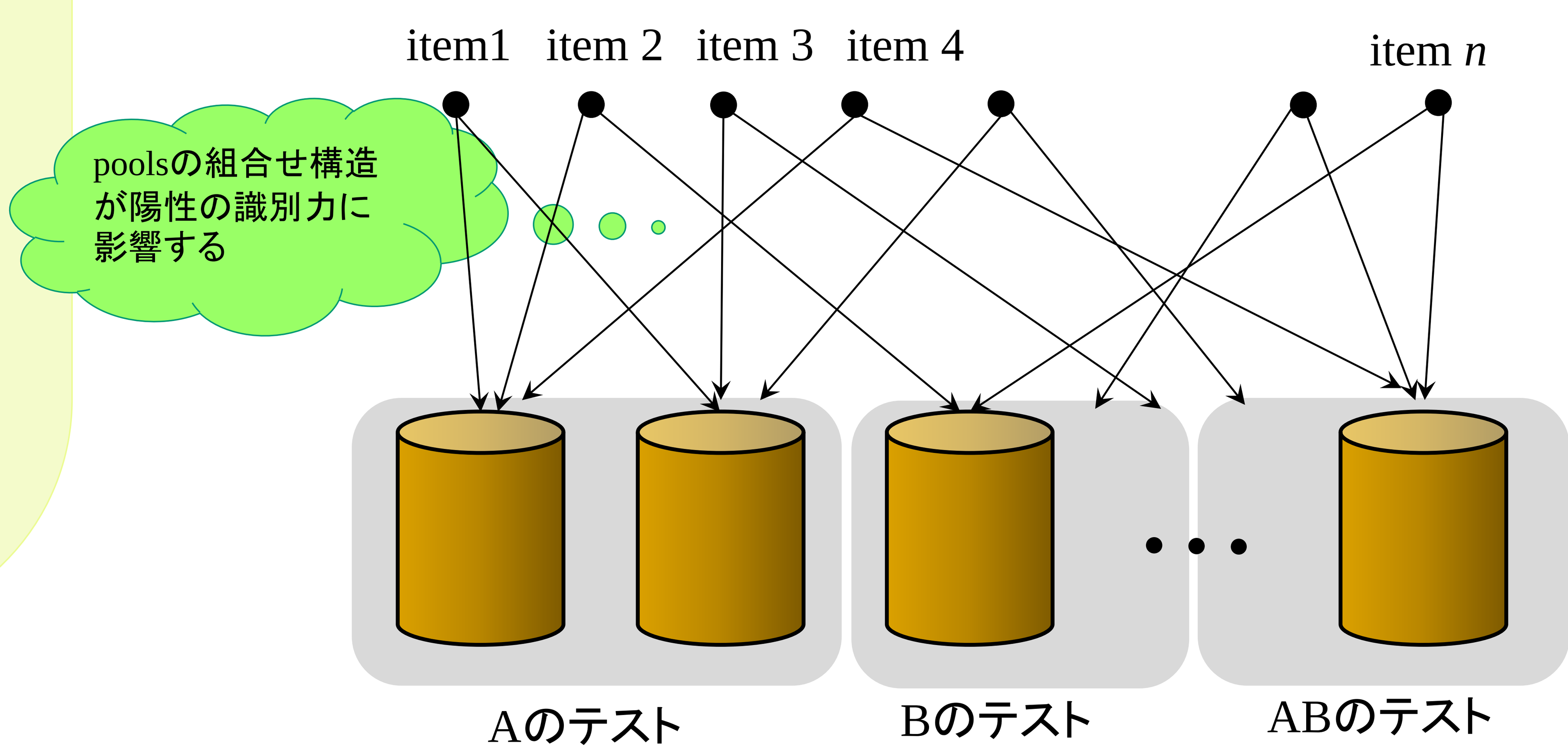
グループテストの研究は
　　　　　最適組合せ構造＋最適アルゴリズムの構築
の融合研究

4．2種類の陽性に対するグループテスト

2種類の陽性がある場合のグループテスト

○A, Bの2種類の陽性があるとき，A, Bそれぞれに対してグループテストを行うと1種類の場合の2倍のテスト回数が必要．

○poolsを3つに分け，(i)Aのテスト，(ii)Bのテスト，(iii)AとBのいずれかに陽性であれば陽性反応を示すテストの3通りのテストを行い，テスト回数を削減したい．



検査結果　2　0　0　1

検査結果の表示：negative(0), weak positive(1), positive(2), strong positive(3)

$$X_j^A=\begin{cases}1 & j\text{がAについてpositive}\\ 0 & j\text{がAについてnegative}\end{cases}$$

$$X_j^B=\begin{cases}1 & j\text{がBについてpositive}\\ 0 & j\text{がBについてnegative}\end{cases}$$

$$\boldsymbol{X^A}=(X_1^A,X_2^A,\dots,X_n^A)$$

$$\boldsymbol{X^B}=(X_1^B,X_2^B,\dots,X_n^B)$$

$$S_i=\begin{cases}0 & \text{pool } i \text{ がnegative}\\ 1 & \text{pool } i \text{ がweak positive}\\ 2 & \text{pool } i \text{ がmedium positive}\\ 3 & \text{pool } i \text{ が positive}\end{cases}$$

$$\boldsymbol{S}=(S_1,S_2,\dots,S_m)$$

計算量$O(2^n)$
効率的な事後確率計算アルゴリズムが必要

検査結果$\boldsymbol{S}=\boldsymbol{s}$が得られたもとでitem j がA, Bに対してpositiveである周辺事後確率 $P\left(X_j^A=a,X_j^B=b \middle| \boldsymbol{S}=\boldsymbol{s}\right)$ $(j=1,2,\dots,n)$を求めたい

○上記の目的のために
1. 効率の良いプーリングデザイン（2部グラフの組合せ構造）の構成
   ⇒　有限アフィン幾何を用いた構成法を提案した
2. 観測値から個々のitemがA，Bそれぞれに対して陽性である事後確率を計算するアルゴリズムの開発
   ⇒　Belief Propagation＋MCMCアルゴリズムを開発した

今後，上記のプーリングデザイン，アルゴリズムの評価とMCMCアルゴリズムの高速化の試みが必要