

2種類の陽性に対するグループテスト

松島 裕康[†] 田島 友祐[‡] 盧 暁南[§] 神保 雅一[¶]
 滋賀大学[†] 滋賀大学[‡] 岐阜大学[§] 統計数理研究所[¶]

1 はじめに

グループテストは、多数の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、集団の中から少ないテスト回数で陽性検体を識別する場合に適用される。N 種の各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、複数の検体(items)を混ぜ合わせたグループ(pools)に対してテストを行うことにより、テスト回数を削減することができる。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。

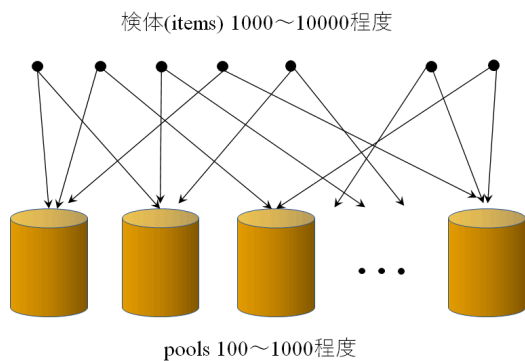


図1 グループテストの概念図

グループテストではテストの効率を高めるために、多くの検体の中からできるだけ少ない

pool 数で目的の検体を見出すことができる実験(pooling design)を計画することが重要であり、組合せ論的な研究がされている[1].

検体 c_j が pool G_i に含まれるとき $x_{ij} = 1$ そうでないとき $x_{ij} = 0$ とし、 $m \times N$ 行列 $M = (x_{ij})$ を pooling design と呼ぶ。陽性が1種類の場合に、pooling design M の任意の $t (\leq d)$ 列の論理和がすべて異なるとき、 M は d -separable であるという。また、 M の任意の d 列の論理和が他の1列を含まないとき、 M は d -disjunct であるという。 d -disjunct であれば d -separable であり、 \bar{d} -separable であれば、テスト結果に誤りがなく、 d 個以下の陽性検体はすべて正しく識別できる。

しかし、一般にこのような検査では、結果が陽性/陰性の2値ではなく多値であったり、テスト結果に偽陽性(FP)/偽陰性(FN)が生じる可能性を排除することはできない。そのため、FP, FN の確率を加味して陽性/陰性の判定を行う必要があり、pool に対するテスト結果から各検体の陽性確率を推定する Belief Propagation(BP) などを用いたアルゴリズムも開発されている[3, 4]。BP によるアルゴリズムを用いる場合に、pooling design がなす2部グラフに短いサイクルが多数存在すると陽性確率の推定値の誤差が大きくなる。

2 2種類の陽性が存在する場合のグループテスト

本研究では、A,B の2種類の陽性が存在する場合を考える。A,B のそれぞれに対して個別にグループテストを実施するとテスト回数が1種

Estimation of Posterior Probability in Group Test for Two Kind Positive

[†] Hiroyasu Matsushima, Shiga University

[‡] Yusuke Tajima, Shiga University

[§] Xiao-Nan Lu, Gifu University

[¶] Masakazu Jimbo, The Institute of Statistical Mathematics

類の場合に比べて2倍になるが、図2のように、A,Bのそれぞれに対するテストを行うpoolと、‘ABに対するテスト’（すなわち、AとBのいずれかに陽性であればテスト結果が陽性となるテスト）を行うpoolを混在させることにより、テスト回数を削減することができる。

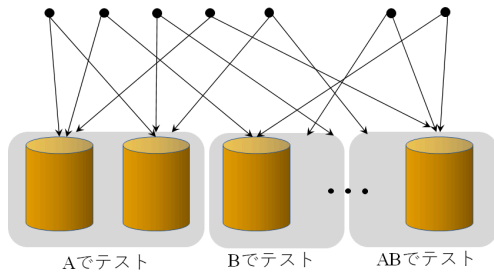


図2 2種類の陽性に対するグループテスト

2.1 2種類の陽性を識別する pooling design

2種類の陽性がある場合、pooling design M の行を、

$$M = \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix}, M_A = \begin{pmatrix} M_1 \\ M_3 \end{pmatrix}, M_B = \begin{pmatrix} M_2 \\ M_3 \end{pmatrix}$$

と置く。そして、 M_1, M_2 の行 (pool) ではそれぞれ、A,Bに対するテストを、 M_3 の行 (pool) では、ABに対するテストを行うとする。[2]は、 $(2, \bar{d})$ -separable という概念を定義し、FP, FN が存在しないとき、すなわち、テスト結果に誤りがないとき、pooling design M について M_A, M_B が \bar{d} -separable で M_1, M_2 が $(d-1)$ -disjunct であれば、 $|D_A \cup D_B| \leq d$ なる任意の2種類の陽性を正しく識別できることを示している。ただし、 D_A, D_B はそれぞれ、A,Bに対して陽性の検体の集合である。

有限アフィン幾何 $AG(n, q)$ の平行な拡張平面と1点ずつ共通部分を持つ直線の集合を用いて次の定理を示すことができる。

定理 1 q を素数べきとするとき、 $m_1 = m_2 = sq^{n-1}, m_3 = tq^{n-1}, N = q^{2(n-1)}, s+t \leq q$ とすると、 M_1, M_2, M_3 がそれぞれ、 $m_1 \times N, m_2 \times N, m_3 \times N$ の $(2, \bar{s})$ -separable pooling design が構成できる。

2.2 偽陽性・偽陰性を考慮したアルゴリズムとその識別力の評価

FP, FN の確率を $p(1|0), p(0|1)$ とし、A,B に対する陽性率を p_A, p_B とする。 M を定理1で構成した pooling design とし、得られたテスト結果に対して、[4]によるBPアルゴリズムを用いて、 M_A, M_B のそれぞれの pooling design について、A,Bそれぞれに対して陽性である事後確率を求める。図3は、FP, FN の確率をそれぞれ $p(1|0) = 0.01, p(0|1) = 0.03$ とし、 $p_A = p_B = 0.001, 0.003, 0.005$ 、定理1において、 $q = 7, n = 3, s = 2, t = 3$ とした pooling design を用いた場合に、各検体が陽性である事後確率を大きい順に並べるシミュレーションを1,000回繰り返したとき、すべての真の陽性検体が r_θ 位までに入る確率のグラフである。

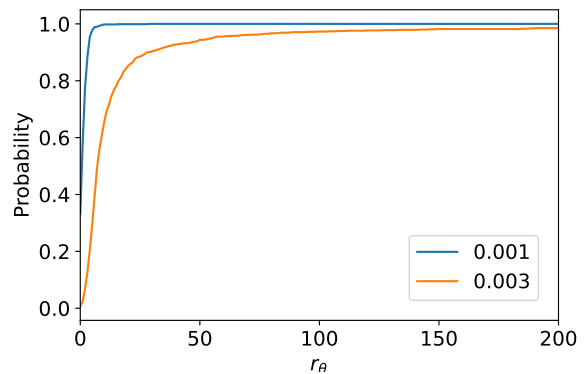


図3 真の陽性検体が r_θ 位内に入る確率

参考文献

- [1] DU, D.-Z. and HWANG, F. K. *Pooling Designs and Nonadaptive Group Testing*, WORLD SCIENTIFIC (2006).
- [2] LU, X.-N. Pooling designs for non-adaptive group testing with multiple variants of positives (manuscript in preparation).
- [3] SEJDINOVIC, D. and JOHNSON, O. Note on noisy group testing: Asymptotic bounds and belief propagation reconstruction, 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton) (2010).
- [4] UEHARA, H. and JIMBO, M. A Positive Detecting Code and Its Decoding Algorithm for DNA Library Screening, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **6**, 4 (2009), 652–666.