

SSE-DP-2023-5

2 種類の陽性に対するグループテストのための
BP と MCMC の hybrid アルゴリズム

松島 裕康¹, 田島 友祐¹, 盧 暁南², 神保 雅一^{1,3}

¹ 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

² 岐阜大学, 工学部 電気電子・情報工学科

³ 統計数理研究所, 大学統計教員育成センター

SSE-DP(ディスカッションペーパー・シリーズ)は以下のサイトから無料で入手可能です。

<https://stat-expert.ism.ac.jp/training/discussionpaper/>

このディスカッションペーパーは、関係者の討論に資するための未定稿の段階にある草稿である。著者の承諾なしに引用・複写することは差し控えられたい。

Belief propagation and MCMC hybrid algorithms for group testing with two kinds of defectives

Hiroyasu Matsushima¹, Yusuke Tajima¹, Xiao-Nan Lu², Masakazu Jimbo^{1,3}

¹Data Science and AI Innovation Research Promotion Center, Shiga University

²Department of Electrical, Electronic and Computer Engineering, Gifu University

³Center for Training Professors in Statistics, The Institute of Mathematics

Abstract

Testing n objectives one by one requires n tests, but the number of defective objective is often small. When multiple objectives can be tested of a pool, if the test result is negative, it can be determined that all objectives in its pool are negative at one time. Or, if the test result for that pool is positive, at least one or more of the objectives in that pool is positive. Thus, testing each pool that made by combining multiple objectives is called group testing. Using a group test, the posterior probability that each specimen is defective can be calculated from the test results of a much smaller number of pools than the total number of objectives. However, when making a positive/negative determination, the probability of false positives/false negatives in each test must be considered. Therefore, for this purpose, algorithms such as Belief Propagation (BP) and Markov Chain Monte Carlo (MCMC) are employed.

In this report, we develop and evaluate BP and MCMC algorithms for a combinatorial group test design that reduces the number of tests when there are two types of defectives.

2種類の陽性に対するグループテストのための BP と MCMC の hybrid アルゴリズム

松島 裕康¹ 田島 友祐¹ 盧 暁南² 神保 雅一^{1,3}

¹ 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

² 岐阜大学, 工学部 電気電子・情報工学科

³ 統計数理研究所, 大学統計教員育成センター

1. グループテスト

n 個の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, グループテストは n 回より少ないテスト回数で陽性検体を識別するために用いられる. 具体的には, 複数の検体を1つにまとめたプールに対してテストを行ない, 結果が陰性であれば, そのプールに含まれる全ての検体が陰性であることを1回で判定できる. また, 陽性であれば, その中のいずれかの検体が陽性である. 様々なプールの集合をプーリングデザインと呼び, その組合せ構造により, グループテストの効率が左右される. 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のような陽性検体の割合が小さい (0.0001~0.01 程度) 場合に有効である. グループテストの結果から各検体が陽性である事後確率を計算することができるが, 各テストにおける false positive (FP)/negative (FN) 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通りがあるが1回のテストに時間がかかる場合には, 大量のプールを一度にテストして陽性検体を識別する nonadaptive な手法が用いられる [2].

2. 2種類の陽性を含むグループテスト

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種類の場合に比べて2倍になることが想定される. これに対し, 図1のように, A, B のそれぞれに対するテストを行うプールと, 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) を行うプールを混在させることにより, テスト回数を削減する.

検体 c_j が A に対して陽性であるとき, $X_j^A = 1$,

そうでないとき, $X_j^A = 0$ とする. B についても同様に定義する. また, $X_j^{AB} = X_j^A \vee X_j^B$ とする. A, B, AB に対するテストを行うプールの集合をそれぞれ, $\mathcal{G}^A, \mathcal{G}^B, \mathcal{G}^{AB}$ とし, A に対するテストを行うプール $G_i \in \mathcal{G}^A$ に対して, $Z_i^A = \bigvee_{j \in G_i} X_j^A$ とする. また, $\mathbf{Z}^A = (Z_i^A | G_i \in \mathcal{G}^A)$ とする. B, AB についても同様に定義する. また, $G_i \in \mathcal{G}^A$ の観測値 s_i^A の値は 0, 1 の2値とし, 各テストにおける FP, FN の確率 $p(1|0), p(0|1)$ は既知とする.

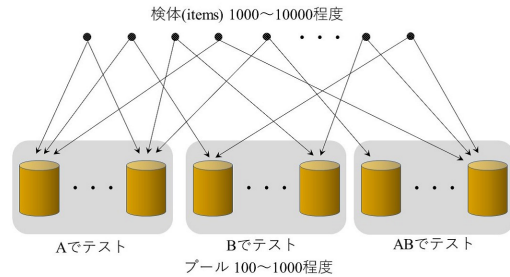


図 1: 2種類の陽性に対するグループテスト

A, B に対する陽性率を p_A, p_B とし, 観測値 s^A, s^B, s^{AB} が得られたもとで各検体 c_j が A, B に対して陽性である周辺事後確率

$$\Pr(X_j^A = a, X_j^B = b | s^A, s^B, s^{AB}) \quad (1)$$

を, $a, b = 0, 1$ の各場合について求めて陽性/陰性を識別したい.

本研究では, そのために Belief Propagation および MCMC を用いたアルゴリズムを開発し, その効率を評価する.

3. Belief Propagation による推定の評価

Belief Propagation (BP) を用いた事後確率の推定を評価するため, FP, FN の確率と陽性確率に従った擬似的なグループテストの結果を生成し, そ

れに対するBPの性能を評価する。陽性の検体については、A、Bそれぞれ10個ずつに固定し、陽性の検体は個数に従ってランダムに割り振った。FP、FNの確率をそれぞれ、 $p(1|0) = 0.01$, $p(0|1) = 0.03$ とし、 $p_A = p_B = 0.002$ として、A、B、ABについてそれぞれ、98回、98回、147回検査する pooling design を用いて、シミュレーションを1000回繰り返した。各シミュレーションにおいて、上記の設定に従いグループテストの観測値が決定される。用いた pooling design はFP、FNが存在しない場合にA、Bそれぞれ4個体程度までは正しく判定されることが保証されているが、本シミュレーションの陽性数はそれぞれ10個体と多く、BPのアルゴリズムでの識別限界に近い。

これらの各グループテストの結果に対してBPを用いて、事後確率を推定した結果を表1に示す。表1は、各検体が陽性である事後確率を大きい順に並べたとき、すべての真の陽性検体が95%、99%の確率で何位までに入るかを示した表である。

陽性検体を含む確率	99%	95%
順位	A:105 B: 97	A:26 B:23

表 1: すべての真の陽性検体が含まれる順位

4. MCMC との hybrid アルゴリズム

BP アルゴリズムは高速だが、陽性検体の数が多くなると、プーリングデザインの2部グラフのサイクルが原因で、BPで求められる周辺事後確率と真の値との乖離が大きくなったり、BPが収束しないことが生じる [4]。これに対し、MCMC(Markov Chain Monte Carlo)法を用いて周辺事後確率を求めることが考えられる [3]。しかし、MCMC法はBPに比べて計算時間がかかる欠点がある。

本研究では、BP アルゴリズムをスクリーニングアルゴリズムとして使い、BPの計算によって陽性である周辺事後確率が一定以下の検体を削除して、残った検体とそれに対する pooling design および観測結果を用いて、MCMC アルゴリズムにより各検体の周辺事後確率を求める hybrid アルゴリズムを提案する。

ここではMCMC アルゴリズムとして、Gibbs samplerを採用する。まず、事前確率 p_A, p_B に応

じて、ランダムに陽性検体の集合 $\mathcal{P}_A^{(0)}, \mathcal{P}_B^{(0)}$ を定め、各ステップ t で検体番号 $j = 1, \dots, n$ について順に、 c_j を $\mathcal{P}_A^{(t)}, \mathcal{P}_B^{(t)}$ に含めるか否かに関する推移確率を求め、それに応じて乱数を発生させて、 $\mathcal{P}_A^{(t)}, \mathcal{P}_B^{(t)}$ の要素を変化させる。これを一定回数繰り返して、各検体が $\mathcal{P}_A^{(\cdot)}, \mathcal{P}_B^{(\cdot)}$ に属した比率で事後確率を推定する。

BPによるスクリーニングを用いて陽性の確率が非常に小さい検体を除き、MCMCにより事後確率を計算することにより、陽性検体が想定以上に多く、BPでの収束値お呼び収束性に問題がある場合にも、計算時間を短縮してMCMCで真の事後確率の良い近似値を求めることができる。実際、あるシミュレーションでは、A、Bそれぞれ10個ずつの陽性を含む2401個の検体に対してBPによりスクリーニングを行った結果、アルゴリズムは収束しなかったが、常に事後確率が小さい検体を除くと187個の検体が残った。これらの検体に対して、事前確率を $p_A = p_B = 0.1$ として、MCMCを実行した。事前確率を0.1とした場合も、0.05の場合も事後確率の順位にはほとんど影響せず、Aでは19位まで、Bでは15位までにすべての陽性検体が収まった。この結果と表1を比較すると、BPのみの場合に比べて、MCMCを併用する方が、より識別力(スクリーニング力)が高くなっていることがわかる。

参考文献

- [1] Aldridge, M., Johnson, O., Scarlett, J., “Group testing: An Information Theory Perspective”, Trends Commu. Infor. Theory, Vol. 15, pp.196-392, 2019.
- [2] Du D.-Z., Hwang F., *Combinatorial Group Testing and its Applications*, pp.336, World Scientific, 1999.
- [3] Knill E. et al., “Interpretation of pooling experiments using the Markov Chain Monte Carlo method”, J. Comp. Biol., **33**, pp.395-406, 1996.
- [4] Uehara, H., Jimbo, M., “A Positive Detecting Code and Its Decoding Algorithm for DNA Library Screening”, IEEE/ACM Trans. Comp. Biol. Bioinfo., Vol.6, No.4, pp.652-666, 2009.

概要

本研究の目的

2種類の陽性検体が含まれたグループテストにおいて、各陽性検体を確率伝搬法(BP)とMCMCで効率的に識別(あるいはスクリーニング)

アプローチ

BPを用いて検体の陽性確率の推定から絞り込み

→ 絞り込んだ陽性検体の候補についてMCMCによる推定

結果

- BPによるスクリーニングが難しい場合に、プーリングデザインの識別能力を大きく超えなければBPとMCMCによるハイブリッドアルゴリズムが有効

グループテスト

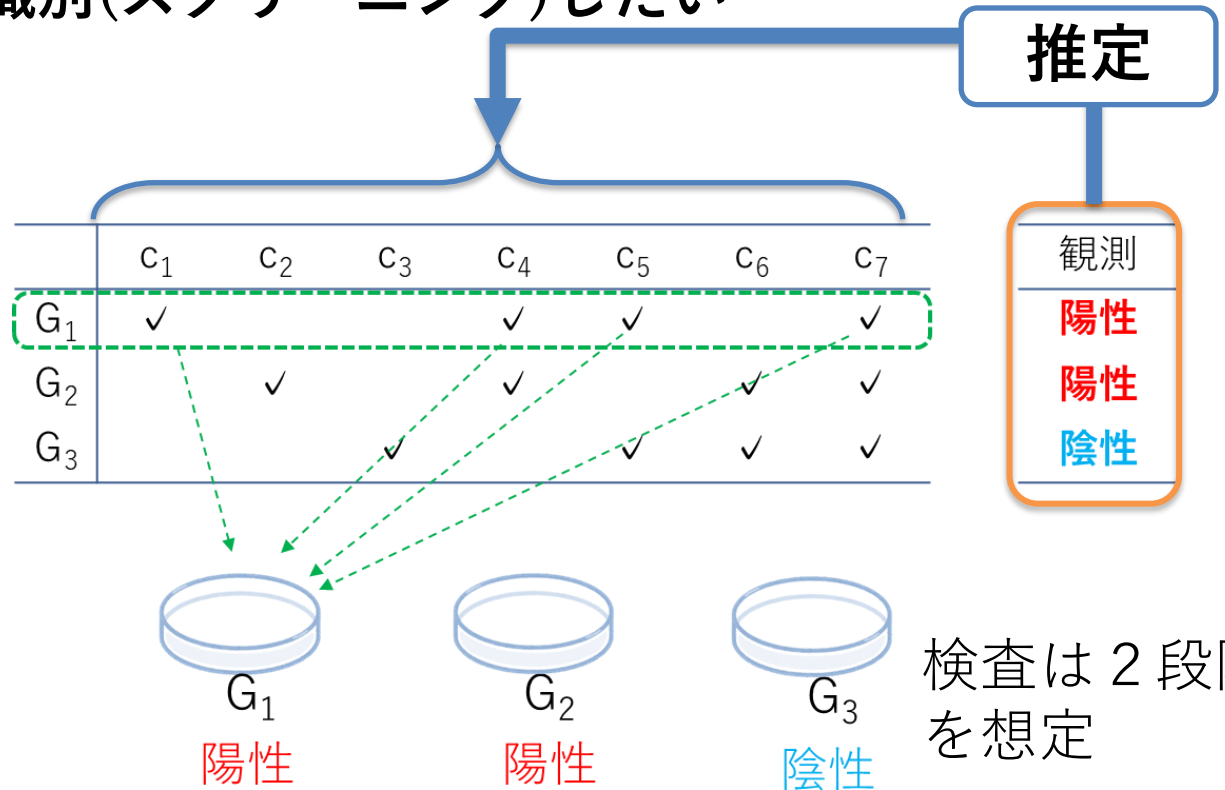
- 対象検体を混ぜ合わせて（組み合わせて）プールを作って検査することで**検査数を削減**する
- 混ぜ合わせたプールを複数用意して検査することで、陽性の**検体を識別（スクリーニング）**したい

検体の集合:

$$C = \{c_1, c_2, c_3, \dots, c_n\}$$

poolsの集合:

$$G = \{G_1, G_2, G_3, \dots, G_m\}$$



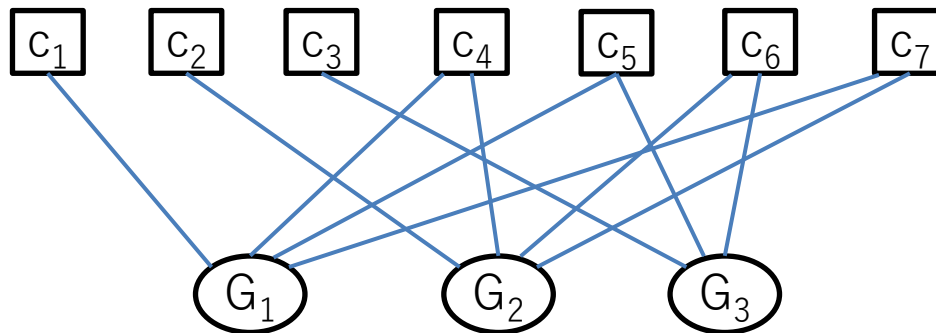
Pooling Design

C と G の包含関係は2部グラフあるいは結合行列で書ける。
この組合せ構造を **pooling design** という

検体の集合: $C = \{c_1, c_2, c_3, \dots, c_n\}$

poolsの集合: $G = \{G_1, G_2, G_3, \dots, G_m\}$

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
G_1	1	0	0	1	1	0	1
G_2	0	1	0	1	0	1	1
G_3	0	0	1	0	1	1	0



グループテストの目的

$$X_j = \begin{cases} 1 & \text{検体 } c_j \text{ が陽性} \\ 0 & \text{検体 } c_j \text{ が陰性} \end{cases} \quad \text{プール } G_i \text{ の状態: } Z_i = \bigvee_{j \in G_i} X_j,$$

$$\text{観測値: } S_i = \begin{cases} 0 & \text{negative} \\ 1 & \text{positive} \end{cases} \quad \begin{array}{l} \text{事前確率: } P(X_j = 1) = \theta \\ \text{(有病率)} \end{array}$$

(あるいは多値)

周辺事後確率: $P(X_j = 1 | S_1 = s_1, S_2 = s_2, \dots, S_m = s_m)$
を求めたい

$$\begin{aligned} & P(X_j = 1 | S_1 = s_1, S_2 = s_2, \dots, S_m = s_m) & \mathbf{S} &= (S_1, \dots, S_m) \\ & = \frac{1}{C} \sum_{\substack{j'=1 \\ j' \neq j}}^n P(\mathbf{X} = \mathbf{x}) P(\mathbf{S} = \mathbf{s} | \mathbf{X} = \mathbf{x}) & \mathbf{X} &= (X_1, \dots, X_n) \end{aligned}$$

← 2^{n-1} 個の和 (計算量が膨大)

→ 確率伝搬法(BP)を用いて周辺事後確率の近似値を求める

グループテストにおける確率伝搬法を用いた推定

関連研究

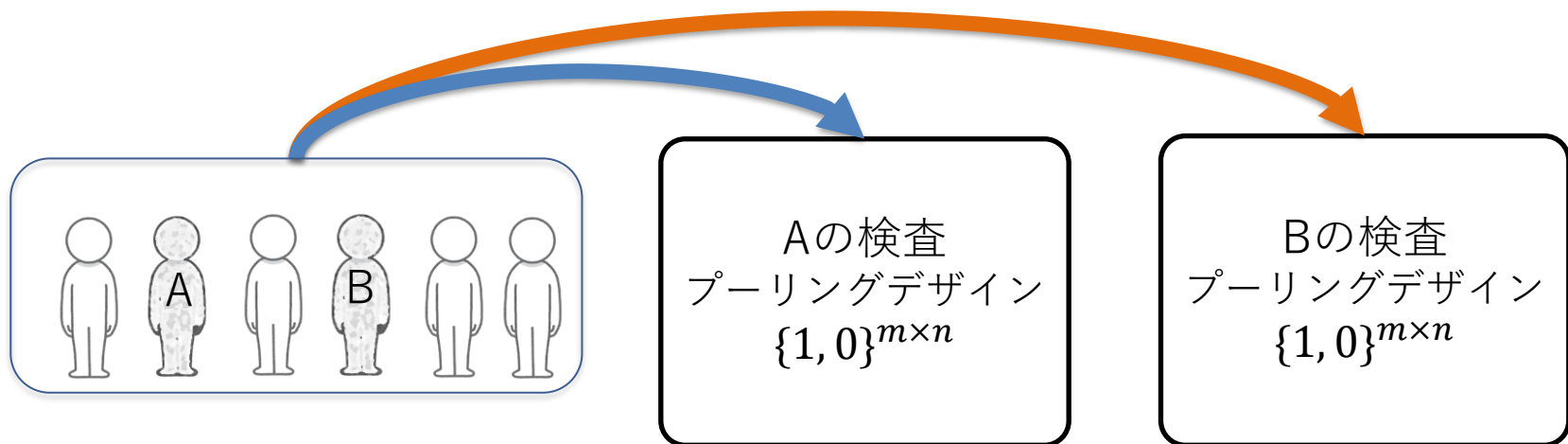
- H. Uehara, M. Jimbo(2009), A Positive Detecting Code and its Decoding Algorithm for DNA Library Screening, IEEE/ACM Transactions on Computational Biology and Bioinformatics 5(4), 652-666.
- H. Uehara, M. Jimbo(2009) A positive detecting algorithm for DNA library screening based on CCCP, J. Japan Statist. Soc. 39, 89-109.
- T. Kanamori, H. Uehara, M. Jimbo(2012), Pooling design and bias correction in DNA library screening. J. Stat. Theory Practice 6(1), 220-238.
- D. Sejdinovic and O. T. Johnson(2010), Note on noisy group testing: Asymptotic bounds and belief propagation reconstruction. In 48th Annual Allerton Conference on Communication, Control, and Computing. 998-1003.
- A. Sakata(2021). Active pooling design in group testing based on Bayesian posterior prediction. Physical Review E, 103, p. 022110
- A. Sakata and Y. Kabashima(2021). Decision theoretic cutoff and ROC analysis for Bayesian optimal group testing. arXiv: 2110.10877
- L. Ciampiconi, B. Ghosh, J. Scarlett, and K. S. Meel, “A MaxSAT-Based Framework for Group Testing”, AAAI, vol.34, no.06, pp.10144-10152, 2020.

陽性が2種類の場合

特定したい陽性の種類が2種類の場合において、
検査数を抑えつつ、それぞれの陽性の検体を特定したい

例) コロナとインフルエンザ，オミクロン株とデルタ株， etc

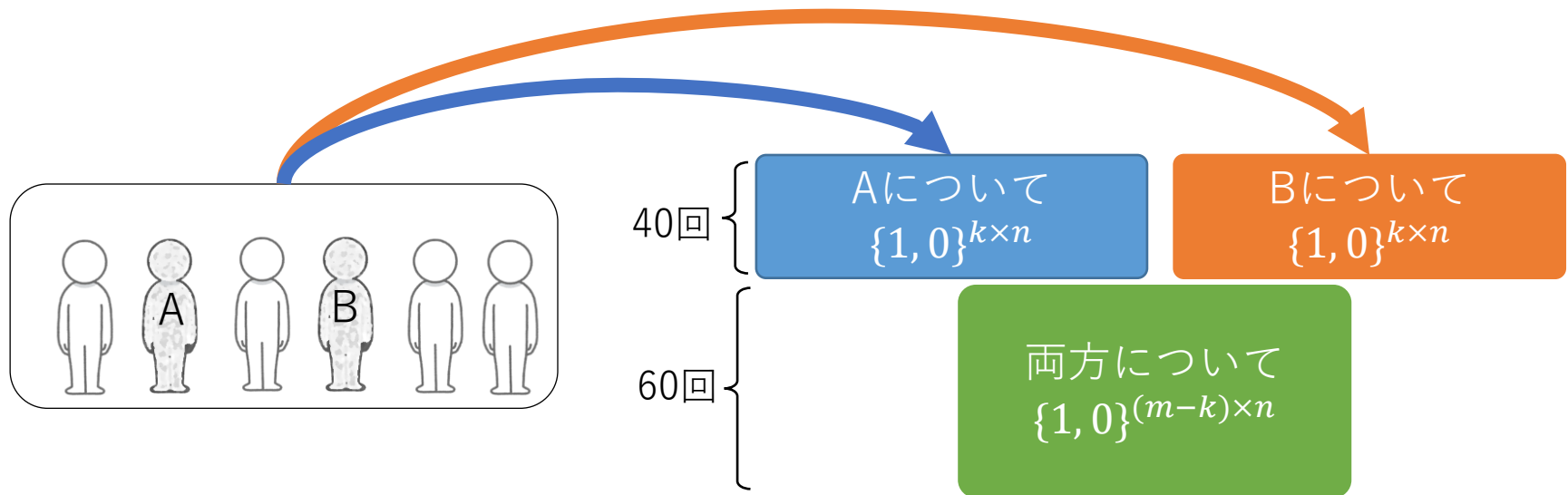
単純に考えると，検査数が倍になる



陽性が2種類の場合

プーリングデザインのプールを，種類ごとの固有の検査(m_A, m_B 回)と**共通の検査**(m_{AB} 回)に分けて使用

共通の検査は種類の区別はできないが，いずれかに陽性であれば陽性反応が現れる



例) $m_A = m_B = 40$, $m_{AB} = 60$ としたとき，
Aについて40回，Bについて40回，
共通部分について60回の**合計140回**

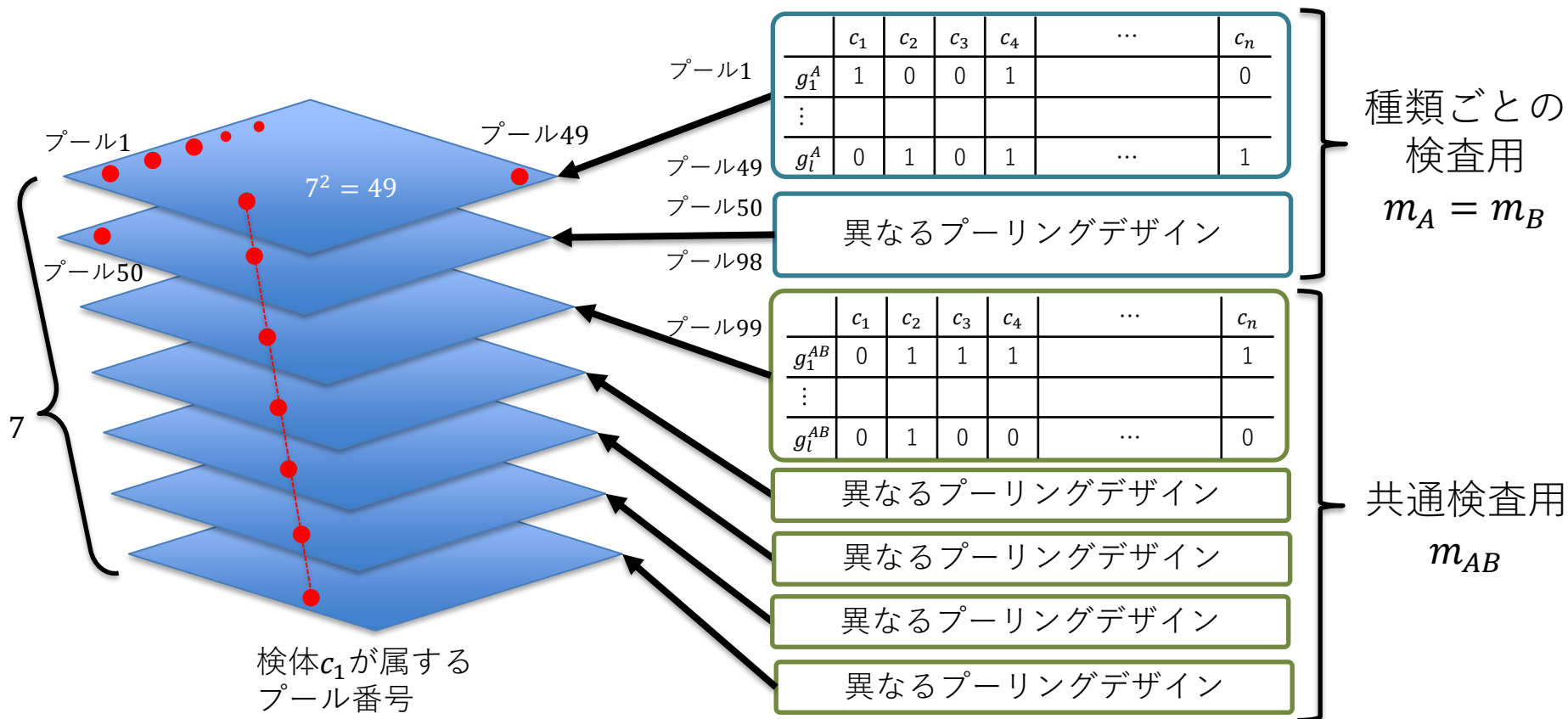
← 共通

プーリングデザイン

有限アフィン幾何を用いたプーリングデザインの作成

(2, \bar{s}) - separableなプーリングデザインを構成

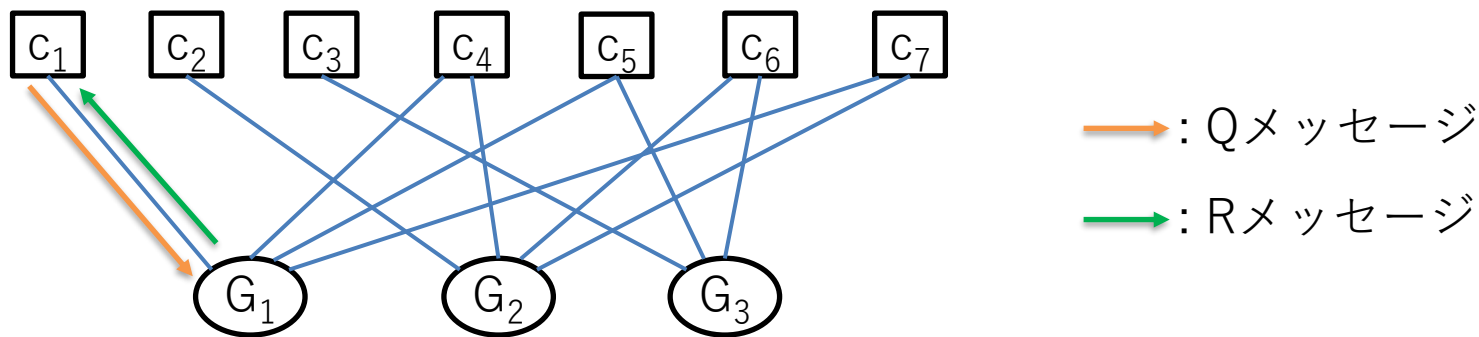
→ 検体は1つのクラス(図中の□で囲まれた表)の中で, 必ず2回は
 どれかのグループ (g) に含まれる



確率伝搬法 (Belief propagation: BP)

グラフィカルモデル上の推論アルゴリズム

2部グラフの上でメッセージをやり取りして、
周辺事後確率の近似値を求める手法



- 2部グラフにloopがなければ正しい値に収束する
- 検査結果を満たす陽性検体の組の候補が複数存在し、2部グラフに短いloopがたくさんあるとBPが収束しない
あるいは、収束値の真の周辺事後確率からの偏りが大きくなる

BPによる絞り込みとMCMC

グループテストの結果

	c_1	c_2	c_3	c_4	...	c_n	S
g_1^A	1	0	0	1		0	0
\vdots							
g_i^A	0	1	0	1	...	1	1

異なるプーリングデザイン

	c_1	c_2	c_3	c_4	...	c_n	S
g_1^{AB}	0	1	1	1		1	1
\vdots							
g_i^{AB}	0	1	0	0	...	0	0

異なるプーリングデザイン

異なるプーリングデザイン

異なるプーリングデザイン

BPによる推定

周辺事後確率が高い順に並び替え,
 θ_{rank} 位以内の検体を抽出

	c_1	c_2	c_3	c_4		c_m
g_1	1	1	1	0		0
\vdots						
g_i	0	1	0	1		1

*直接MCMCにかけると
膨大な計算時間を要する

MCMC
による推定

陽性の検体を
さらに絞り込み

MCMC

ギブスサンプリング法によるMCMCアルゴリズムで陽性の確率を推定する

1. 陽性A,Bについて事前確率に従い, $x_j^A(0) = 0,1$, $x_j^B(0) = 0,1$ をランダムに発生させる (あるいはBPによる推定値に基づいて発生)

$$P(x_j^A(t) = 1) = \theta_A, \quad P(x_j^B(t) = 1) = \theta_B$$

$$\begin{array}{l} \text{各検体の初期状態} \\ (t = 0) \end{array} \quad \mathbf{x}(0) = \begin{pmatrix} x_1^A(0) & \dots & x_n^A(0) \\ x_1^B(0) & \dots & x_n^B(0) \end{pmatrix}$$

2. 各検体 c_j について,
現在の状態 $(x_j^A, x_j^B) = (1,1), (0,1), (1,0), (0,0)$ に遷移する確率の比を求め,
その比に従い乱数を発生させ遷移先を決定

3. 2.を t 回繰り返す

$$\hat{\mathbf{x}} = \frac{1}{T_0} \sum \mathbf{x}(t) \text{が安定するまで捨て, (バーンイン)}$$

$$\hat{\mathbf{x}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}(t) \text{を各検体の周辺事後確率ベクトルとする}$$

シミュレーション実験

2種類の陽性が存在する集団の中から、**陽性の確率が高いすべての検体をスクリーニングする**（実用上はその後個別検査でFP削減）

実験のパラメータ

- 検体の数: $n = 2401$

- poolの数:

(プールの大きさ)

Aについて
 $\{1, 0\}^{k \times n}$

Bについて
 $\{1, 0\}^{k \times n}$

両方について
 $\{1, 0\}^{(m-k) \times n}$

(1). $m_A = m_B = 49, m_{AB} = 245$ ($m_{total} = 343$)

(2). $m_A = m_B = 98, m_{AB} = 147$ ($m_{total} = 343$)

(3). $m_A = m_B = 49, m_{AB} = 294$ ($m_{total} = 392$)

(4). $m_A = m_B = 98, m_{AB} = 245$ ($m_{total} = 441$)

(5). $m_A = m_B = 147, m_{AB} = 196$ ($m_{total} = 490$)

*各ケースで1000回実験

- 有病率: $\theta_A = \theta_B = 0.002$ (BP), $\theta_A = \theta_B = 0.1$ (MCMC),

- 絞り込む順位: $\theta_{rank} = 50$,

- 検査結果のFN, FPの確率[†]

$$P(S_g = 0 | Z_g = 0) = 0.99$$

$$P(S_g = 0 | Z_g = 1) = 0.03$$

$$P(S_g = 1 | Z_g = 0) = 0.01$$

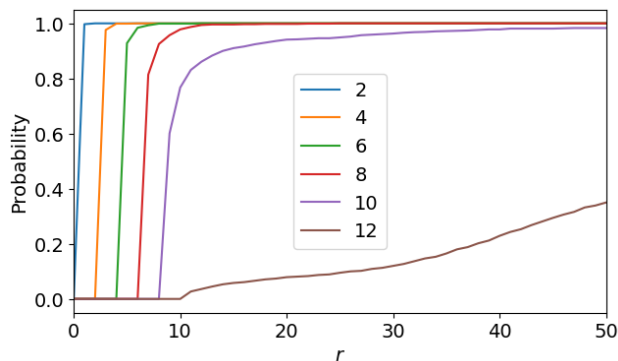
$$P(S_g = 1 | Z_g = 1) = 0.97$$

[†]引用: <https://www.cap.org/member-resources/articles/how-good-are-covid-19-sars-cov-2-diagnostic-pcr-tests#:~:text=Summary,and%2098%2D99%25%20specificity>

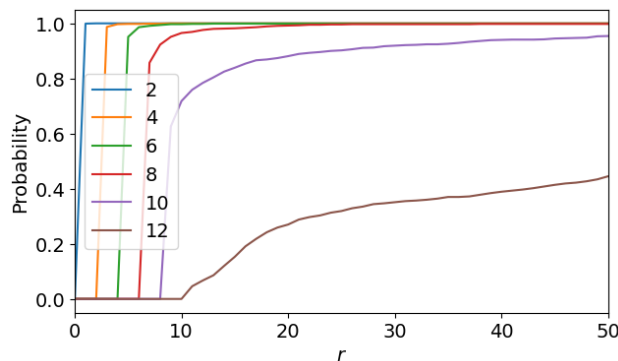
BPの結果: 確率順による累積確率

横軸は順位 (すべての陽性検体が r 位以内に含まれる確率)

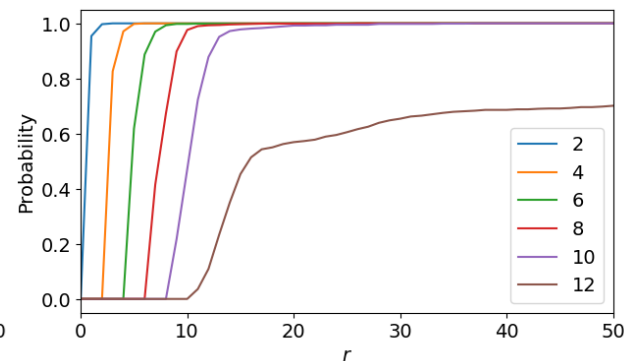
(1) $A=B=49, AB=245$



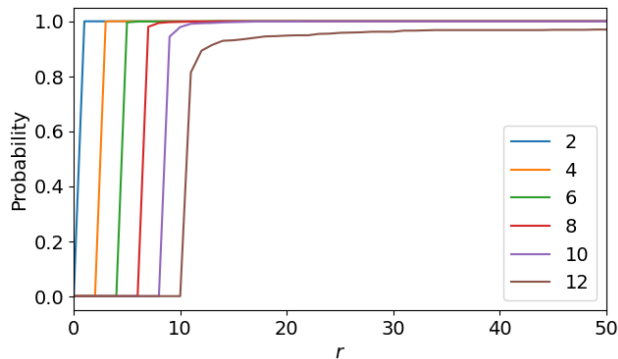
(2) $A=B=98, AB=147$



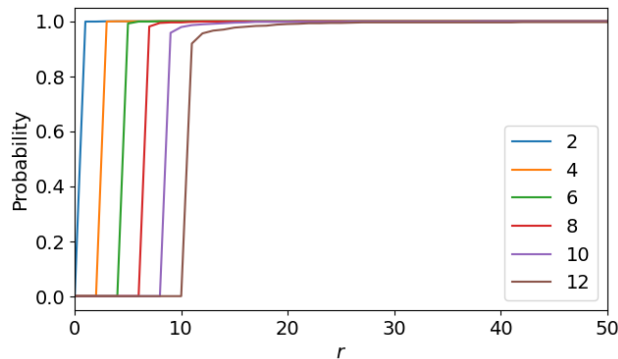
(3) $A=B=49, AB=294$



(4) $A=B=98, AB=245$



(5) $A=B=147, AB=196$



陽性数による差を見るため、
陽性検体数は
2, 4, 6, 8, 10, 12とした

(有病率 θ_A, θ_B に応じて決めてない)

BPの結果: 高確率で陽性を検出できる閾値(順位)

すべての陽性の検体が含まれている順位

→ 確率の高い順に下表の順位までを閾値とすると, 95% or 99%の確率で陽性の検体が全て含まれる

用いたプーリングデザインでFP, FNがない場合の識別限界



各陽性の数	2		4		6		8		10		12	
	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%
(1) $m_A = m_B = 49$ $m_{AB} = 245$	A:2 B:2	A:2 B:2	A:4 B:4	A:4 B:4	A:7 B:7	A:7 B:7	A:13 B:15	A:10 B:10	A:95 B:85	A:27 B:29	A:599 B:661	A:240 B:257
(2) $m_A = m_B = 98$ $m_{AB} = 147$	A:2 B:2	A:2 B:2	A:5 B:4	A:4 B:4	A:8 B:11	A:6 B:6	A:20 B:19	A:10 B:10	A:97 B:80	A:50 B:44	A:404 B:394	A:156 B:181
(3) $m_A = m_B = 49$ $m_{AB} = 294$	A:3 B:3	A:2 B:3	A:6 B:6	A:5 B:5	A:9 B:9	A:8 B:8	A:12 B:14	A:11 B:11	A:21 B:19	A:14 B:15	A:453 B:307	A:112 B:100
(4) $m_A = m_B = 98$ $m_{AB} = 245$	A:2 B:2	A:2 B:2	A:4 B:4	A:4 B:4	A:6 B:6	A:6 B:6	A:9 B:9	A:8 B:8	A:12 B:12	A:11 B:11	A:100 B:104	A:24 B:18
(5) $m_A = m_B = 147$ $m_{AB} = 196$	A:2 B:2	A:2 B:2	A:4 B:4	A:4 B:4	A:6 B:6	A:6 B:6	A:9 B:9	A:8 B:8	A:14 B:13	A:10 B:10	A:21 B:18	A:13 B:13

■ : 概ね識別ができています (FPを少し含む)

■ : 真の陽性の数の2倍程度の順位までに真の陽性が高い確率で含まれる (スクリーニング良好)

BPの結果: 高確率で陽性を検出できる閾値(順位)

すべての陽性の検体が含まれている順位

→ 確率の高い順に下表の順位までを閾値とすると, 95% or 99%の確率で陽性の検体が全て含まれる

各陽性の数	2		4		6		8		10		12	
	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%
(1) $m_A = m_B = 49$ $m_{AB} = 245$	A:2 B:2	A:2 B:2	A:4 B:4	A:4 B:4	A:7 B:7	A:7 B:7	A:13 B:15	A:10 B:10	A:95 B:85	A:27 B:29	A:599 B:661	A:240 B:257
(2) $m_A = m_B = 98$ $m_{AB} = 147$	A:2 B:2	A:2 B:2	A:5 B:4	A:4 B:4	A:8 B:11	A:6 B:6	A:20 B:19	A:10 B:10	A:97 B:80	A:50 B:44	A:404 B:394	A:156 B:181
(3) $m_A = m_B = 49$ $m_{AB} = 294$	A:3 B:3	A:2 B:3	A:6 B:6	A:5 B:5	A:9 B:9	A:8 B:8	A:12 B:14	A:11 B:11	A:21 B:19	A:14 B:15	A:453 B:307	A:112 B:100
(4) $m_A = m_B = 98$ $m_{AB} = 245$	A:2 B:2	A:2 B:2	A:4 B:4	A:4 B:4	A:6 B:6	A:6 B:6	A:9 B:9	A:8 B:8	A:12 B:12	A:11 B:11	A:100 B:104	A:24 B:18
(5) $m_A = m_B = 147$ $m_{AB} = 196$	A:2 B:2	A:2 B:2	A:4 B:4	A:4 B:4	A:6 B:6	A:6 B:6	A:9 B:9	A:8 B:8	A:14 B:13	A:10 B:10	A:21 B:18	A:13 B:13

ケースをピックアップしてMCMCによる推定を検討

→ A, Bについてそれぞれ50位までをスクリーニングし,
それらについてMCMCによる周辺事後確率を求める

MCMCによる真の陽性検体の確率と順位

BPの結果, A, Bについてそれぞれ50位までの検体を取り出し, 計約100検体についてMCMCで確率を推定

真の陽性検体の中でワースト順位が20位を超えるケースをピックアップ

→ 123ケース中79ケースにおいて, MCMCによる推定による真の陽性検体のワースト順位が改善した

良くなった一例 (3) A,Bのテスト: 98, ABのテスト: 147, 陽性の検体数: 10

陽性A

id	BP		MCMC	
	rank	Prob	rank	prob
233	56	0.242999	11	0.6560
310	8	0.996883	3	0.9975
399	7	0.997173	5	0.9905
1290	5	0.999803	4	0.9915
1481	1	0.999908	1	0.9995
1542	6	0.997443	1	0.9995
1629	3	0.999856	7	0.9885
1641	2	0.999904	8	0.9120
1688	55	0.332162	6	0.9890
1895	4	0.999812	13	0.3600

陽性B

id	BP		MCMC	
	rank	prob	rank	prob
265	2	0.999906	2	0.9990
475	3	0.999904	1	0.9995
502	46	0.335500	7	0.9750
1201	16	0.931539	2	0.9990
1243	1	0.999919	5	0.9985
1285	38	0.496782	8	0.9515
1510	6	0.998726	10	0.9310
1869	7	0.997314	6	0.9760
1910	8	0.996783	9	0.9465
2270	4	0.999896	2	0.9990

真の陽性検体の確率と順位

前頁のシミュレーションについて、
BPを用いて、A, Bについて50位までの検体を取り出し、
MCMCで確率を推定

MCMCでBPの順位が改善できなかったケースが、123ケース中
25ケースあった

- プーリングデザインの識別境界を大きく超えている
- FP, FN による影響

 より識別力の高いプーリングデザインを用いる必要がある

例えば

- (1) のデザインでは、陽性数がA,Bそれぞれで8個以内
- (2) のデザインでは、陽性数がA,Bそれぞれで10個以内

まとめ

本研究の目的

2種類の陽性検体が含まれたグループテストにおいて、各陽性検体をBPとMCMCで効率的に識別(あるいはスクリーニング)

アプローチ

BPを用いて検体の陽性確率の推定から絞り込み

→ 絞り込んだ陽性検体の候補についてMCMCによる推定

結果

- 陽性検体数が8以下の場合、どのプーリングデザインでも、BPのみで陽性検体数の2倍以内の順位の検体までみれば、99%以上の確率で全ての陽性検体をスクリーニングできる
- BPによる絞り込みから、MCMCで更に識別が向上するケースがある
- BPおよびMCMCでも困難なケースも存在
 - 必要に応じて識別力の高いプーリングデザインの利用、FP・FNの確率の影響など、今後、検証する必要がある

今後の課題

- 検体数とプール数の関係
 - θ_A, θ_B と FN, FP の確率が与えられたとき, プール数 m_A, m_B, m_{AB} をどの程度にする必要があるか ($m < n$)
 - プーリングデザインについて
 - θ_A, θ_B に応じたプーリングデザインの組み合わせ的構成法
 - 2種類の場合, 種類ごとの部分と共通部分のプーリングデザインの分割の仕方の検討
 - 第2段階のテストを行う検体の順位のしきい値の決定法
 - 与えられた FN の確率に対するしきい値を求めたい
 - FN だけでなく A を B, B を A と誤判別する確率の評価
 - FN, FP の設定
例えば PCR 検査の場合
 - Specificity (特異度, TN) は総じて 98%-99.9% と高い.
 - Sensitivity (感度, TP) 検査方法や環境によって違う. 85-86%, 97% など
- [参考サイト]
- <https://www.uptodate.com/contents/covid-19-diagnosis>
- <https://www.cap.org/member-resources/articles/how-good-are-covid-19-sars-cov-2-diagnostic-pcr-tests#:~:text=Summary,and%2098%2D99%25%20specificity>