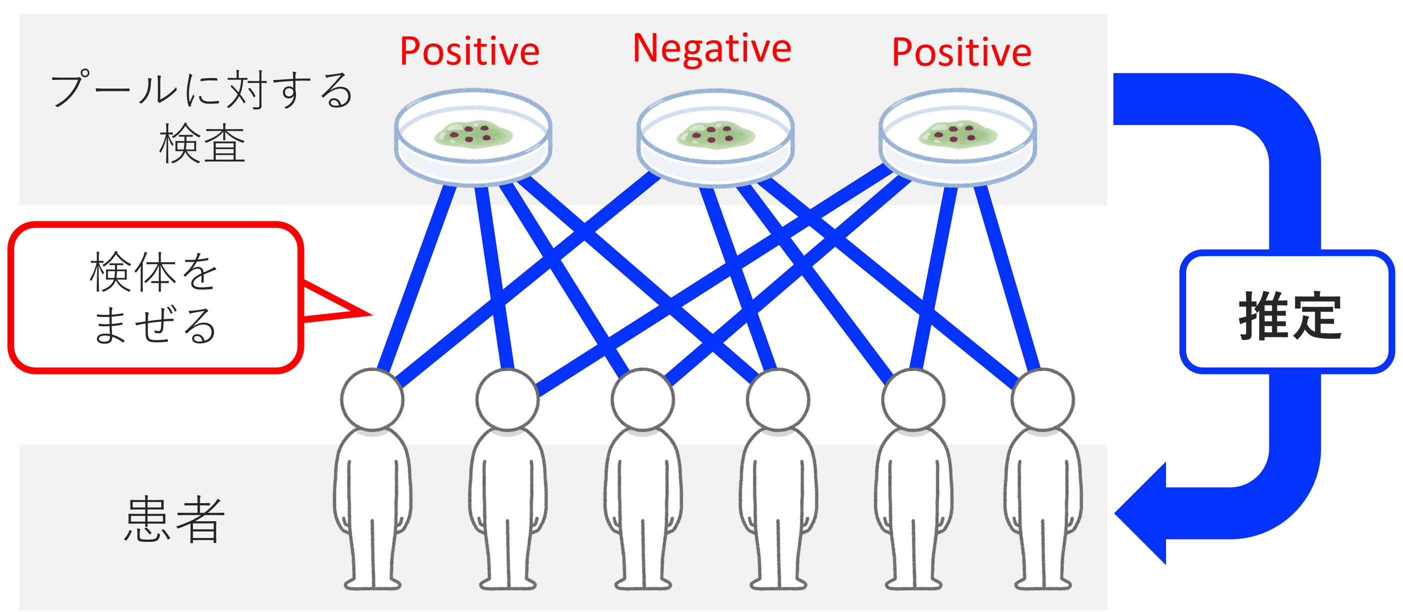
2022年6月17日　統計数理研究所　オープンハウス

ベイズ推定を用いたグループテストによる検査エラー修正

坂田　綾香　　数理・推論研究系　准教授

■ グループテストとは



- 患者から採取した検体を混ぜて検査する方法
  - プール法とも呼ばれる
- プール数＜患者数とする
  - プールに対するテスト回数は患者数より少ない
- プールに対する検査から患者の状態を特定する必要がある
  - さまざまな推定法が提案されてきた

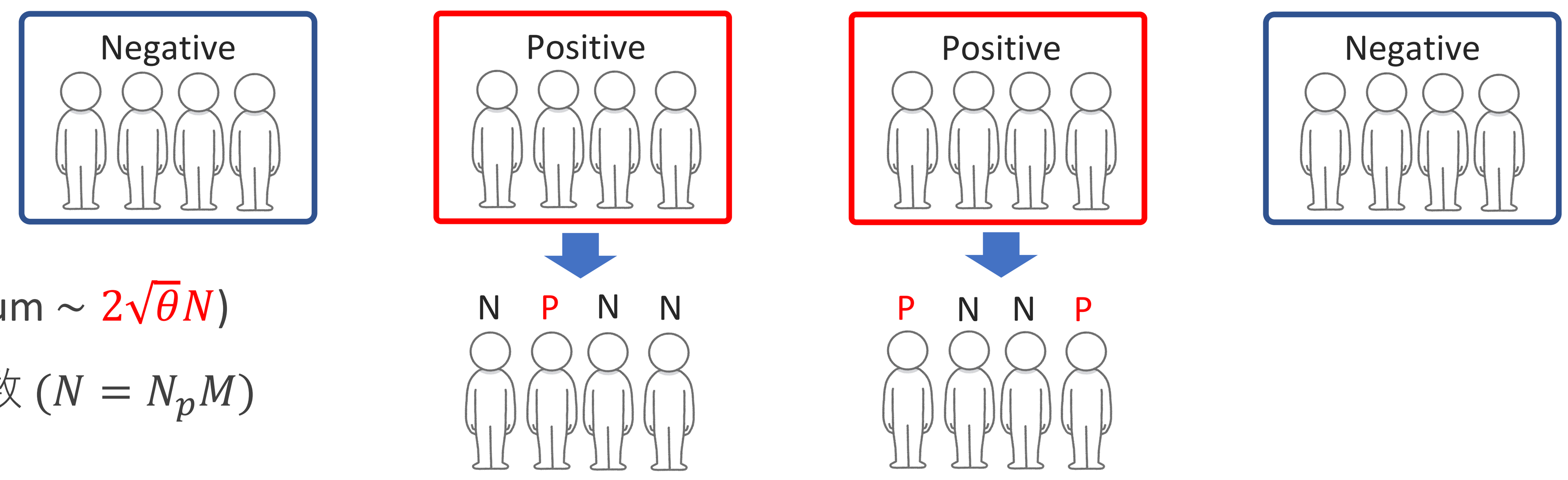
◆ **決定論的な推定の例：二段階推定**

- 第一段階: 適当に作ったプールに対して検査する
- 第二段階: 陽性プールに含まれる患者を再検査する

→ 期待される検査数: $M + \{1-(1-\theta)^{N_p}\}MN_p$ (at minimum ~ $2\sqrt{\theta}N$)

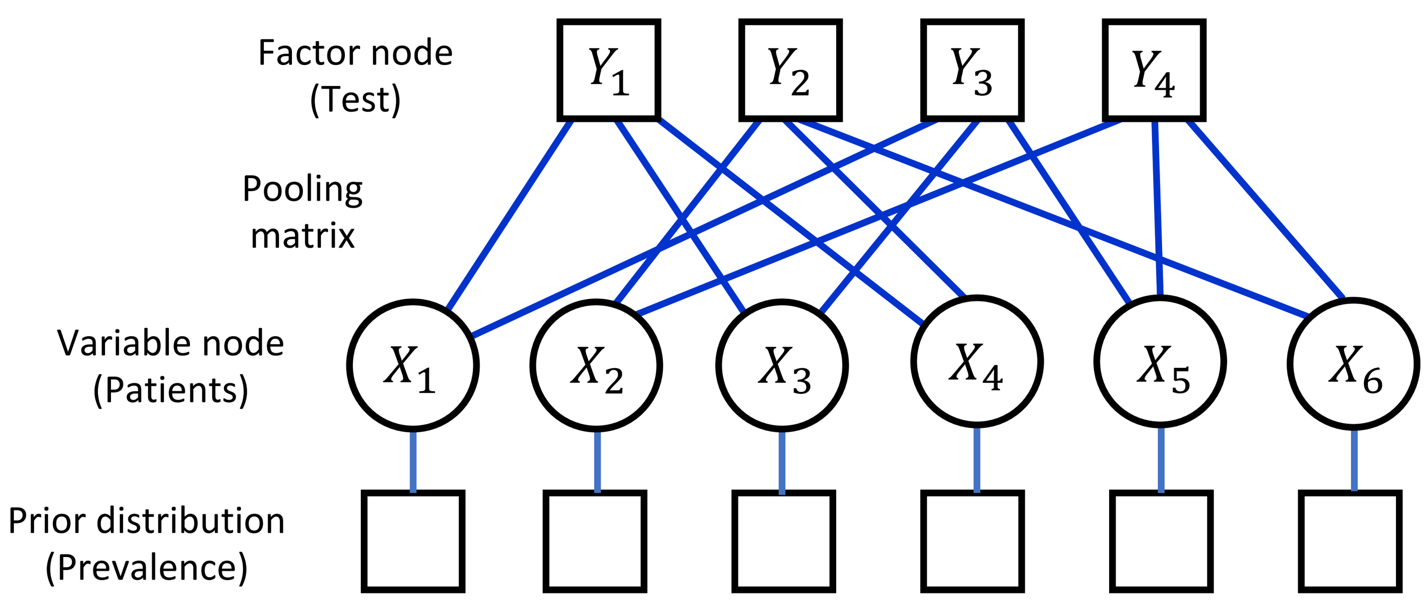
- M: プール数, θ: 有病率, N_p: プールサイズ, N: 患者数 ($N = N_pM$)

✓ 欠点：検査エラーを修正することができない



Dorfman, Ann. Math. Statist. (1943)

■ グループテストにおけるベイズ推定



因子グラフによるグループテストの表現

- 検査結果$\boldsymbol{Y}$の確率モデルを導入
  - プールの真の状態がある確率で反転する、と仮定する
    - 真陽性確率 p_{TP}, 偽陽性確率p_{FP}を用いる
  - プールの真の状態：一人でも陽性者がいれば陽性、それ以外は陰性
- 患者の状態$\boldsymbol{X}$を、$\boldsymbol{Y}$の確率モデルのパラメータと見做して推定する
  - 患者の状態は「陽性」「陰性」の二状態とする
- 周辺事後確率から患者の状態を推定
  - グループテストを因子グラフで表現し、確率伝搬法により周辺事後確率を推定する

Sakata, J. Phys. Soc. Jpn. 89 (2020)

■ カットオフ値の決定

- ベイズ推定の問題点：周辺事後確率は連続変数
  - 連続変数から「陽性」「陰性」を決めなくてはならない
- カットオフ値を導入する必要がある
  - リスク関数を用いて決める
    - 偽陽性が与える損失をλ_{FP}, 偽陰性が与える損失をλ_{FN}とする
- リスク関数Rの定義：$R = \lambda_{\mathrm{FN}}\mathrm{FN} + \lambda_{\mathrm{FP}}\mathrm{FP}$
  - FN(偽陰性率), FP(偽陽性率)はカットオフの関数
  - リスク関数を最小化するようにカットオフを決める

リスク関数最小の意味で最適なカットオフ

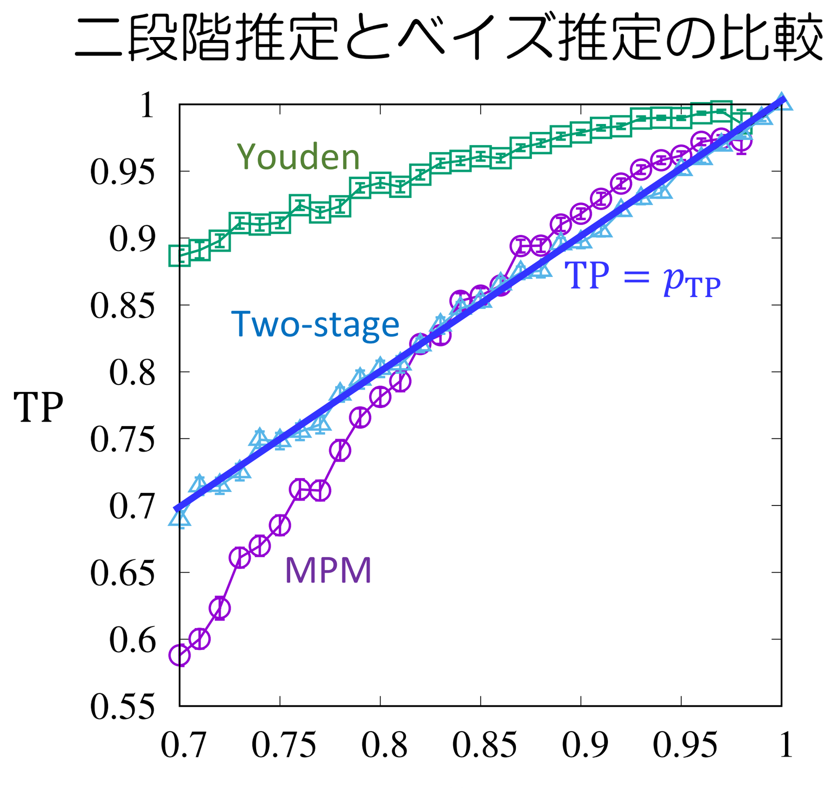
$$\hat{x}_i^*(\boldsymbol{y}) = \mathbb{I}\left(\rho_i(\boldsymbol{y}) > \frac{\theta\lambda_{\mathrm{FP}}}{\lambda_{\mathrm{FN}}(1-\theta)+\lambda_{\mathrm{FP}}\theta}\right)$$

- ρ_i: i番目の患者の周辺事後確率

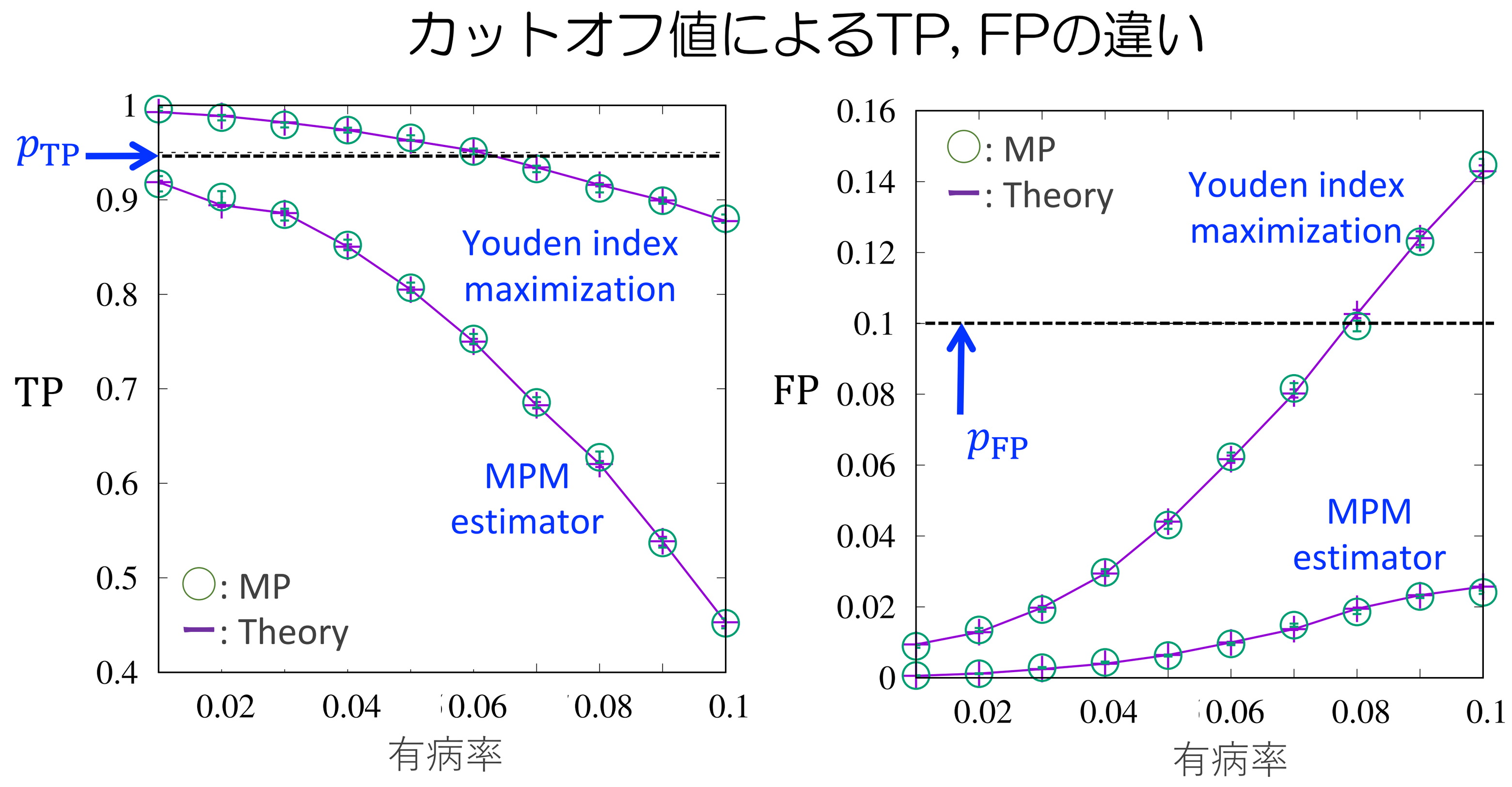
- Youden index 最大化：$\lambda_{\mathrm{FN}} = \lambda_{\mathrm{FP}} = 0.5$
  - ✓ カットオフ ＝ 有病率
    - 偽陽性、偽陰性をバランスよく修正
- MPM推定量： カットオフ0.5
  - ✓ $\lambda_{\mathrm{FN}} = \theta,\ \lambda_{\mathrm{FP}} = 1-\theta$
    - $\theta < 0.5$では偽陽性の修正を優先する

Sakata and Kabashima, arXiv:2110.10877

二段階推定とベイズ推定の比較



カットオフ値によるTP, FPの違い



- プールの作り方を工夫することも可能
- 能動学習によりプールを逐次的に構成すると誤り訂正能力が上がる

カットオフを適切に選ぶと偽陰性が修正される

Sakata, Phys. Rev. E 103, 022110 (2021)