Statistical physics for Bayesian statistical decision theory: an application to group testing

Department of Statistical Inference & Mathematics / Research Center for Statistical Machine Learning

The Institute of Statistical Mathematics

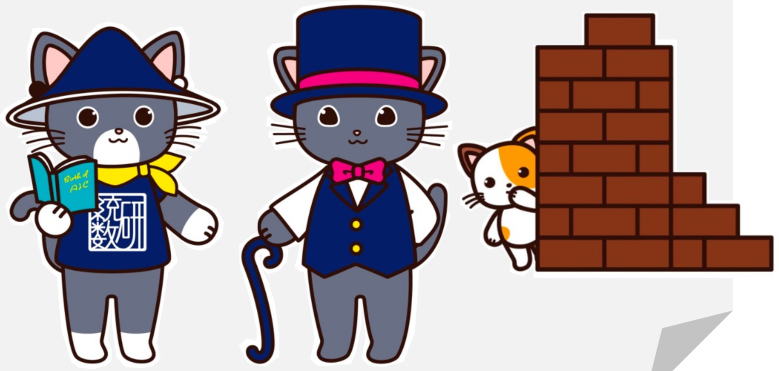
Ayaka Sakata

Special thanks to Yoshiyuki Kabashima (UT) and Yukito Iba (ISM)

Ayaka Sakata (坂田綾香)

- Department of Statistical Inference & Mathematics, The Institute of Statistical Mathematics

- **- 2008.3** Master's student, Graduate School of Arts & Sciences, University of Tokyo
  - Supervisor: Kunihiko Kaneko
- **- 2011.3** PhD student, Graduate School of Arts & Sciences, University of Tokyo
  - Supervisor: Koji Hukushima
- **- 2014.3** JSPS Fellow (PD) @ Tokyo Tech (Kabashima Group)
- **- 2015.3** SPDR @ RIKEN Theoretical Biology Group
- **- 2020.3** Assistant Professor @ ISM
- **2019.10 -** JST PRESTO Researcher
- **2020.4 -** Associate Professor @ ISM



Outline

- Introduction
  - Statistical physics and Bayesian inference
    - Sparse estimation
  - Group testing
    - Bayesian inference for group testing

Outline

- Introduction
  - Statistical physics and Bayesian inference
    - Sparse estimation
  - Group testing
    - Bayesian inference for group testing
- Our contributions
  - Bayesian statistical decision for group testing
    - Make a diagnose as an optimal “action”
  - Algorithm for actual inference in group testing: message passing
  - Related topics in Bayesian group testing

Statistical Physics and Bayesian Inference

Randomness and data

Bayesian inference

- Purpose: Model the generative process of given data $\boldsymbol{y}$

Bayesian inference

- Purpose: Model the generative process of given data $\boldsymbol{y}$
  - $\boldsymbol{y}$ is considered to be a realized value of a random variable.

Bayesian inference

- Purpose: Model the generative process of given data $\boldsymbol{y}$
  - $\boldsymbol{y}$ is considered to be a realized value of a random variable.
- Assumption: The generative process is governed by a parameter $\boldsymbol{x}$.

Bayesian inference

- Purpose: Model the generative process of given data $\boldsymbol{y}$
  - $\boldsymbol{y}$ is considered to be a realized value of a random variable.
- Assumption: The generative process is governed by a parameter $\boldsymbol{x}$.
  - The distribution of $\boldsymbol{x}$ is the center for the modeling.

Bayesian inference

- Purpose: Model the generative process of given data $\boldsymbol{y}$
  - $\boldsymbol{y}$ is considered to be a realized value of a random variable.
- Assumption: The generative process is governed by a parameter $\boldsymbol{x}$
  - The distribution of $\boldsymbol{x}$ is the center for the modeling.

- Definition:

  A Bayesian statistical model is made of a parametric statistical model (likelihood), $f(\boldsymbol{y}|\boldsymbol{x})$, and a prior distribution on the parameters, $\phi(\boldsymbol{x})$.

Bayesian inference

- Purpose: Model the generative process of given data $\boldsymbol{y}$
  - $\boldsymbol{y}$ is considered to be a realized value of a random variable.
- Assumption: The generative process is governed by a parameter $\boldsymbol{x}$
  - The distribution of $\boldsymbol{x}$ is the center for the modeling.

- Definition:

  A Bayesian statistical model is made of a parametric statistical model (likelihood), $f(\boldsymbol{y}|\boldsymbol{x})$, and a prior distribution on the parameters, $\phi(\boldsymbol{x})$.

Posterior distribution $$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{x})\phi(\boldsymbol{x})}{Z(\boldsymbol{y})}$$

Posterior distribution

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x}) \phi(\boldsymbol{x})$$

Posterior distribution

Bayesian inference

- $\boldsymbol{y}$: Data
- $\boldsymbol{x}$: Model parameter

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x})\phi(\boldsymbol{x})$$

Posterior distribution

Bayesian inference

- $\boldsymbol{y}$: Data
- $\boldsymbol{x}$: Model parameter

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x})\phi(\boldsymbol{x})$$

Statistical physics

- $\boldsymbol{y}$: Quenched randomness
- $\boldsymbol{x}$: Dynamical variables

Posterior distribution

Bayesian inference

- $\boldsymbol{y}$: Data
- $\boldsymbol{x}$: Model parameter

Statistical physics

- $\boldsymbol{y}$: Quenched randomness
- $\boldsymbol{x}$: Dynamical variables

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x})\phi(\boldsymbol{x})$$

Prior distribution

$\exp(-\beta h)$

External field

Posterior distribution

Bayesian inference

- $\boldsymbol{y}$: Data
- $\boldsymbol{x}$: Model parameter

Statistical physics

- $\boldsymbol{y}$: Quenched randomness
- $\boldsymbol{x}$: Dynamical variables

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{y})} \overbrace{\underbrace{f(\boldsymbol{y}|\boldsymbol{x})}_{\substack{\exp(-\beta H(\boldsymbol{x};\boldsymbol{y})) \\ \text{Hamiltonian}}}}^{\text{Likelihood}} \overbrace{\underbrace{\phi(\boldsymbol{x})}_{\substack{\exp(-\beta h) \\ \text{External field}}}}^{\text{Prior distribution}}$$

Posterior distribution

Bayesian inference

- $\boldsymbol{y}$: Data
- $\boldsymbol{x}$: Model parameter

Statistical physics

- $\boldsymbol{y}$: Quenched randomness
- $\boldsymbol{x}$: Dynamical variables

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x})\phi(\boldsymbol{x})$$

Probability distribution of data

Likelihood

Prior distribution

Partition function

$\exp(-\beta H(\boldsymbol{x};\boldsymbol{y}))$

Hamiltonian

$\exp(-\beta h)$

External field

Posterior distribution

Bayesian inference

- $\boldsymbol{y}$: Data
- $\boldsymbol{x}$: Model parameter

Statistical physics

- $\boldsymbol{y}$: Quenched randomness
- $\boldsymbol{x}$: Dynamical variables

Probability distribution of data

Likelihood

Prior distribution

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x})\phi(\boldsymbol{x})$$

Partition function

$\exp(-\beta H(\boldsymbol{x};\boldsymbol{y}))$

Hamiltonian

$\exp(-\beta h)$

External field

- Point estimates = Thermal average
  - Posterior mean: $\langle x_i \rangle = \sum_{\boldsymbol{x}} x_i \, P(\boldsymbol{x}|\boldsymbol{y})$
  - Maximum a posteriori estimator : $\hat{x}_i = \max_{x_i} P(\boldsymbol{x}|\boldsymbol{y})$ … Ground state

Statistical physics as Bayesian inference

- References
  - Mézard, Parisi, Virasoro, “Spin-glass theory and beyond” (1987)
  - Iba, “The Nishimori Line and Bayesian statistics”, J Phys A (1999)
  - Nishimori, “Statistical Physics of Spin Glasses and Information Processing: An Introduction” (2001)
    - In Japanese, 西森「スピングラスと情報統計力学」(1999)
- Applications
  - Coding theory
  - Learning theory
  - Computational science
  - Signal processing
  - Statistics

Statistical physics as Bayesian inference

- References
  - Mézard, Parisi, Virasoro, “Spin-glass theory and beyond” (1987)
  - Iba, “The Nishimori Line and Bayesian statistics”, J Phys A (1999)
  - Nishimori, “Statistical Physics of Spin Glasses and Information Processing: An Introduction” (2001)
    - In Japanese, 西森「スピングラスと情報統計力学」(1999)

- Applications
  - Coding theory
  - Learning theory
  - Computational science
  - Signal processing
  - Statistics

Statistical-physics-based studies

- Phase transition in learning/inference
- Development of algorithms
- Analysis for algorithms

Sparse estimation

Sparse estimation

- Assumption: The model parameter is sparse.

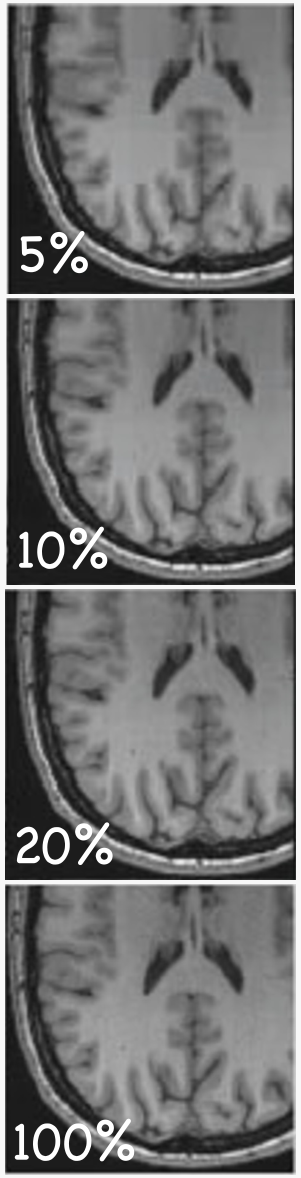
Sparse estimation

- Assumption: The model parameter is sparse.
  - Sparse: There exist many zero components.

Sparse estimation

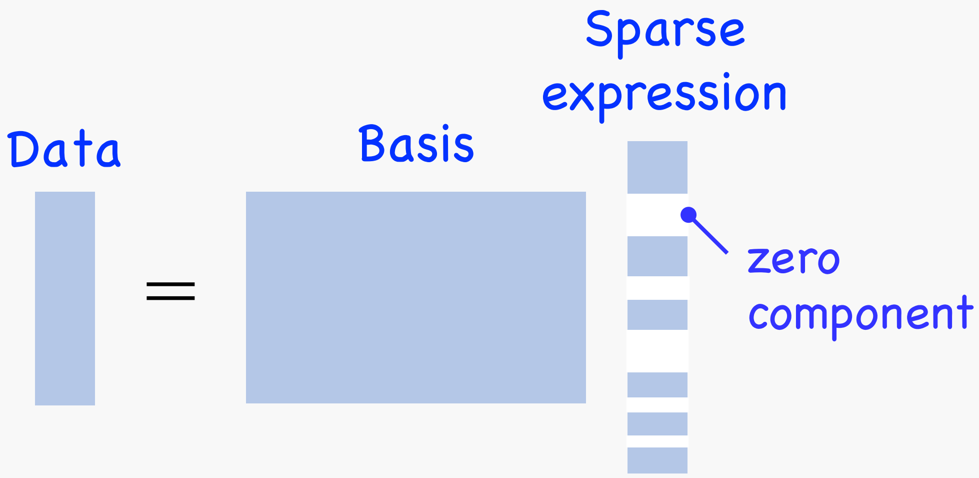
- Assumption: The model parameter is sparse.
  - Sparse: There exist many zero components.
    - The assumption is based on the empirical knowledge that natural data is sparse on some basis.

Reconstructed images using $x\%$-top components on the discrete cosine basis [Lustig et al. (2007)]

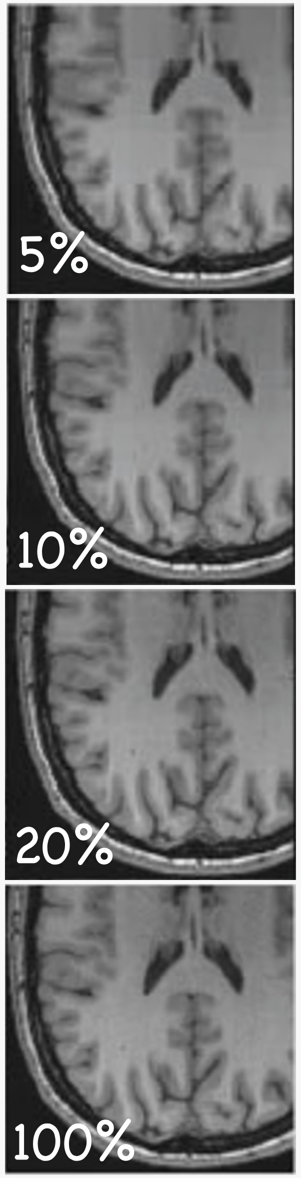


Sparse estimation

- Assumption: The model parameter is sparse.
  - Sparse: There exist many zero components.
    - The assumption is based on the empirical knowledge that natural data is sparse on some basis.
  - Sparsity assumption reduces the effective dimension of variables to be estimated
    - Compressed sensing (signal processing), LASSO (statistics)



Reconstructed images using $x\%$-top components on the discrete cosine basis [Lustig et al. (2007)]



Group Testing

A sparse estimation problem for discrete variables

Group testing

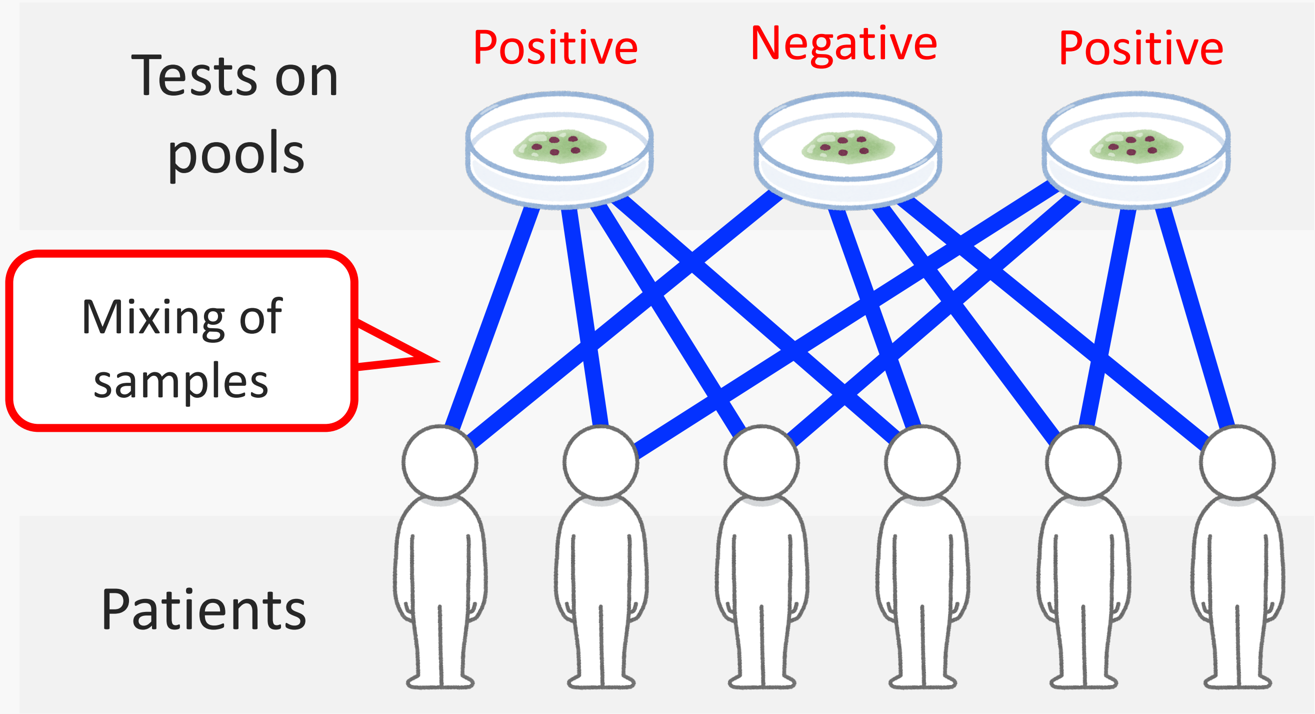
Perform tests on pools consist of mixed samples to

- Reduce the number of tests
- Correct test errors

Group testing

Perform tests on pools consist of mixed samples to

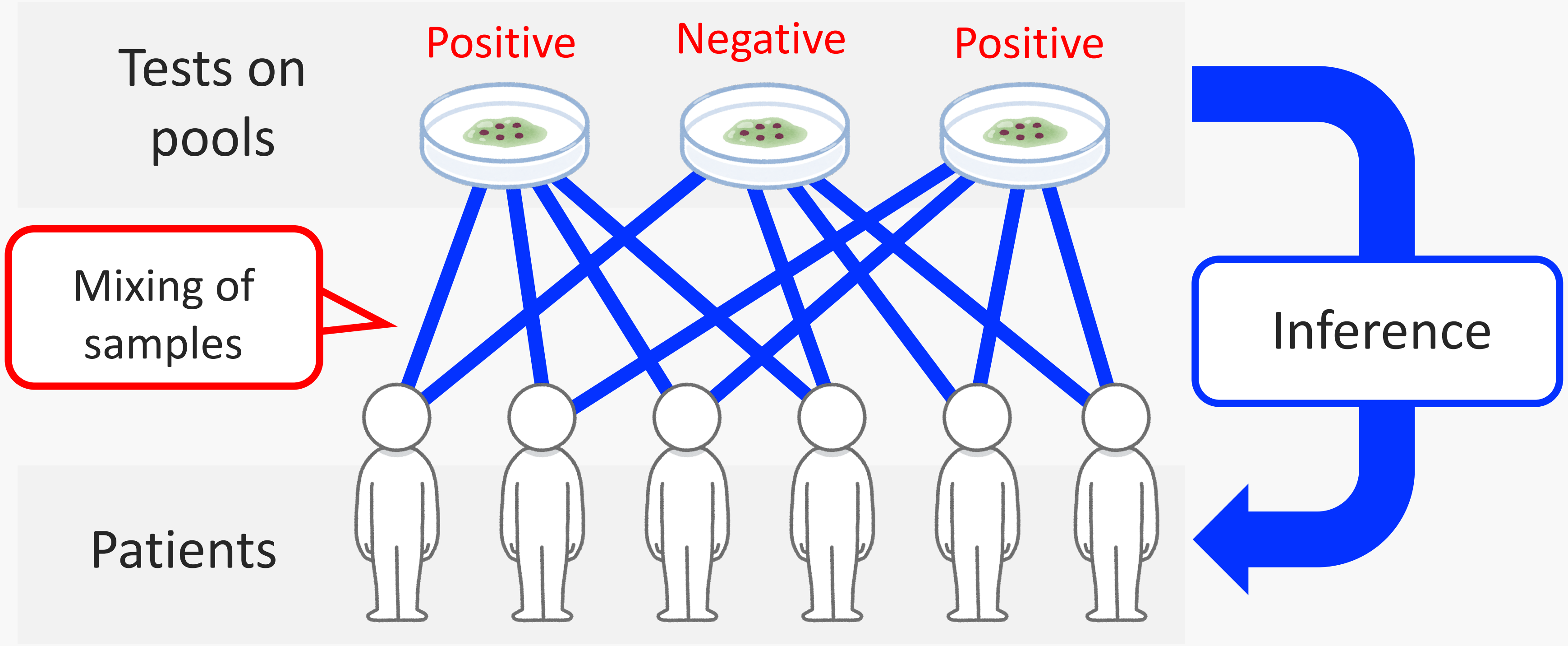
- Reduce the number of tests
- Correct test errors



Group testing

Perform tests on pools consist of mixed samples to

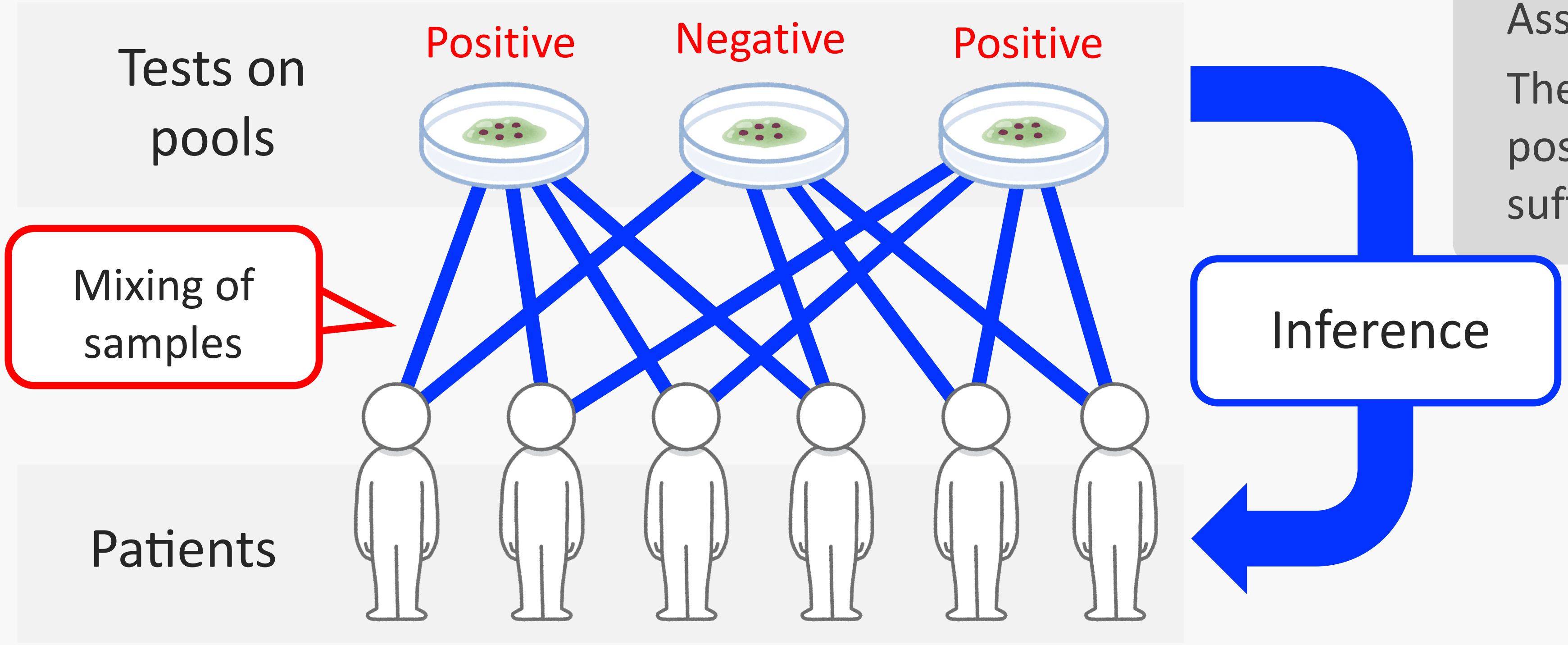
- Reduce the number of tests
- Correct test errors



Group testing

Perform tests on pools consist of mixed samples to

- Reduce the number of tests
- Correct test errors



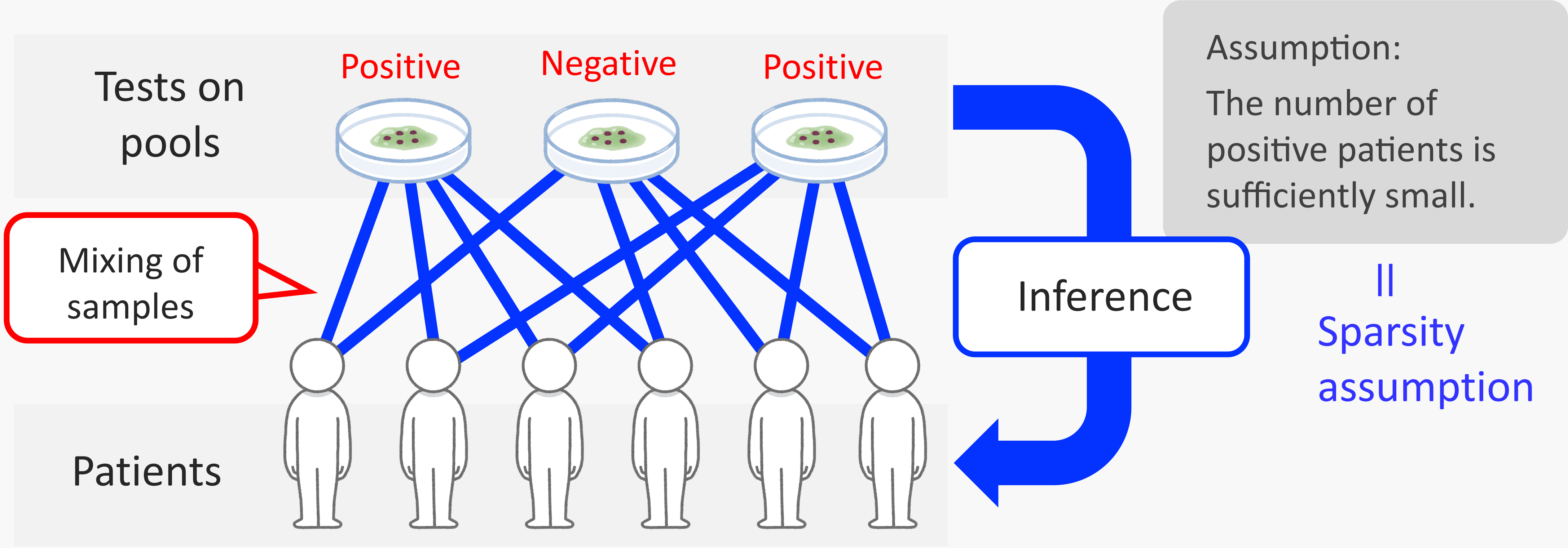
Assumption:

The number of positive patients is sufficiently small.

Group testing

Perform tests on pools consist of mixed samples to

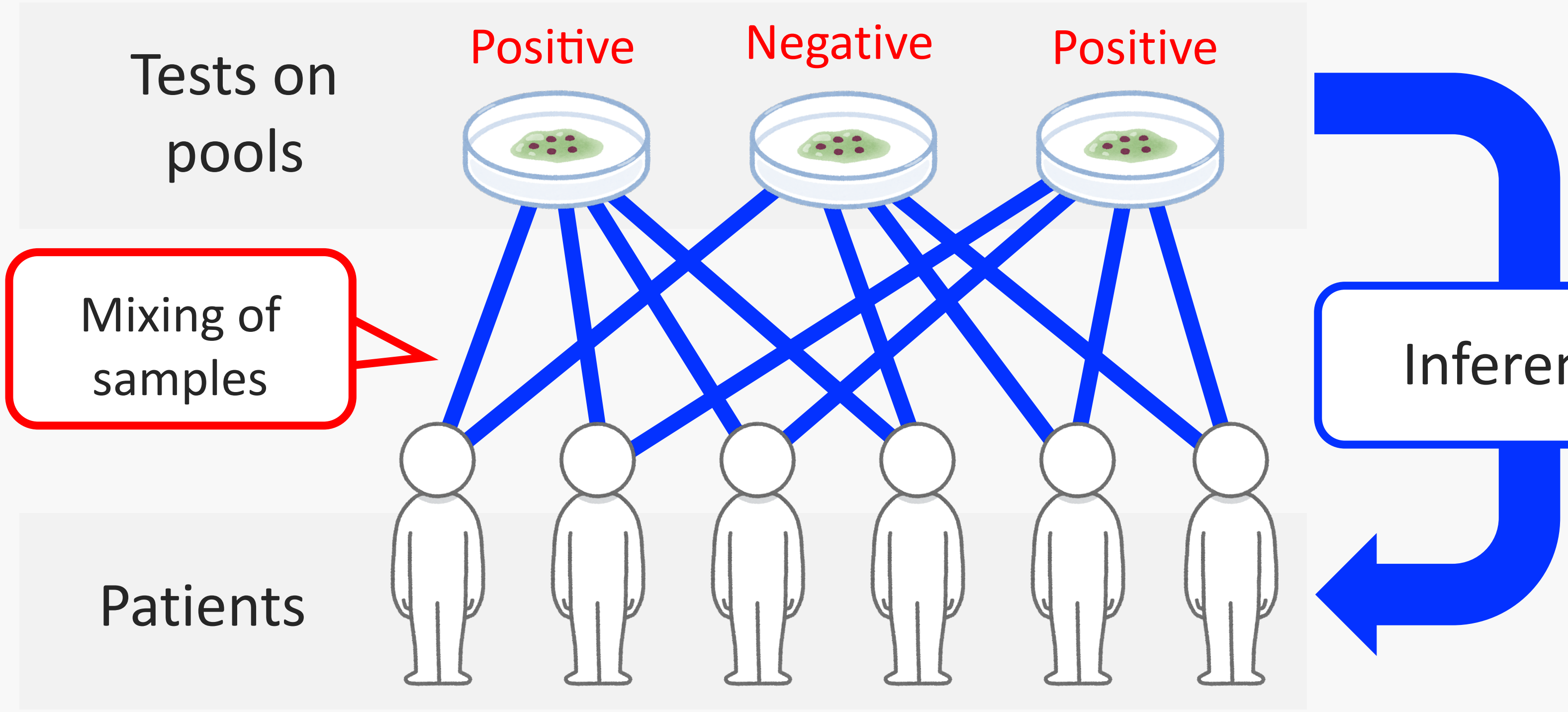
- Reduce the number of tests
- Correct test errors



Group testing

Perform tests on pools consist of mixed samples to

- Reduce the number of tests
- Correct test errors



Assumption:
The number of positive patients is sufficiently small.

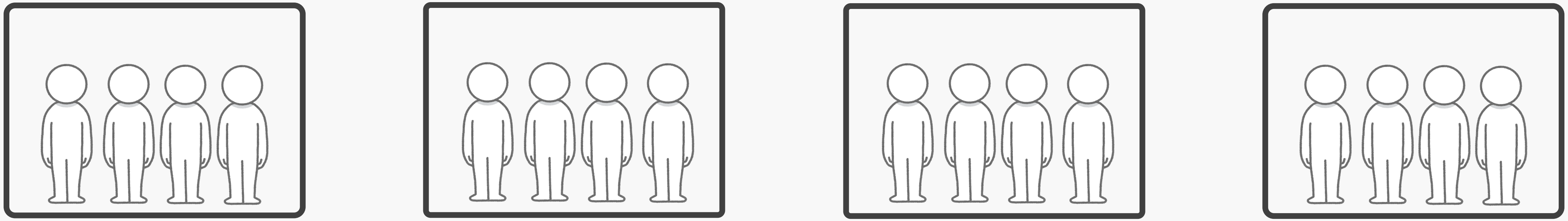
||
Sparsity assumption

Aldridge et al., "Group Testing: An Information Theory Perspective" (2019)

Deterministic approach: Two-stage testing

Dorfman, Ann. Math. Statist. (1943)

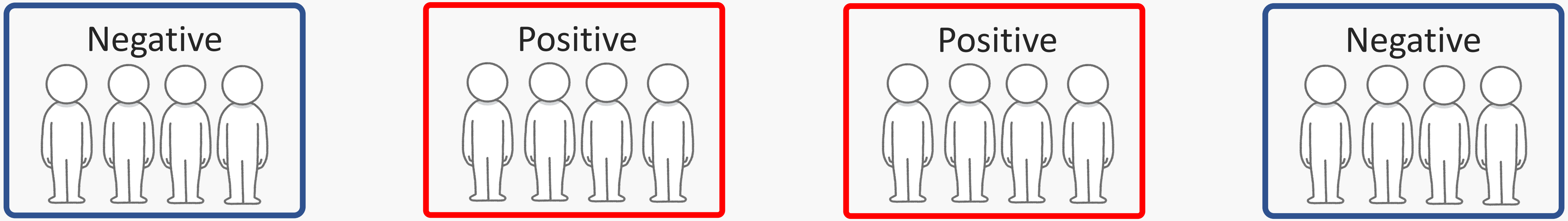
Deterministic approach: Two-stage testing



- First stage: Perform tests on random pools

Dorfman, Ann. Math. Statist. (1943)

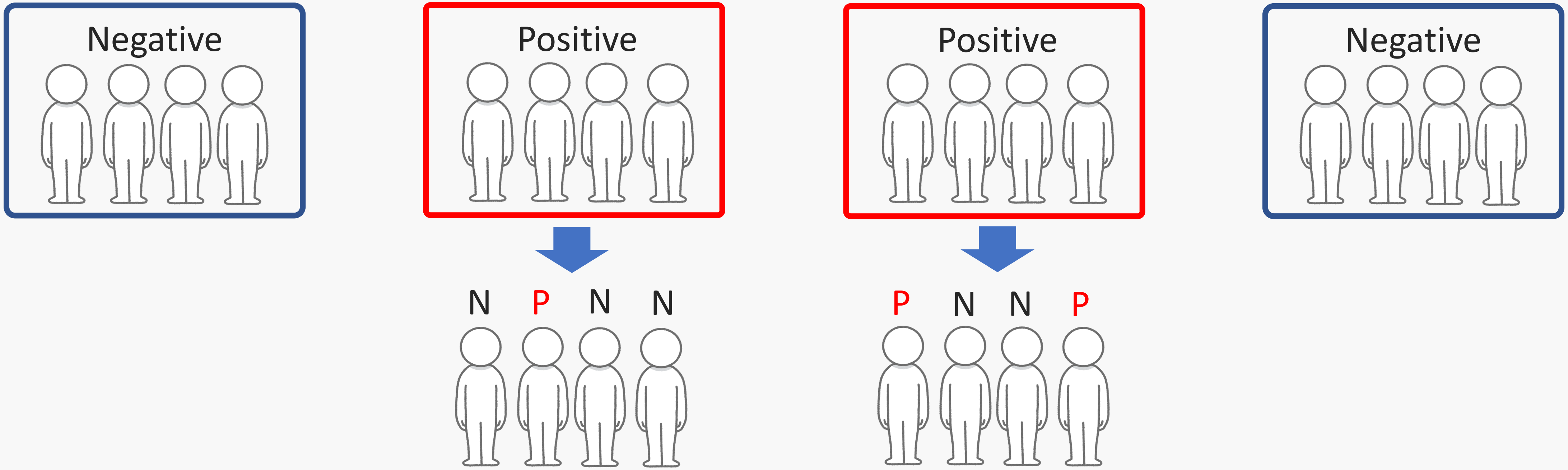
Deterministic approach: Two-stage testing



- First stage: Perform tests on random pools

Dorfman, Ann. Math. Statist. (1943)

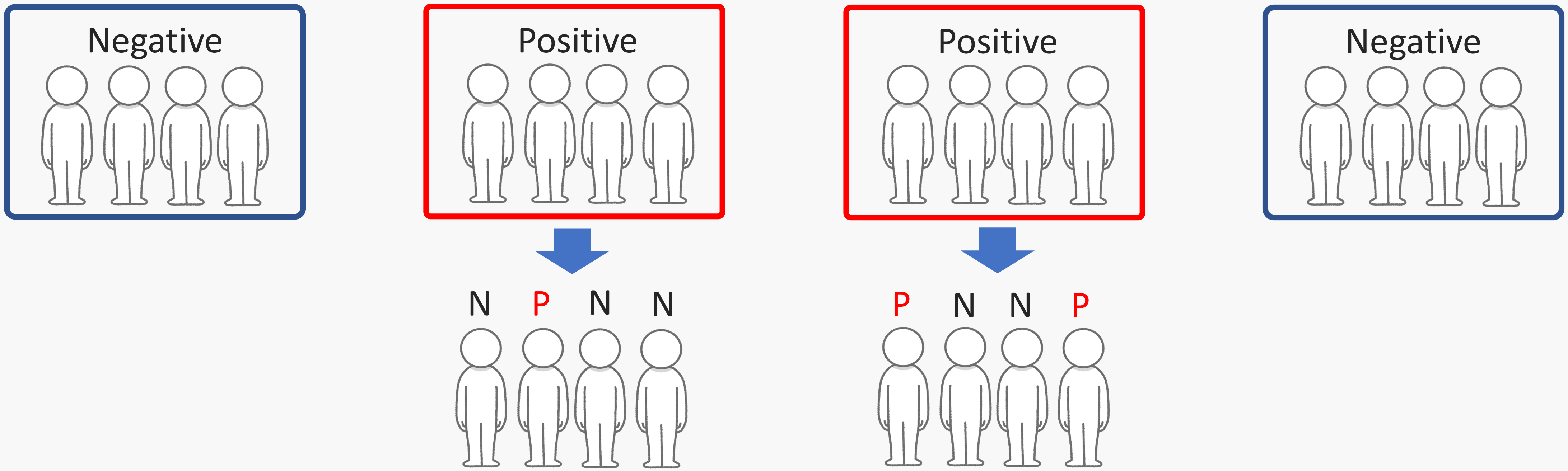
Deterministic approach: Two-stage testing



- First stage: Perform tests on random pools
- Second stage: Test patients in positive pools

Dorfman, Ann. Math. Statist. (1943)

Deterministic approach: Two-stage testing



- First stage: Perform tests on random pools
- Second stage: Test patients in positive pools
- Expected number of tests: $M + \{1 - (1-\theta)^{N_p}\} M N_p$ (at minimum ~ $2\sqrt{\theta}N$)
  - M: Number of pools , θ: Prevalence (fraction of positive patients), N_p: Pool size
  - N: Number of patients ($N = N_p M$)

Dorfman, Ann. Math. Statist. (1943)

Weakness of deterministic approaches

Weakness of deterministic approaches

- Not robust to test errors
  - In deterministic approaches, the possibility of false test results is not considered.

Weakness of deterministic approaches

- Not robust to test errors
  - In deterministic approaches, the possibility of false test results is not considered.

- In practice, testing cannot be error-free.

Weakness of deterministic approaches

- Not robust to test errors
  - In deterministic approaches, the possibility of false test results is not considered.

- In practice, testing cannot be error-free.
  - Pre-analytical error: inappropriate sample transport, insufficient number of samples
  - Analytical error: calibration error, systematic error, random error
  - Post-analytical error: incorrect calculation

[Teshome et al., Journal of Multidisciplinary Healthcare (2020)]

Weakness of deterministic approaches

● Not robust to test errors

- In deterministic approaches, the possibility of false test results is not considered.

■ In practice, testing cannot be error-free.

- Pre-analytical error: inappropriate sample transport, insufficient number of samples
- Analytical error: calibration error, systematic error, random error
- Post-analytical error: incorrect calculation

[Teshome et al., Journal of Multidisciplinary Healthcare (2020)]

◆ Modeling of test output considering both patients‘ states and test errors will be useful for robust inference.

Bayesian Inference for Group Testing

Modeling of outputs of tests performed on pools

Basic idea

Basic idea

- Consider the test result $\boldsymbol{y} \in \{0,1\}^M$ as observed data

Basic idea

- Consider the test result $\boldsymbol{y} \in \{0,1\}^M$ as observed data
  - y_μ: Test results of μ-th pool (1 = positive, 0 = negative)

Basic idea

- Consider the test result $\boldsymbol{y} \in \{0,1\}^M$ as observed data
  - y_μ: Test results of μ-th pool (1 = positive, 0 = negative)

  → The state of patients corresponds to the underlying model parameter $\boldsymbol{x}^{(0)} \in \{0,1\}^N$.
    - $x_i^{(0)}$: True state of i-th patient (1 = positive, 0 = negative)

Basic idea

- Consider the test result $\boldsymbol{y} \in \{0,1\}^M$ as observed data
  - y_μ: Test results of μ-th pool (1 = positive, 0 = negative)

  → The state of patients corresponds to the underlying model parameter $\boldsymbol{x}^{(0)} \in \{0,1\}^N$.
    - $x_i^{(0)}$: True state of i-th patient (1 = positive, 0 = negative)

- Infer $\boldsymbol{x}$ from the test results $\boldsymbol{y}$

  = Model the generative process of $\boldsymbol{y}$ using $\boldsymbol{x}$ and set appropriate $\boldsymbol{x}$

Basic idea

- Consider the test result $\boldsymbol{y} \in \{0,1\}^M$ as observed data
  - y_μ: Test results of μ-th pool (1 = positive, 0 = negative)

  → The state of patients corresponds to the underlying model parameter $\boldsymbol{x}^{(0)} \in \{0,1\}^N$.

    - $x_i^{(0)}$: True state of i-th patient (1 = positive, 0 = negative)

- Infer $\boldsymbol{x}$ from the test results $\boldsymbol{y}$

  = Model the generative process of $\boldsymbol{y}$ using $\boldsymbol{x}$ and set appropriate $\boldsymbol{x}$

  - The pooling matrix $\boldsymbol{F} \in \{0,1\}^{M\times N}$ is known.
    - $F_{\mu i} = 1$: i-th patient is in μ-th pool.
    - $F_{\mu i} = 0$: i-th patient is not in μ-th pool.

Basic idea

- Consider the test result $\boldsymbol{y} \in \{0,1\}^M$ as observed data
  - y_μ: Test results of μ-th pool (1 = positive, 0 = negative)

  → The state of patients corresponds to the underlying model parameter $\boldsymbol{x}^{(0)} \in \{0,1\}^N$.
    - $x_i^{(0)}$: True state of i-th patient (1 = positive, 0 = negative)

- Infer $\boldsymbol{x}$ from the test results $\boldsymbol{y}$

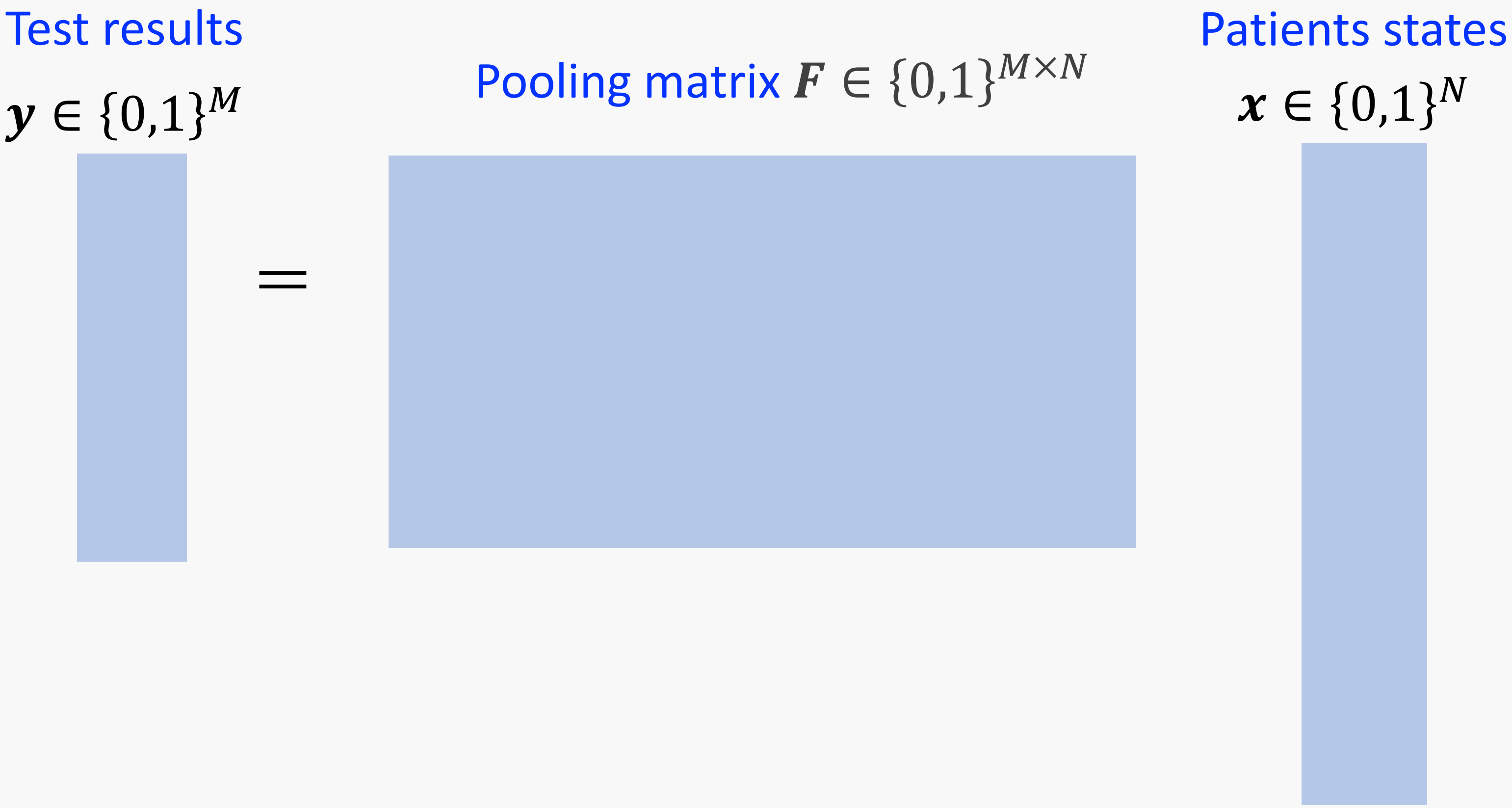
  = Model the generative process of $\boldsymbol{y}$ using $\boldsymbol{x}$ and set appropriate $\boldsymbol{x}$
  - The pooling matrix $\boldsymbol{F} \in \{0,1\}^{M\times N}$ is known.
    - $F_{\mu i} = 1$: i-th patient is in μ-th pool.
    - $F_{\mu i} = 0$: i-th patient is not in μ-th pool.

There are two kinds of group testing.

- Non-adaptive GT: $\boldsymbol{F}$ is fixed in advance.
- Adaptive GT: $\boldsymbol{F}$ is sequentially designed.

Matrix representation

Matrix representation



Matrix representation

Test results

$\boldsymbol{y} \in \{0,1\}^M$

$=$

Pooling matrix $\boldsymbol{F} \in \{0,1\}^{M \times N}$

Patients states

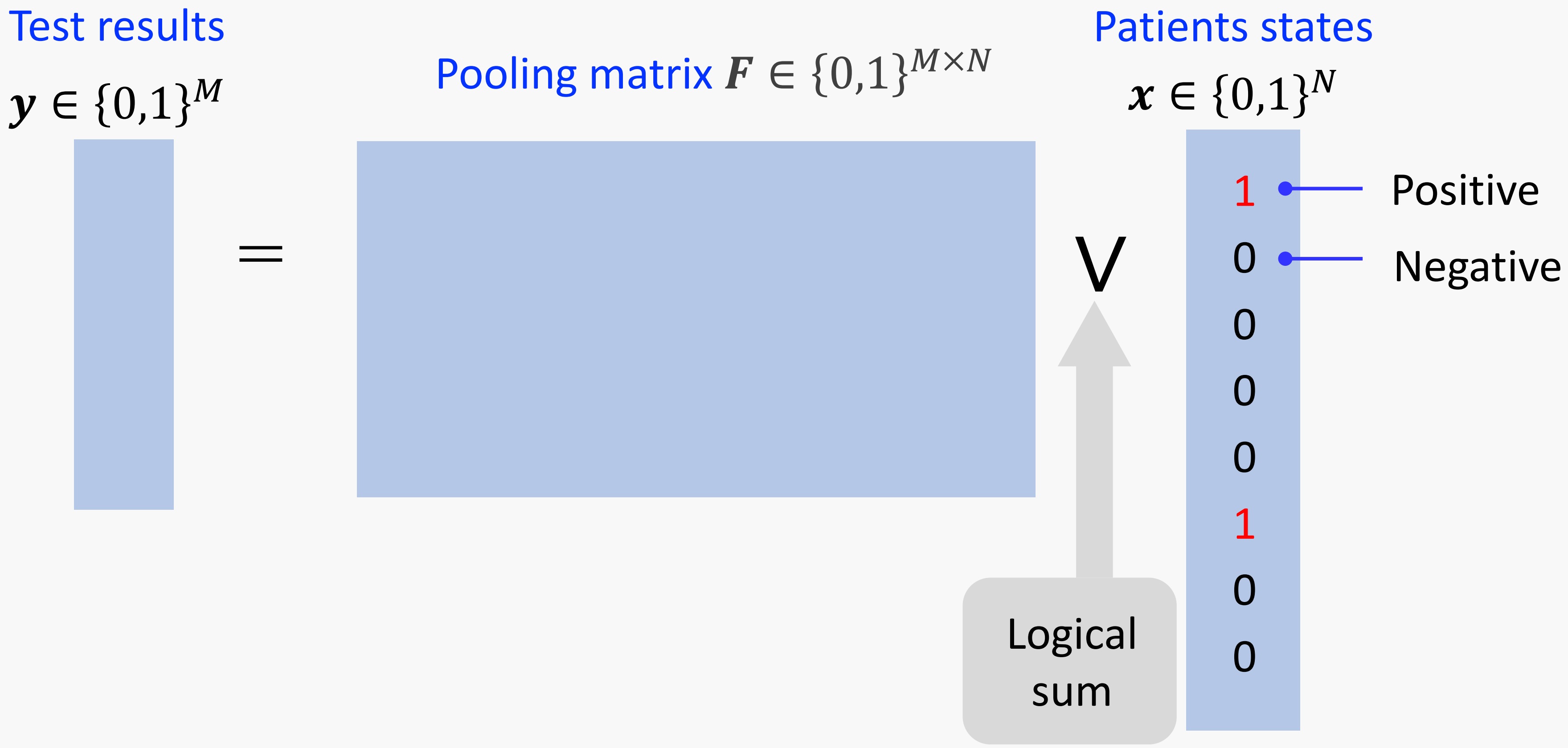
$\boldsymbol{x} \in \{0,1\}^N$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Positive

Negative

Matrix representation



Matrix representation

Test results

$\boldsymbol{y} \in \{0,1\}^M$

Pooling matrix $\boldsymbol{F} \in \{0,1\}^{M \times N}$

Patients states

$\boldsymbol{x} \in \{0,1\}^N$

$$\begin{bmatrix} 0 \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{bmatrix} \vee \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Positive

Negative

Logical sum

Matrix representation

Test results

$\boldsymbol{y} \in \{0,1\}^M$

Pooling matrix $\boldsymbol{F} \in \{0,1\}^{M \times N}$

Patients states

$\boldsymbol{x} \in \{0,1\}^N$

$$
\begin{bmatrix} 0 \\ \\ \\ \\ \end{bmatrix}
=
\begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{bmatrix}
\vee
\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}
$$

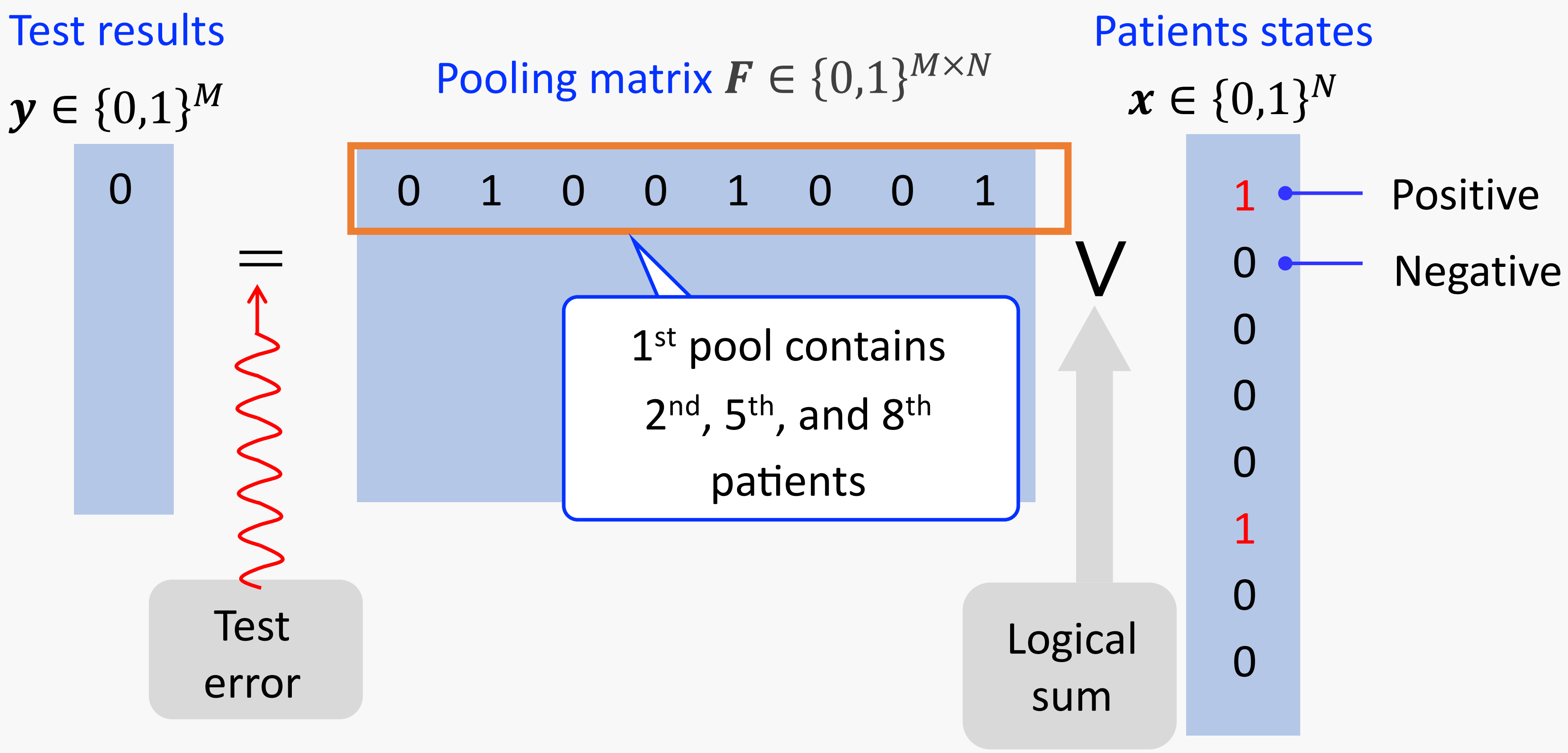
1st pool contains 2nd, 5th, and 8th patients

Logical sum

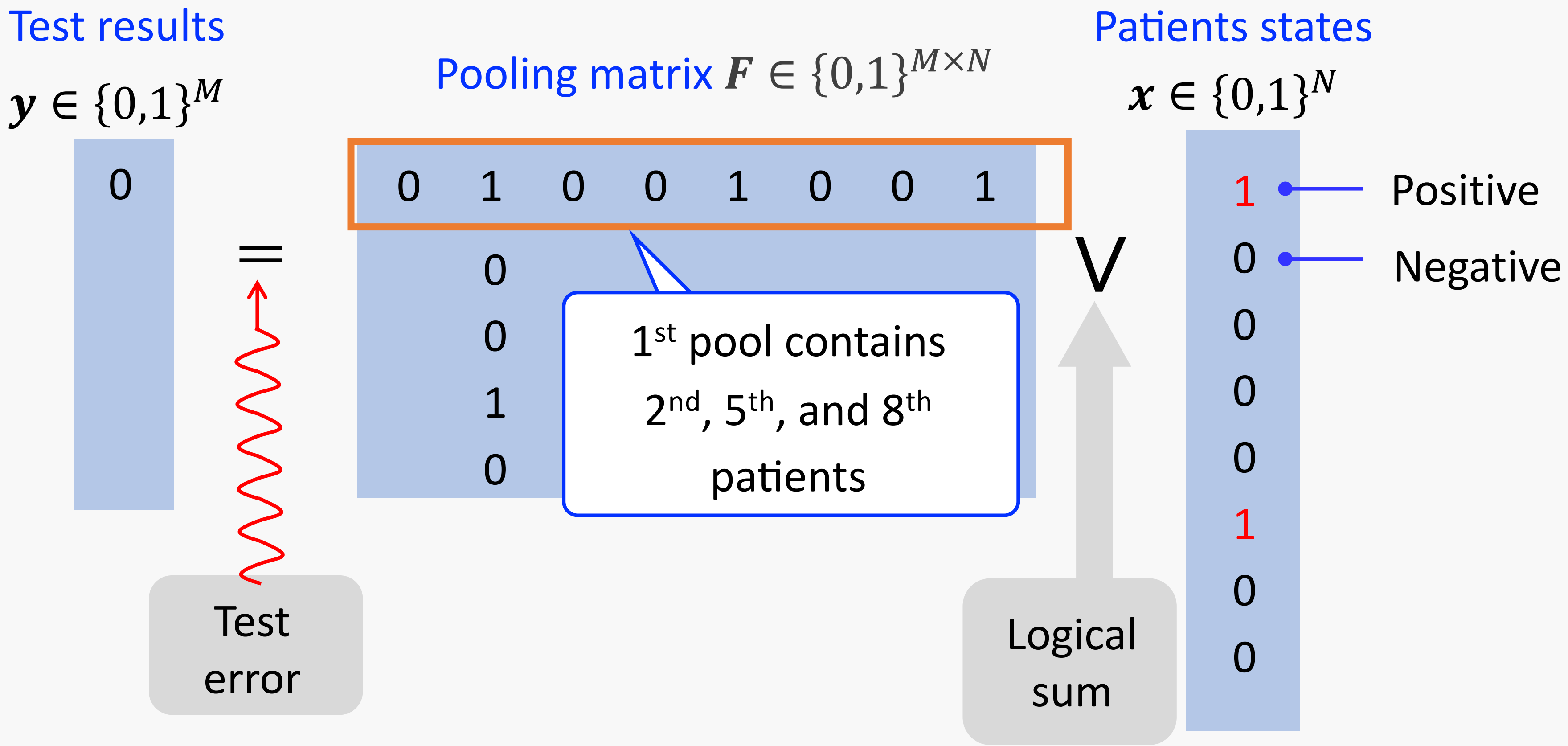
1 — Positive

0 — Negative

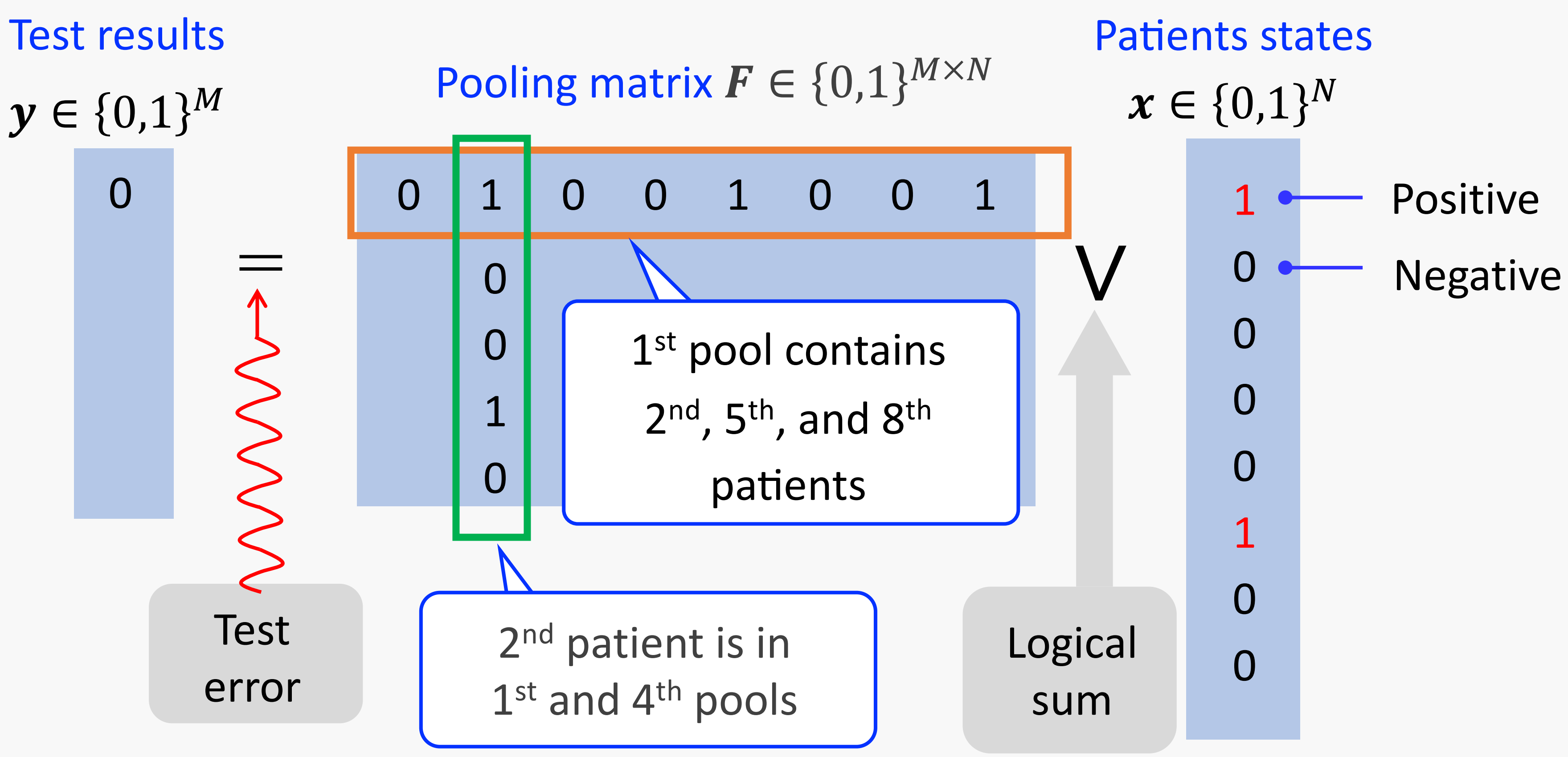
Matrix representation



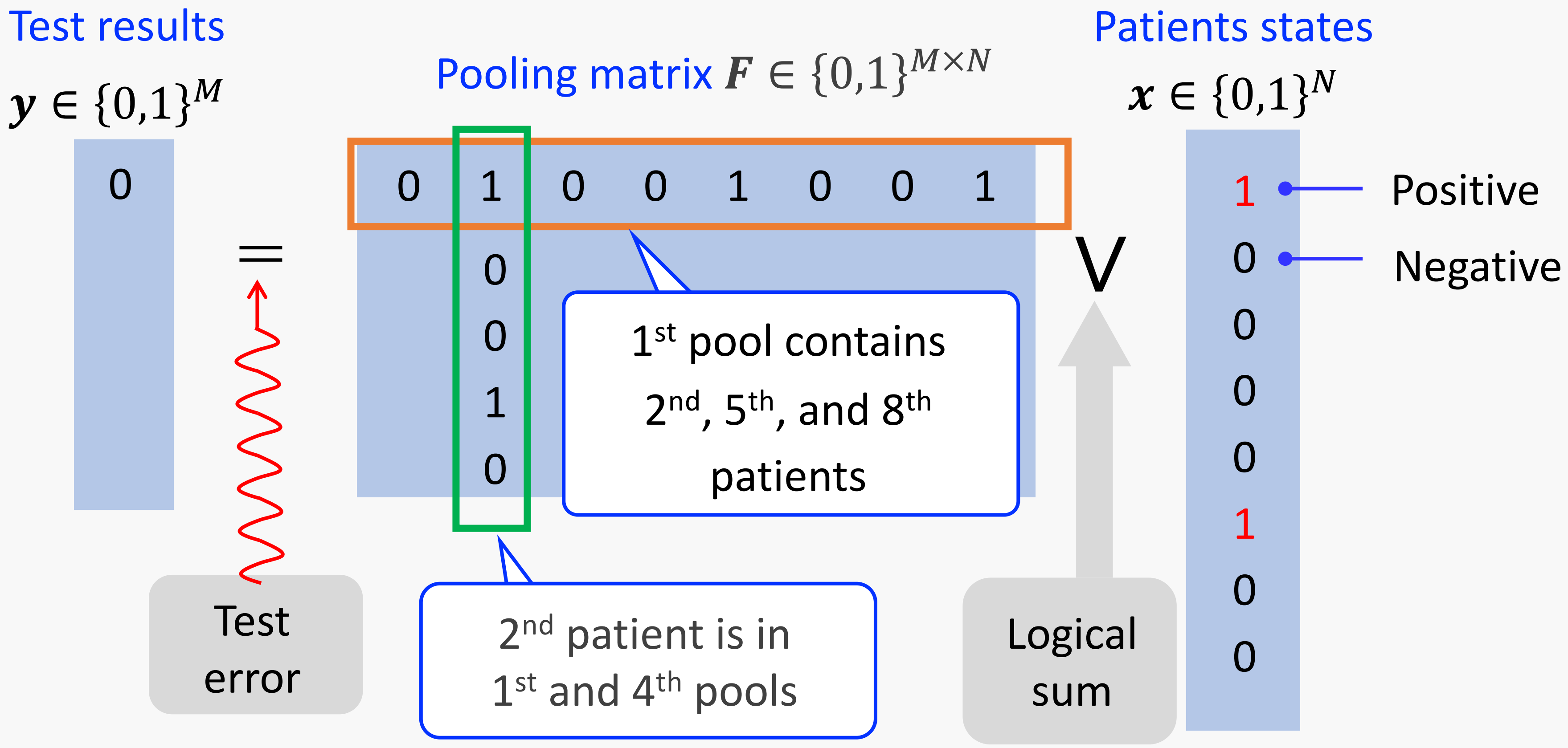
Matrix representation



Matrix representation

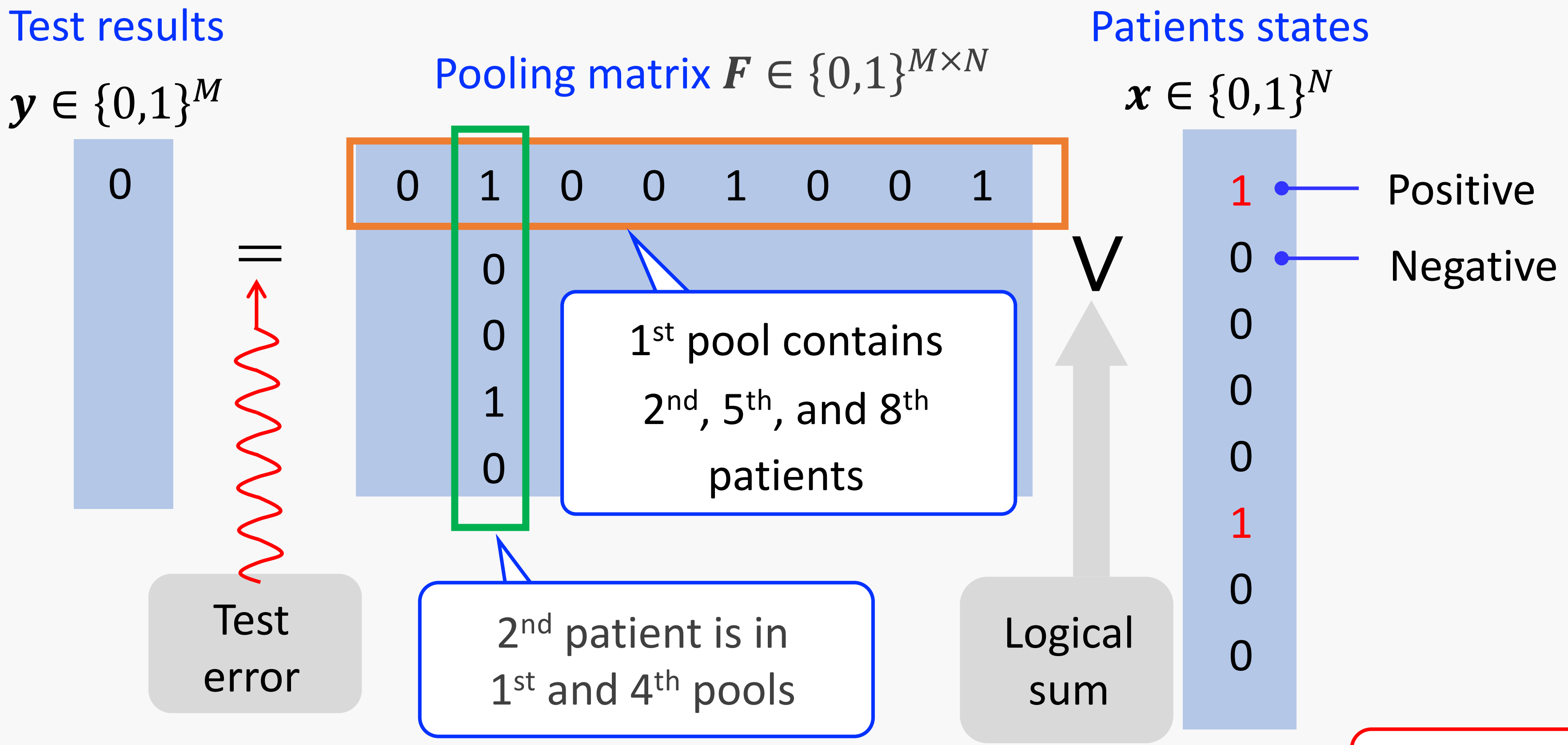


Matrix representation



- Number of tests (pools) M is smaller than number of patients N.
- Identification of $\boldsymbol{x}$ from $\boldsymbol{y}$ is an underdetermined problem.
  - Identification can be achieved if prevalence θ is sufficiently small.

Matrix representation



- Number of tests (pools) M is smaller than number of patients N.
- Identification of $\boldsymbol{x}$ from $\boldsymbol{y}$ is an underdetermined problem.
  - Identification can be achieved if prevalence θ is sufficiently small.

Important model parameters

- M/N
- θ

Modeling of test output

Modeling of test output

- Assumed generative process for test results $\boldsymbol{y}$ (likelihood)

$$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F}) = \prod_{\nu=1}^{M}\Big[\{p_{\mathrm{TP}}y_\nu + (1-p_{\mathrm{TP}})(1-y_\nu)\}T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) + \\ + \{p_{\mathrm{FP}}y_\nu + (1-p_{\mathrm{FP}})(1-y_\nu)\}\Big(1-T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big)\Big)\Big]$$

- $\widetilde{\boldsymbol{F}}_\nu$: ν-th row vector of $\boldsymbol{F}$

Modeling of test output

- Assumed generative process for test results $\boldsymbol{y}$ (likelihood)

$$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F}) = \prod_{\nu=1}^{M}\Big[\{p_{\mathrm{TP}}y_\nu + (1-p_{\mathrm{TP}})(1-y_\nu)\}T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) + \\ + \{p_{\mathrm{FP}}y_\nu + (1-p_{\mathrm{FP}})(1-y_\nu)\}\Big(1 - T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big)\Big)\Big]$$

- $\widetilde{\boldsymbol{F}}_\nu$: ν-th row vector of $\boldsymbol{F}$

  - $T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) = \vee_i F_{\nu i}x_i$ ：True state of ν-th pool ($\vee$: logical sum)
    - Pools with at least one positive patient are positive

Modeling of test output

- Assumed generative process for test results $\boldsymbol{y}$ (likelihood)

$$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F}) = \prod_{\nu=1}^{M}\Big[\{p_{\mathrm{TP}}y_\nu + (1-p_{\mathrm{TP}})(1-y_\nu)\}T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) + \\ + \{p_{\mathrm{FP}}y_\nu + (1-p_{\mathrm{FP}})(1-y_\nu)\}\Big(1-T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big)\Big)\Big]$$

- $\widetilde{\boldsymbol{F}}_\nu$: ν-th row vector of $\boldsymbol{F}$

  - $T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) = \vee_i F_{\nu i}x_i$ ：True state of ν-th pool ($\vee$: logical sum)
    - Pools with at least one positive patient are positive
  - Parameters
    - p_{TP} ：True positive probability
    - p_{FP} ：False positive probability

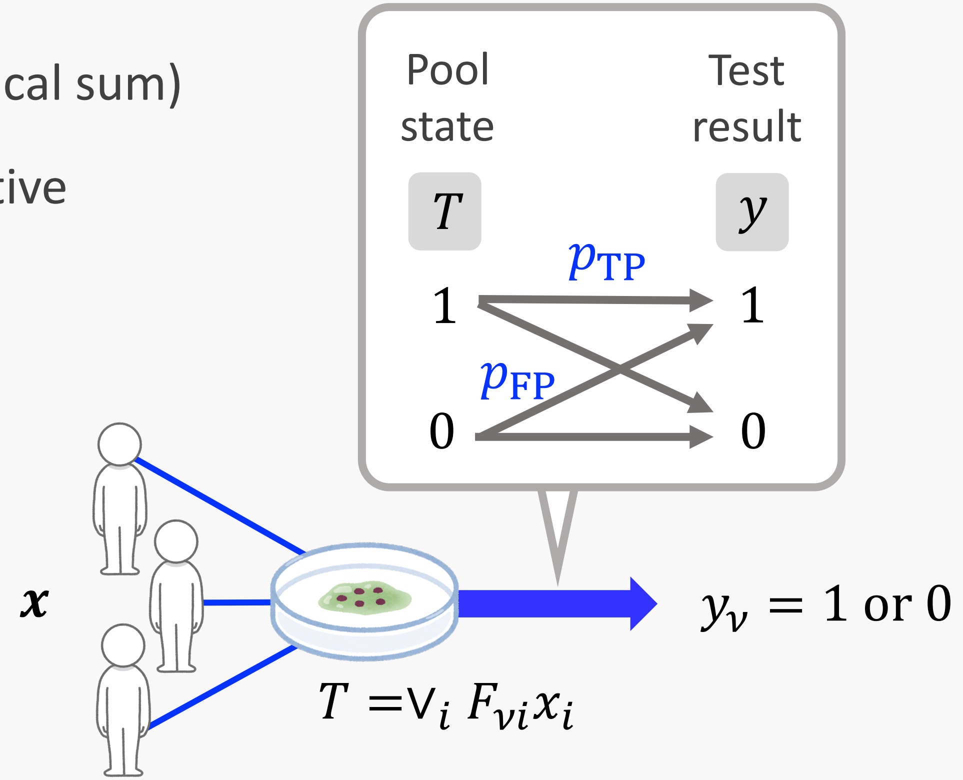
Modeling of test output

- Assumed generative process for test results $\boldsymbol{y}$ (likelihood)

$$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F}) = \prod_{\nu=1}^{M}\Big[\{p_{\mathrm{TP}}y_\nu + (1-p_{\mathrm{TP}})(1-y_\nu)\}T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) + \\ + \{p_{\mathrm{FP}}y_\nu + (1-p_{\mathrm{FP}})(1-y_\nu)\}\Big(1-T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big)\Big)\Big]$$

  - $T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) = \vee_i F_{\nu i}x_i$: True state of ν-th pool ($\vee$: logical sum)
    - Pools with at least one positive patient are positive
  - Parameters
    - p_{TP} : True positive probability
    - p_{FP} : False positive probability

- $\widetilde{\boldsymbol{F}}_\nu$: ν-th row vector of $\boldsymbol{F}$



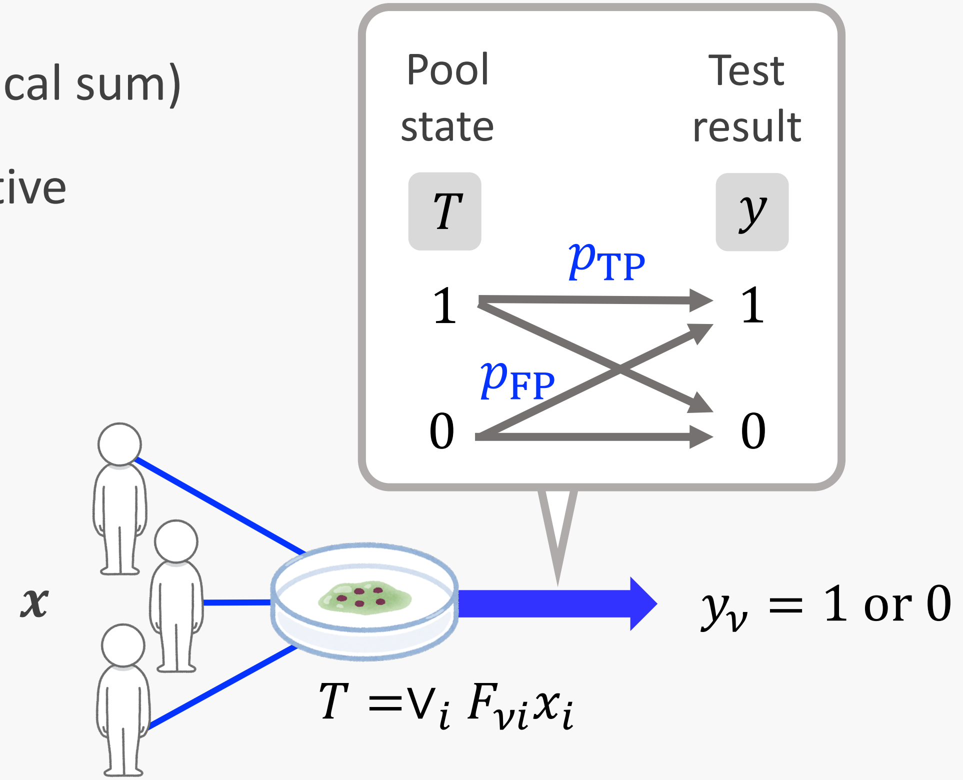
Modeling of test output

- Assumed generative process for test results $\boldsymbol{y}$ (likelihood)

$$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F}) = \prod_{\nu=1}^{M}\Big[\{p_{\mathrm{TP}}y_\nu + (1-p_{\mathrm{TP}})(1-y_\nu)\}T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) + \{p_{\mathrm{FP}}y_\nu + (1-p_{\mathrm{FP}})(1-y_\nu)\}\Big(1-T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big)\Big)\Big]$$

- $\widetilde{\boldsymbol{F}}_\nu$: ν-th row vector of $\boldsymbol{F}$

  - $T\big(\boldsymbol{x},\widetilde{\boldsymbol{F}}_\nu\big) = \vee_i F_{\nu i}x_i$: True state of ν-th pool ($\vee$: logical sum)
    - Pools with at least one positive patient are positive
  - Parameters
    - p_{TP} : True positive probability
    - p_{FP} : False positive probability
  - Assumption
    - Tests are independent.



Posterior distribution

Posterior distribution

- Prior distribution

$$\phi(\boldsymbol{x}) = \prod_{i=1}^{N} \{(1-\theta)(1-x_i) + \theta x_i\}, \qquad \theta \text{ : Prevalence}$$

Posterior distribution

- Prior distribution

$$\phi(\boldsymbol{x}) = \prod_{i=1}^{N} \{(1-\theta)(1-x_i) + \theta x_i\}, \qquad \theta \text{ : Prevalence}$$

- Assumptions
  - Independence between patients
  - Common prior probability

Posterior distribution

- Prior distribution

$$\phi(\boldsymbol{x}) = \prod_{i=1}^{N} \{(1-\theta)(1-x_i) + \theta x_i\}, \qquad \theta \text{ : Prevalence}$$

  - Assumptions
    - Independence between patients
    - Common prior probability

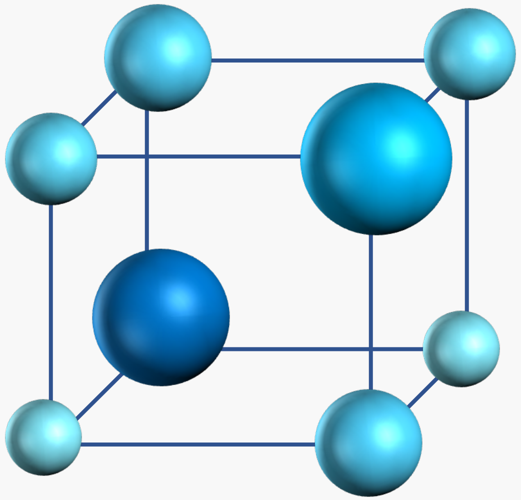
- Posterior distribution

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x}), \qquad \text{where} \quad Z(\boldsymbol{y}) = \sum_{\boldsymbol{x}} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$

Problem caused by Bayesian inference

Problem caused by Bayesian inference

- Basis of inference is the posterior distribution.

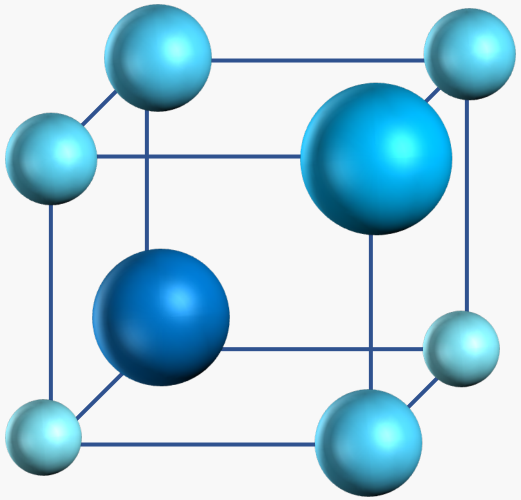


Posterior distribution
on $\{0,1\}^N$

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$

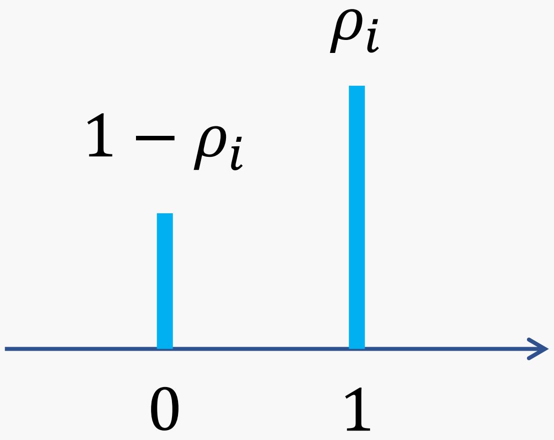
Problem caused by Bayesian inference

- Basis of inference is the posterior distribution.
  - ✦ Here, we focus on the posterior marginal probability $\rho_i \in [0,1]$.



Posterior distribution on $\{0,1\}^N$

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$



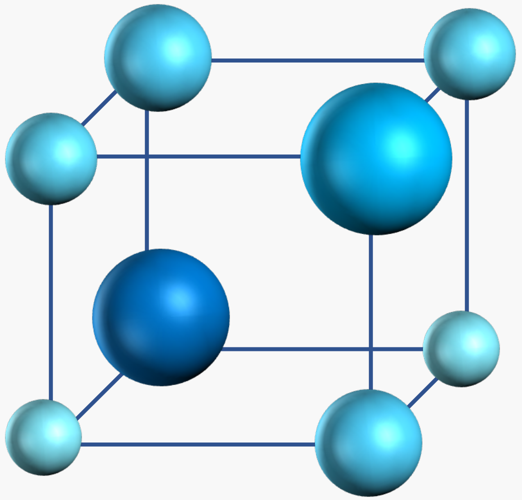
Posterior marginal probability for each patient

$$\rho_i = \frac{1}{Z(\boldsymbol{y})} \sum_{\boldsymbol{x}} x_i\, f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$

Problem caused by Bayesian inference

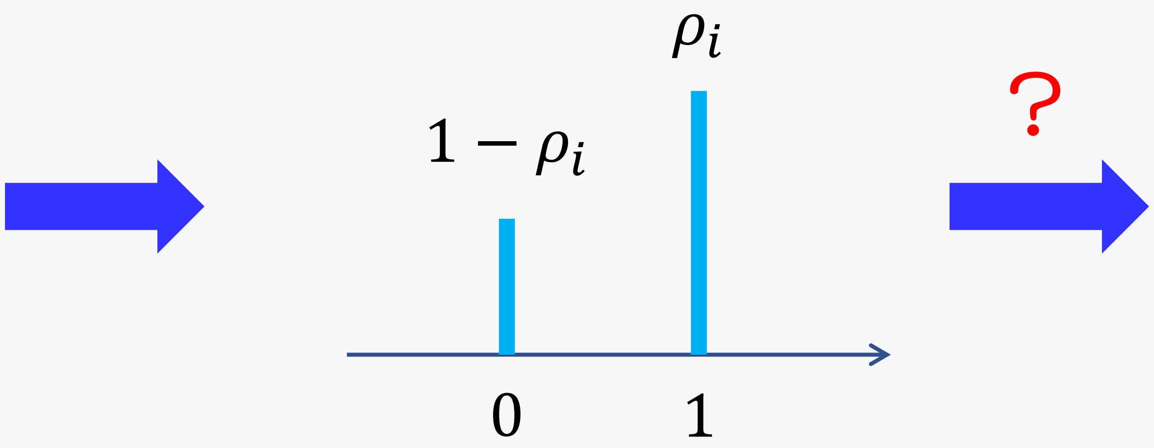
- Basis of inference is the posterior distribution.
  - ✦ Here, we focus on the posterior marginal probability $\rho_i \in [0,1]$.

We need a map from $\rho_i \in [0,1]$ to $\hat{x}_i \in \{0,1\}$ to determine patients' states.



Posterior distribution on $\{0,1\}^N$

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$



Posterior marginal probability for each patient

$$\rho_i = \frac{1}{Z(\boldsymbol{y})} \sum_{\boldsymbol{x}} x_i f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$

Determine patient's state $\hat{x}_i \in \{0,1\}$

Bayesian Statistical Decision for group testing

Maximization of expected utility

Sakata & Kabashima, arXiv:2110.10877

(submitted to IEEE Transaction on Information Theory)

Decision-making

Decision-making

- Action: a strategy that is being considered

Decision-making

- Action: a strategy that is being considered
  - ✦ In the Bayesian group testing problem:
    - The possible actions are to make diagnoses as $\hat{x}_i = 1$ or $\hat{x}_i = 0$.

Decision-making

- Action: a strategy that is being considered
    - ✦ In the Bayesian group testing problem:
        - The possible actions are to make diagnoses as $\hat{x}_i = 1$ or $\hat{x}_i = 0$.
        - We need to choose one of the actions based on the posterior marginal probability ρ_i..

Decision-making

- Action: a strategy that is being considered
  - ✦ In the Bayesian group testing problem:
    - The possible actions are to make diagnoses as $\hat{x}_i = 1$ or $\hat{x}_i = 0$.
    - We need to choose one of the actions based on the posterior marginal probability ρ_i.
    - In medical statistics, ρ_i corresponds to the diagnostic variable.

Decision-making

- Action: a strategy that is being considered
  - ✦ In the Bayesian group testing problem:
    - The possible actions are to make diagnoses as $\hat{x}_i = 1$ or $\hat{x}_i = 0$.
    - We need to choose one of the actions based on the posterior marginal probability ρ_i.
    - In medical statistics, ρ_i corresponds to the diagnostic variable.
- Utility: a reward for the action

Decision-making

- Action: a strategy that is being considered
  - ✦ In the Bayesian group testing problem:
    - The possible actions are to make diagnoses as $\hat{x}_i = 1$ or $\hat{x}_i = 0$.
    - We need to choose one of the actions based on the posterior marginal probability ρ_i.
    - In medical statistics, ρ_i corresponds to the diagnostic variable.
- Utility: a reward for the action
- Utility function: total reward considering the patients population

Decision-making

- Action: a strategy that is being considered
  - ✦ In the Bayesian group testing problem:
    - The possible actions are to make diagnoses as $\hat{x}_i = 1$ or $\hat{x}_i = 0$.
    - We need to choose one of the actions based on the posterior marginal probability ρ_i.
    - In medical statistics, ρ_i corresponds to the diagnostic variable.
- Utility: a reward for the action
- Utility function: total reward considering the patients population

Optimal action is defined as that which maximizes the expected utility.

Expectation is explained later.

Utility function = relative desirability of outcomes

Utility function = relative desirability of outcomes

Utilities for action on single patient

Target hypothesis (Patient's true state)

Action (Diagnosis)	Positive	Negative
Positive	u_{TP}	u_{FP}
Negative	u_{FN}	u_{TN}

Utility function = relative desirability of outcomes

Utilities for action on single patient

Target hypothesis (Patient's true state)

Action (Diagnosis)	Positive	Negative
Positive	u_{TP}	u_{FP}
Negative	u_{FN}	u_{TN}

We set

- $u_{\mathrm{TP}} > u_{\mathrm{FN}}$
- $u_{\mathrm{TN}} > u_{\mathrm{FP}}$

Utility function = relative desirability of outcomes

Utilities for action on single patient

Target hypothesis (Patient's true state)

Action (Diagnosis)	Positive	Negative
Positive	u_{TP}	u_{FP}
Negative	u_{FN}	u_{TN}

We set

- $u_{\mathrm{TP}} > u_{\mathrm{FN}}$
- $u_{\mathrm{TN}} > u_{\mathrm{FP}}$

- Definition: Utility function for patient population

$$U = u_{\mathrm{TP}}\mathrm{TP} + u_{\mathrm{FN}}\mathrm{FN} + u_{\mathrm{FP}}\mathrm{FP} + u_{\mathrm{TN}}\mathrm{TN}$$

Utility function = relative desirability of outcomes

Utilities for action on single patient

Target hypothesis (Patient's true state)

Action (Diagnosis)	Positive	Negative
Positive	u_{TP}	u_{FP}
Negative	u_{FN}	u_{TN}

We set
- $u_{\mathrm{TP}} > u_{\mathrm{FN}}$
- $u_{\mathrm{TN}} > u_{\mathrm{FP}}$

- Definition: Utility function for patient population

$$U = u_{\mathrm{TP}}\mathrm{TP} + u_{\mathrm{FN}}\mathrm{FN} + u_{\mathrm{FP}}\mathrm{FP} + u_{\mathrm{TN}}\mathrm{TN}$$

- TP, $\mathrm{FN} = 1 - \mathrm{TP}$, FP, and $\mathrm{TN} = 1 - \mathrm{FP}$ are functions of action $\hat{x}_i$ and true parameter $x_i^{(0)}$.

$$\mathrm{TP}\left(x_i^{(0)}, \hat{x}_i(\boldsymbol{y})\right) = \frac{1}{N\theta}\sum_{i=1}^{N} x_i^{(0)}\hat{x}_i(\boldsymbol{y}), \qquad \mathrm{FP}\left(x_i^{(0)}, \hat{x}_i(\boldsymbol{y})\right) = \frac{1}{N(1-\theta)}\sum_{i=1}^{N}\left(1 - x_i^{(0)}\right)\hat{x}_i(\boldsymbol{y})$$

Utility function → risk function

Utility function → risk function

- $\mathrm{TP} + \mathrm{FN} = 1$ and $\mathrm{FP} + \mathrm{TN} = 1$, hence

$$\begin{aligned} U\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}\right) &= u_{\mathrm{TP}}\mathrm{TP} + u_{\mathrm{FN}}\mathrm{FN} + u_{\mathrm{FP}}\mathrm{FP} + u_{\mathrm{TN}}\mathrm{TN} \\ &= (u_{\mathrm{FN}} - u_{\mathrm{TP}})\mathrm{FN} + (u_{\mathrm{FP}} - u_{\mathrm{TN}})\mathrm{FP} + u_{\mathrm{TP}} + u_{\mathrm{TN}}. \end{aligned}$$

Utility function → risk function

- $\mathrm{TP} + \mathrm{FN} = 1$ and $\mathrm{FP} + \mathrm{TN} = 1$, hence

$$\begin{aligned} U\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}\right) &= u_{\mathrm{TP}}\mathrm{TP} + u_{\mathrm{FN}}\mathrm{FN} + u_{\mathrm{FP}}\mathrm{FP} + u_{\mathrm{TN}}\mathrm{TN} \\ &= (u_{\mathrm{FN}} - u_{\mathrm{TP}})\mathrm{FN} + (u_{\mathrm{FP}} - u_{\mathrm{TN}})\mathrm{FP} + u_{\mathrm{TP}} + u_{\mathrm{TN}}. \end{aligned}$$

- We define the risk function as

$$R\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}; \boldsymbol{\lambda}\right) = \lambda_{\mathrm{FN}}\mathrm{FN}\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}\right) + \lambda_{\mathrm{FP}}\mathrm{FP}\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}\right).$$

  - Loss caused by false positives or negatives
    - $\lambda_{\mathrm{FN}} = u_{\mathrm{TP}} - u_{\mathrm{FN}}\ (> 0)$ … False negative loss
    - $\lambda_{\mathrm{FP}} = u_{\mathrm{TN}} - u_{\mathrm{FP}}\ (> 0)$ … False positive loss

Utility function → risk function

- $\mathrm{TP} + \mathrm{FN} = 1$ and $\mathrm{FP} + \mathrm{TN} = 1$, hence

$$\begin{aligned} U\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}\right) &= u_{\mathrm{TP}}\mathrm{TP} + u_{\mathrm{FN}}\mathrm{FN} + u_{\mathrm{FP}}\mathrm{FP} + u_{\mathrm{TN}}\mathrm{TN} \\ &= (u_{\mathrm{FN}} - u_{\mathrm{TP}})\mathrm{FN} + (u_{\mathrm{FP}} - u_{\mathrm{TN}})\mathrm{FP} + u_{\mathrm{TP}} + u_{\mathrm{TN}}. \end{aligned}$$

- We define the risk function as

$$R\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}; \boldsymbol{\lambda}\right) = \lambda_{\mathrm{FN}}\mathrm{FN}\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}\right) + \lambda_{\mathrm{FP}}\mathrm{FP}\left(\widehat{\boldsymbol{x}}, \boldsymbol{x}^{(0)}\right).$$

  - Loss caused by false positives or negatives
    - $\lambda_{\mathrm{FN}} = u_{\mathrm{TP}} - u_{\mathrm{FN}}$ (> 0) … False negative loss
    - $\lambda_{\mathrm{FP}} = u_{\mathrm{TN}} - u_{\mathrm{FP}}$ (> 0) … False positive loss

- Maximization of the utility function = minimization of the risk function

Bayes risk and optimal action

Bayes risk and optimal action

- Definition: Optimal action $\widehat{\boldsymbol{x}}^{*}$ minimizes the Bayes risk

$$\overline{R}[\widehat{\boldsymbol{x}};\boldsymbol{\lambda}]=\sum_{\boldsymbol{x}^{(0)}}\sum_{\boldsymbol{y}}p\left(\boldsymbol{y}\,|\boldsymbol{x}^{(0)},\boldsymbol{F}\right)\varphi\left(\boldsymbol{x}^{(0)}\right)R\left(\boldsymbol{x}^{(0)},\widehat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda}\right)$$

as

$$\widehat{\boldsymbol{x}}^{*}=\min_{\widehat{\boldsymbol{x}}\in\Omega}\overline{R}[\widehat{\boldsymbol{x}};\boldsymbol{\lambda}]$$

  - $\boldsymbol{y}$: Test results
  - $\boldsymbol{x}^{(0)}$: Patients' true states
  - $\widehat{\boldsymbol{x}}(\boldsymbol{y})$: Action (estimated patients' states) $\in\{0,1\}^{N}$
  - $p\left(\boldsymbol{y}\,|\boldsymbol{x}^{(0)},\boldsymbol{F}\right)\varphi\left(\boldsymbol{x}^{(0)}\right)$: True generative process of $\boldsymbol{y}$ and $\boldsymbol{x}^{(0)}$ (unknown)
  - Ω: Set of possible functions that map $[0,1]\rightarrow\{0,1\}$

Bayes risk and optimal action

- Definition: Optimal action $\widehat{\boldsymbol{x}}^*$ minimizes the Bayes risk

$$\overline{R}[\widehat{\boldsymbol{x}};\boldsymbol{\lambda}] = \sum_{\boldsymbol{x}^{(0)}}\sum_{\boldsymbol{y}} p\left(\boldsymbol{y}\,|\boldsymbol{x}^{(0)},\boldsymbol{F}\right)\varphi\left(\boldsymbol{x}^{(0)}\right)R\left(\boldsymbol{x}^{(0)},\widehat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda}\right)$$

as

$$\widehat{\boldsymbol{x}}^* = \min_{\widehat{\boldsymbol{x}}\in\Omega} \overline{R}[\widehat{\boldsymbol{x}};\boldsymbol{\lambda}]$$

  - $\boldsymbol{y}$: Test results
  - $\boldsymbol{x}^{(0)}$: Patients' true states
  - $\widehat{\boldsymbol{x}}(\boldsymbol{y})$: Action (estimated patients' states) $\in \{0,1\}^N$
  - $p\left(\boldsymbol{y}\,|\boldsymbol{x}^{(0)},\boldsymbol{F}\right)\varphi\left(\boldsymbol{x}^{(0)}\right)$: True generative process of $\boldsymbol{y}$ and $\boldsymbol{x}^{(0)}$ (unknown)
  - Ω: Set of possible functions that map $[0,1] \rightarrow \{0,1\}$

■ Bayes risk is unobservable because we do not know $p\left(\boldsymbol{y}\,|\boldsymbol{x}^{(0)},\boldsymbol{F}\right)\varphi\left(\boldsymbol{x}^{(0)}\right)$.

Posterior risk in Bayesian optimal setting

Posterior risk in Bayesian optimal setting

- Definition: Posterior risk

$$\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y});\lambda) = \sum_{\boldsymbol{x}} P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) R(\boldsymbol{x},\hat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda})$$

  - $P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F})$: Posterior distribution (known)

Posterior risk in Bayesian optimal setting

- Definition: Posterior risk

$$\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y});\lambda)=\sum_{\boldsymbol{x}}P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F})R(\boldsymbol{x},\hat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda})$$

  - $P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F})$: Posterior distribution (known)

■ Posterior risk is observable.

Posterior risk in Bayesian optimal setting

- Definition: Posterior risk

$$\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y});\lambda) = \sum_{\boldsymbol{x}} P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) R(\boldsymbol{x},\hat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda})$$

  - $P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F})$: Posterior distribution (known)

  - Posterior risk is observable.

- Theorem:

  In the Bayesian optimal setting, posterior risk coincides with Bayes risk with expectation:

$$\overline{R}[\hat{\boldsymbol{x}};\boldsymbol{\lambda}] = E_{\boldsymbol{y}}\left[\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y});\lambda)\right].$$

  - $E_{\boldsymbol{y}}[\dots]$: expectation of $\boldsymbol{y}$ according to the true generative process

Bayesian optimal setting

Bayesian optimal setting

- Definition:

  When the assumed model is equivalent to the true model $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})$, the setting is said to be Bayesian optimal.

$$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \{0,1\}^N, \forall \boldsymbol{y} \in \{0,1\}^M$$

Bayesian optimal setting

- Definition:

  When the assumed model is equivalent to the true model $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})$, the setting is said to be Bayesian optimal.

$$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x}), \qquad \forall\boldsymbol{x} \in \{0,1\}^N, \forall\boldsymbol{y} \in \{0,1\}^M$$

- Under the Bayesian optimal setting, the posterior probability is given by

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})}{Z(\boldsymbol{y})}, \qquad Z(\boldsymbol{y}) = \sum_{\boldsymbol{x}} p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})$$

Substituting the posterior probability into the posterior risk yields

$$E_{\boldsymbol{y}}\left[\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda})\right] = \sum_{\boldsymbol{y}} P(\boldsymbol{y}) \sum_{\boldsymbol{x}} \frac{1}{Z(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x}) R(\boldsymbol{x},\hat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda})$$

Bayesian optimal setting

- Definition:

  When the assumed model is equivalent to the true model $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})$, the setting is said to be Bayesian optimal.

$$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x}), \qquad \forall\boldsymbol{x} \in \{0,1\}^N, \forall\boldsymbol{y} \in \{0,1\}^M$$

- Under the Bayesian optimal setting, the posterior probability is given by

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})}{Z(\boldsymbol{y})}, \qquad Z(\boldsymbol{y}) = \sum_{\boldsymbol{x}} p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})$$

Substituting the posterior probability into the posterior risk yields

$$E_{\boldsymbol{y}}\left[\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda})\right] = \sum_{\boldsymbol{y}} P(\boldsymbol{y}) \sum_{\boldsymbol{x}} \frac{1}{Z(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x}) R(\boldsymbol{x},\hat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda})$$

$P(\boldsymbol{y}) = Z(\boldsymbol{y})$

$$= \sum_{\boldsymbol{y}} \sum_{\boldsymbol{x}} p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x}) R(\boldsymbol{x},\hat{\boldsymbol{x}}(\boldsymbol{y});\boldsymbol{\lambda})$$ …Bayes risk

Posterior risk minimization

Posterior risk minimization

- Definition:

  Action $\widehat{\boldsymbol{x}}^{**}$ minimizes posterior risk under the Bayesian optimal setting

Posterior risk minimization

- Definition:
  Action $\hat{\boldsymbol{x}}^{**}$ minimizes posterior risk under the Bayesian optimal setting

- Theorem:
  $\hat{\boldsymbol{x}}^{**}$ is the optimal action that minimizes Bayes risk.

Posterior risk minimization

- Definition:

  Action $\hat{\boldsymbol{x}}^{**}$ minimizes posterior risk under the Bayesian optimal setting

- Theorem:

  $\hat{\boldsymbol{x}}^{**}$ is the optimal action that minimizes Bayes risk.

■ Proof :

By definition, $\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y}); \boldsymbol{\lambda}) \geq \hat{R}(\hat{\boldsymbol{x}}^{**}(\boldsymbol{y}); \boldsymbol{\lambda})$ holds for any $\hat{\boldsymbol{x}}(\boldsymbol{y})$.

Expectation with respect to $\boldsymbol{y}$ on both sides leads to $\overline{R}[\hat{\boldsymbol{x}}; \boldsymbol{\lambda}] \geq \overline{R}[\hat{\boldsymbol{x}}^{**}; \boldsymbol{\lambda}]$.

- $\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y}); \boldsymbol{\lambda})$: Posterior risk under the Bayesian optimal setting
- $\overline{R}[\hat{\boldsymbol{x}}; \boldsymbol{\lambda}]$: Bayes risk

Posterior risk minimization

- Definition:
  Action $\hat{\boldsymbol{x}}^{**}$ minimizes posterior risk under the Bayesian optimal setting

- Theorem:
  $\hat{\boldsymbol{x}}^{**}$ is the optimal action that minimizes Bayes risk.

- Proof :
  By definition, $\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y}); \boldsymbol{\lambda}) \geq \hat{R}(\hat{\boldsymbol{x}}^{**}(\boldsymbol{y}); \boldsymbol{\lambda})$ holds for any $\hat{\boldsymbol{x}}(\boldsymbol{y})$.
  Expectation with respect to $\boldsymbol{y}$ on both sides leads to $\overline{R}[\hat{\boldsymbol{x}}; \boldsymbol{\lambda}] \geq \overline{R}[\hat{\boldsymbol{x}}^{**}; \boldsymbol{\lambda}]$.
  - $\hat{R}(\hat{\boldsymbol{x}}(\boldsymbol{y}); \boldsymbol{\lambda})$: Posterior risk under the Bayesian optimal setting
  - $\overline{R}[\hat{\boldsymbol{x}}; \boldsymbol{\lambda}]$: Bayes risk

- Using posterior risk, which is observable, the optimal action is obtained.

Cutoff-based estimator

Cutoff-based estimator

- Theorem:

The optimal action is given by a cutoff-based function as

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

- ρ_i: Posterior marginal probability of i-th patient under Bayesian optimal setting
- θ: Prevalence
- $\lambda_{\mathrm{FP}}, \lambda_{\mathrm{FN}}$: Loss caused by false positives or negatives
- $\mathbb{I}(a)$: Indicator function (1 if a is true, 0 otherwise)

Cutoff-based estimator

- Theorem:

The optimal action is given by a cutoff-based function as

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

- ρ_i: Posterior marginal probability of i-th patient under Bayesian optimal setting
- θ: Prevalence
- $\lambda_{\mathrm{FP}}, \lambda_{\mathrm{FN}}$: Loss caused by false positives or negatives
- $\mathbb{I}(a)$: Indicator function (1 if a is true, 0 otherwise)

■ Proof: Minimization of the posterior risk under the constraint $\hat{x}_i \in \{0,1\}$.

Cutoff-based estimator

- Theorem:

The optimal action is given by a cutoff-based function as

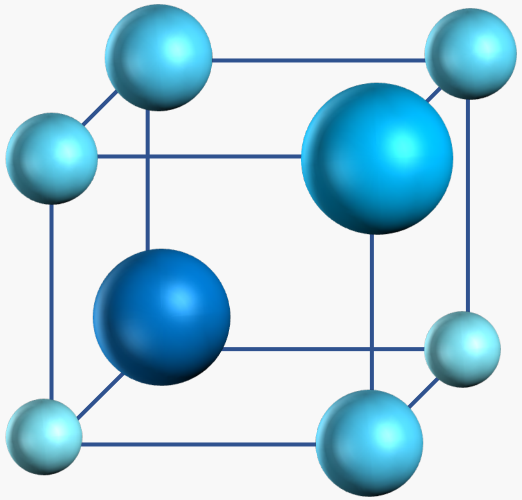
$$\hat{x}_i^*(\boldsymbol{y}) = \mathbb{I}\left(\rho_i(\boldsymbol{y}) > \frac{\theta\lambda_{\mathrm{FP}}}{\lambda_{\mathrm{FN}}(1-\theta)+\lambda_{\mathrm{FP}}\theta}\right).$$

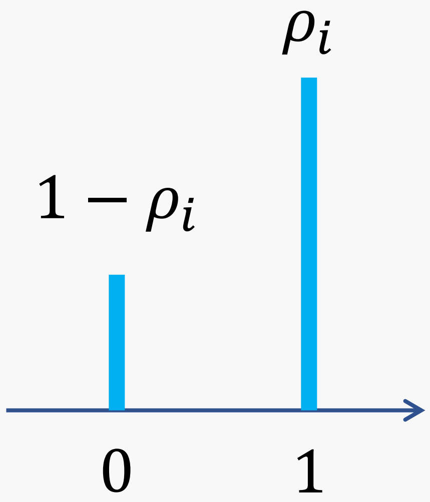
- ρ_i: Posterior marginal probability of i-th patient under Bayesian optimal setting
- θ: Prevalence
- $\lambda_{\mathrm{FP}}, \lambda_{\mathrm{FN}}$: Loss caused by false positives or negatives
- $\mathbb{I}(a)$: Indicator function (1 if a is true, 0 otherwise)

■ Proof: Minimization of the posterior risk under the constraint $\hat{x}_i \in \{0,1\}$.

■ Cutoff-based action is the optimal for the map $[0,1] \to \{0,1\}$.

Summary: optimal action





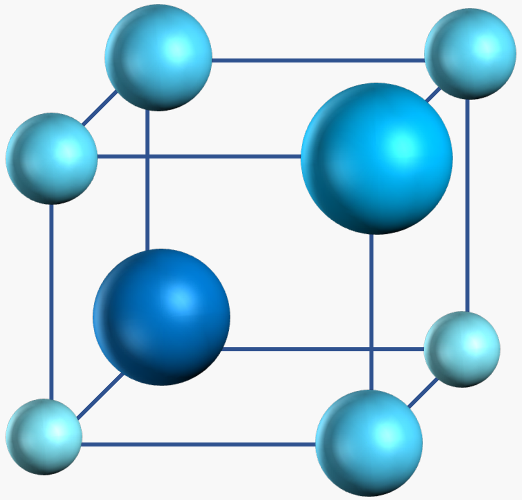
Posterior distribution
on $\{0,1\}^N$

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$

Posterior marginal probability
for each patient

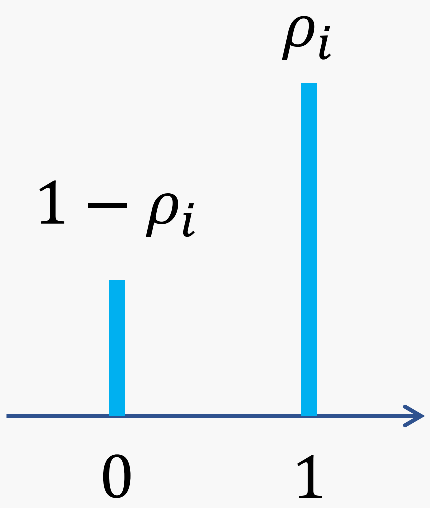
$$\rho_i = \frac{1}{Z(\boldsymbol{y})} \sum_{\boldsymbol{x}} x_i f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$

Summary: optimal action



Posterior distribution on $\{0,1\}^N$

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$



Posterior marginal probability for each patient

$$\rho_i = \frac{1}{Z(\boldsymbol{y})} \sum_{\boldsymbol{x}} x_i\, f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$

Setting loss λ_{FP} and λ_{FN}

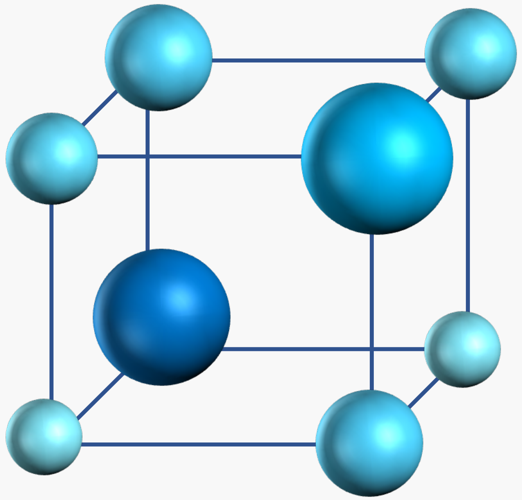
$$\hat{x}_i^*(\boldsymbol{y}) = \mathbb{I}\left(\rho_i(\boldsymbol{y}) > \frac{\theta\lambda_{\mathrm{FP}}}{\lambda_{\mathrm{FN}}(1-\theta)+\lambda_{\mathrm{FP}}\theta}\right)$$

- θ: Prevalence

Summary: optimal action

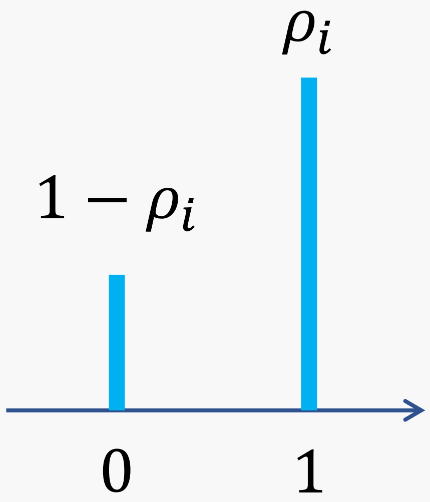
The appropriateness of marginalization as the diagnostic variable is mathematically supported by the area under the curve.

Details are shown in arXiv:2110.10877.



Posterior distribution on $\{0,1\}^N$

$$P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F}) = \frac{1}{Z(\boldsymbol{y})} f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$



Posterior marginal probability for each patient

$$\rho_i = \frac{1}{Z(\boldsymbol{y})} \sum_{\boldsymbol{x}} x_i f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x})$$

Setting loss λ_{FP} and λ_{FN}

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

- θ: Prevalence

Unified view of well-known actions

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

Unified view of well-known actions

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

- ✦ Maximum posterior marginal (MPM) estimator
  - Defined by $\hat{x}_i = \mathbb{I}(\rho_i > 0.5)$

Unified view of well-known actions

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

- ✦ Maximum posterior marginal (MPM) estimator
  - Defined by $\hat{x}_i = \mathbb{I}(\rho_i > 0.5)$
  - Corresponds to Bayes risk minimization at the loss $\lambda_{\mathrm{FP}} = 1-\theta,\ \lambda_{\mathrm{FN}} = \theta$

Unified view of well-known actions

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

✦ Maximum posterior marginal (MPM) estimator

- Defined by $\hat{x}_i = \mathbb{I}(\rho_i > 0.5)$
- Corresponds to Bayes risk minimization at the loss $\lambda_{\mathrm{FP}} = 1-\theta,\ \lambda_{\mathrm{FN}} = \theta$
- When $\theta < 0.5$, the decrease of FP is preferred in the action.

Unified view of well-known actions

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

✦ Maximum posterior marginal (MPM) estimator

- Defined by $\hat{x}_i = \mathbb{I}(\rho_i > 0.5)$
- Corresponds to Bayes risk minimization at the loss $\lambda_{\mathrm{FP}} = 1-\theta,\ \lambda_{\mathrm{FN}} = \theta$
- When $\theta < 0.5$, the decrease of FP is preferred in the action.

! MPM estimator for group testing, which is efficient at small θ, immediately leads to a decrease of FP rather than FN.

Unified view of well-known actions

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

✦ Maximum posterior marginal (MPM) estimator

- Defined by $\hat{x}_i = \mathbb{I}(\rho_i > 0.5)$
- Corresponds to Bayes risk minimization at the loss $\lambda_{\mathrm{FP}} = 1-\theta,\ \lambda_{\mathrm{FN}} = \theta$
- When $\theta < 0.5$, the decrease of FP is preferred in the action.

> **!** MPM estimator for group testing, which is efficient at small θ, immediately leads to a decrease of FP rather than FN.

✦ Youden index maximization: frequently used in medical statistics

- Defined by risk minimization at the loss $\lambda_{\mathrm{FP}} = \lambda_{\mathrm{FN}} = 0.5$.
- Corresponds to the action $\hat{x}_i = \mathbb{I}(\rho_i > \theta)$.

Usefulness of Bayesian optimal setting

Usefulness of Bayesian optimal setting

- Bayesian optimal setting gives the upper bounds.

Usefulness of Bayesian optimal setting

- Bayesian optimal setting gives the upper bounds.
  - The Bayesian optimal setting is not realistic.
    - $f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})$ does not generally hold.

Usefulness of Bayesian optimal setting

Bayesian optimal setting gives the upper bounds.

- The Bayesian optimal setting is not realistic.
  - $f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{F})\phi(\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{F})\varphi(\boldsymbol{x})$ does not generally hold.
- Bayesian optimal setting can describe possible optimal decision.
  - Optimal action is given by the posterior marginal probability.
  - Optimal action leads minimal Bayes risk.

Usefulness of Bayesian optimal setting

Bayesian optimal setting gives the upper bounds.

- The Bayesian optimal setting is not realistic.
  - $f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})$ does not generally hold.
- Bayesian optimal setting can describe possible optimal decision.
  - Optimal action is given by the posterior marginal probability.
  - Optimal action leads minimal Bayes risk.
- From the Bayesian optimal setting, we can obtain practical guides.
  - There is no gain to introduce GT in the parameter region where GT under the Bayesian optimal setting is not efficient.

Usefulness of Bayesian optimal setting

■ Bayesian optimal setting gives the upper bounds.

- The Bayesian optimal setting is not realistic.
  - $f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\phi(\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{F})\varphi(\boldsymbol{x})$ does not generally hold.
- Bayesian optimal setting can describe possible optimal decision.
  - Optimal action is given by the posterior marginal probability.
  - Optimal action leads minimal Bayes risk.
- From the Bayesian optimal setting, we can obtain practical guides.
  - There is no gain to introduce GT in the parameter region where GT under the Bayesian optimal setting is not efficient.

● Inequalities bounded by the Bayesian optimal setting are equivalent to those that hold on the Nishimori line in spin-glass theory.

See Iba, JPA (1999) & Nishimori (2001)

Algorithm for Actual Inference in Group Testing

Graphical representation and message passing

A Sakata, J Phys Soc Jpn 89, 084001 (2020)

Approximation by message passing

- We know that

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

  is the optimal in terms of Bayes risk minimization.

Approximation by message passing

- We know that

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

  is the optimal in terms of Bayes risk minimization.

- The remaining problem is the calculation of
  the posterior marginal probability $\rho_i(\boldsymbol{y})$.

Approximation by message passing

- We know that

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

  is the optimal in terms of Bayes risk minimization.

- The remaining problem is the calculation of the posterior marginal probability $\rho_i(\boldsymbol{y})$.

- The cost for the exact computation is $O(\exp(N))$.

Approximation by message passing

- We know that

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

is the optimal in terms of Bayes risk minimization.

- The remaining problem is the calculation of the posterior marginal probability $\rho_i(\boldsymbol{y})$.

- The cost for the exact computation is $O(\exp(N))$.

→ Approximation: message passing on the graphical representation (computational cost is a polynomial of N).

Graphical representation: factor graph

- 2-body spin-glass model

$$P(\boldsymbol{\sigma}) \propto \exp\left(\beta \sum_{(i,j)} J_{ij}\sigma_i\sigma_j\right)$$

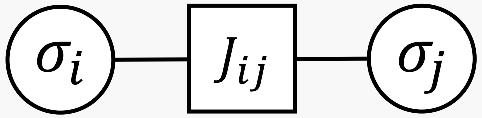
$$\equiv \prod_{(i,j)} \psi_{ij}(\sigma_i, \sigma_j; J_{ij})$$

Graphical representation: factor graph

- 2-body spin-glass model

$$P(\boldsymbol{\sigma}) \propto \exp\left(\beta \sum_{(i,j)} J_{ij}\sigma_i\sigma_j\right)$$

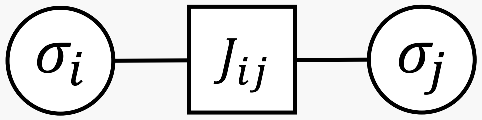
$$\equiv \prod_{(i,j)} \psi_{ij}(\sigma_i, \sigma_j; J_{ij})$$



Graphical representation: factor graph

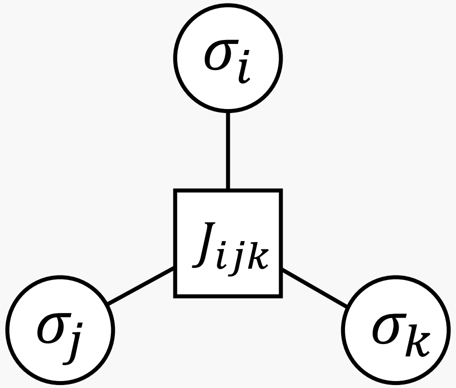
- 2-body spin-glass model

$$P(\boldsymbol{\sigma}) \propto \exp\left(\beta \sum_{(i,j)} J_{ij}\sigma_i\sigma_j\right) \equiv \prod_{(i,j)} \psi_{ij}(\sigma_i, \sigma_j; J_{ij})$$



- 3-body spin-glass model

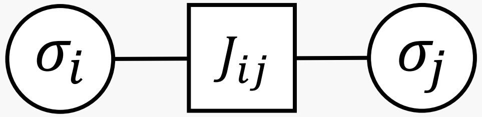
$$P(\boldsymbol{\sigma}) \propto \exp\left(\beta \sum_{(i,j,k)} J_{ijk}\sigma_i\sigma_j\sigma_k\right) \equiv \prod_{(i,j,k)} \psi_{ijk}(\sigma_i, \sigma_j, \sigma_k; J_{ijk})$$



Graphical representation: factor graph

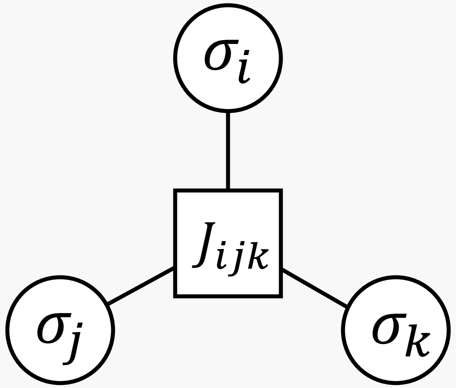
- 2-body spin-glass model

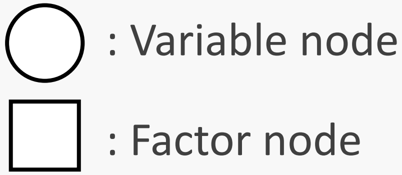
$$P(\boldsymbol{\sigma}) \propto \exp\left(\beta \sum_{(i,j)} J_{ij}\sigma_i\sigma_j\right)$$
$$\equiv \prod_{(i,j)} \psi_{ij}(\sigma_i, \sigma_j; J_{ij})$$



- 3-body spin-glass model

$$P(\boldsymbol{\sigma}) \propto \exp\left(\beta \sum_{(i,j,k)} J_{ijk}\sigma_i\sigma_j\sigma_k\right)$$
$$\equiv \prod_{(i,j,k)} \psi_{ijk}(\sigma_i, \sigma_j, \sigma_k; J_{ijk})$$

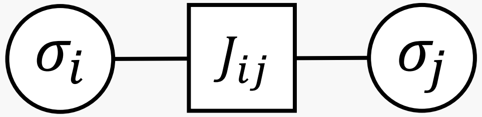




Graphical representation: factor graph

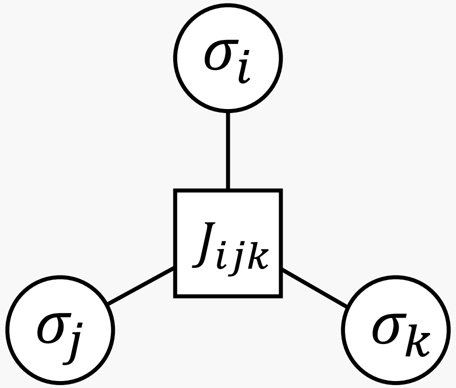
- 2-body spin-glass model

$$P(\boldsymbol{\sigma}) \propto \exp\left(\beta \sum_{(i,j)} J_{ij}\sigma_i\sigma_j\right)$$
$$\equiv \prod_{(i,j)} \psi_{ij}(\sigma_i, \sigma_j; J_{ij})$$



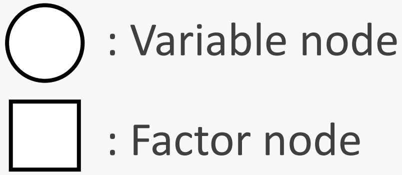
- 3-body spin-glass model

$$P(\boldsymbol{\sigma}) \propto \exp\left(\beta \sum_{(i,j,k)} J_{ijk}\sigma_i\sigma_j\sigma_k\right)$$
$$\equiv \prod_{(i,j,k)} \psi_{ijk}(\sigma_i, \sigma_j, \sigma_k; J_{ijk})$$



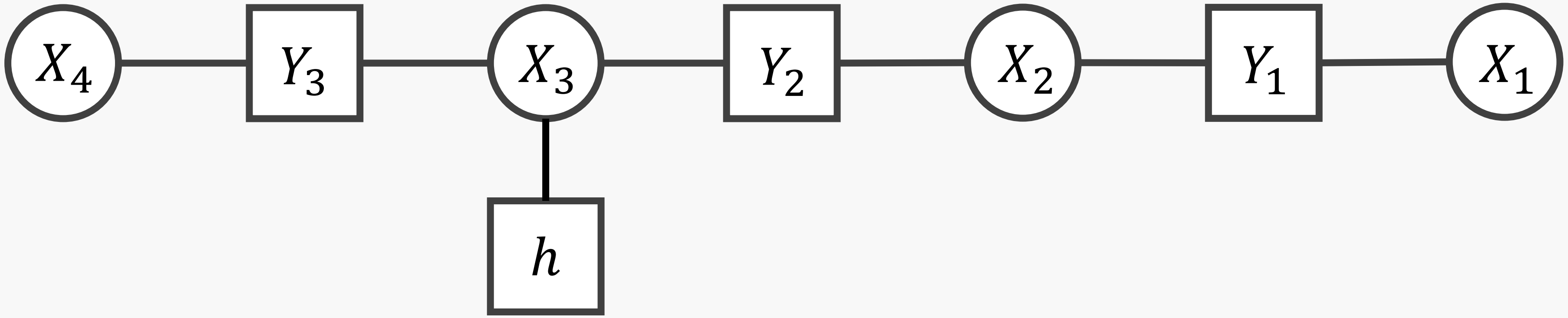
Procedures for message passing:

- Product of factors ψ
- Summing out the variables



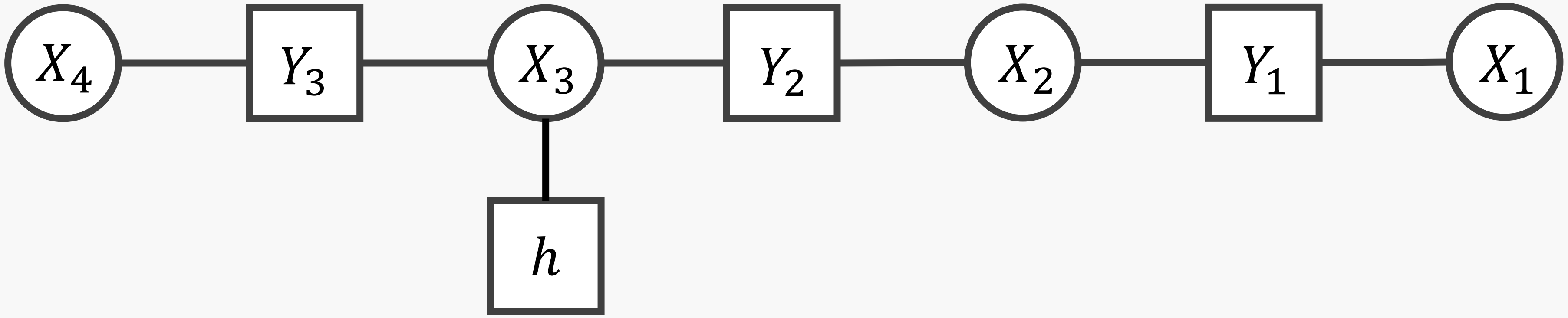
Sum-product on factor graph

- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3} \sum_{X_2} \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y}) P_h(X_3)$ on the graph



Sum-product on factor graph

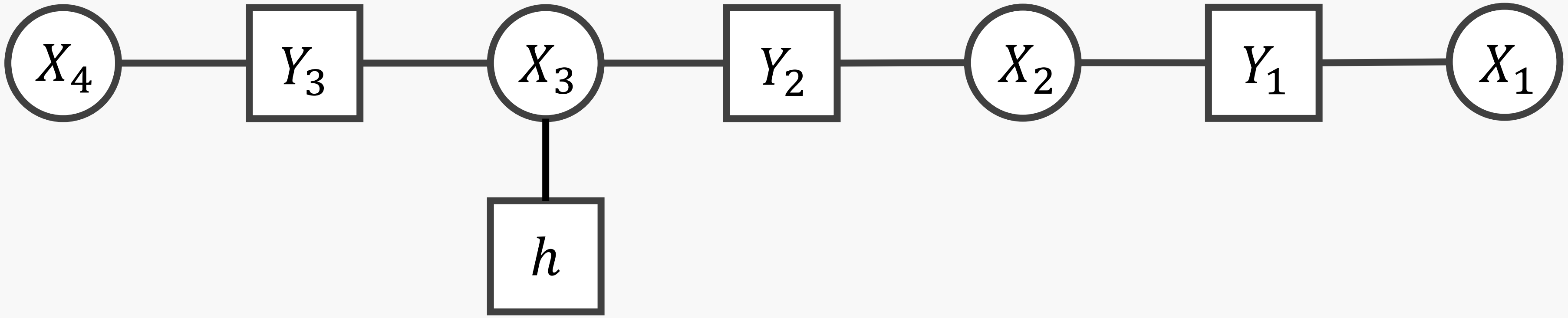
- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3} \sum_{X_2} \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y}) P_h(X_3)$ on the graph



$$P(X_4|\boldsymbol{Y}) = \sum_{X_3} \psi_{34}(X_3, X_4; Y_3)\phi(X_3, h_3) \sum_{X_2} \psi_{23}(X_2, X_3; Y_2) \sum_{X_1} \psi_{12}(X_1, X_2; Y_1)$$

Sum-product on factor graph

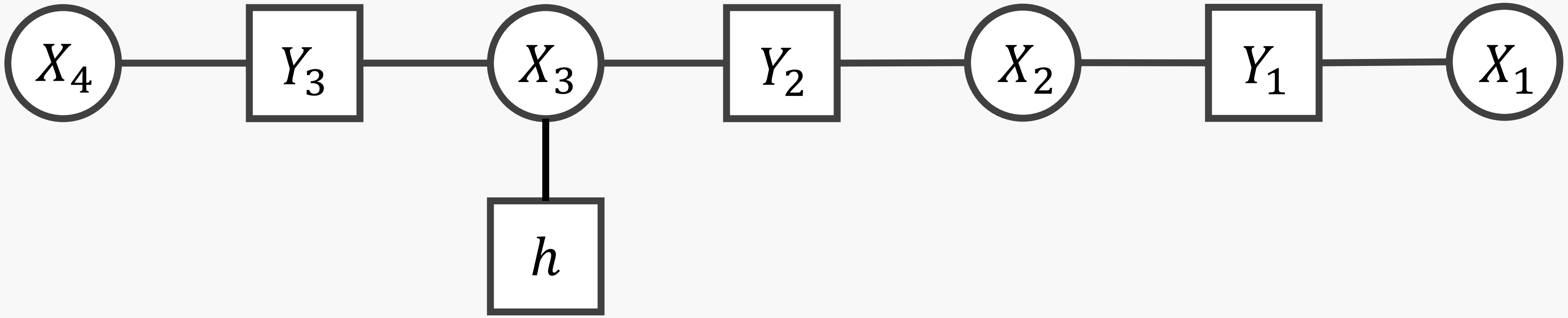
- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3} \sum_{X_2} \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y}) P_h(X_3)$ on the graph



$$P(X_4|\boldsymbol{Y}) = \sum_{X_3} \psi_{34}(X_3, X_4; Y_3)\phi(X_3, h_3) \sum_{X_2} \psi_{23}(X_2, X_3; Y_2) \sum_{X_1} \psi_{12}(X_1, X_2; Y_1)$$

Sum-product on factor graph

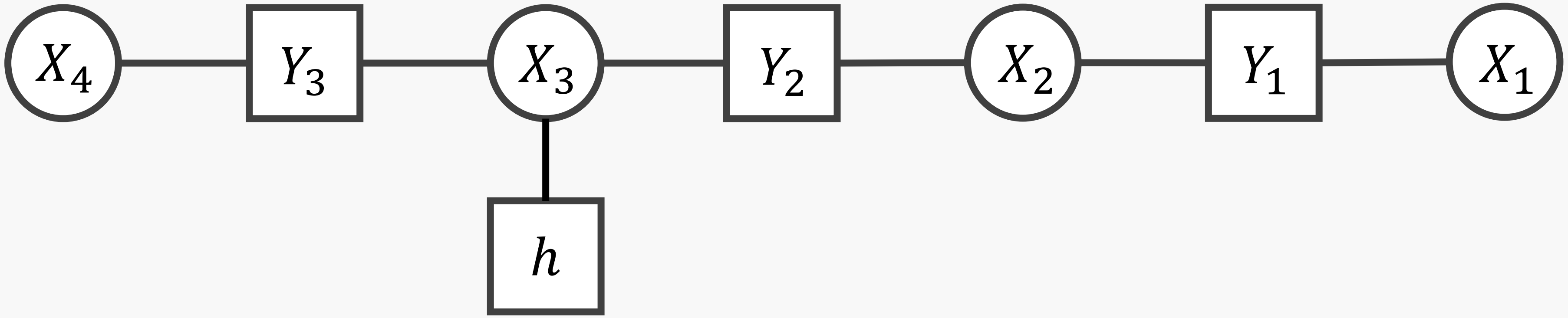
- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3} \sum_{X_2} \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y}) P_h(X_3)$ on the graph



$$P(X_4|\boldsymbol{Y}) = \sum_{X_3} \psi_{34}(X_3, X_4; Y_3)\phi(X_3, h_3) \sum_{X_2} \psi_{23}(X_2, X_3; Y_2) \sum_{X_1} \psi_{12}(X_1, X_2; Y_1)$$

Sum-product on factor graph

- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3}\sum_{X_2}\sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})P_h(X_3)$ on the graph

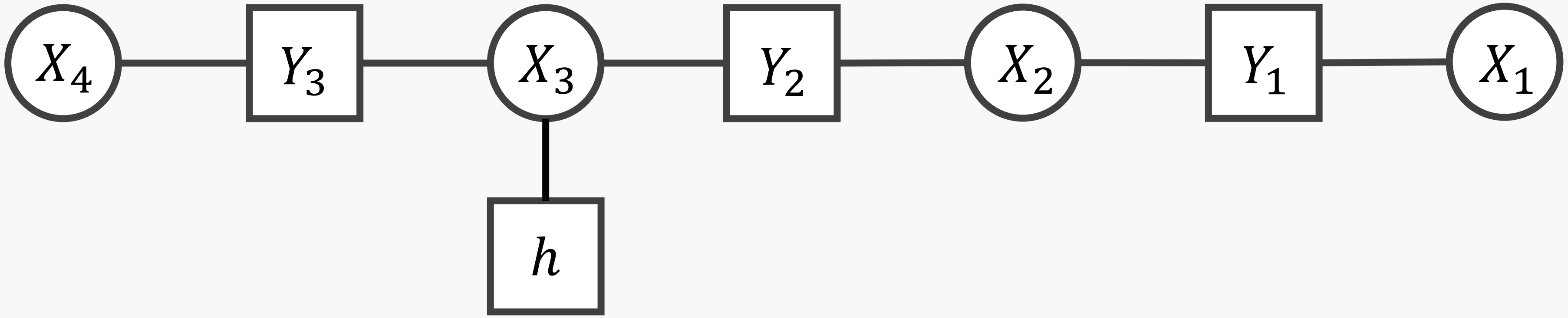


$$P(X_4|\boldsymbol{Y}) = \sum_{X_3}\psi_{34}(X_3,X_4;Y_3)\phi(X_3,h_3)\sum_{X_2}\psi_{23}(X_2,X_3;Y_2)\underbrace{\sum_{X_1}\psi_{12}(X_1,X_2;Y_1)}_{\tilde{m}_{1\to 2}(X_2)}$$

- $\tilde{m}_{\mu\to i}(X_i)$: Message from μ-th factor node to i-th variable node (input message)

Sum-product on factor graph

- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3}\sum_{X_2}\sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})P_h(X_3)$ on the graph

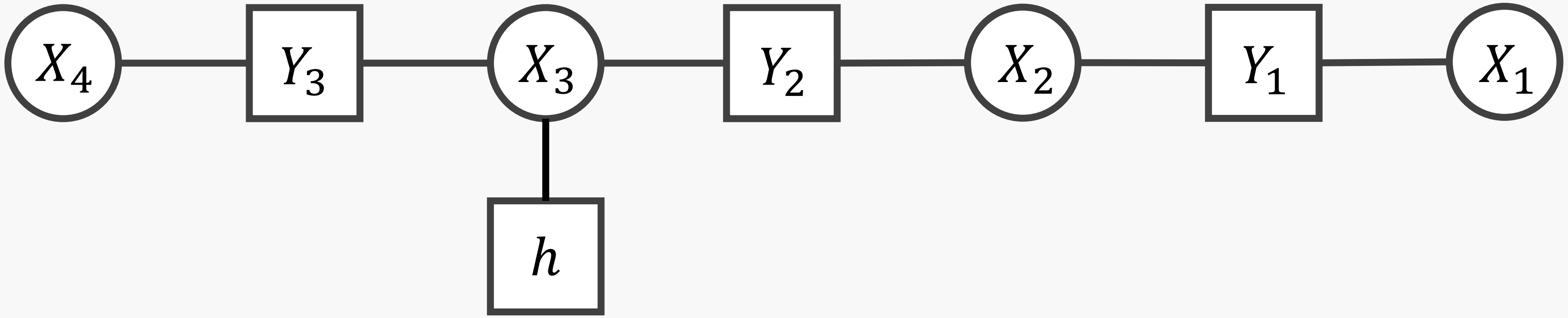


$$P(X_4|\boldsymbol{Y}) = \sum_{X_3}\psi_{34}(X_3, X_4; Y_3)\phi(X_3, h_3)\underbrace{\sum_{X_2}\psi_{23}(X_2, X_3; Y_2)\underbrace{\sum_{X_1}\psi_{12}(X_1, X_2; Y_1)}_{\tilde{m}_{1\to 2}(X_2)}}_{\tilde{m}_{2\to 3}(X_3)}$$

- $\tilde{m}_{\mu\to i}(X_i)$: Message from μ-th factor node to i-th variable node (input message)

Sum-product on factor graph

- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3}\sum_{X_2}\sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})P_h(X_3)$ on the graph

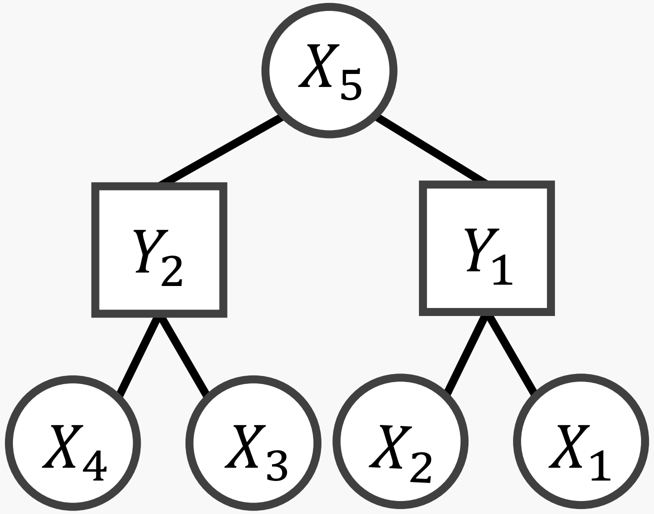


$$P(X_4|\boldsymbol{Y}) = \underbrace{\sum_{X_3} \psi_{34}(X_3,X_4;Y_3)\phi(X_3,h_3) \underbrace{\sum_{X_2}\psi_{23}(X_2,X_3;Y_2)\underbrace{\sum_{X_1}\psi_{12}(X_1,X_2;Y_1)}_{\tilde{m}_{1\to2}(X_2)}}_{\tilde{m}_{2\to3}(X_3)}}_{\tilde{m}_{3\to4}(X_4)}$$

- $\tilde{m}_{\mu\to i}(X_i)$: Message from μ-th factor node to i-th variable node (input message)
  - ✓ Equivalent to the transfer matrix method

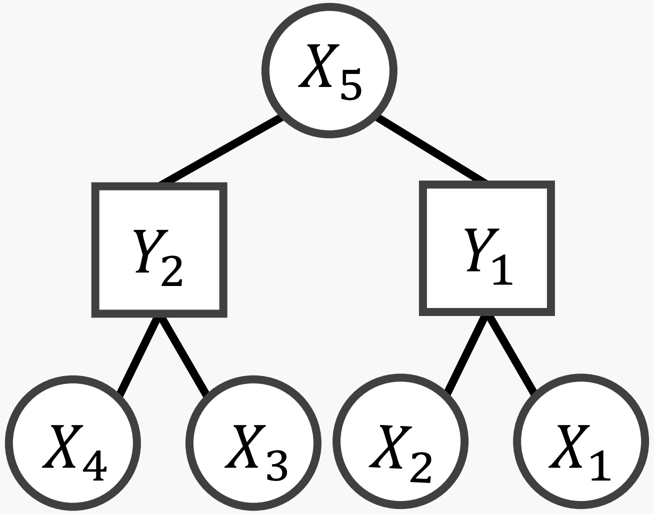
Sum-product on factor graph

- Example: Calculate $P(X_5|\boldsymbol{Y}) = \sum_{X_4} \cdots \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})$ on the graph



Sum-product on factor graph

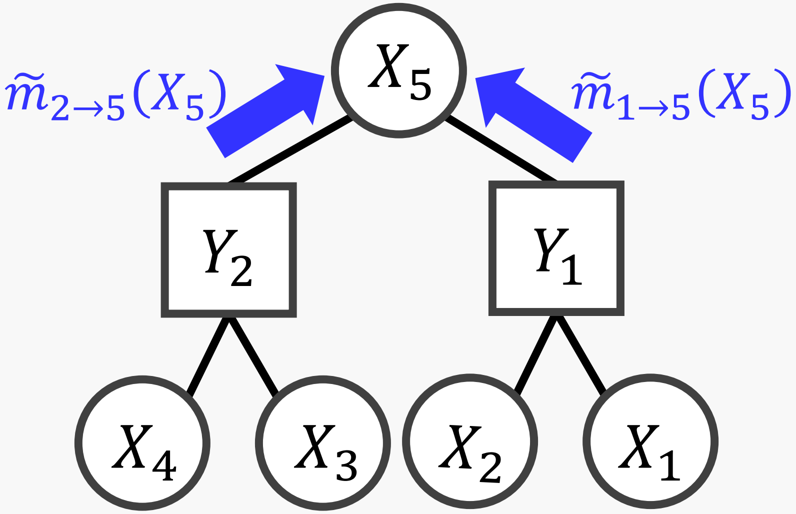
- Example: Calculate $P(X_5|\boldsymbol{Y}) = \sum_{X_4} \cdots \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})$ on the graph



$$P(X_5|\boldsymbol{Y}) = \sum_{X_3,X_4} \psi_2(X_3, X_4, X_5; Y_2) \sum_{X_1,X_2} \psi_1(X_1, X_2, X_5; Y_1)$$

Sum-product on factor graph

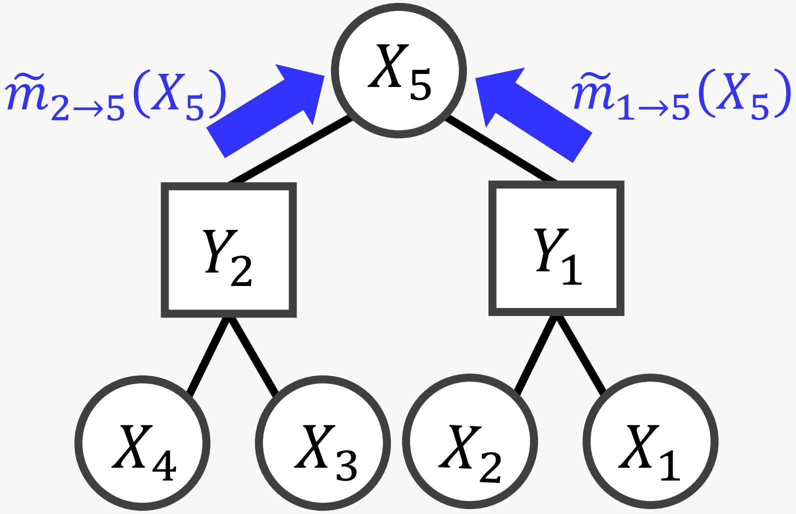
- Example: Calculate $P(X_5|\boldsymbol{Y}) = \sum_{X_4} \cdots \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})$ on the graph



$$P(X_5|\boldsymbol{Y}) = \underbrace{\sum_{X_3,X_4} \psi_2(X_3,X_4,X_5;Y_2)}_{=\tilde{m}_{2\to5}(X_5)} \underbrace{\sum_{X_1,X_2} \psi_1(X_1,X_2,X_5;Y_1)}_{=\tilde{m}_{1\to5}(X_5)}$$

Sum-product on factor graph

- Example: Calculate $P(X_5|\boldsymbol{Y}) = \sum_{X_4} \cdots \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})$ on the graph

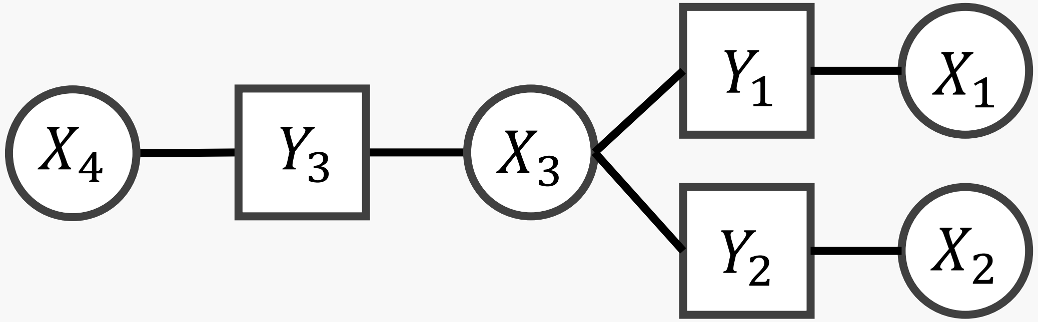


$$P(X_5|\boldsymbol{Y}) = \underbrace{\sum_{X_3,X_4} \psi_2(X_3,X_4,X_5;Y_2)}_{=\tilde{m}_{2\to5}(X_5)} \underbrace{\sum_{X_1,X_2} \psi_1(X_1,X_2,X_5;Y_1)}_{=\tilde{m}_{1\to5}(X_5)}$$

■ Marginal distribution is given by the product of the input messages.

Sum-product on factor graph

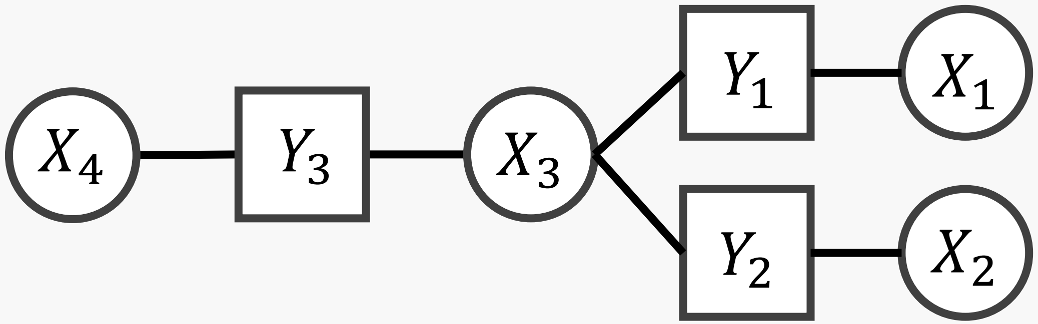
- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3} \sum_{X_2} \sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})$ on the graph



$$P(X_4|\boldsymbol{Y}) = \sum_{X_3} \psi_3(X_3, X_4; Y_3) \sum_{X_2} \psi_2(X_2, X_3; Y_2) \sum_{X_1} \psi_1(X_1, X_3; Y_1)$$

Sum-product on factor graph

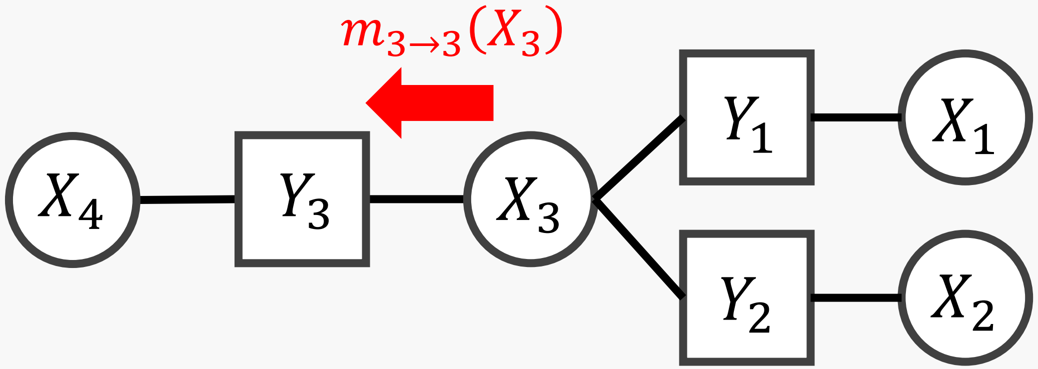
- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3}\sum_{X_2}\sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})$ on the graph



$$P(X_4|\boldsymbol{Y}) = \underbrace{\sum_{X_3}\psi_3(X_3,X_4;Y_3)\underbrace{\sum_{X_2}\psi_2(X_2,X_3;Y_2)}_{\tilde{m}_{2\to3}(X_3)}\underbrace{\sum_{X_1}\psi_1(X_1,X_3;Y_1)}_{\tilde{m}_{1\to3}(X_3)}}_{\tilde{m}_{3\to4}(X_4)}$$

Sum-product on factor graph

- Example: Calculate $P(X_4|\boldsymbol{Y}) = \sum_{X_3}\sum_{X_2}\sum_{X_1} P(\boldsymbol{X}|\boldsymbol{Y})$ on the graph

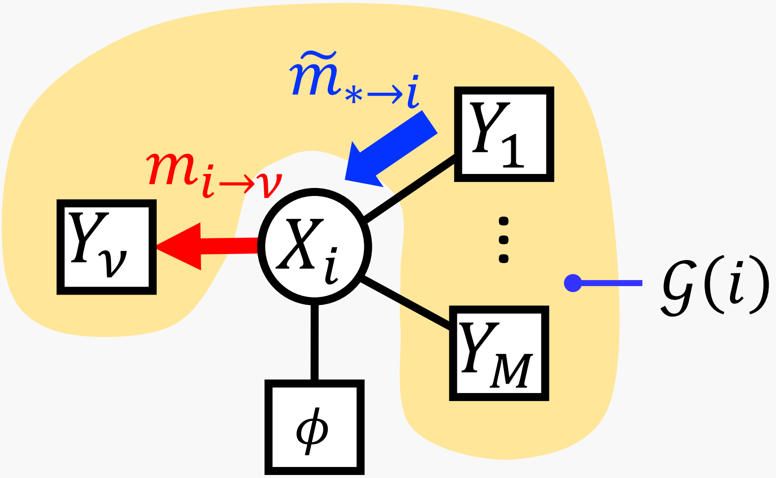


$$P(X_4|\boldsymbol{Y}) = \underbrace{\sum_{X_3}\psi_3(X_3,X_4;Y_3)\underbrace{\underbrace{\sum_{X_2}\psi_2(X_2,X_3;Y_2)}_{\tilde{m}_{2\to3}(X_3)}\underbrace{\sum_{X_1}\psi_1(X_1,X_3;Y_1)}_{\tilde{m}_{1\to3}(X_3)}}_{m_{3\to3}(X_3)}}_{\tilde{m}_{3\to4}(X_4)}$$

- $m_{i\to\mu}(X_i)$: Message from i-th variable node to μ-th factor node (output message)

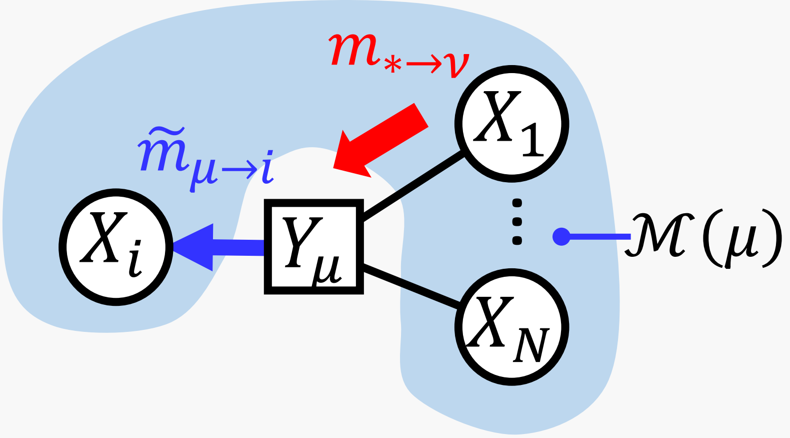
General form of message passing

- $\mathcal{M}(\mu)$: Set of variables in the interaction Y_μ
- $\mathcal{G}(i)$: Set of multi-body interactions that X_i belongs to
- $A \setminus B$: Set of the elements in A except B



$$\tilde{m}_{\mu\to i}(X_i) = \frac{1}{\tilde{Z}_{\mu\to i}} \sum_{\boldsymbol{X}\setminus X_i} \psi_\mu(\boldsymbol{X}|Y_\mu) \prod_{j\in\mathcal{M}(\mu)\setminus i} m_{j\to\mu}(X_j)$$

$$m_{i\to\nu}(X_i) = \frac{1}{Z_{i\to\nu}} \phi(X_i) \prod_{\gamma\in\mathcal{G}(i)\setminus\nu} \tilde{m}_{\gamma\to i}(X_i)$$



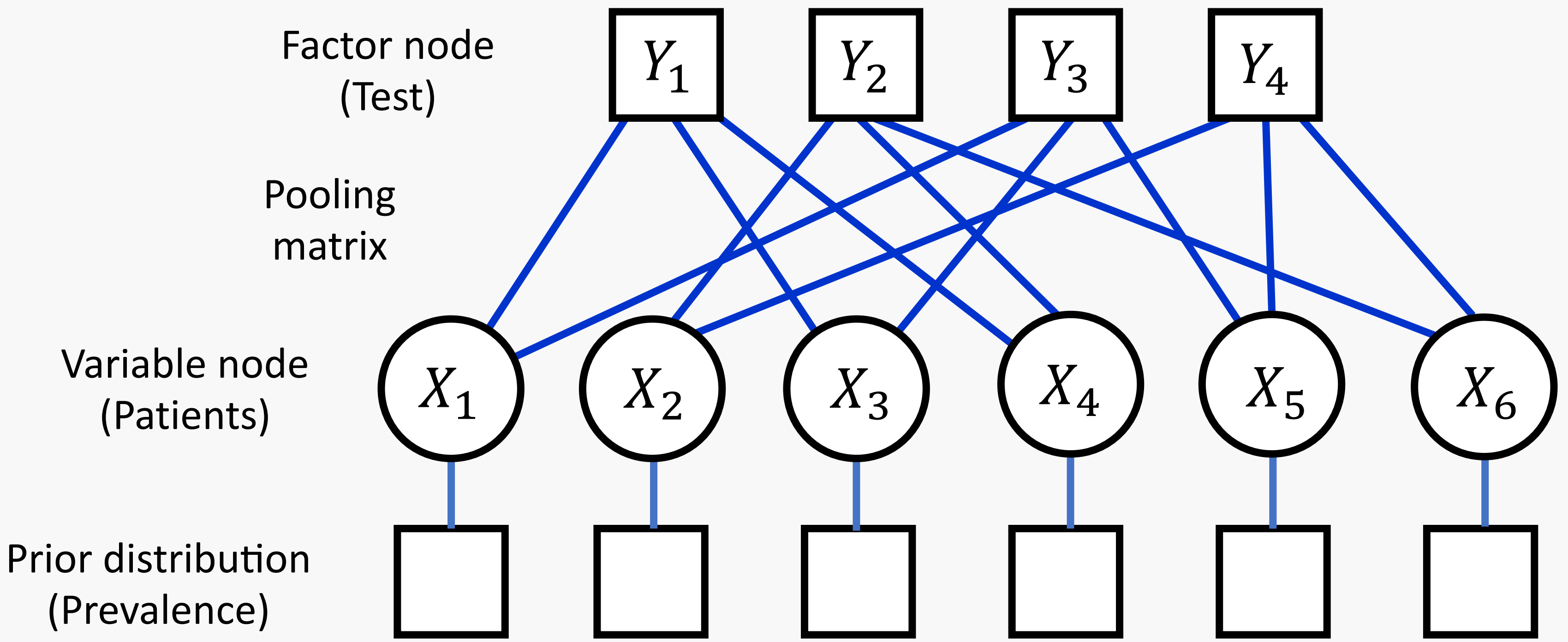
Marginal distribution:

$$P_i(X_i) = \frac{1}{Z_i} \phi(X_i) \prod_{\gamma\in\mathcal{G}(i)} \tilde{m}_{\gamma\to i}(X_i)$$

When the graph does not have any loops, the computation is exact.

Factor graph for group testing

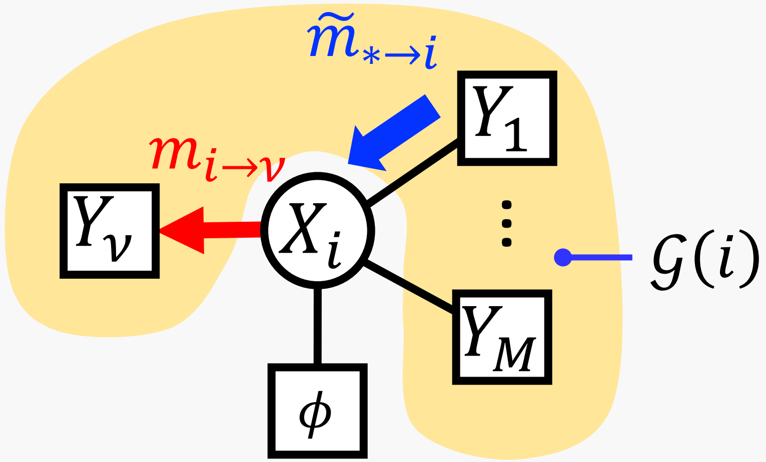
Example:
- Each patient belongs to 2 pools
- Each pool contains 3 patients

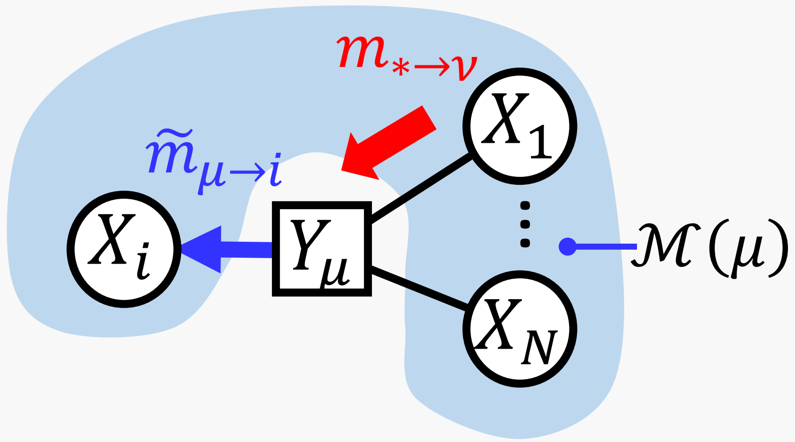


Factor : $\psi_\mu = \{p_{\mathrm{TP}} y_\mu + (1 - p_{\mathrm{TP}})(1 - y_\mu)\} T(\boldsymbol{x}, \widetilde{\boldsymbol{F}}_\mu) + \{p_{\mathrm{FP}} y_\mu + (1 - p_{\mathrm{FP}})(1 - y_\mu)\} \left(1 - T(\boldsymbol{x}, \widetilde{\boldsymbol{F}}_\mu)\right)$

Meaning of messages in group testing

- $\mathcal{M}(\mu)$: Set of the patients in μ-th pool
- $\mathcal{G}(i)$: Set of pools that i-th patient belong to



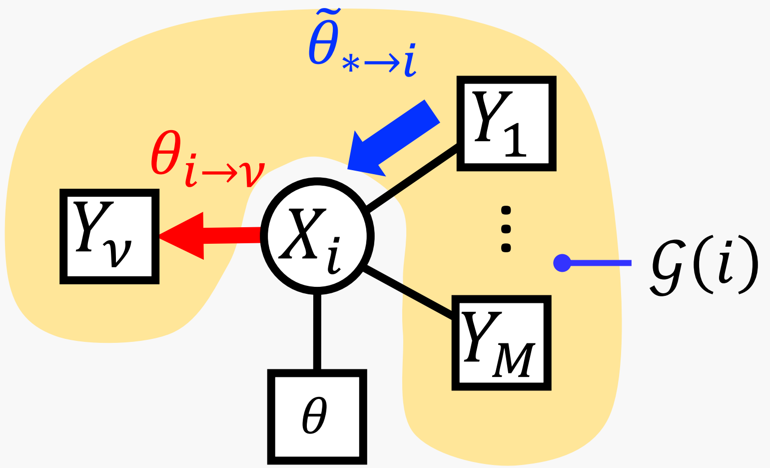


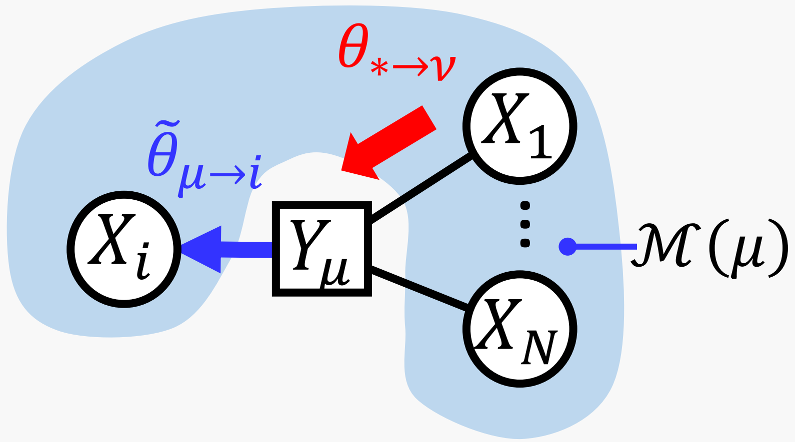
Meaning of messages in group testing

- $\mathcal{M}(\mu)$: Set of the patients in μ-th pool
- $\mathcal{G}(i)$: Set of pools that i-th patient belong to

● Variables are binary, hence the messages can be represented by Bernoulli variables.

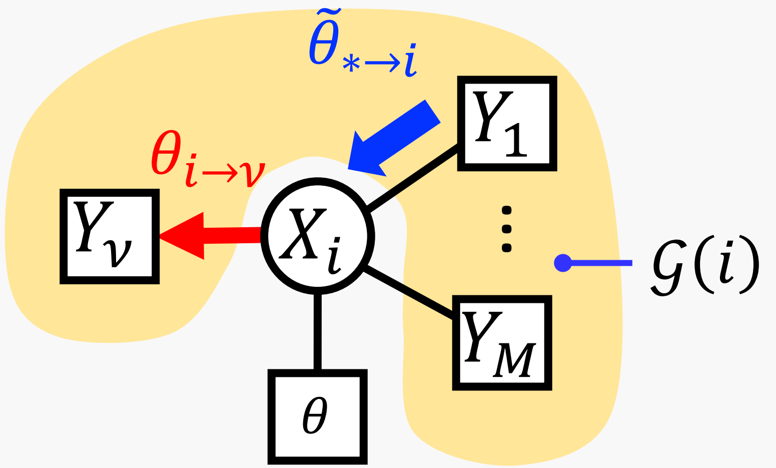
- $\tilde{m}_{\mu\to i}(X_i) = \mathrm{Bernoulli}(\tilde{\theta}_{\mu\to i}),\ m_{i\to\mu}(X_i) = \mathrm{Bernoulli}(\theta_{i\to\mu})$





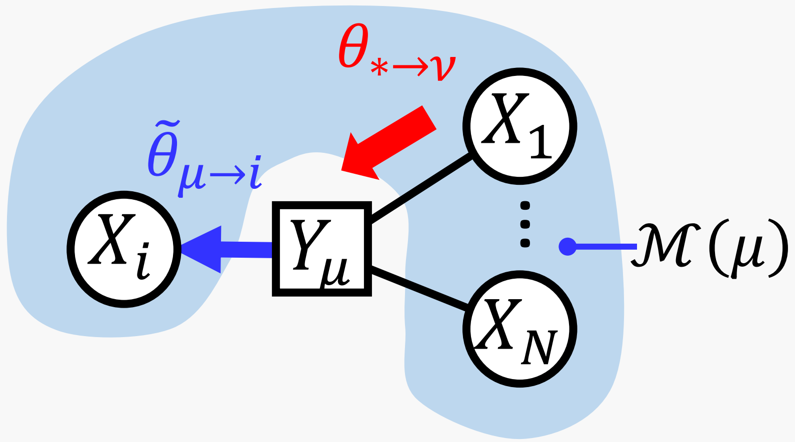
Meaning of messages in group testing

- $\mathcal{M}(\mu)$: Set of the patients in μ-th pool
- $\mathcal{G}(i)$: Set of pools that i-th patient belong to



- Variables are binary, hence the messages can be represented by Bernoulli variables.
  - $\tilde{m}_{\mu\to i}(X_i) = \text{Bernoulli}(\tilde{\theta}_{\mu\to i}),\ m_{i\to\mu}(X_i) = \text{Bernoulli}(\theta_{i\to\mu})$

- Meaning of messages
  - $\tilde{\theta}_{\mu\to i}$: Probability of positive for i-th patient after performing tests on μ-th pool $(\mu \in \mathcal{G}(i))$
  - $\theta_{i\to\mu}$: Probability of positive for i-th patient before performing test on μ-th pool $(\mu \in \mathcal{G}(i))$



Message passing for group testing

- Message from variable to factor : $$\theta_{i\to\mu} = \frac{\theta \prod_{\nu\in\mathcal{G}(i)\setminus\mu} \tilde{\theta}_{\nu\to i}}{Z_{i\to\mu}}$$
- Message from factor to variable : $$\tilde{\theta}_{\mu\to i} = \frac{U_\mu}{\tilde{Z}_{\mu\to i}}$$

where $U_\mu = p_{\mathrm{TP}} Y_\mu + (1-p_{\mathrm{TP}})(1-Y_\mu)$, $\quad W_\mu = p_{\mathrm{FP}} Y_\mu + (1-p_{\mathrm{FP}})(1-Y_\mu)$

$$Z_{i\to\mu} = \theta \prod_{\nu\in\mathcal{G}(i)\setminus\mu} \tilde{\theta}_{\nu\to i} + (1-\theta) \prod_{\nu\in\mathcal{G}(i)\setminus\mu} (1-\tilde{\theta}_{\nu\to i})$$

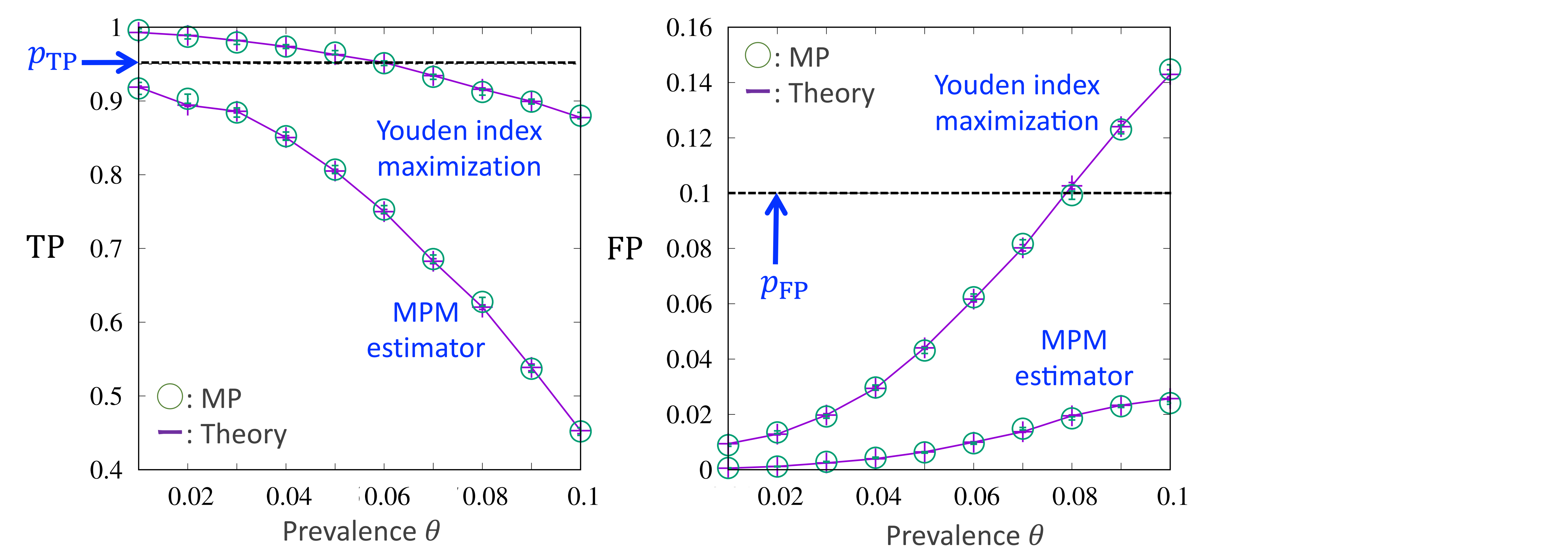
$$\tilde{Z}_{\mu\to i} = U_\mu + U_\mu \left(1 - \prod_{j\in\mathcal{M}(\mu)\setminus i} (1-\theta_{j\to\mu})\right) + W_\mu \prod_{j\in\mathcal{M}(\mu)\setminus i} (1-\theta_{j\to\mu})$$

- Marginal distribution : $$\rho_i = \frac{\theta \prod_{\mu\in\mathcal{G}(i)} \tilde{\theta}_{\mu\to i}}{\theta \prod_{\mu\in\mathcal{G}(i)} \tilde{\theta}_{\mu\to i} + (1-\theta) \prod_{\mu\in\mathcal{G}(i)} (1-\tilde{\theta}_{\mu\to i})}$$

- θ: Prevalence
- $\mathcal{G}(i)$: Pools that include i-th patient
- $\mathcal{M}(\mu)$: Patients in μ-th pool

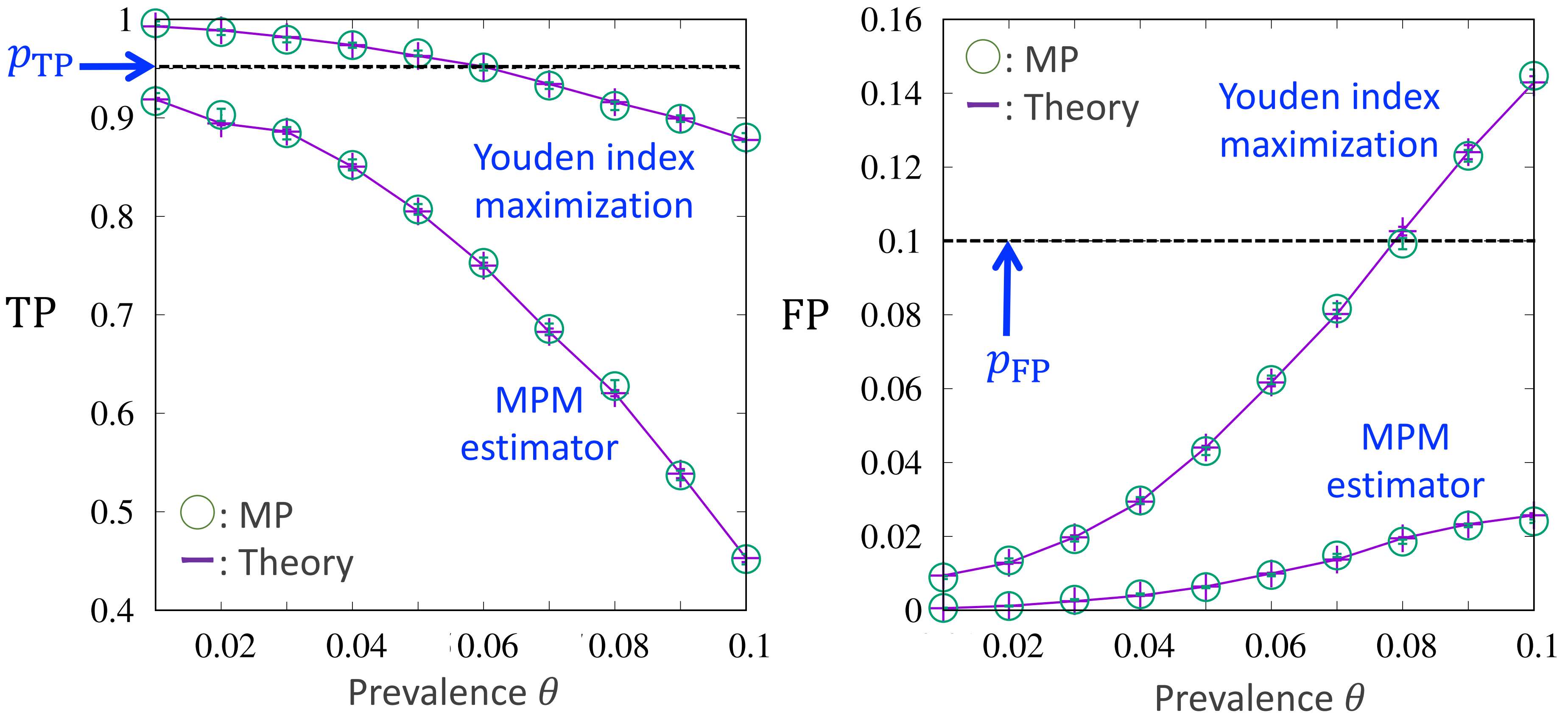
Error correction by group testing

- $N = 1000, M/N = 0.5, p_{\mathrm{TP}} = 0.95, p_{\mathrm{FP}} = 0.1$, pool size of 10 for Bayesian optimal setting
  - M: Number of tests (pools), N: Number of patients



Error correction by group testing

- $N = 1000, M/N = 0.5, p_{\mathrm{TP}} = 0.95, p_{\mathrm{FP}} = 0.1$, pool size of 10 for Bayesian optimal setting
  - M: Number of tests (pools), N: Number of patients



- Youden index maximization:

  $\lambda_{\mathrm{FN}} = \lambda_{\mathrm{FP}} = 1/2$

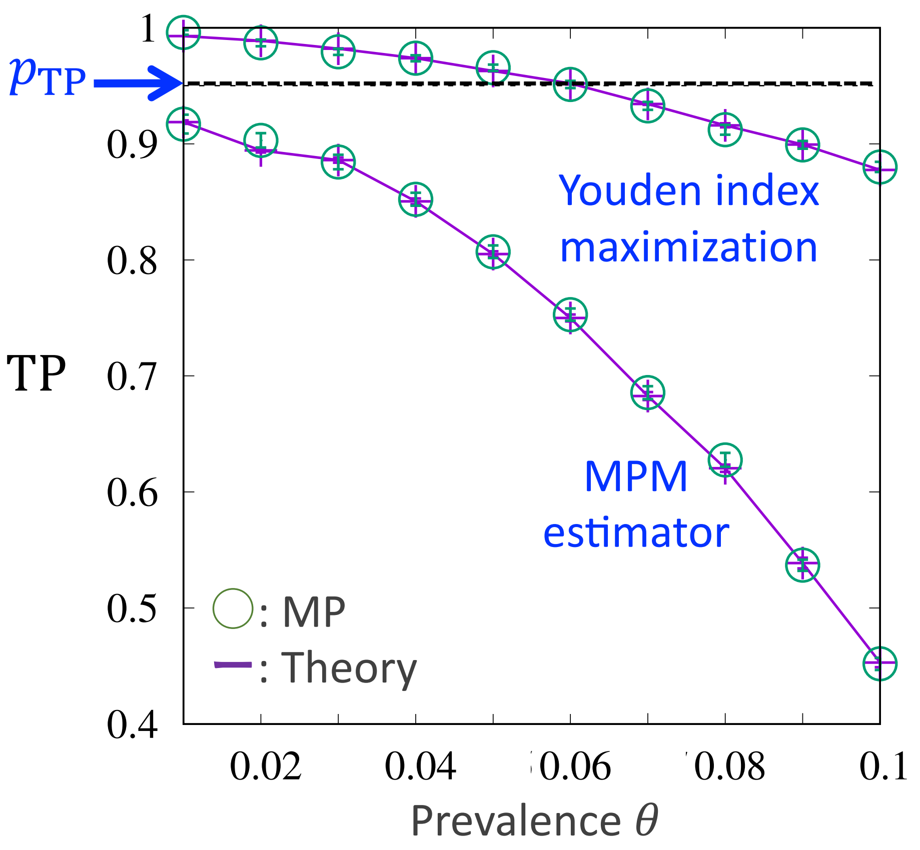
- MPM estimator :

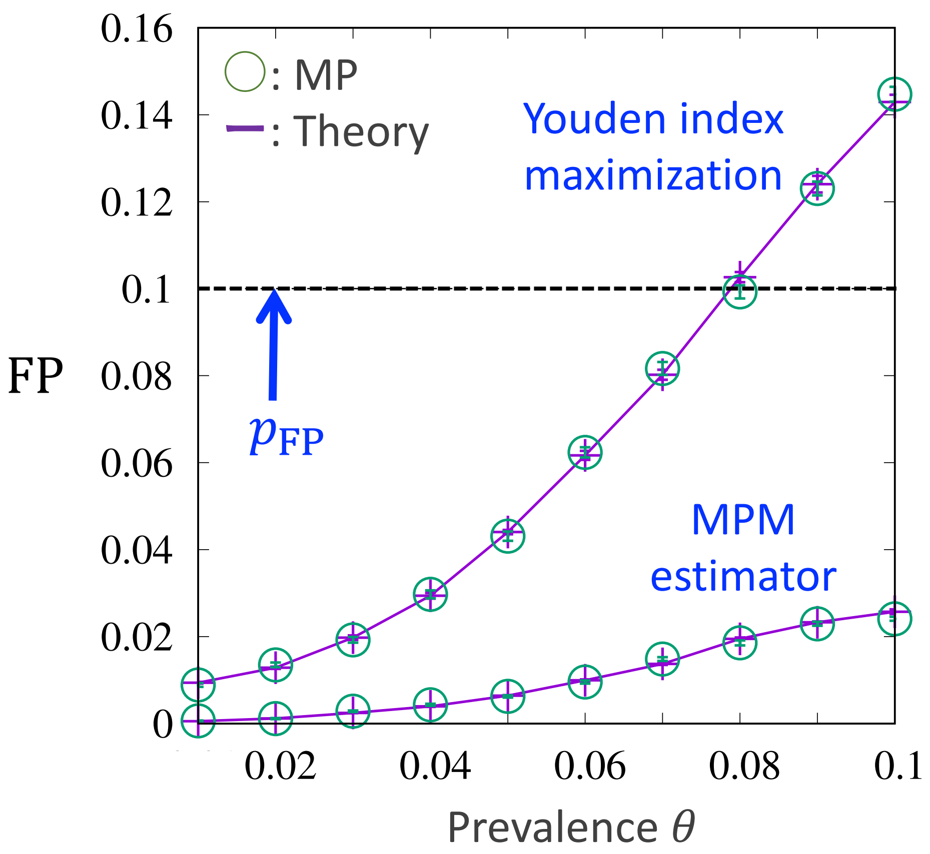
  $\lambda_{\mathrm{FN}} = 1 - \theta$

  $\lambda_{\mathrm{FP}} = \theta$

Error correction by group testing

- $N = 1000, M/N = 0.5, p_{\mathrm{TP}} = 0.95, p_{\mathrm{FP}} = 0.1$, pool size of 10 for Bayesian optimal setting
  - M: Number of tests (pools), N: Number of patients





- Youden index maximization:

$$\lambda_{\mathrm{FN}} = \lambda_{\mathrm{FP}} = 1/2$$

- MPM estimator :

$$\lambda_{\mathrm{FN}} = 1 - \theta$$
$$\lambda_{\mathrm{FP}} = \theta$$

■ $\mathrm{TP} > p_{\mathrm{TP}}$:

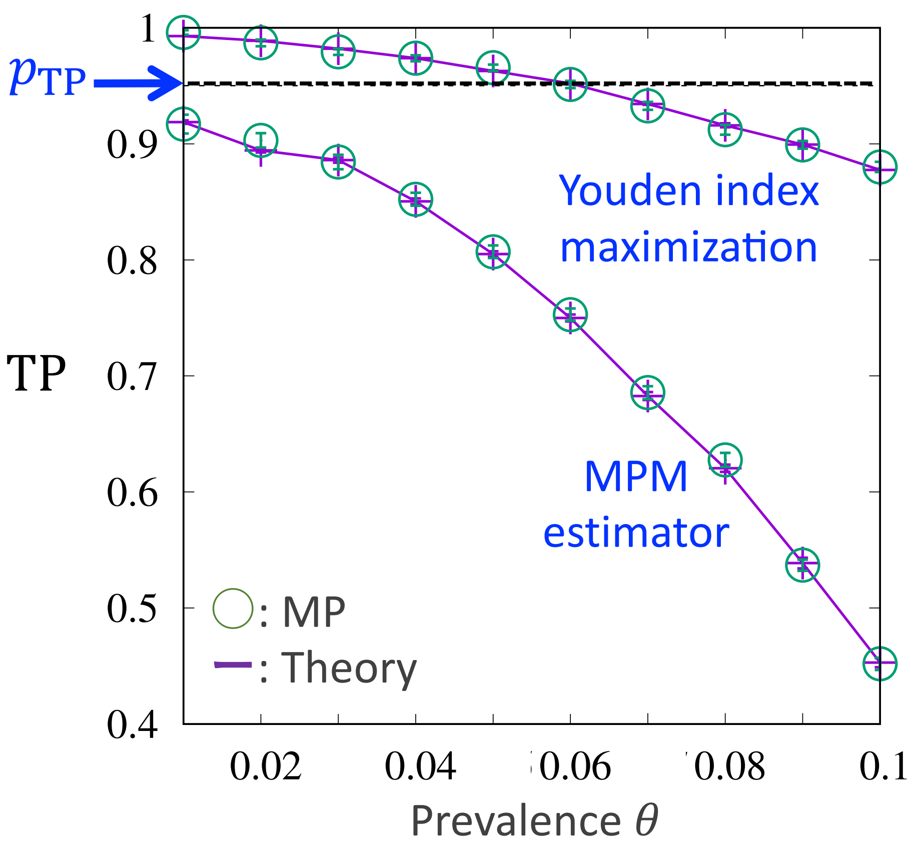
FN is corrected by GT.

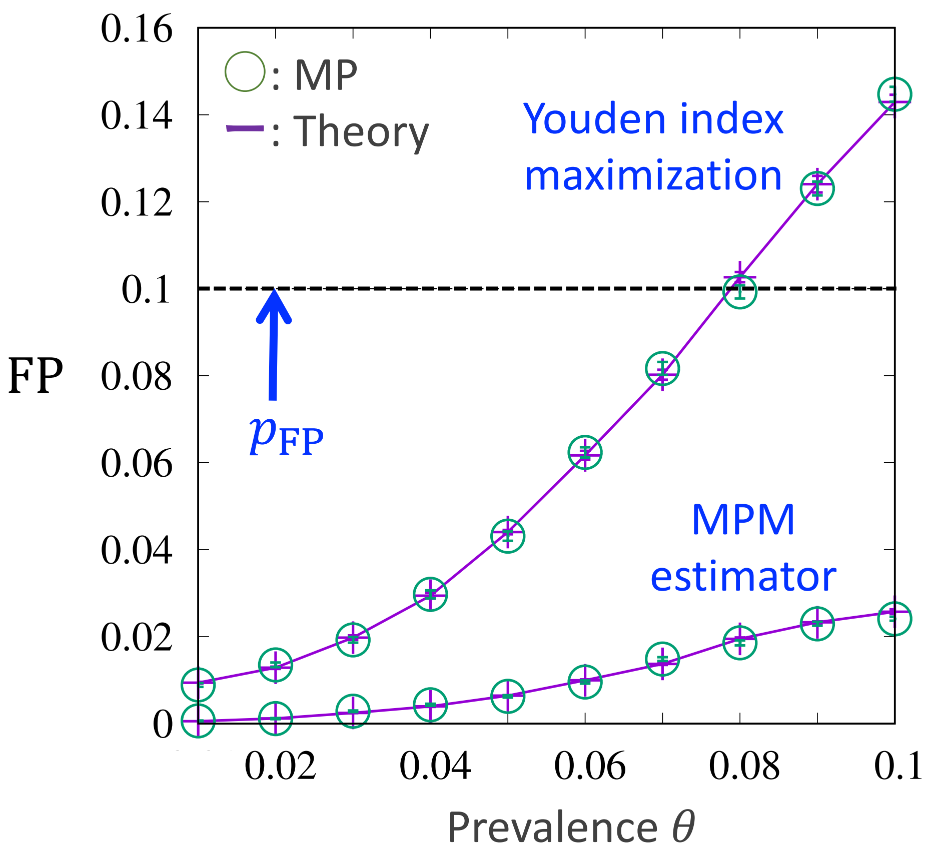
■ $\mathrm{FP} < p_{\mathrm{FP}}$:

FP is corrected by GT.

Error correction by group testing

- $N = 1000, M/N = 0.5, p_{\mathrm{TP}} = 0.95, p_{\mathrm{FP}} = 0.1$, pool size of 10 for Bayesian optimal setting
  - M: Number of tests (pools), N: Number of patients





- Youden index maximization:
  $$\lambda_{\mathrm{FN}} = \lambda_{\mathrm{FP}} = 1/2$$
- MPM estimator :
  $$\lambda_{\mathrm{FN}} = 1 - \theta$$
  $$\lambda_{\mathrm{FP}} = \theta$$

■ $\mathrm{TP} > p_{\mathrm{TP}}$:

FN is corrected by GT.

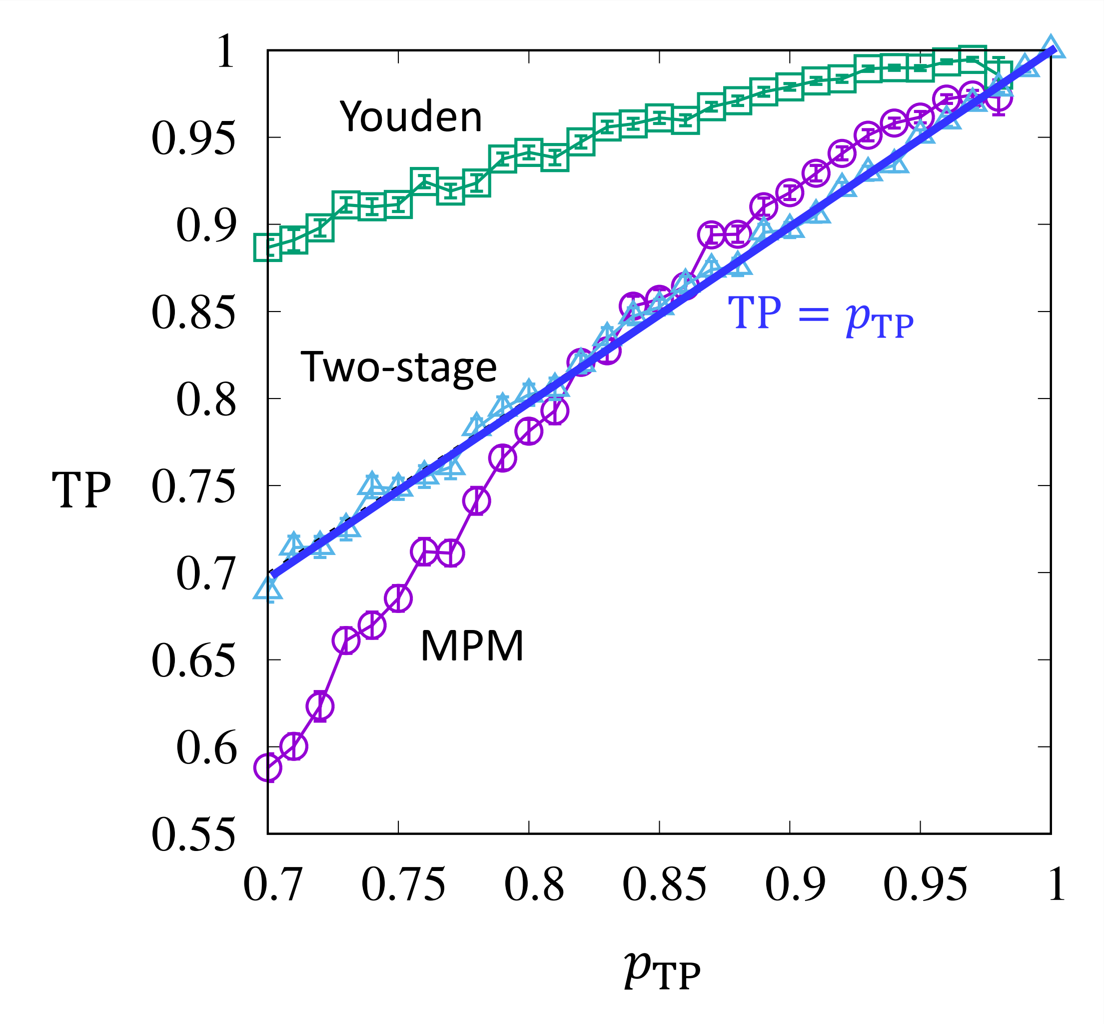
■ $\mathrm{FP} < p_{\mathrm{FP}}$:

FP is corrected by GT.

For the derivation of theoretical lines, see arXiv:2110.10877

Comparison with two-stage testing

- $N = 1000, M/N = 0.5, \theta = 0.05, p_{\mathrm{FP}} = 0$, pool size of 10 for Bayesian optimal setting



- Bayesian inference + optimal action outperform two-stage testing with smaller number of tests using appropriate cutoffs.

  ✓ In two-stage testing,
  the number of the 1st stage is 500,
  and the total number of tests is about 700.

Correspondence with replica method

Replica method

- Analytical method for obtaining partition function and thermal expectation in random systems such as spin-glass model
  - Techniques for averaging over quenched randomness

$$E_{\boldsymbol{J}}[\ln Z(\boldsymbol{J})] \to \lim_{n\to 0+} \frac{\partial}{\partial n} E_{\boldsymbol{J}}[Z^n(\boldsymbol{J})]$$

- $\boldsymbol{J}$: quenched randomness
- $Z(\boldsymbol{J})$: partition function

Replica method

- Analytical method for obtaining partition function and thermal expectation in random systems such as spin-glass model
  - Techniques for averaging over quenched randomness

$$E_{\boldsymbol{J}}[\ln Z(\boldsymbol{J})] \to \lim_{n\to 0+} \frac{\partial}{\partial n} E_{\boldsymbol{J}}[Z^n(\boldsymbol{J})]$$

- $\boldsymbol{J}$: quenched randomness
- $Z(\boldsymbol{J})$: partition function

- In the case of group testing, quenched randomness corresponds to patients' true states $\boldsymbol{x}^{(0)}$, test results $\boldsymbol{y}$, and pooling matrix $\boldsymbol{F}$.

- We can obtain the typical performance of Bayesian group testing by replica method.

$$E_{\boldsymbol{y},\boldsymbol{F},\boldsymbol{x}^{(0)}}[\mathrm{TP}], \qquad E_{\boldsymbol{y},\boldsymbol{F},\boldsymbol{x}^{(0)}}\left[\sum_{\boldsymbol{x}} x_i\, P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{F})\right] \text{ etc.}$$

Message passing → Replica method

- Theorem:

  The empirical distribution of messages at $N \to \infty$ characterizes the saddle point of the replica symmetric free energy.

- Empirical distributions :

$$\hat{p}_V^+(\theta_V) \equiv \lim_{N\to\infty} \frac{1}{N\theta C} \sum_{i=1}^{N} \sum_{\nu\in\mathcal{G}(i)} x_i^{(0)} \delta(\theta_{i\to\nu} - \theta_V)$$ …output messages from positive patients

$$\hat{p}_V^-(\theta_V) \equiv \lim_{N\to\infty} \frac{1}{N(1-\theta)C} \sum_{i=1}^{N} \sum_{\nu\in\mathcal{G}(i)} \left(1 - x_i^{(0)}\right) \delta(\theta_{i\to\nu} - \theta_V)$$ …output messages from negative patients

$$\hat{p}_F^+(\tilde{\theta}_F) \equiv \lim_{N\to\infty} \frac{1}{MK\theta} \sum_{\nu=1}^{M} \sum_{i\in\mathcal{M}(\nu)} x_i^{(0)} \delta\left(\tilde{\theta}_{\nu\to i} - \tilde{\theta}_F\right)$$ …input messages to positive patients

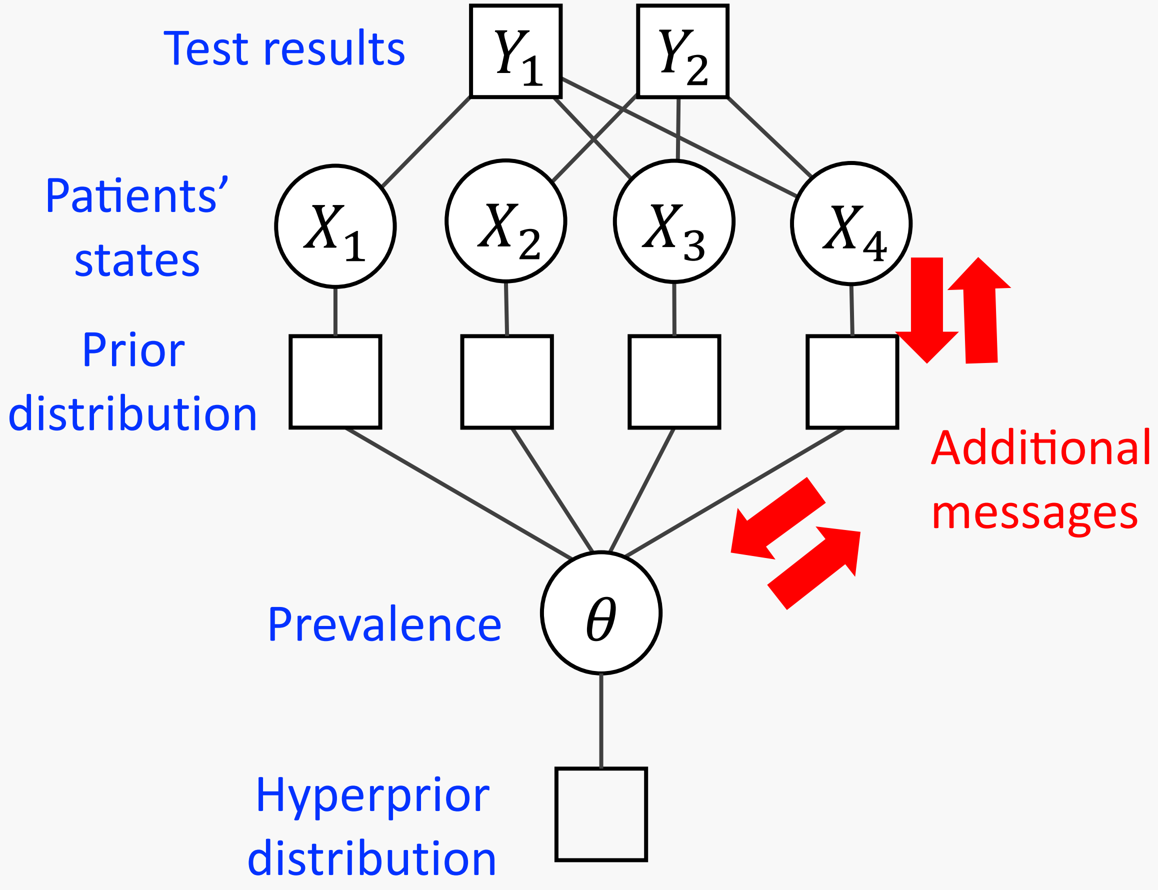
$$\hat{p}_F^-(\tilde{\theta}_F) \equiv \lim_{N\to\infty} \frac{1}{MK\theta} \sum_{\nu=1}^{M} \sum_{i\in\mathcal{M}(\nu)} \left(1 - x_i^{(0)}\right) \delta\left(\tilde{\theta}_{\nu\to i} - \tilde{\theta}_F\right)$$ …input messages to negative patients

Details are shown in arXiv:2110.10877.

Related Topics in Bayesian Group Testing

Estimation of prevalence

Graphical representation of prevalence estimation



- Prevalence is regarded as a dynamical variable.
- Prior distribution $\phi(\theta)$ is regarded as interaction between $\boldsymbol{X}$ and θ.
- Hyperprior distribution $\pi(\theta)$ is introduced.
  - We set $\pi(\theta)$ as the beta distribution, which is the conjugate of the Bernoulli distribution.

Details are shown in JPSJ 89, 084001 (2020).

Adaptive pooling + message passing

- Adaptive group testing :

  Sequentially design pools based on test results in previous steps

- From the viewpoint of active learning :

  Take into account the pools that can output uncertain test results

  - Uncertain = The posterior distribution cannot describe the test results

- Uncertainty measure: Posterior predictive distribution

$$p\left(Y'\middle|\boldsymbol{Y},\boldsymbol{F},\widetilde{\boldsymbol{F}}\right)=\sum_{X} f\left(Y'\middle|\boldsymbol{X},\widetilde{\boldsymbol{F}}\right)P(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{F})$$

- New pool $\widetilde{\boldsymbol{F}}^*$ is given by $\widetilde{\boldsymbol{F}}^* = \underset{F}{\operatorname{argmax}}\, \mathcal{S}\left(\boldsymbol{Y},\boldsymbol{F},\widetilde{\boldsymbol{F}}\right)$

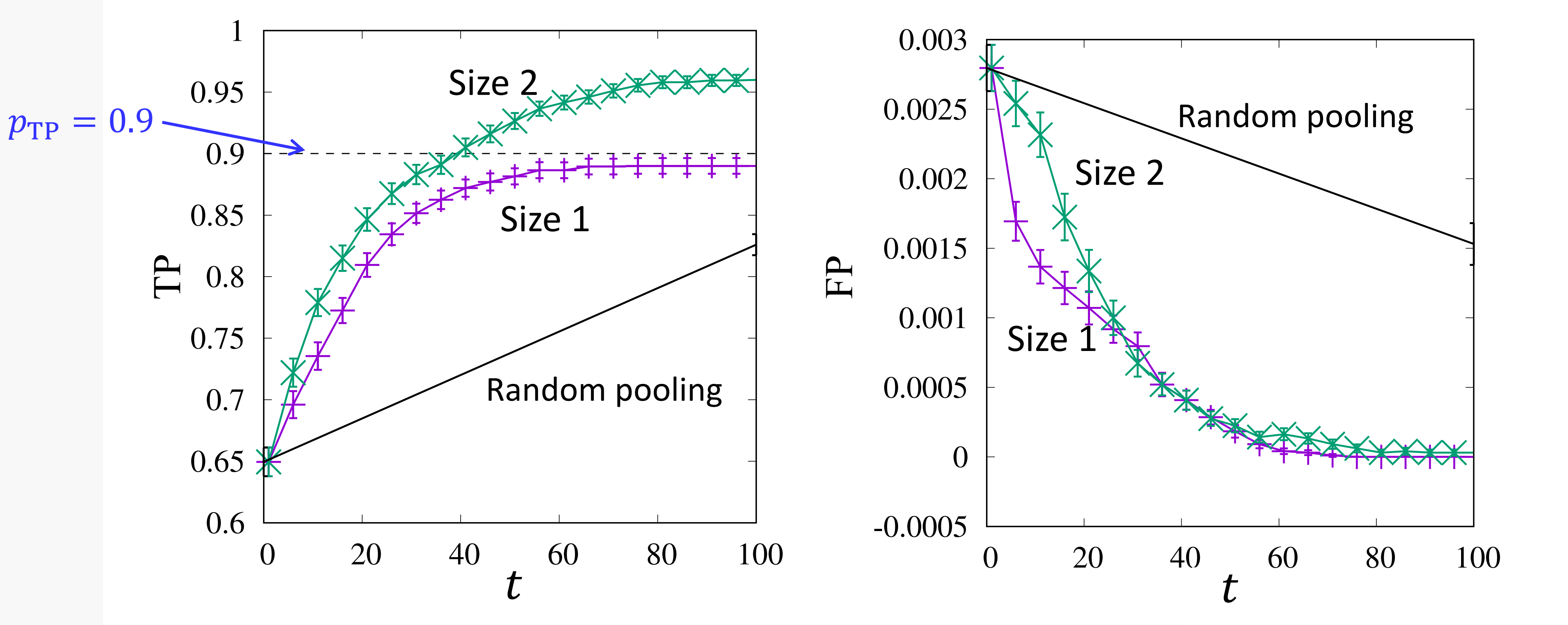
  where $\mathcal{S}\left(\boldsymbol{Y},\boldsymbol{F},\widetilde{\boldsymbol{F}}\right) = -\sum_{Y'} p\left(Y'\middle|\boldsymbol{Y},\boldsymbol{F},\widetilde{\boldsymbol{F}}\right)\ln p\left(Y'\middle|\boldsymbol{Y},\boldsymbol{F},\widetilde{\boldsymbol{F}}\right)$

- $\boldsymbol{F}$: Existing pools
- $\boldsymbol{Y}$: Obtained test results
- $\widetilde{\boldsymbol{F}}$: Candidate new pool
- Y': New test results

Details are shown in PRE 103, 022110 (2021).

Dependence on number of tests

- $N = 1000, \theta = 0.02, p_{\mathrm{TP}} = 0.9, p_{\mathrm{FP}} = 0.05,$ Cutoff = 0.5
- t times adaptive testing on pools with size of 1 or 2 after 300 random tests



Adaptive pooling reduces the number of tests required for TP > p_{TP}.

Summary

- Bayesian inference for group testing
  - A variant of the sparse estimation problem with discrete variables
  - Consideration as a statistical physics model
- Identification of patients' states as a Bayesian decision problem
  - Optimal action is given by the Bayesian optimal setting
- Approximate calculation of the posterior marginal probability

Summary

- Bayesian inference for group testing
  - A variant of the sparse estimation problem with discrete variables
  - Consideration as a statistical physics model
- Identification of patients' states as a Bayesian decision problem
  - Optimal action is given by the Bayesian optimal setting
- Approximate calculation of the posterior marginal probability

The discussion in this talk is not restricted to group testing.

Let's map real problems to statistical physics and enjoy statistical inference!