スパース線形回帰に対する半解析的ブートストラップ法

樺島祥介
東京大学大学院理学系研究科
知の物理学研究センター＆物理学専攻
共同研究者：小渕智之（京大），高橋昂（東工大）

自己紹介

樺島祥介（かばしま　よしゆき）

- 福岡県生まれ，その後大阪で育つ
- 1989　京都大学理学部卒（主に物理学）
- 1993　同大学院物理学第一専攻博士後期課程中退
- 1993　奈良女子大学理学部物理学科助手
- 1996　東京工業大学大学院総合理工学研究科知能システム科学専攻講師
  その後，同助教授，教授
- 2016　同大学情報理工学院数理・計算科学系教授（改組に伴う配置転換）
- 2020　東京大学大学院理学系研究科　知の物理学研究センター教授

研究内容：統計力学と情報科学の境界領域の研究．具体的な研究対象は，誤り訂正符号，暗号，CDMAマルチユーザ検出，圧縮符号，圧縮センシング，ランダム行列，機械学習，スピングラスモデルなど（に現れる協力現象とそれを理学的に解明または工学的に利用するための数理）

動機：統計力学の方法論的可能性に関心

概要

- ブートストラップ法とは
- スパース線形回帰とは
- スパース線形回帰に対する半解析的ブートストラップ法
  - レプリカ法
  - 確率伝搬法（belief propagation: BP）
  - 期待値伝搬法（expectation propagation: EP）

  スピングラス理論
  由来の計算法
- まとめ

機械学習と信頼性

- 機械学習:与えられたデータ

$$D^M = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_M, y_M)\}$$

  の背後にある入出力関係を表現するように，仮定した関係 $y = f(\boldsymbol{x}; \boldsymbol{w})$ の範囲内でパラメータ $\boldsymbol{w}$ を調整する．

- 論点
  - 構造・表現設計: $y = f(\boldsymbol{x}; \boldsymbol{w})$ の関数形をどう与えるか？
    - 深層ネット，BM，RBM，GP，SVM，リカレントネット，リザバー計算，etc.
      - 近年，盛んに研究されている．
  - 学習アルゴリズム開発:どうやって，最良パラメータを探索するか？
    - 最急降下法，貪欲法，凸緩和，確率推論，etc.
      - 近年，盛んに研究されている．
  - 信頼性評価:得られたパラメータはどのくらい良いのか？
    - CV，情報量基準，経験ベイズ法，ブートストラップ法，etc.
      - 今後，盛んに研究されるようになる（はず）．

ブートストラップ法とは

- Efron (1981) : 与えられた単一のデータセットから，学習によって得られたパラメータの信頼性を評価する数値的方法.
  - 単一のデータセットからエラーバーの評価が可能！

ブートストラップ法の概要

与えられたデータセット: $D^M = \left\{ (\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_M, y_M) \right\}$

経験分布: $P_{\mathrm{emp}}(\boldsymbol{x}, y) = \dfrac{1}{M} \sum_{\mu=1}^{M} \delta(\boldsymbol{x} - \boldsymbol{x}_\mu) \delta(y - y_\mu)$

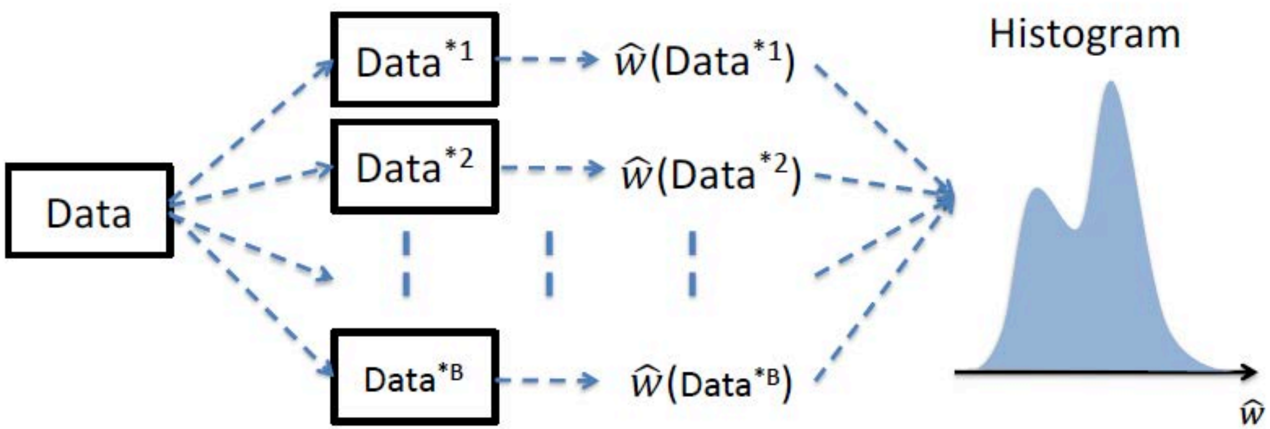
以下を何度も繰り返し，学習により得られるパラメータの分布を求める

1. ブートストラップサンプルの生成

$$P_{\mathrm{emp}}(\boldsymbol{x}, y) \Rightarrow D^{*M_{\mathrm{B}}} = \left\{ (\boldsymbol{x}_1^*, y_1^*), (\boldsymbol{x}_2^*, y_2^*), \ldots, (\boldsymbol{x}_{M_{\mathrm{B}}}^*, y_{M_{\mathrm{B}}}^*) \right\}$$

2. パラメータ推定／学習

$$D^{*M_{\mathrm{B}}} \Rightarrow \hat{\boldsymbol{w}}\left(D^{*M_{\mathrm{B}}} \right)$$



ブートストラップ法の利点と欠点

- 利点
  - 汎用性
    - 独立同分布の仮定さえ成り立てば，任意のデータ，任意の推定・学習法にも使える
  - アルゴリズムの簡便さ
    - 経験分布からリサンプリングしてパラメータ推定することを繰り返すだけ

- 欠点
  - 計算量的負荷
    - 数値的にパラメータ推定する場合は計算量的に辛い．
  - 理論的根拠
    - 母集団の統計とブートストラップサンプルの統計との関係は一般に非自明．

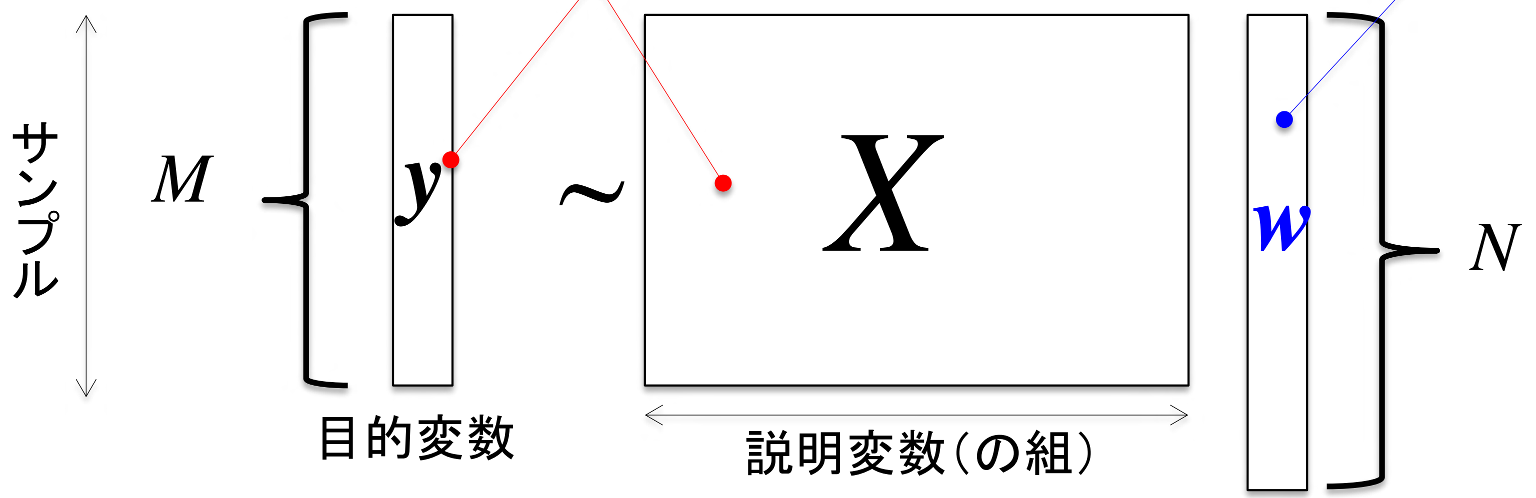
本日のお話

- スパース線形回帰に対し，レプリカ法と平均場近似を組み合わせることで，ブートストラップを実際に実行することなしに，ブートストラップ統計を近似的に評価する方法を開発する.
  - 必要計算量は1サンプルに対する従来の推定法と同程度.
    → 計算量的負荷の大幅な軽減.

スパース線形回帰とは

<u>線形回帰／信号復元</u>

データ $D=\left\{\left(\boldsymbol{x}_\mu, y_\mu\right)\right\}_{\mu=1}^{M}$ $\xrightarrow{\text{推定}}$ パラメータ $\boldsymbol{w}$



すべての説明変数が目的変数に関与しているとは限らない→「変数選択」が必要

<u>l_1罰則項を用いた変数選択（LASSO: Ishikawa (1994), Tibshirani (1996)）</u>

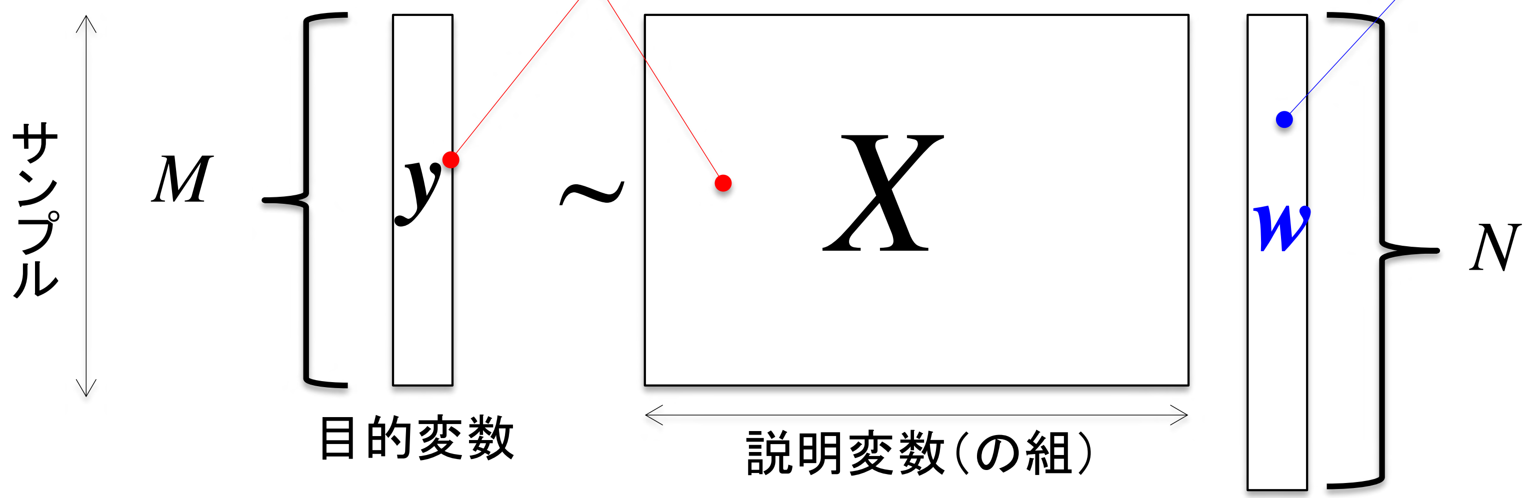
以下のコスト関数の最小化により，不要な変数は削減される（ゼロになる）

$$E_{\mathrm{LASSO}}\left(\boldsymbol{w} \middle| \lambda, D\right)=\frac{1}{2}\sum_{\mu=1}^{M}\left(y_\mu-\sum_{i=1}^{N}x_{\mu i}w_i\right)^2+\lambda\sum_{i=1}^{N}\left|w_i\right|$$

スパース線形回帰とは

<u>線形回帰／信号復元</u>

データ $D=\left\{\left(\boldsymbol{x}_\mu, y_\mu\right)\right\}_{\mu=1}^{M}$ $\xrightarrow{\text{推定}}$ パラメータ $\boldsymbol{w}$



すべての説明変数が目的変数に関与しているとは限らない→「変数選択」が必要

<u>l_1罰則項を用いた変数選択（LASSO: Ishikawa (1994), Tibshirani (1996)）</u>

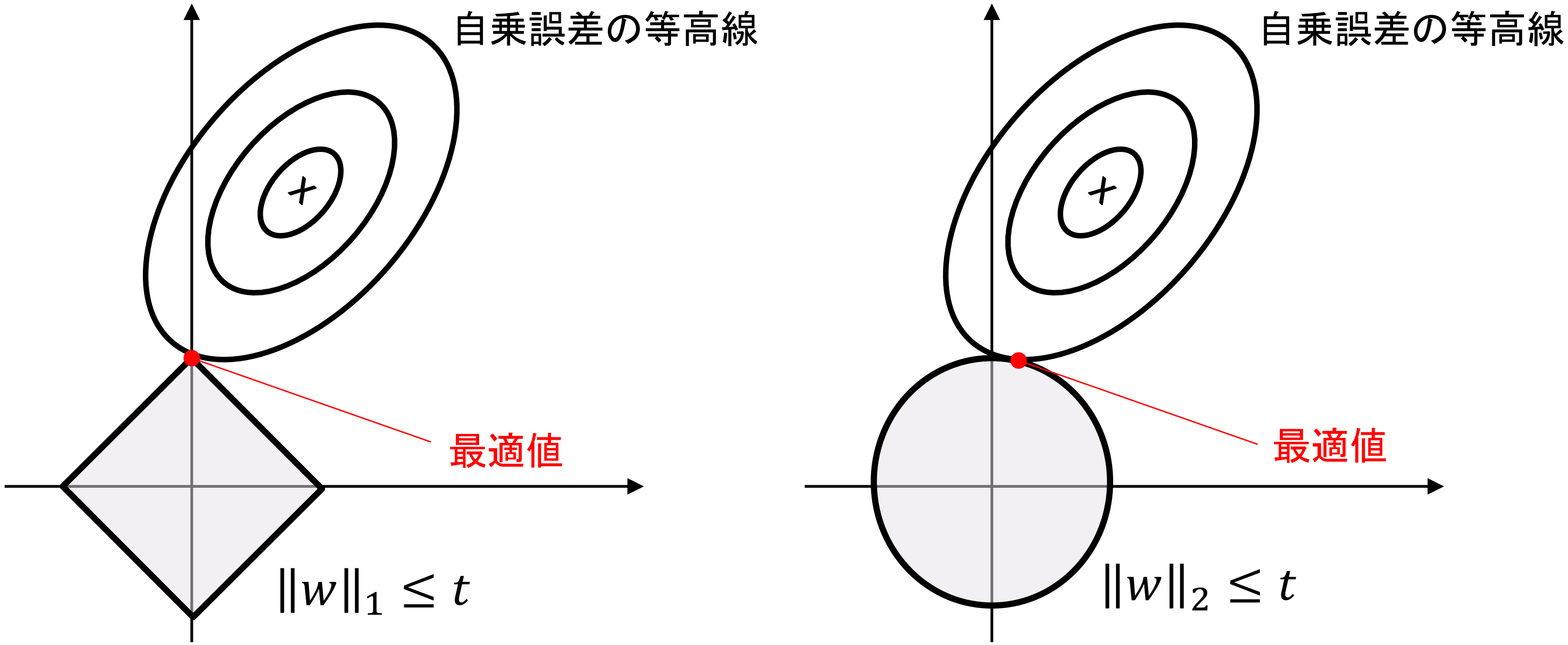
深層学習が「予測」を目的とした機械学習であるのに対し，スパース（線形）回帰は
「仕組みの理解」を目的としてもちいられることが多い

$$E_{\mathrm{LASSO}}\left(\boldsymbol{w}\middle|\lambda, D\right)=\frac{1}{2}\sum_{\mu=1}^{M}\left(y_\mu-\sum_{i=1}^{N}x_{\mu i}w_i\right)^2+\lambda\sum_{i=1}^{N}\left|w_i\right|$$

l_1罰則項でなぜスパース推定ができるのか？

⇒制約付き最小化問題として考える

$\min \frac{1}{2}\sum_{\mu=1}^{M}\|y_\mu - x_\mu \cdot w\|_2^2$ subj. to $\|w\|_1 \leq t$　　$\min \frac{1}{2}\sum_{\mu=1}^{M}\|y_\mu - x_\mu \cdot w\|_2^2$ subj. to $\|w\|_2 \leq t$

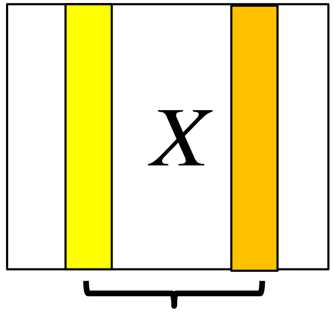


最適値は軸上に位置する
⇒　スパース推定

最適値は一般的な点
⇒　スパース推定にならない

共線性とLASSO

$\boldsymbol{x}_i \qquad \boldsymbol{x}_j \simeq c_1 \boldsymbol{x}_i + c_0$



線形従属に近い

- 線形回帰では，説明変数間の相関が強いと
推定結果の信頼性が低下する ⇒ *共線性(colinearity)*

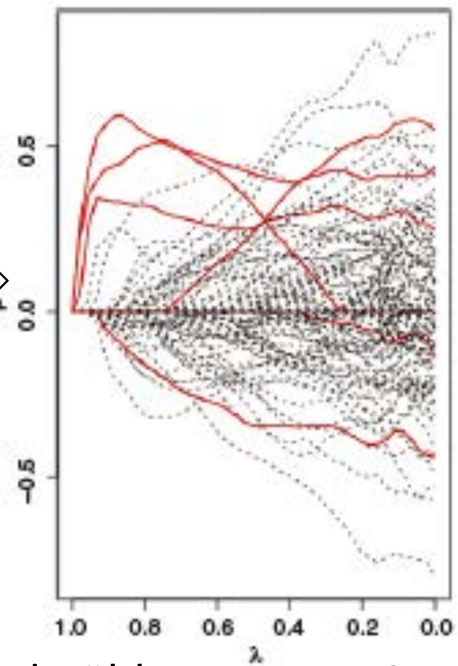
$$\min_{\boldsymbol{w}} \left\{ \frac{1}{2} \| \boldsymbol{y} - X\boldsymbol{w} \| \right\} \Rightarrow \hat{\boldsymbol{w}} = \underline{\left(X^\top X \right)^{-1}} X^\top \boldsymbol{y}$$

ランク落ちに近くなり不安定化

- LASSOは共線性に比較的強いとされている
<u>ただし，それにも限度がある</u>

Meinshausen and Buhlmann (2010)より
・枯草菌のビタミンB2生産速度115サンプル
・4088種の遺伝子発現量データでLASSO回帰
・（有意な）6種類以外の遺伝子はランダムシャッフル

$$\hat{\boldsymbol{w}}^{\mathrm{LASSO}} = \arg\min_{\boldsymbol{w}} \left\{ \frac{1}{2} \| \boldsymbol{y} - X\boldsymbol{w} \|_2^2 + \lambda \sum_{i=1}^{N} |w_i| \right\}$$



Meinshausen and Bühlmann, J. R. Statist. Soc. B 72, 417 (2010)より

ブートストラップを利用した安定化

- この問題に対し，Bach (2008)，Meinshausen and Buhlmann (2010) はブートストラップを用いることで安定的な変数選択ができることを示している

Stability selection (SS): ブートストラップ (+ **λ** のランダム化)

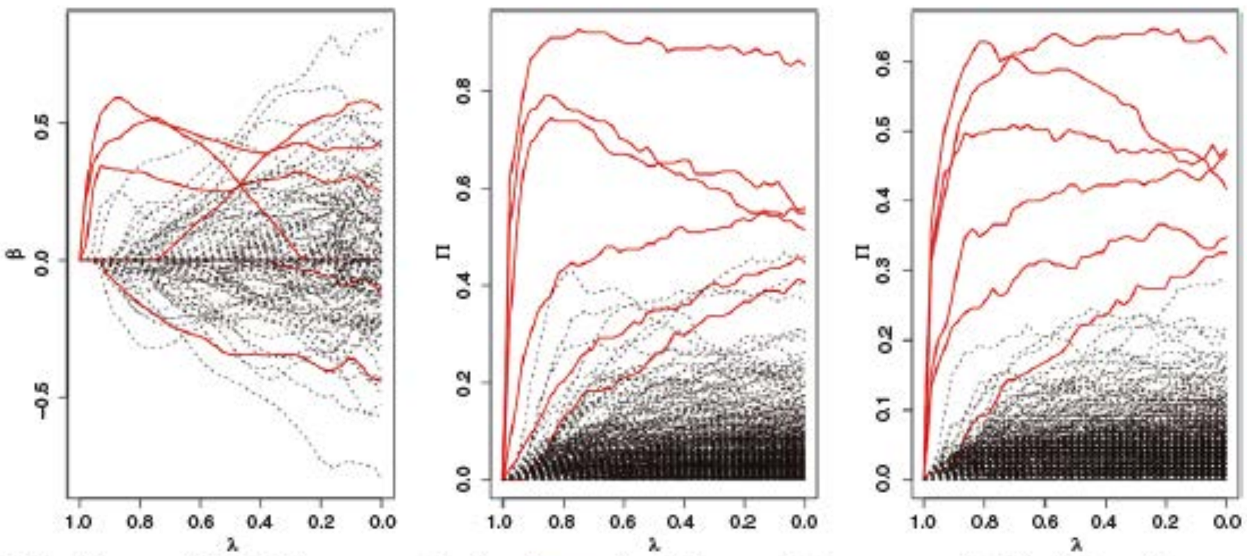


Fig. 1. (a) Lasso path for the vitamin gene expression data set (——, paths of six non-permuted genes; -------, paths of the 4082 permuted genes; selecting a model with all six unpermuted genes invariably means selecting a large number of irrelevant noise variables), (b) stability path of the lasso (the first four variables chosen with stability selection are truly non-permuted variables) and (c) stability path for the randomized lasso with weakness $\alpha = 0.2$, introduced in Section 3.1 (now all six non-permuted variables are chosen before any noise variable enters the model)

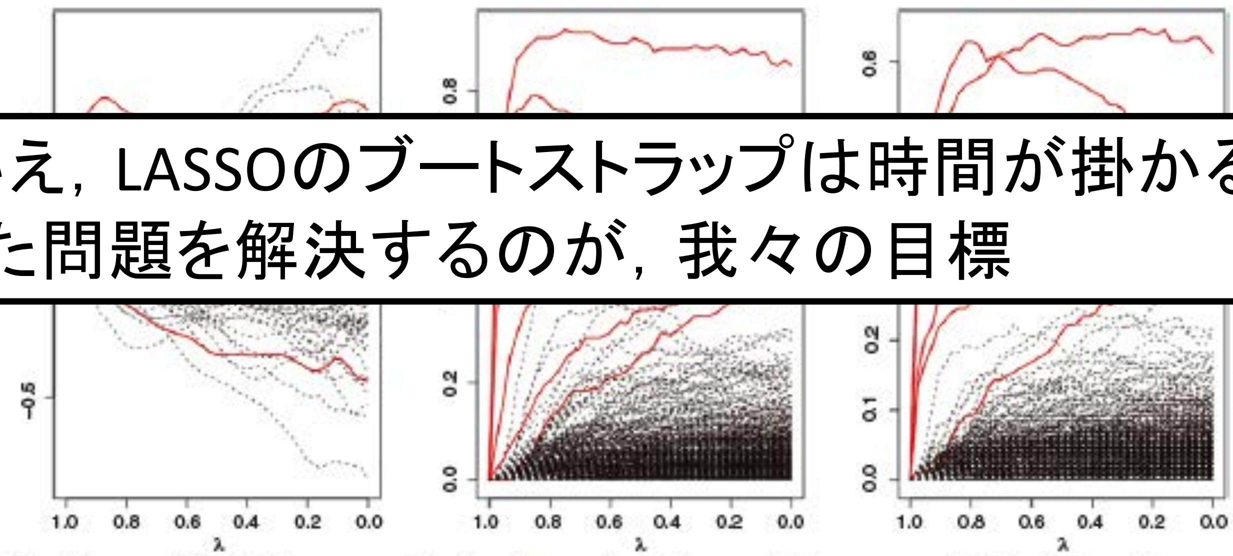
SSを使うと（中右）ノイズ解に埋もれなくなる

Meinshausen and Bühlmann, J. R. Statist. Soc. B 72, 417 (2010)より

ブートストラップを利用した安定化

- この問題に対し，Bach (2008)，Meinshausen and Buhlmann (2010) はブートストラップを用いることで安定的な変数選択ができることを示している

Stability selection (SS): ブートストラップ (+ **λ** のランダム化)



とはいえ，LASSOのブートストラップは時間が掛かる
こうした問題を解決するのが，我々の目標

SSを使うと（中右）ノイズ解に埋もれなくなる

Fig. 1. (a) Lasso path for the vitamin gene expression data set (——, paths of six non-permuted genes; -------, paths of the 4082 permuted genes; selecting a model with all six unpermuted genes invariably means selecting a large number of irrelevant noise variables), (b) stability path of the lasso (the first four variables chosen with stability selection are truly non-permuted variables) and (c) stability path for the randomized lasso with weakness $\alpha = 0.2$, introduced in Section 3.1 (now all six non-permuted variables are chosen before any noise variable enters the model)

Meinshausen and Bühlmann, J. R. Statist. Soc. B 72, 417 (2010)より

LASSOに関するブートストラップ

ブートストラップサンプルは各データが何回リサンプリングされたかで記述できる

0　　2　　1　　1

例）　$D^4 = \left\{ (\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), (\boldsymbol{x}_3, y_3), (\boldsymbol{x}_4, y_4) \right\}$

ブートストラップ

$$D^{*4} = \left\{ (\boldsymbol{x}_2, y_2), (\boldsymbol{x}_3, y_3), (\boldsymbol{x}_2, y_2), (\boldsymbol{x}_4, y_4) \right\}$$

$\Leftrightarrow$　$\boldsymbol{c} = (c_1, c_2, c_3, c_4) = (0,2,1,1)$　リサンプリングされた回数

リサンプリングされた回数は多項分布（～ポアソン分布の積）

$$P(\boldsymbol{c}) = \frac{M_{\mathrm{B}}!}{c_1! c_2! \ldots c_{M_{\mathrm{B}}}!} \left(\frac{1}{M} \right)^{M_{\mathrm{B}}} \simeq \prod_{\mu=1}^{M_{\mathrm{B}}} e^{-M_{\mathrm{B}}/M} \frac{(M_{\mathrm{B}}/M)^{c_\mu}}{c_\mu!}$$

＃比較的単純な分布で表現できる

統計力学的定式化

- 最小化問題⇔ゼロ温度の統計力学

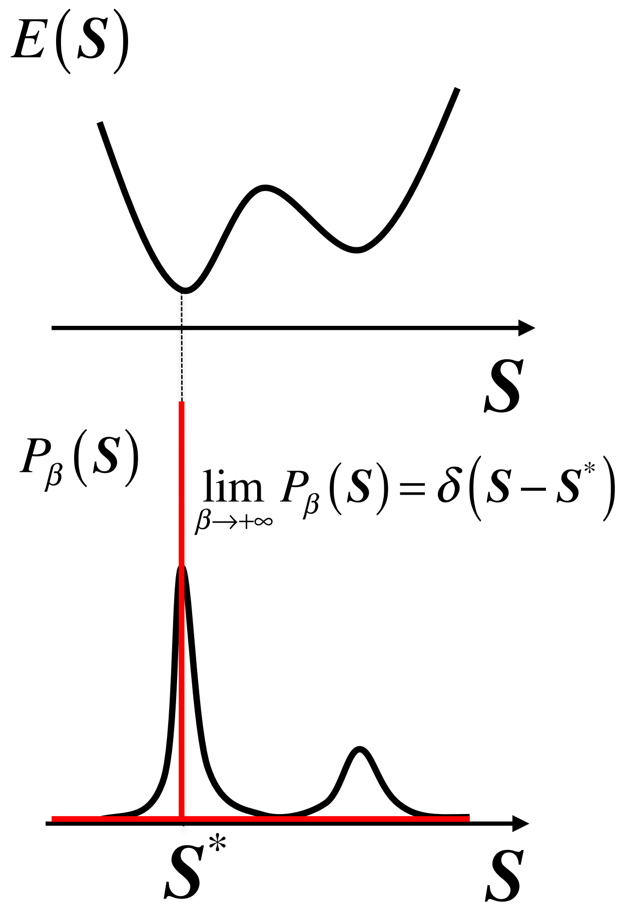
コスト関数⇔エネルギー関数

$$E_{\mathrm{LASSO}}^{\mathrm{BS}}\left(\boldsymbol{w}\middle|\boldsymbol{c},D,\lambda\right)=\frac{1}{2}\sum_{\mu=1}^{M_{\mathrm{B}}}c_{\mu}\left(y_{\mu}-\sum_{i=1}^{N}x_{\mu i}w_{i}\right)^{2}+\lambda\sum_{i=1}^{N}\left|w_{i}\right|$$

LASSO推定量⇔基底状態

$$\hat{\boldsymbol{w}}^{\mathrm{LASSO}}\left(\boldsymbol{c},D,\lambda\right)=\arg\min_{\boldsymbol{w}}\left\{E_{\mathrm{LASSO}}^{\mathrm{BS}}\left(\boldsymbol{w}\middle|\boldsymbol{c},D,\lambda\right)\right\}$$

$$=\lim_{\beta\to\infty}\frac{1}{Z\left(\boldsymbol{c},D,\lambda,\beta\right)}\int\boldsymbol{w}e^{-\beta E_{\mathrm{LASSO}}^{\mathrm{BS}}\left(\boldsymbol{w}\middle|\boldsymbol{c},D,\lambda\right)}d\boldsymbol{w}$$



⇩

分配関数 $Z\left(\boldsymbol{c},D,\lambda,\beta\right)\triangleq\int e^{-\beta E_{\mathrm{LASSO}}^{\mathrm{BS}}\left(\boldsymbol{w}\middle|\boldsymbol{c},D,\lambda\right)}d\boldsymbol{w}$ の評価に帰着

スピングラス問題⇔ブートストラップ

- スピングラス問題

ランダム結合

$$P(\mathbf{J}) = \prod_{\langle ij \rangle} P(J_{ij})$$

エネルギー関数

$$H(\mathbf{S}|\mathbf{J}) = -\sum_{\langle ij \rangle} J_{ij} S_i S_j - h \sum_i S_i$$

熱平均

$$\langle S_i \rangle = \sum_S S_i \frac{e^{-\beta H(\mathbf{S}|\mathbf{J})}}{Z(\mathbf{J},\beta)}$$

配位平均

$$\left[\langle S_i \rangle^p \right] = \int \langle S_i \rangle^p P(\mathbf{J}) d\mathbf{J}$$

$$(p = 1,2,\ldots)$$

- ブートストラップ

ブートストラップサンプル

$$P(\boldsymbol{c}) = \frac{M_{\mathrm{B}}!}{c_1! c_2! \ldots c_{M_{\mathrm{B}}}!} \left(\frac{1}{M} \right)^{M_{\mathrm{B}}} \simeq \prod_{\mu=1}^{M_{\mathrm{B}}} e^{-M_{\mathrm{B}}/M} \frac{(M_{\mathrm{B}}/M)^{c_\mu}}{c_\mu !}$$

コスト関数

$$E_{\mathrm{LASSO}}^{\mathrm{BS}}(\boldsymbol{w}|\boldsymbol{c}, D, \lambda) = \frac{1}{2} \sum_{\mu=1}^{M_{\mathrm{B}}} c_\mu \left(y_\mu - \sum_{i=1}^{N} x_{\mu i} w_i \right)^2 + \lambda \sum_{i=1}^{N} |w_i|$$

LASSO解

$$\hat{\mathbf{w}}^{\mathrm{LASSO}} = \lim_{\beta \to \infty} \int \frac{\mathbf{w} e^{-\beta E_{\mathrm{LASSO}}^{\mathrm{BS}}(\mathbf{w}|\mathbf{c}, D, \lambda)}}{Z(\mathbf{c}, D, \lambda, \beta)} d\mathbf{w}$$

ブートストラップ平均

$$\left[\left(\hat{w}_i^{\mathrm{LASSO}} \right)^p \right]_{\mathbf{c}} = \sum_{\mathbf{c}} \left(\hat{w}_i^{\mathrm{LASSO}} \right)^p P(\mathbf{c})$$

$$(p = 1,2,\ldots)$$

スピングラス問題⇔ブートストラップ

- スピングラス問題

ランダム結合

$$P(\mathbf{J}) = \prod_{\langle ij \rangle} P(J_{ij})$$

エネルギー関数

$$H(\mathbf{S}|\mathbf{J}) = -\sum J_{ij} S_i S_j - h \sum S_i$$

配位平均

$$\left[\langle S_i \rangle^p \right] = \int \langle S_i \rangle^p P(\mathbf{J}) d\mathbf{J} \quad (p = 1, 2, \ldots)$$

- ブートストラップ

ブートストラップサンプル

$$P(\boldsymbol{c}) = \frac{M_{\mathrm{B}}!}{c_1! c_2! \ldots c_{M_{\mathrm{B}}}!} \left(\frac{1}{M} \right)^{M_{\mathrm{B}}} \simeq \prod_{\mu=1}^{M_{\mathrm{B}}} e^{-M_{\mathrm{B}}/M} \frac{(M_{\mathrm{B}}/M)^{c_\mu}}{c_\mu !}$$

コスト関数

$$E_{\mathrm{LASSO}}^{\mathrm{BS}}(\boldsymbol{w}|\boldsymbol{c}, D, \lambda) = \frac{1}{2} \sum_{\mu=1}^{M_{\mathrm{B}}} c_\mu \left(y_\mu - \sum_{i=1}^{N} x_{\mu i} w_i \right)^2 + \lambda \sum_{i=1}^{N} |w_i|$$

ブートストラップ平均

$$\left[\left(\hat{w}_i^{\mathrm{LASSO}} \right)^p \right]_{\mathbf{c}} = \sum_{\mathbf{c}} \left(\hat{w}_i^{\mathrm{LASSO}} \right)^p P(\mathbf{c}) \quad (p = 1, 2, \ldots)$$

ブートストラップ平均の評価に
スピングラスの解析法が使えるのではないか？！

レプリカ法

- 配位平均(ブートストラップ平均)を求めることは難しい
  - 根本的な原因は，配位平均を取る量の「分母」に分配関数があること

$$\left[\langle O\rangle^{k}\right]=\int\prod_{(ij)}dJ_{ij}P\left(J_{ij}\right)\left(\frac{\operatorname*{Tr}_{\mathbf{S}}O(\mathbf{S})e^{-\beta H(\mathbf{S}|\mathbf{J})}}{\operatorname*{Tr}_{\mathbf{S}}e^{-\beta H(\mathbf{S}|\mathbf{J})}}\right)^{k}\qquad Z_{\beta}(\mathbf{J})=\operatorname*{Tr}_{\mathbf{S}}e^{-\beta H(\mathbf{S}|\mathbf{J})}$$

- この困難は$n\geq k\in\mathbb{N}$に対する「拡張された平均」では解消される

$$\left[\langle O\rangle^{k}\right]_{n}\triangleq\frac{\left[Z_{\beta}^{n}(J)\langle O\rangle^{k}\right]}{\left[Z_{\beta}^{n}(J)\right]}=\frac{\int\prod_{(ij)}dJ_{ij}P\left(J_{ij}\right)\left(\operatorname*{Tr}_{\mathbf{S}}e^{-\beta H(\mathbf{S}|\mathbf{J})}\right)^{n-k}\left(\operatorname*{Tr}_{\mathbf{S}}O(\mathbf{S})e^{-\beta H(\mathbf{S}|\mathbf{J})}\right)^{k}}{\int\prod_{(ij)}dJ_{ij}P\left(J_{ij}\right)\left(\operatorname*{Tr}_{\mathbf{S}}e^{-\beta H(\mathbf{S}|\mathbf{J})}\right)^{n}}$$

別々に配位平均を求めればよいので“分母”の問題は生じない

レプリカ法

- 拡張された平均＝n個の結合「レプリカ系」の同時分布に関する平均
  - 配位平均を先に取り，レプリカ系の同時分布を作る

$$P_\beta\left(\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n\right) \triangleq \frac{\int\prod_{(ij)} dJ_{ij} P\left(J_{ij}\right)\exp\left(-\sum_{a=1}^{n}\beta H\left(\mathbf{S}^a \middle| \mathbf{J}\right)\right)}{\int\prod_{(ij)} dJ_{ij} P\left(J_{ij}\right) Z_\beta^n(\mathbf{J})}$$

$$\left[\langle O\rangle^k\right]_n = \mathop{\mathrm{Tr}}_{\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n} P_\beta\left(\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n\right) O\left(\mathbf{S}^1\right) O\left(\mathbf{S}^2\right)\cdots O\left(\mathbf{S}^k\right)$$

$$H\left(\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n\right) \triangleq -\frac{1}{\beta}\ln\left(\int\prod_{(ij)} dJ_{ij} P\left(J_{ij}\right)\exp\left(-\sum_{a=1}^{n}\beta H\left(\mathbf{S}^a \middle| \mathbf{J}\right)\right)\right)$$

と定義すれば，これはレプリカ系に対するカノニカル分布の平均（熱平均）

$\Rightarrow$ 通常の統計力学の知見が使える！

レプリカ法

- 以下の手順(レプリカトリック)で(熱平均量の)配位平均を求める

1. $n=1,2,\ldots\in\mathbb{N}$に対し，レプリカ系での平均を$n$の関数としてもとめる

$$\left[\langle O\rangle^k\right]_n=\operatorname*{Tr}_{\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n} P_\beta\left(\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n\right)O\left(\mathbf{S}^1\right)O\left(\mathbf{S}^2\right)\cdots O\left(\mathbf{S}^k\right)$$

2. この表現を$n\to 0$に外挿し，所望の配位平均を評価する

$$\left[\langle O\rangle^k\right]_n\triangleq\frac{\left[Z^n(J)\langle O\rangle^k\right]}{\left[Z^n(J)\right]}=\frac{\int\prod_{(ij)}dJ_{ij}P\left(J_{ij}\right)\left(\operatorname*{Tr}_{\mathbf{S}}e^{-\beta H(\mathbf{S}|\mathbf{J})}\right)^{n-k}\left(\operatorname*{Tr}_{\mathbf{S}}O(\mathbf{S})e^{-\beta H(\mathbf{S}|\mathbf{J})}\right)^k}{\int\prod_{(ij)}dJ_{ij}P\left(J_{ij}\right)\left(\operatorname*{Tr}_{\mathbf{S}}e^{-\beta H(\mathbf{S}|\mathbf{J})}\right)^n}$$

$$\xrightarrow{n\to 0}\int\prod_{(ij)}dJ_{ij}P\left(J_{ij}\right)\left(\frac{\operatorname*{Tr}_{\mathbf{S}}O(\mathbf{S})e^{-\beta H(\mathbf{S}|\mathbf{J})}}{\operatorname*{Tr}_{\mathbf{S}}e^{-\beta H(\mathbf{S}|\mathbf{J})}}\right)^k=\left[\langle O\rangle^k\right]$$

レプリカ法

- 通常，分配関数の対数（自由エネルギー，キュムラント母関数）に関する配位平均の評価として象徴的に表現されることが多い

1. $n = 1,2,\ldots \in \mathbb{N}$ に対し，分配関数のn次モーメントをnの関数としてもとめる

$$\left[Z^n(\mathbf{J}) \right] = \mathop{\mathrm{Tr}}_{\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n} \int \prod_{\langle ij \rangle} dJ_{ij} P\left(J_{ij}\right) \exp\left(-\beta H\left(\mathbf{S} \middle| \mathbf{J}\right)\right)$$

2. この表現を$n \to 0$に外挿し，自由エネルギーの配位平均を評価する

$$\left[\ln Z(\mathbf{J}) \right] = \lim_{n \to 0} \frac{\partial}{\partial n} \ln \left[Z^n(\mathbf{J}) \right]$$

- ただし，実際的な問題では，分配関数の評価は計算量的に困難

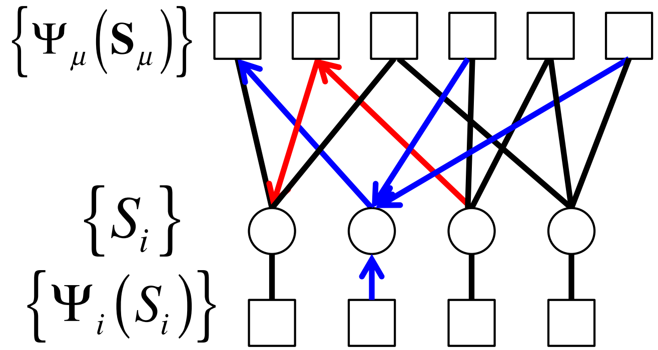
$\Rightarrow$ 適当な平均場近似を使って評価する

確率伝搬法(belief propagation: BP)

確率伝搬法＝転送行列法の一般化

$$P(\mathbf{S})=\frac{1}{Z}\prod_{\mu}\Psi_{\mu}\left(\mathbf{S}_{\mu}\right)\times\prod_{i}\Psi_{i}\left(S_{i}\right)$$

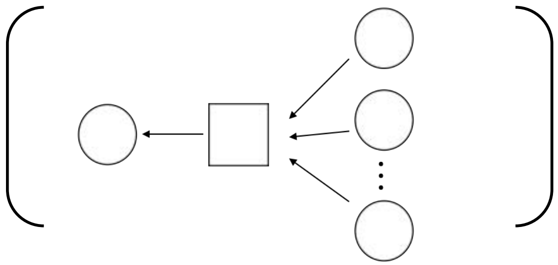
変数の依存関係を2部グラフで表現



- 次の関数方程式を反復法で解く

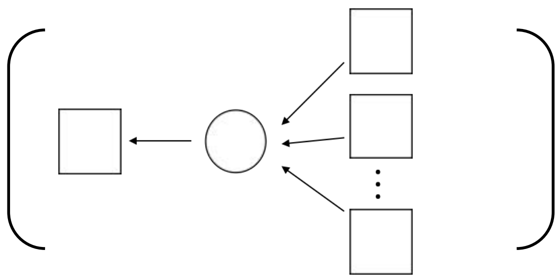
水平ステップ

$$m_{\mu\to i}\left(S_i\right)=\alpha_{\mu\to i}\sum_{\mathbf{S}_{\mu}\backslash S_i}\Psi_{\mu}\left(\mathbf{S}_{\mu}\right)\prod_{j\in\partial\mu\backslash i}m_{j\to\mu}\left(S_j\right)$$



垂直ステップ

$$m_{i\to\mu}\left(S_i\right)=\alpha_{i\to\mu}\Psi_{i}\left(S_i\right)\prod_{\nu\in\partial i\backslash\mu}m_{\nu\to i}\left(S_i\right)$$



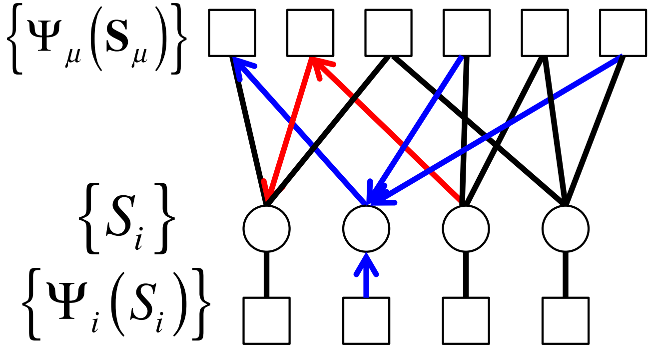
確率伝搬法(belief propagation: BP)

- 自由エネルギーの近似値

$$\ln Z \simeq \sum_{\mu} \ln\left(\sum_{\mathbf{S}_\mu} \Psi_\mu(\mathbf{S}_\mu) \prod_{i\in\partial\mu} m_{i\to\mu}(S_i) \right)$$

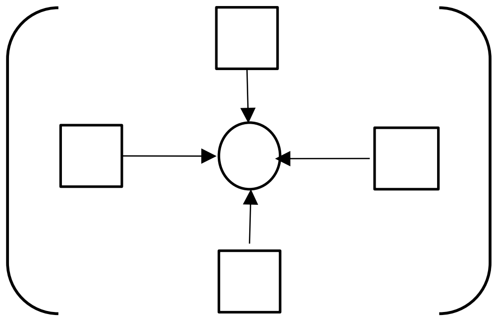
$$-\sum_{\mu}\sum_{i\in\partial\mu} \ln\left(\sum_{S_i} m_{\mu\to i}(S_i) m_{i\to\mu}(S_i) \right) + \sum_i \ln\left(\sum_{S_i} \Psi_i(S_i) \prod_{\mu\in\partial i} m_{\mu\to i}(S_i) \right)$$

変数の依存関係を2部グラフで表現



- 周辺分布の近似値

$$P_i(S_i) = \sum_{\mathbf{S}\backslash S_i} P(\mathbf{S}) \simeq \alpha_i \Psi_i(S_i) \prod_{\mu\in\partial i} m_{\mu\to i}(S_i)$$



#グラフにループがなければ厳密な評価

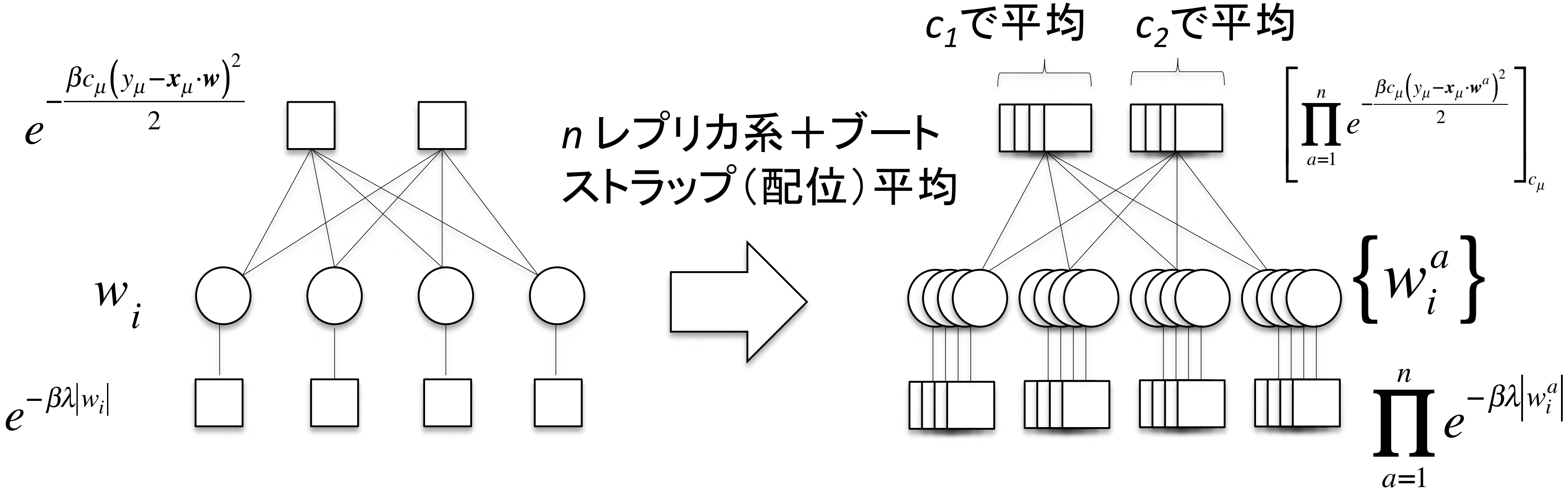
BP＋レプリカ法

$$P(\boldsymbol{w}|\boldsymbol{c}) \propto \prod_{\mu=1}^{M} e^{-\frac{\beta c_\mu \left(y_\mu - \boldsymbol{x}_\mu \cdot \boldsymbol{w}\right)^2}{2}} \times \prod_{i=1}^{N} e^{-\beta\lambda|w_i|}$$

ブートストラップデータに対するボルツマン因子

n レプリカ系＋ブートストラップ（配位）平均

$$P\left(\boldsymbol{w}^1, \boldsymbol{w}^2, \ldots, \boldsymbol{w}^n\right) = \mathop{\mathrm{Tr}}_{\boldsymbol{c}} \left(\prod_{a=1}^{n} P\left(\boldsymbol{w}^a \middle| \boldsymbol{c}\right) \right) P(\boldsymbol{c}) \propto \prod_{\mu=1}^{M_{\mathrm{B}}} \left[\prod_{a=1}^{n} e^{-\frac{\beta c_\mu \left(y_\mu - \boldsymbol{x}_\mu \cdot \boldsymbol{w}^a\right)^2}{2}} \right]_{c_\mu} \times \prod_{i=1}^{N} \prod_{a=1}^{n} e^{-\beta\lambda_i \left|w_i^a\right|}$$



BP＋レプリカ法

- 水平ステップ

$$m_{\mu\to i}\left(w_i^1,w_i^2,\ldots,w_i^n\right)=\alpha_{\mu\to i}\int\left[\prod_{a=1}^{n}e^{-\frac{\beta c_\mu\left(y_\mu-\boldsymbol{x}_\mu\cdot\boldsymbol{w}^a\right)^2}{2}}\right]_{c_\mu}\prod_{j\neq i}m_{j\to\mu}\left(w_j^1,w_j^2,\ldots,w_j^n\right)\prod_{j\neq i,a}dw_j^a$$

- 垂直ステップ

$$m_{i\to\mu}\left(w_i^1,w_i^2,\ldots,w_i^n\right)=\alpha_{i\to\mu}\prod_{a=1}^{n}e^{-\beta\lambda\left|w_i^a\right|}\prod_{\nu\neq\mu}m_{\nu\to i}\left(w_i^1,w_i^2,\ldots,w_i^n\right)$$

- 周辺分布

$$b_i\left(w_i^1,w_i^2,\ldots,w_i^n\right)=\alpha_i\prod_{a=1}^{n}e^{-\beta\lambda\left|w_i^a\right|}\prod_{\mu=1}^{M}m_{\mu\to i}\left(w_i^1,w_i^2,\ldots,w_i^n\right)$$

レプリカ対称仮定

- レプリカ対称性（＝レプリカ添字の入れ替えに対する不変性）を仮定し，中心極限定理を利用すると，メッセージを以下の形にパラメトライズすることができる

$$
\begin{aligned}
& m_{i\to\mu}\left(w_i^1, w_i^2, \ldots, w_i^n\right) \\
& \propto \exp\left(-\frac{\beta A_{i\to\mu}}{2}\sum_{a=1}^{n}\left(w_i^a\right)^2 + \frac{\beta^2 C_{i\to\mu}}{2}\left(\sum_{a=1}^{n} w_i^a\right)^2 + \beta B_{i\to\mu}\sum_{a=1}^{n} w_i^a - \beta\lambda\sum_{a=1}^{n}\left|w_i^a\right|\right) \\
& = \int \frac{dz e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \prod_{a=1}^{n} \exp\left(\beta\left(-\frac{A_{i\to\mu}}{2}\left(w_i^a\right)^2 + \left(B_{i\to\mu} + \sqrt{C_{i\to\mu}}z - \operatorname{sgn}\left(w_i^a\right)\right)w_i^a\right)\right)
\end{aligned}
$$

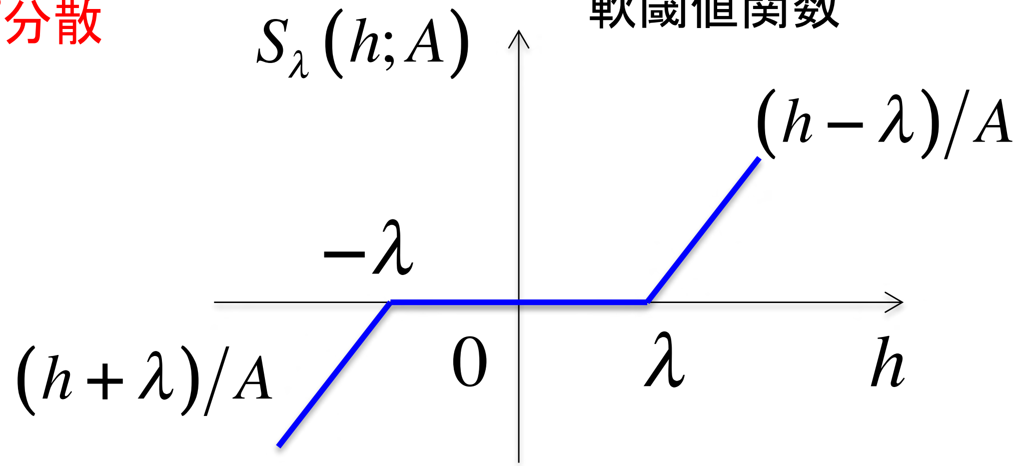
- $n \to 0, \beta \to \infty$ の極限でパラメータの更新式を導出する.
- 加えて，“１つ抜き”の影響を摂動的に扱う

導出されたアルゴリズム

ブートストラップ平均　　ブートストラップ分散

$$
\begin{cases}
w_i = \int S_\lambda\left(B_i + \sqrt{C_i} z; A_i\right) Dz \\
\chi_i = \int \frac{\partial}{\partial B_i} S_\lambda\left(B_i + \sqrt{C_i} z; A_i\right) Dz \\
W_i = \int \left(S_\lambda\left(B_i + \sqrt{C_i} z; A_i\right)\right)^2 Dz - w_i^2
\end{cases}
$$

軟閾値関数

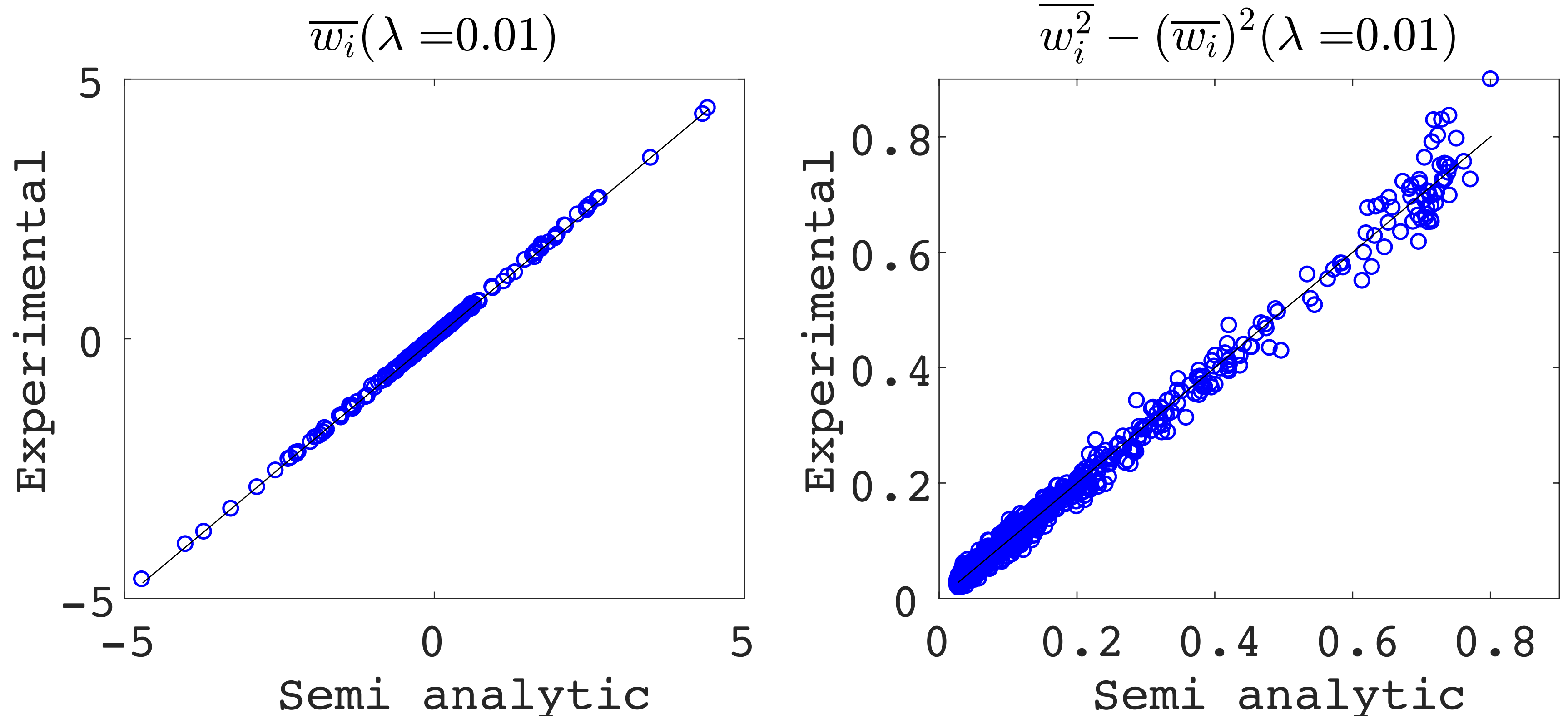


$$
\begin{cases}
V_\mu = \sum_{i=1}^{N} x_{\mu i}^2 \chi_i \\
U_\mu = \sum_{i=1}^{N} x_{\mu i}^2 W_i \\
a_\mu = \left[\frac{c}{1+cV_\mu}\right]_c \left(y_\mu - \sum_{i=1}^{N} x_{\mu i} w_i + V_\mu a_\mu\right) \\
R_\mu = a_\mu \Big/ \left[\frac{c}{1+cV_\mu}\right]_c
\end{cases}
$$

$$
\begin{cases}
A_i = \sum_{\mu=1}^{M_\mathrm{B}} x_{\mu i}^2 \left[\frac{c}{1+cV_\mu}\right]_c \\
B_i = \sum_{\mu=1}^{M_\mathrm{B}} x_{\mu i} a_\mu + A_i w_i \\
C_i = \sum_{\mu=1}^{M_\mathrm{B}} x_{\mu i}^2 \left(U_\mu \left[\left(\frac{c}{1+cV_\mu}\right)^2\right]_c + R_\mu^2 \left(\left[\left(\frac{c}{1+cV_\mu}\right)^2\right]_c - \left[\frac{c}{1+cV_\mu}\right]_c^2 \right) \right)
\end{cases}
$$

提案法と1000回のブートストラップ実験との比較

- N=1000, M=500, 真の非ゼロパラメータ数 K=200
  ノイズの分散σ^2=0.01, L_1罰則項の強さλ=0.01



1/1000の計算量でほぼ正確なブートストラップ統計量が得られる！

Stability Selection (SS)への応用

Meinshausen and Bühlmann, J. R. Statist. Soc. B 72, 417 (2010)

- ℓ_1正則化付き線形回帰（LASSO）

$$\hat{\boldsymbol{w}}(D_M, \boldsymbol{\lambda}) = \arg\min_{\boldsymbol{w}} \left\{ \frac{1}{2} \sum_{\mu=1}^{M} \left(y_\mu - \boldsymbol{x}_\mu^\top \boldsymbol{w} \right)^2 + \sum_{i=1}^{N} \lambda_i |w_i| \right\}$$

- Stability Selection (SS)：LASSOにもとづいた変数選択法
  - 基本アイデア
    1. ブートストラップ＋$\boldsymbol{\lambda}$のランダム化　→　推定値の分布を計算
    2. 分布から変数 w_i が非ゼロの確率 Π_i（陽性確率）を求める
    3. Π_i が大きければ有意，そうでなければ棄却
  - 実装
    - $$P(\boldsymbol{\lambda}; p, a) = \prod_{i=1}^{N} \{ p\delta(\lambda_i - \lambda/a) + (1-p)\delta(\lambda_i - \lambda) \}$$
    - $$M_B = \tau M$$

人工データ

- 生成モデル

$$\boldsymbol{y} = X\boldsymbol{w}_0 + \boldsymbol{\xi}$$

$$\boldsymbol{w}_0 \sim \prod_i \{\rho_0 \mathcal{N}(\cdot|0, 1/\rho_0) + (1-\rho_0)\delta(w_{0i})\} \ (\rho_0 = 0.2)$$

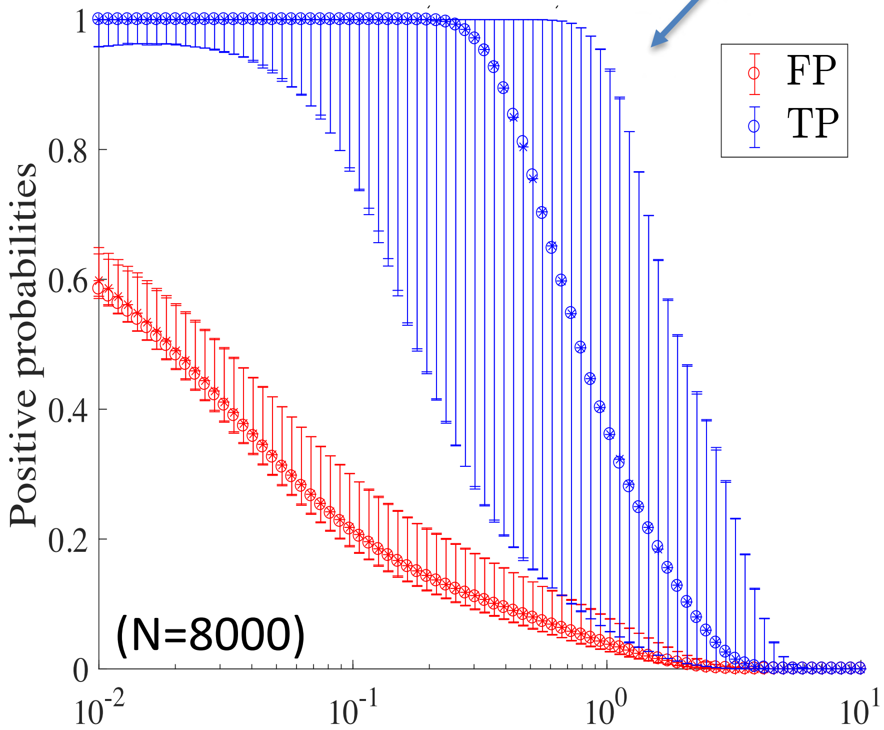
$$\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \sigma_\xi^2 I), \quad X_{\mu i} \sim \mathcal{N}(0, 1/N)$$

- SSのパラメータ: a=τ=p=0.5に固定
- 半解析的 vs 数値的リサンプリング
  – 半解析的：AMPRをMatlabでコーディング（not fast）
  – 数値的
    - Glmnetを使用
    - N_{samp}=1000サンプルから陽性確率を評価

人工データ

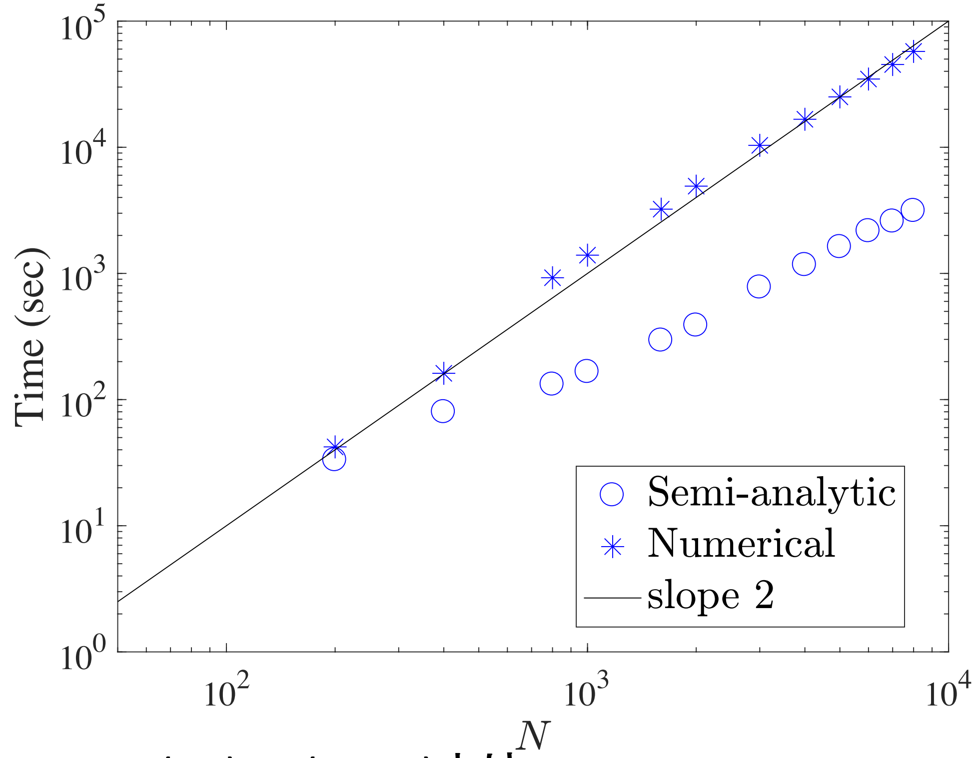
数値・半解析の同時プロット(ほぼ完全に一致)

Stability path



計算時間

M=2N



Obuchi and Kabashima, JMLR 20 (70), 1 (2019)より

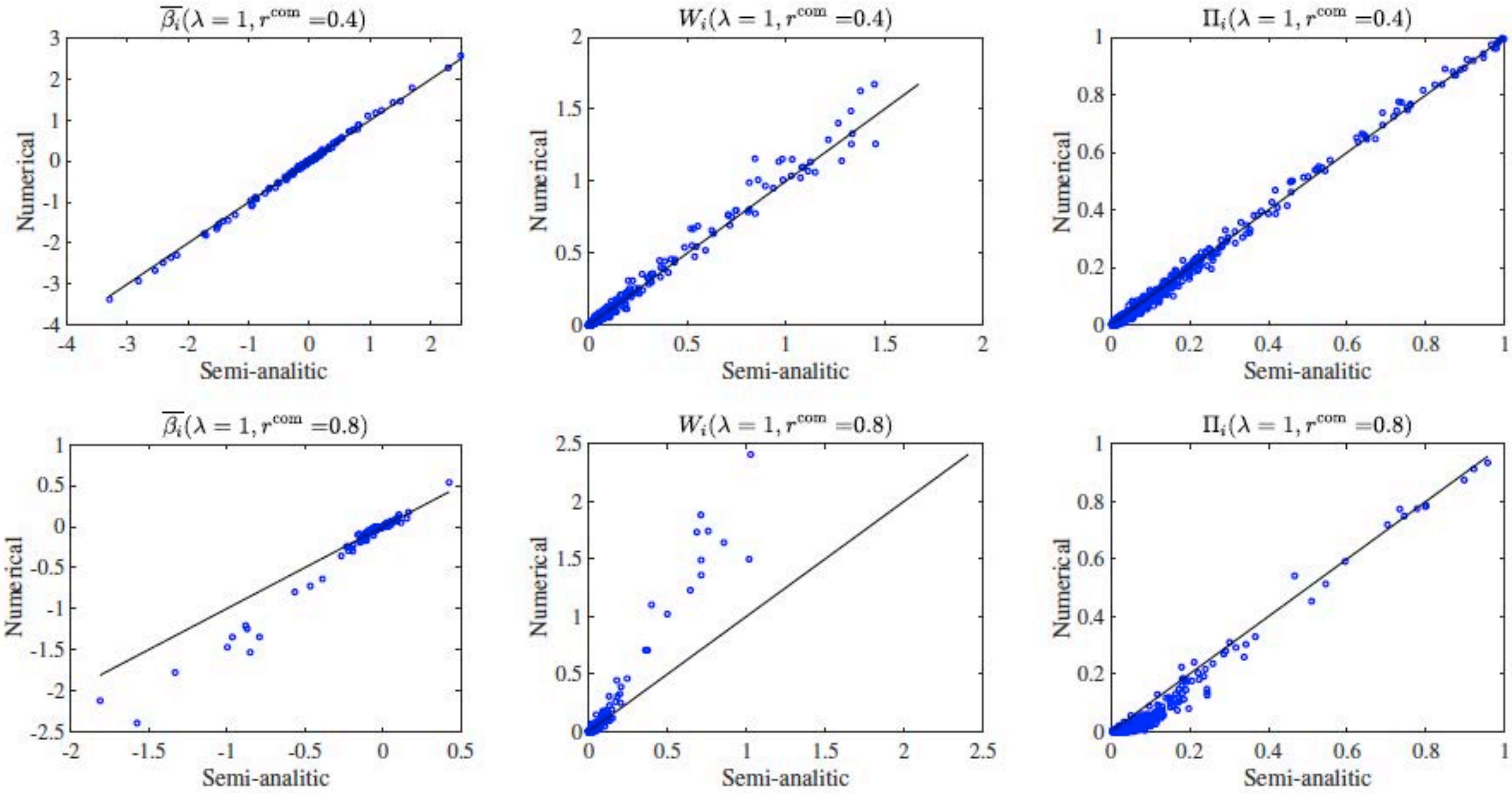
TP: サポート変数集合 $\{\Pi_i(\lambda)\}_{i\in S}$ のq=14,86パーセンタイル点

FP: 非サポート変数集合 $\{\Pi_i(\lambda)\}_{i\in S^c}$ のq=14,86パーセンタイル点

人工データ

説明変数間の相関の影響

計画行列（説明変数の行列）X を各列間に相関が生じるように設計



マイルドな相関 $(R \sim 0.16)$

強い相関 $(R \sim 0.64)$

Obuchi and Kabashima, JMLR 20 (70), 1 (2019)より

マイルドな相関：近似精度はそれほど悪化しない

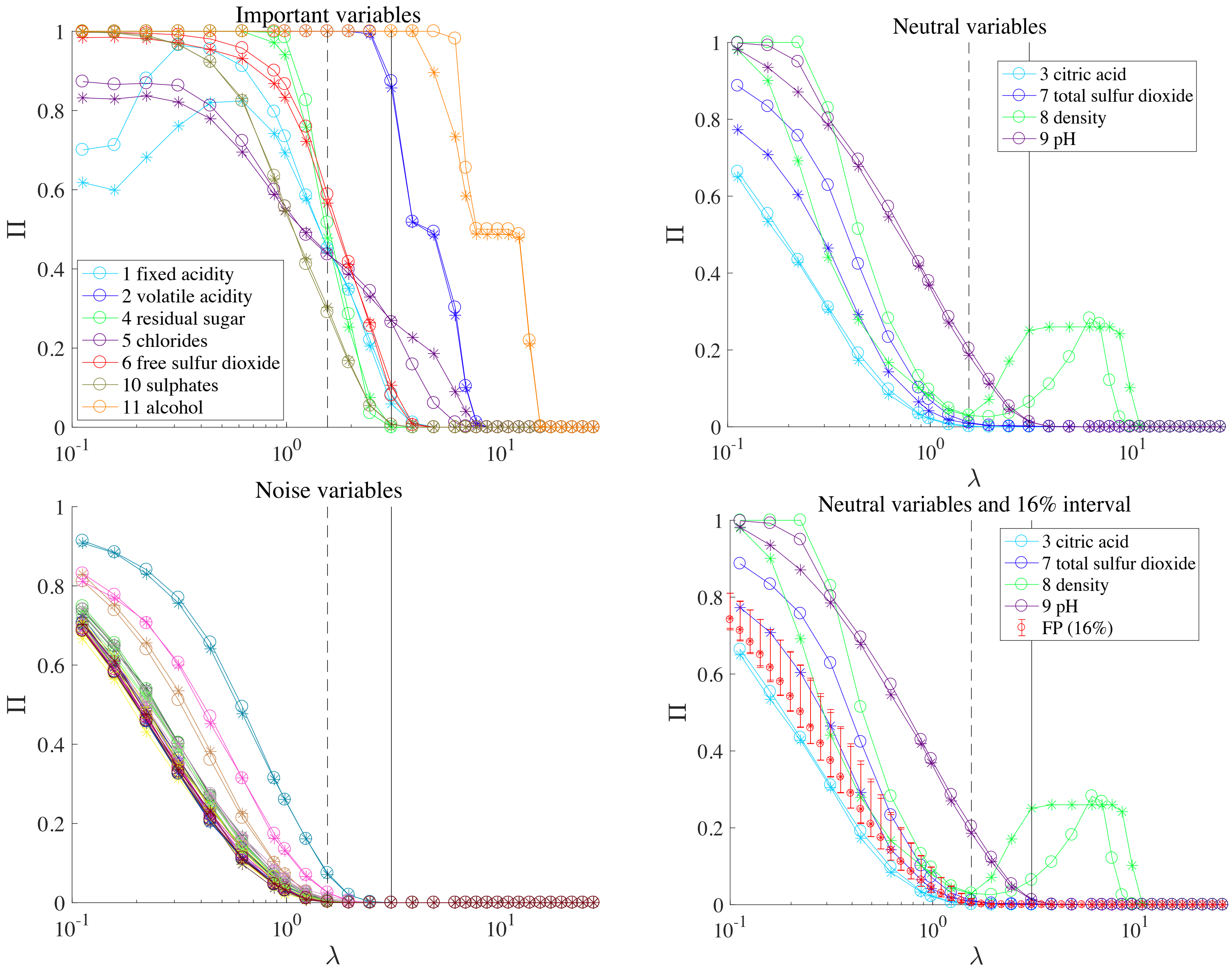
強い相関： 近似精度がかなり悪化

ワインデータ

- Wine quality dataset
  - UCI Machine learning repository から
  - 目的変数：ワインの味10段階評価
    - プロによるブラインドチェック，M=4898（白ワインのみ）
  - 説明変数：密度，酸度，糖度などの化学的性質，N=11

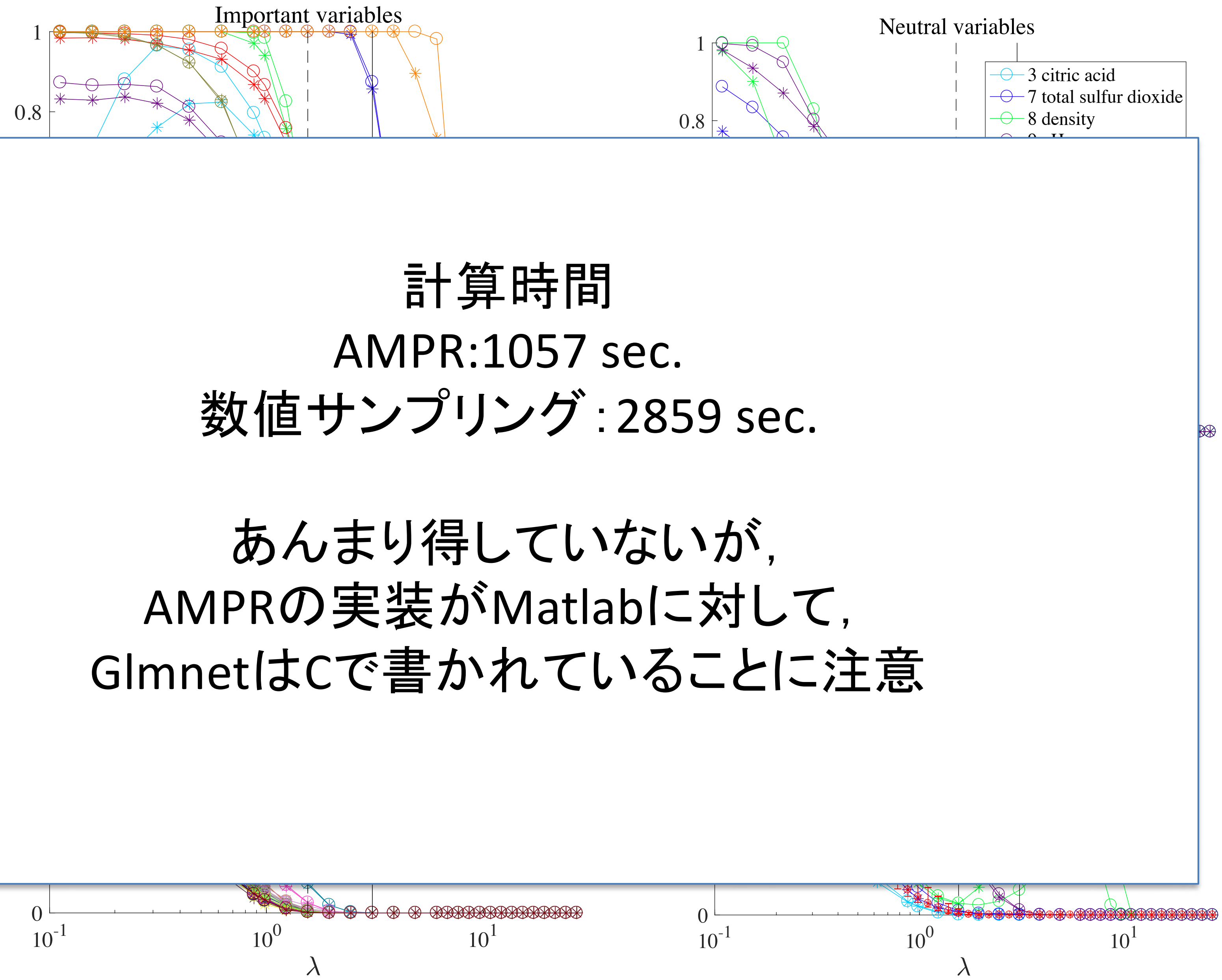
- ランダム説明変数（dummy variable）をN_{dummy}=689個追加
  - zero-meanガウシアンからのi.i.d.: $x_{\mu i} \sim \mathcal{N}(0, 1/N)$
  - 信頼区間の参照値として位置づけ
    - Dummy variableのStability path分布を利用

ワインデータ



Obuchi and Kabashima, JMLR 20 (70), 1 (2019)より

ワインデータ



計算時間
AMPR:1057 sec.
数値サンプリング : 2859 sec.

あんまり得していないが，
AMPRの実装がMatlabに対して，
GlmnetはCで書かれていることに注意

Obuchi and Kabashima, JMLR 20 (70), 1 (2019)より

BP＋レプリカ法に関するリマーク

- 計算コストは$O(NM)$/update
  - BP近似（AMP）による1サンプルのLASSO推定と同じオーダー
  - ブートストラップ回数倍の計算量削減
- 変数間の相関があまり強くなければ良い近似
  - 人工データ

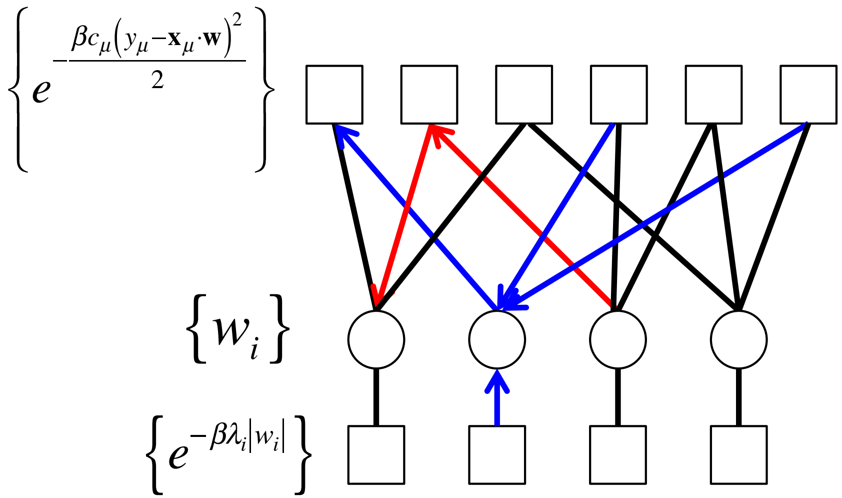
- ただし，やはり変数間の相関が強くなると近似精度は悪化
  - 人工データ
  - ワインデータの1 fixed acidity（酒石酸濃度）, 7 total sulfur dioxide（総亜硫酸濃度）, 8 density（密度） はややズレが大きい

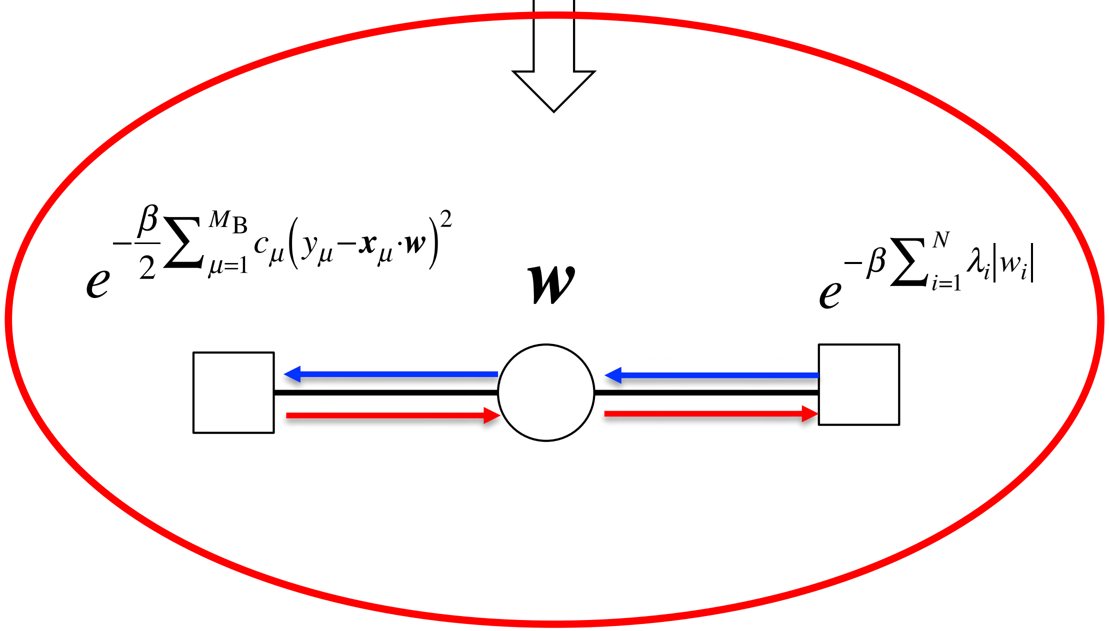
計算コストが増えてもよいので，近似精度を上げられないか

期待値伝搬法(expectation propagation: EP)

- 確率伝搬法は説明変数間の相関が強いと精度が悪化する
  ⇒ 関数のかたまりを1つのノードとして扱うことで相関を取り入れる
  - Expectation propagation (Minka 2001, Opper and Winther 2006, Rangan et al 2017)

$$P(\boldsymbol{w}|\boldsymbol{c}) = \frac{1}{Z} e^{-\frac{\beta}{2}\sum_{\mu=1}^{M_B} c_\mu (y_\mu - \boldsymbol{x}_\mu \cdot \boldsymbol{w})^2} \times e^{-\beta \sum_{i=1}^{N} \lambda_i |w_i|}$$

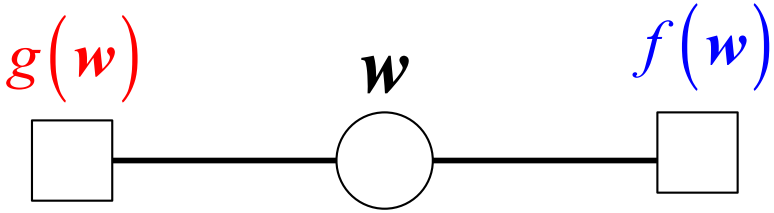




- グラフにループがないのでこのまま計算できれば厳密評価
- 残念ながら，多次元ガウス×ラプラスは計算困難

期待値伝搬法(expectation propagation: EP)

- 計算困難はファクトライズされた
  ガウス分布にもとづく近似で解消



$$P(\boldsymbol{w}|\boldsymbol{c}) = \frac{1}{Z}\underbrace{e^{-\frac{\beta}{2}\sum_{\mu=1}^{M_{\mathrm{B}}} c_\mu \left(y_\mu - \boldsymbol{x}_\mu\cdot\boldsymbol{w}\right)^2}}_{g(\boldsymbol{w})} \times \underbrace{e^{-\beta\sum_{i=1}^{N} \lambda_i |w_i|}}_{f(\boldsymbol{w})}$$

$$\propto g(\boldsymbol{w})\exp\left(-\frac{1}{2}\sum_{i=1}^{N}\beta\Lambda_i^{\mathrm{G}} w_i^2 + \sum_{i=1}^{N}\beta\gamma_i^{\mathrm{G}} w_i\right) \Longleftarrow \text{計算可能(Gaussian)}$$

$$\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\beta\Lambda_i^{\mathrm{F}} w_i^2 + \sum_{i=1}^{N}\beta\gamma_i^{\mathrm{F}} w_i\right) f(\boldsymbol{w}) \Longleftarrow \text{計算可能(Factorized)}$$

$$\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(\beta\Lambda_i^{\mathrm{G}} + \beta\Lambda_i^{\mathrm{F}}\right) w_i^2 + \sum_{i=1}^{N}\left(\beta\gamma_i^{\mathrm{G}} + \beta\gamma_i^{\mathrm{F}}\right) w_i\right) \Longleftarrow \text{計算可能}$$

(Gaussian & Factorized)

Moment matching: ３つの表現で２次までのモーメントが一致するように
$\left\{\Lambda_i^{\mathrm{G}}, \gamma_i^{\mathrm{G}}\right\}, \left\{\Lambda_i^{\mathrm{F}}, \gamma_i^{\mathrm{F}}\right\}$ を定める

EP + レプリカ法

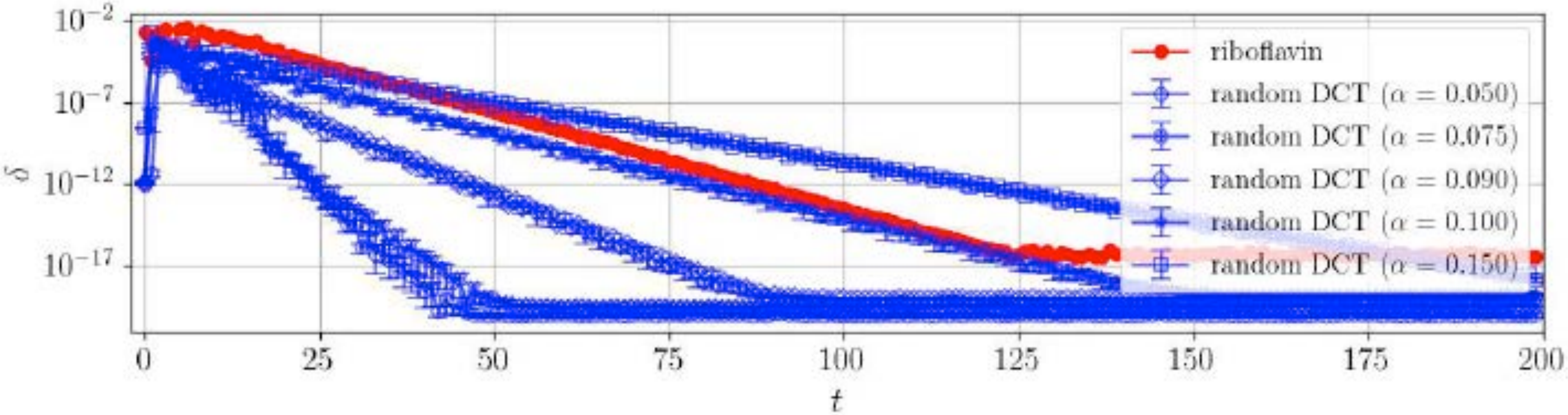
- n レプリカ系をEPで近似し，レプリカ対称仮定の下$n \to 0$としてアルゴリズムを導出する.
  - 煩雑なので，式の表現は省略
  - 詳細は Takahashi and YK, JSTAT (2020) 093402 を参照のこと

- 計算コストは$O(\min(M,N)^3 + MN)$/update
  - 逆行列計算が必要
    - 大きなサイズのデータには使いにくい
  - 更に近似を入れて，特異値分解を一度行うだけで確率伝搬法と同程度の$O(MN)$/updateまで計算コストを落とすことも可能

- 人工データと枯草菌のビタミンB2生産速度のデータに適用してみた

EP + レプリカ法

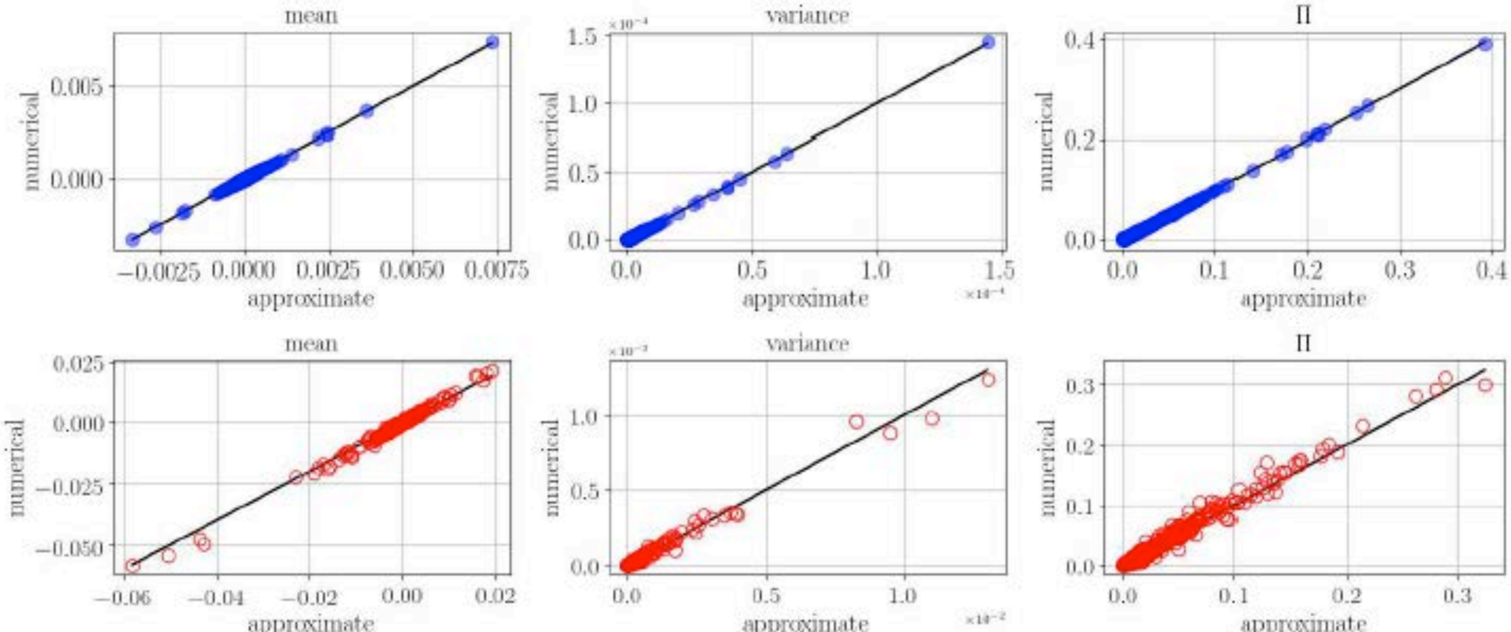
反復に対して指数関数的に収束



人工データ，実データともEP＋レプリカ法が直接的ブートストラップを高い精度で近似

人工データ

枯草菌のビタミンB2生産速度



全ての図：Takahashi and Kabashima, arXiv:1905.09545 (2019)より

まとめ

- LASSO推定に関し，ブートストラップ統計量をブートストラップを実施することなしに，近似的に評価する方法を開発
  - 確率伝搬法＋レプリカ法
    - 計算コスト $O(MN)$/update，変数間の相関にやや弱い
  - 期待値伝搬法＋レプリカ法
    - 計算コスト $O(\min(M,N)^3 + MN)$/update，変数間の相関に強い
  - これらに限らず別の平均場近似法を使うことも可能

- 今度の課題
  - LASSO以外の推定法についても同様のアルゴリズムを開発
  - 生成モデルを仮定し，ブートストラップ法の有用性の根拠を分析
  - 実データ解析への応用
  - ブートストラップ統計量の他の有用な使い方を開発

ご清聴ありがとうございました

- 参考文献

  - T. Obuchi and YK, Semi-Analytic Resampling in Lasso, Journal of Machine Learning Research 20 (70) pp. 1-33 (2019)
  - T. Takahashi and YK, Replicated Vector Approximate Message Passing For Resampling Problem, arXiv:1905.09545 (2019)
  - T. Takahashi and YK, Semi-analytic approximate stability selection for correlated data in generalized linear models, Journal of Statistical Mechanics: Theory and Experiment (2020) no. 9, 093402(1-36)