セルワイズな外れ値に対してロバストなスパースグラフィカルモデリング

片山翔太

東京工業大学

データ科学特別セミナー at 大阪大学
2018.6.13

Outline

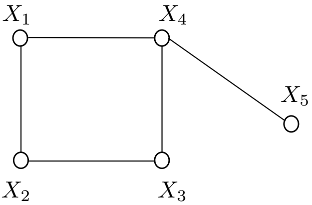
- グラフィカルモデルとは
  - 無向独立グラフと条件付き独立性
  - Hammersley-Clifford Theorem
- ガウスグラフィカルモデル
  - スパースモデリング
  - 外れ値への対処 (row-wise)
- セルワイズ外れ値への対処 †
- Discussion

† Joint work with Prof.Fujisawa and Prof.Drton (2018, *Stat*, vol.7)

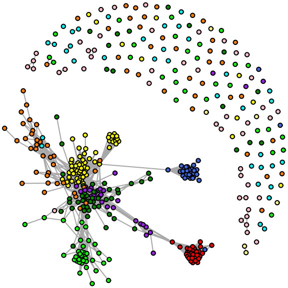
グラフィカルモデル

確率変数間の「依存関係」をグラフで表現したもの

- グラフ $G = (V, E)$; 頂点集合 V と辺集合 E (無向)
- 頂点と確率変数 $\{X_v : v \in V\}$ が対応
- 辺 (u, v) の有無が X_u, X_v 間の「依存関係」の有無に対応



$V = \{1, 2, 3, 4, 5\}$
$E = \{(1, 2), (1, 4), (2, 3), (3, 4), (4, 5)\}$



「依存関係」とは

ペアワイズマルコフ

確率変数 $\{X_v : v \in V\}$ がグラフ $G = (V, E)$ に関してペアワイズマルコフであるとは，次が成立つときのことをいう:

$$X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}} \quad \text{for all } (u, v) \notin E$$

- 依存関係 $=$ 条件付き独立性
- どうやって確認? $\Rightarrow$ Hammersley-Clifford Theorem

Remark

- 他にも局所 or 大域マルコフと呼ばれる関係性もある
- 詳しくは Lauritzen (1996) など

Hammersley-Clifford Theorem

確率変数 $\boldsymbol{X} = \{X_v : v \in V\}$ がグラフ $G = (V, E)$ に関してペアワイズマルコフであることと次は同値:

$$f(\boldsymbol{x}) = \prod_{C \in \mathcal{C}} \phi_C(\boldsymbol{x}_C), \quad \boldsymbol{x} \in \mathbb{R}^{|V|}$$

ここで $\phi_C : \mathbb{R}^{|C|} \to \mathbb{R}$ であり，$\mathcal{C}$ は G の全ての (極大) クリークの集合

- $\mathcal{C}$ の定義：$C \in \mathcal{C}$ if $(u, v) \in E$ for all $u, v \in C$
- 極大の意味：他のクリークに含まれないこと

Remark

- $\{\phi_C : C \in \mathcal{C}\}$ は一意に定まらない
- ϕ_C が密度的な意味を持つ必要はない

Examples

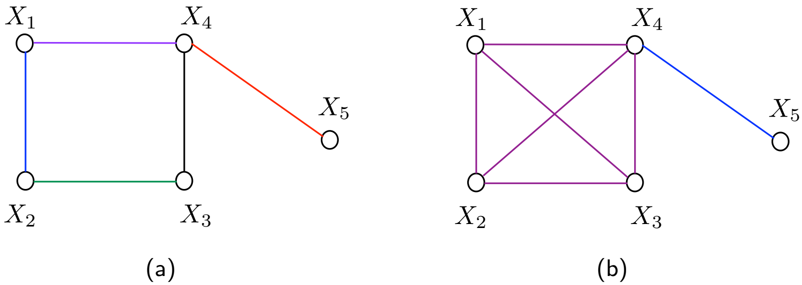


Figure (a)

- $f(\boldsymbol{x}) = \phi_{12}(x_1, x_2)\phi_{14}(x_1, x_4)\phi_{23}(x_2, x_3)\phi_{34}(x_3, x_4)\phi_{45}(x_4, x_5)$

Figure (b)

- $f(\boldsymbol{x}) = \phi_{1234}(x_1, x_2, x_3, x_4)\phi_{45}(x_4, x_5)$

ガウスグラフィカルモデル

Goal：データに適合するグラフ G の推定

- 密度 $f(\boldsymbol{x})$ の情報が必要
- 一般的に推定が困難 $\Rightarrow$ 分布族の特定 or グラフ構造に制約

ガウスグラフィカルモデル

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^p|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2} \sum_{u,v=1}^{p} \omega_{uv} x_u x_v \right), \quad \boldsymbol{x} \in \mathbb{R}^p$$

- 正規分布 $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ の密度．ただし $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} = (\omega_{uv})$
- 平均が $\boldsymbol{\mu} \neq \mathbf{0}$ でも OK

$$\boxed{\omega_{uv} = \omega_{vu} = 0 \Leftrightarrow (u,v) \notin E}$$

Examples

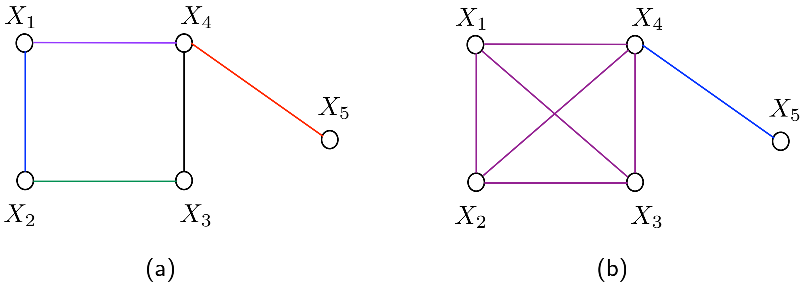


Figure (a)

- $f(\boldsymbol{x}) \propto e^{-\frac{\omega_{12}+\omega_{21}}{2}x_1x_2} e^{-\frac{\omega_{14}+\omega_{41}}{2}x_1x_4} \cdots e^{-\frac{\omega_{45}+\omega_{54}}{2}x_4x_5}$

Figure (b)

- $f(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2}\sum_{u,v=1}^{4}\omega_{uv}x_ux_v\right)\exp\left\{\frac{1}{2}(\omega_{45}+\omega_{54})x_4x_5\right\}$

ガウスグラフィカルモデル

Goal $\Rightarrow$ 精度行列 $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ の推定

- 特に精度行列 $\boldsymbol{\Omega}$ の 0 要素を推定
- スパースモデリングに帰着!
  - 半自動的に幾つかのパラメータを 0 に推定可能

Famous methods

- Node-wise regression (Meinshausen and Buhlmann, 2006)
- Glasso (Friedman et al., 2008)
- CLIME (Cai et al., 2011)

スパースガウスグラフィカルモデリング

Lasso

- $\boldsymbol{y} \in \mathbb{R}^n$: 結果変数, $\boldsymbol{Z} \in \mathbb{R}^{n \times p}$: 説明変数行列
- $(\boldsymbol{y}, \boldsymbol{Z})$ は共に中心化しておく

$$\text{Lasso}: \operatorname*{argmin}_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2n} \|\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1$$

- 次のようにも書ける

$$\operatorname*{argmin}_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2} \boldsymbol{\theta}^T \left(\frac{\boldsymbol{Z}^T \boldsymbol{Z}}{n} \right) \boldsymbol{\theta} - \left(\frac{\boldsymbol{Z}^T \boldsymbol{y}}{n} \right)^T \boldsymbol{\theta} + \lambda \|\boldsymbol{\theta}\|_1$$

  - $\boldsymbol{Z}^T \boldsymbol{Z}/n$: 説明変数間の共分散行列
  - $\boldsymbol{Z}^T \boldsymbol{y}/n$: 結果変数と各説明変数の共分散

Node-wise regression

Anderson (2003)

If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then

$$X_j \,|\, \boldsymbol{X}_{\backslash j} \sim N\left(\mu_j + \boldsymbol{\Sigma}_{j,\backslash j}\boldsymbol{\Sigma}_{\backslash j,\backslash j}^{-1}(\boldsymbol{X}_{\backslash j} - \boldsymbol{\mu}_{\backslash j}), \boldsymbol{\Sigma}_{jj} - \boldsymbol{\Sigma}_{j,\backslash j}\boldsymbol{\Sigma}_{\backslash j,\backslash j}^{-1}\boldsymbol{\Sigma}_{\backslash j,j}\right)$$

- 線形回帰問題に帰着

$$X_j = \beta_0^{(j)} + \boldsymbol{X}_{\backslash j}^T \boldsymbol{\beta}^{(j)} + \varepsilon_j, \quad j = 1, 2, \ldots, p$$

where $\beta_0^{(j)} = \mu_j - \boldsymbol{\Sigma}_{j,\backslash j}\boldsymbol{\Sigma}_{\backslash j,\backslash j}^{-1}\boldsymbol{\mu}_{\backslash j}$ and $\boldsymbol{\beta}^{(j)} = \boldsymbol{\Sigma}_{\backslash j,\backslash j}^{-1}\boldsymbol{\Sigma}_{\backslash j,j}$

- 重要な関係式

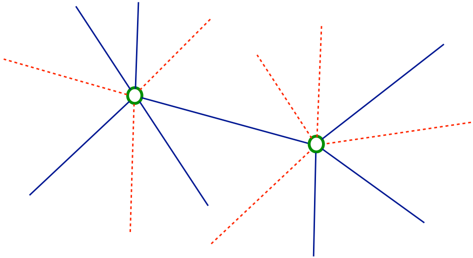
$$\boldsymbol{\Omega}_{\backslash j,j} = -\left(\mathbb{V}(\varepsilon_j)\right)^{-1}\boldsymbol{\beta}^{(j)}$$

$\boldsymbol{\Omega}$ の 0 要素推定 $\Rightarrow$ Lasso $\times$ 次元数

- $\boldsymbol{S}$: 標本共分散行列

$$\hat{\boldsymbol{\beta}}^{(j)} = \mathop{\mathrm{argmin}}_{\boldsymbol{\beta} \in \mathbb{R}^{p-1}} \frac{1}{2} \boldsymbol{\beta}^T \boldsymbol{S}_{\backslash j, \backslash j} \boldsymbol{\beta} - \boldsymbol{S}_{j, \backslash j} \boldsymbol{\beta} + \lambda \|\boldsymbol{\beta}\|_1, \quad j = 1, 2, \ldots, p$$

- 辺集合の推定
  - AND rule : $\hat{E} = \{(i,j) : \hat{\boldsymbol{\beta}}_i^{(j)} \neq 0$ and $\hat{\boldsymbol{\beta}}_j^{(i)} \neq 0\}$
  - OR rule : $\hat{E} = \{(i,j) : \hat{\boldsymbol{\beta}}_i^{(j)} \neq 0$ or $\hat{\boldsymbol{\beta}}_j^{(i)} \neq 0\}$



Glasso (Graphical Lasso)

ℓ_1 制約付き対数尤度を最適化

$$\text{Glasso}: \ \underset{\boldsymbol{\Omega} > \boldsymbol{O}}{\operatorname{argmin}} \ \mathrm{tr}\boldsymbol{S}\boldsymbol{\Omega} - \log|\boldsymbol{\Omega}| + \lambda\|\boldsymbol{\Omega}\|_1$$

辺集合の推定

- 推定量の非 0 要素で OK : $\hat{E} = \{(i,j) : \hat{\boldsymbol{\Omega}}_{ij} \neq 0\}$

最適化アルゴリズム

- Original (Friedman et al., 2008)
  - $\boldsymbol{\Omega}^{-1}$ を更新 $\Leftrightarrow$ 双対問題の解で更新 (Mazumder and Hastie, 2012)
- QUIC (Hsieh et al., 2011)
  - $\boldsymbol{\Omega}$ を直接更新
  - R package `QUIC`

QUIC

2 次近似 + 座標降下法

- $g(\boldsymbol{\Omega}) = \mathrm{tr}\boldsymbol{S}\boldsymbol{\Omega} - \log|\boldsymbol{\Omega}|$
- $\boldsymbol{\Omega}_{t+1} \leftarrow \boldsymbol{\Omega}_t + \alpha\boldsymbol{D}$ と更新．ただし，

$$\boldsymbol{D} = \operatorname*{argmin}_{\boldsymbol{\Delta}} \bar{g}_{\boldsymbol{\Omega}_t}(\boldsymbol{\Delta}) + \lambda\|\boldsymbol{\Omega}_t + \boldsymbol{\Delta}\|_1,$$

$$\bar{g}_{\boldsymbol{\Omega}_t}(\boldsymbol{\Delta}) = g(\boldsymbol{\Omega}_t) + \mathrm{tr}\{(\boldsymbol{S} - \boldsymbol{\Omega}_t^{-1})\boldsymbol{\Delta}\} + \frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}_t^{-1}\boldsymbol{\Delta}\boldsymbol{\Omega}_t^{-1}\boldsymbol{\Delta})$$

- $\boldsymbol{\Delta}$ の最適化は座標降下法

$$\boldsymbol{D}_{ij} = \boldsymbol{D}_{ji} = \operatorname*{argmin}_{\boldsymbol{\Delta}_{ij}=\boldsymbol{\Delta}_{ji}} \bar{g}_{\boldsymbol{\Omega}_t}(\boldsymbol{\Delta}) + \lambda\|\boldsymbol{\Omega}_t + \boldsymbol{\Delta}\|_1,$$

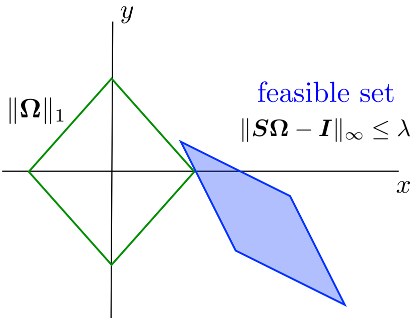
$$\text{given } \{\boldsymbol{\Delta}_{k\ell} : (k,\ell) \neq (i,j),(j,i)\}$$

CLIME

Idea： $\boldsymbol{S\Omega} \simeq \boldsymbol{I}$ なスパース $\boldsymbol{\Omega}$ を見つける

$$\hat{\boldsymbol{\Omega}}^0 = \operatorname{argmin} \|\boldsymbol{\Omega}\|_1 \quad \text{subject to} \quad \|\boldsymbol{S\Omega} - \boldsymbol{I}\|_\infty \le \lambda$$

- 右図は行列 $\begin{pmatrix} x & y \\ y & x \end{pmatrix}$ 上での最適化のイメージ
- 線形計画問題に帰着!



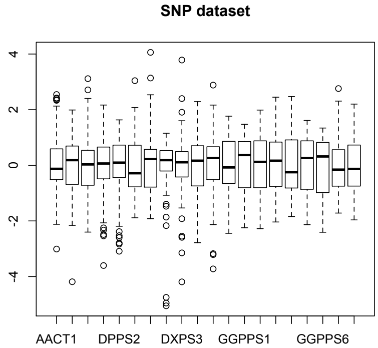
最後に対称化が必要

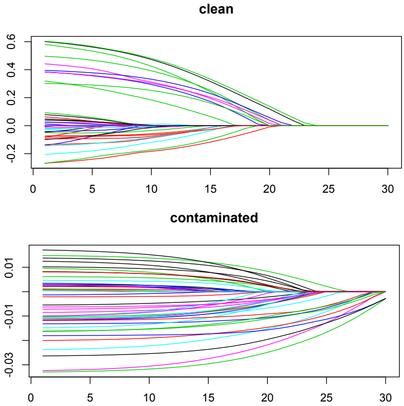
$$\hat{\boldsymbol{\Omega}}^{\text{CLIME}} = \operatorname{argmin} \|\hat{\boldsymbol{\Omega}}^0 - \boldsymbol{\Omega}\|_{\text{F}}^2 \quad \text{on} \quad \{\boldsymbol{\Omega} = \boldsymbol{\Omega}^T\}$$

外れ値への対処

Motivation

- 裾が重いデータが意外と多い
- グラフの推定結果は外れ値に敏感





t-Glasso

正規分布よりも裾の重い分布に変更

- 多変量 t 分布を仮定 (Finegold and Drton, 2011)

$$\boldsymbol{X} = \boldsymbol{\mu} + \frac{1}{\sqrt{\tau}}\boldsymbol{Z}, \quad \boldsymbol{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}), \quad \tau \sim \Gamma(\nu/2, \nu/2)$$

- $\boldsymbol{X}$ のみ観測，それ以外は潜在変数

Remark

- Goal：グラフ構造 $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ の推定
- $\boldsymbol{X}$ のグラフィカルモデルを推定していない

t-Glasso

推定アルゴリズム

- Given $\boldsymbol{\mu}^t, \boldsymbol{\Omega}^t$, update

$$\tau_i^{t+1} \leftarrow \frac{\nu + p}{\nu + \delta(\boldsymbol{x}_i; \boldsymbol{\mu}^t, \boldsymbol{\Omega}^t)},$$
$$\delta(\boldsymbol{x}_i; \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) = (\boldsymbol{x}_i - \boldsymbol{\mu}^t)^T \boldsymbol{\Omega}^t (\boldsymbol{x}_i - \boldsymbol{\mu}^t)$$

- Update

$$\boldsymbol{\mu}^{t+1} \leftarrow \frac{\sum_{i=1}^n \tau_i^{(t+1)} \boldsymbol{x}_i}{\sum_{i=1}^n \tau_i^{t+1}}, \quad \boldsymbol{\Omega}^{t+1} \leftarrow \text{Glasso}(\boldsymbol{S}^{t+1}; \lambda)$$

where

$$\boldsymbol{S}^{t+1} = \frac{1}{n} \sum_{i=1}^n \tau_i^{t+1} (\boldsymbol{x}_i - \boldsymbol{\mu}^{t+1})(\boldsymbol{x}_i - \boldsymbol{\mu}^{t+1})^T$$

γ-Glasso

外れ値の影響を緩和するように Glasso の目的関数を変更

- KL divergence ⇒ γ-divergence (Hirose et al., 2017)

推定アルゴリズム

- Given $\gamma > 0, \boldsymbol{\mu}^t, \boldsymbol{\Omega}^t$, calculate

$$w_i^t = \frac{\exp\left\{-\frac{\gamma}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}^t)^T \boldsymbol{\Omega}^t (\boldsymbol{x}_i - \boldsymbol{\mu}^t)\right\}}{\sum_{k=1}^n \exp\left\{-\frac{\gamma}{2}(\boldsymbol{x}_k - \boldsymbol{\mu}^t)^T \boldsymbol{\Omega}^t (\boldsymbol{x}_k - \boldsymbol{\mu}^t)\right\}}$$

- Update $\boldsymbol{\mu}^{t+1} \leftarrow \sum_{i=1}^n w_i^t \boldsymbol{x}_i$ and

$$\boldsymbol{\Omega}^{t+1} \leftarrow \operatorname*{argmin}_{\boldsymbol{\Omega} > \boldsymbol{O}} \operatorname{tr}(\boldsymbol{S}^{t+1}\boldsymbol{\Omega}) - \frac{1}{1+\gamma} \log|\boldsymbol{\Omega}| + \lambda \|\boldsymbol{\Omega}\|_1$$

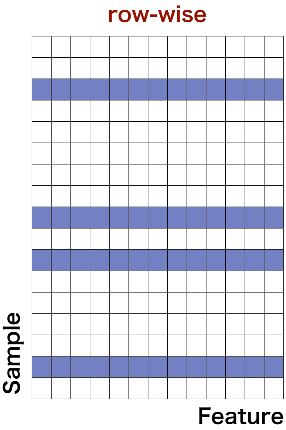
where

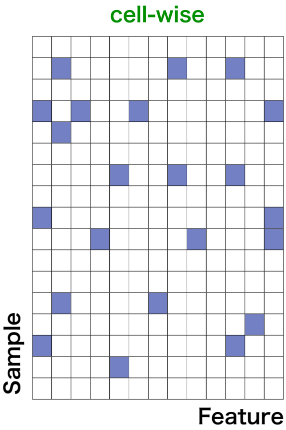
$$\boldsymbol{S}^{t+1} = \sum_{i=1}^n w_i^t (\boldsymbol{x}_i - \boldsymbol{\mu}^{t+1})(\boldsymbol{x}_i - \boldsymbol{\mu}^{t+1})^T$$

外れ値の構造

t-Glasso と γ-Glasso の性質

- 観測ベクトルに重み $\Rightarrow$ row-wise 外れ値の影響を緩和
- cell-wise 外れ値の場合はかなりの観測ベクトルが除外





セルワイズ外れ値への対処

Node-wise regression

$$\hat{\boldsymbol{\beta}}^{(j)} = \mathop{\mathrm{argmin}}_{\boldsymbol{\beta} \in \mathbb{R}^{p-1}} \frac{1}{2} \boldsymbol{\beta}^T \boldsymbol{S}_{\setminus j, \setminus j} \boldsymbol{\beta} - \boldsymbol{S}_{j, \setminus j} \boldsymbol{\beta} + \lambda \|\boldsymbol{\beta}\|_1, \quad j = 1, 2, \dots, p$$

Glasso

$$\mathop{\mathrm{argmin}}_{\boldsymbol{\Omega} > \boldsymbol{O}} \mathrm{tr} \boldsymbol{S} \boldsymbol{\Omega} - \log |\boldsymbol{\Omega}| + \lambda \|\boldsymbol{\Omega}\|_1$$

CLIME

$$\mathrm{argmin} \|\boldsymbol{\Omega}\|_1 \quad \text{subject to} \quad \|\boldsymbol{S}\boldsymbol{\Omega} - \boldsymbol{I}\|_\infty \leq \lambda$$

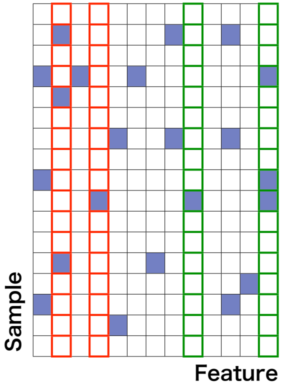
共分散行列 $\boldsymbol{\Sigma}$ がうまく推定できれば良さそう

Goal : $\boldsymbol{\Sigma}$ のロバスト推定 under セルワイズ外れ値

- 得られた $\hat{\boldsymbol{\Sigma}}$ を Node-wise regression, Glasso, CLIME に入力

Katayama, Fujisawa and Drton (2017)

- γ-divergence を用いてペアワイズに $\boldsymbol{\Sigma}$ を推定



モデル

Cell-wise contamination model

- $\boldsymbol{X} \in \mathbb{R}^p$ with

$$\boldsymbol{X} = (\boldsymbol{I} - \boldsymbol{E})\boldsymbol{Y} + \boldsymbol{E}\boldsymbol{H}$$

- $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\boldsymbol{H}$：外れ値 with density $h(\boldsymbol{x})$
- $\boldsymbol{E} = \mathrm{diag}(E_1, \ldots, E_p)$：それぞれが $\{0, 1\}$ を取る 2 値確率変数
- $\mathbb{P}(E_j = 1) = \varepsilon_j$：外れ値の混入率
- もし $E_1 = \cdots = E_p$ ならば row-wise 外れ値

サンプル

- $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ が独立に上記のモデルから生成

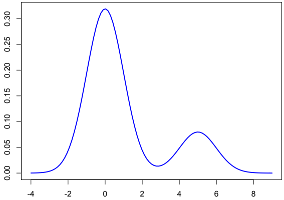
γ-divergence

記号の用意

- f : データの真の (周辺) 密度関数
  - ペアワイズ推定なので考えるのは高々 2 次元
- g_θ : パラメータ θ を持つ正規密度
  - 例えば 1 次元だと $\theta = (\mu_j, \sigma_{jj})$
- θ^* : 真のパラメータ
- 上記の記号を使うと $f = (1-\varepsilon)g_{\theta^*} + \varepsilon h$

やりたいこと

- データから θ^* の推定
- f と g_θ を普通に近づけてもダメ
- γ-divergence を使って近づける!



γ-divergence

- Fujisawa and Eguchi (2008)
- 次の目的関数を最小化 ($\gamma > 0$)：

$$d_\gamma(f_n, g_\theta) = -\frac{1}{\gamma}\log\int f_n(x)g_\theta(x)^\gamma dx + \frac{1}{1+\gamma}\log\int g_\theta(x)^{1+\gamma}dx$$
$$= -\frac{1}{\gamma}\log\sum_{i=1}^n g_\theta(x_i)^\gamma + \frac{1}{1+\gamma}\log\int g_\theta(x)^{1+\gamma}dx$$

- 重要な性質：

$$\hat{\theta}_\gamma = \underset{\theta}{\operatorname{argmin}}\, d_\gamma(f_n, g_\theta) \to \theta^*_\gamma = \underset{\theta}{\operatorname{argmin}}\, d_\gamma(f, g_\theta) \simeq \underset{\theta}{\operatorname{argmin}}\, d_\gamma(g, g_\theta)$$

外れ値の影響を自動的に除去!

トリック

$$
\begin{aligned}
d_\gamma(f, g_\theta) &= -\frac{1}{\gamma}\log\int f g_\theta^\gamma dx + \frac{1}{1+\gamma}\log\int g_\theta^{1+\gamma}dx \\
&= -\frac{1}{\gamma}\log\int \{(1-\varepsilon)g + \varepsilon h\} g_\theta^\gamma dx + \frac{1}{1+\gamma}\log\int g_\theta^{1+\gamma}dx \\
&= -\frac{1}{\gamma}\log\left\{(1-\varepsilon)\int g g_\theta^\gamma dx + \varepsilon\int h g_\theta^\gamma dx\right\} + \frac{1}{1+\gamma}\log\int g_\theta^{1+\gamma}dx \\
&\simeq -\frac{1}{\gamma}\log(1-\varepsilon)\int g g_\theta^\gamma dx + \frac{1}{1+\gamma}\log\int g_\theta^{1+\gamma}dx \\
&\propto d_\gamma(g, g_\theta)
\end{aligned}
$$

$\boxed{\nu(\theta;\gamma) := \int h(x) g_\theta^\gamma(x) dx \text{ が十分小さいときに成立}}$

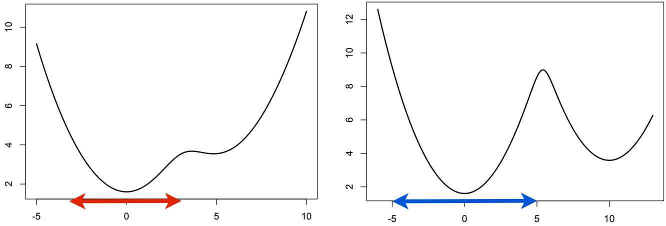
$\nu(\theta;\gamma)=\int h(x)g_\theta^\gamma(x)dx$ は小さくなるか?

- h が g_θ^γ の裾付近にあれば OK
- $h\sim N(\alpha,1)$, $g_\theta\sim N(\theta,1)\Rightarrow \nu(\theta;\gamma)=c_{1,\gamma}\exp\{-c_{2,\gamma}(\alpha-\theta)^2\}$

理想的には

- θ を θ^* 近傍で最適化
  - 良い初期値を使って最適化

$(\theta^*=0,\alpha=5)$ $(\theta^*=0,\alpha=10)$



γ-divergence を用いた共分散行列推定

シンプルな 2 段階推定

- $\{X_{ij}\}_{i=1}^n$ から (μ_j, σ_{jj}) を推定
- 標準化 $Z_{ij} = (X_{ij} - \mu_j)/\sqrt{\sigma_{jj}}$ 後, $\rho_{jk} = \mathrm{Corr}(X_{ij}, X_{ik})$ を推定

Step 1

- $g_{(\mu_j,\sigma_{jj})} \sim N(\mu_j, \sigma_{jj})$

$$(\hat{\mu}_j, \hat{\sigma}_{jj}) = \underset{\mu_j,\sigma_{jj}}{\mathrm{argmin}}\, d_\gamma\left(f_n^{(j)}, g_{(\mu_j,\sigma_{jj})}\right)$$

Step 2

- $g_{\rho_{jk}} \sim$ 相関 ρ_{jk} を持つ 2 次元標準正規分布

$$\hat{\rho}_{jk} = \underset{|\rho_{jk}|<1}{\mathrm{argmin}}\, d_\gamma\left(f_n^{(j,k)}, g_{\rho_{jk}}\right)$$

Output : $\hat{\Sigma}_{jk} = \hat{\sigma}_{jj}^{1/2}\hat{\sigma}_{kk}^{1/2}\hat{\rho}_{jk}$

Step 1の詳細

EM-like algorithm by Fujisawa and Eguchi (2008)

- 以下を収束するまで繰り返す：

$$\mu_j^{t+1} \leftarrow \sum_{i=1}^{n} w_{ij}^t X_{ij}, \quad \sigma_{jj}^{t+1} \leftarrow (1+\gamma) \sum_{i=1}^{n} w_{ij}^t (X_{ij} - \mu_j^{t+1})^2$$

where

$$w_{ij}^t = \frac{\exp\left\{-\frac{\gamma}{2\sigma_{jj}^t}(X_{ij} - \mu_j^t)^2\right\}}{\sum_{\ell=1}^{n} \exp\left\{-\frac{\gamma}{2\sigma_{jj}^t}(X_{\ell j} - \mu_j^t)^2\right\}}$$

- 初期値はロバストなものを選ぶ
    - $\mu_j^0 = \mathrm{Med}(\{X_{ij}\}_{i=1}^n)$
    - $(\sigma_{jj}^0)^{1/2} = \mathrm{Mad}(\{X_{ij}\}_{i=1}^n) = 1.483\,\mathrm{Med}(\{|X_{ij} - \mu_j^0|\}_{i=1}^n)$

Step 2 の詳細

制約を閉区間に変更

$$\underset{|\rho_{jk}| \le R}{\operatorname{argmin}}\, h(\rho_{jk}) := d_\gamma(f_n^{(jk)}, g_{\rho_{jk}}), \quad R < 1$$

以下を収束するまで繰り返す (射影勾配法)

- ϕ : step size

$$\rho_{jk}^{t+1} \leftarrow \begin{cases} \bar{\rho}_{jk}^t, & |\bar{\rho}_{jk}^t| \le R \\ \operatorname{sgn}(\bar{\rho}_{jk}^t) R, & \text{otherwise} \end{cases}$$

where

$$\bar{\rho}_{jk}^t = \rho_{jk}^t - \phi^{-1} \nabla h(\rho_{jk}^t)$$

Remark

- 初期値は $\rho^0 = 0$ で OK (外れ値間に相関がない限り)

行列の負定値性について

Remark

- $\hat{\Sigma}$ は非負定値性が保証されない

Node-wise regression

- $\hat{\Sigma} \geq O \Rightarrow$ 目的関数の凸性を保証

Glasso

- Freedman のアルゴリズム $\Rightarrow \hat{\Sigma} + \lambda \boldsymbol{I} > \boldsymbol{O}$ が必要
- QUIC $\Rightarrow \hat{\Sigma} \geq \boldsymbol{O}$ のときのみ収束保証

CLIME

- $\hat{\Sigma} < \boldsymbol{O} \Rightarrow \hat{\Omega}$ が負の対角要素を持つ
- $\hat{\Sigma} = \boldsymbol{O} \Rightarrow$ 実行可能解が存在しないケースがある

基本的に $\hat{\boldsymbol{\Sigma}} \geq \boldsymbol{O}$ が必要

- 条件を満たすように射影 ($\delta \geq 0$)

$$\min \|\boldsymbol{S} - \hat{\boldsymbol{\Sigma}}\|_{\mathrm{F}} \quad \text{subject to } \boldsymbol{S} \geq \delta \boldsymbol{I}$$

- $\hat{\boldsymbol{\Sigma}}$ を固有値分解して固有値 λ_j を $\max(\lambda_j, \delta)$ で置き換えるだけ

推定手法のまとめ

Goal：グラフ G の推定 $\Leftrightarrow \boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ の推定

- 全ての変数のペア (j,k) に対して $\hat{\boldsymbol{\Sigma}}_{jk}$ を計算
- 行列 $\hat{\boldsymbol{\Sigma}} = (\hat{\boldsymbol{\Sigma}}_{jk})$ を構成し $\hat{\boldsymbol{\Sigma}} \geq \delta \boldsymbol{I}$ となるように射影
- 射影した行列 $\hat{\boldsymbol{\Sigma}}$ を入力として各スパースグラフ推定法を適用

Option：各チューニングパラメータの選択

- Node-wise regression $\Rightarrow$ Stability approach (Liu et al., 2010)
- Glasso と CLIME $\Rightarrow$ クロスバリデーション

$$L(\lambda) = \mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_2 \hat{\boldsymbol{\Omega}}_1(\lambda)\right) - \log\det \hat{\boldsymbol{\Omega}}_1(\lambda)$$

パフォーマンス比較

比較対象

ランクに基づく相関の推定

- $\hat{\Sigma}_{jk} = \hat{\sigma}_{jj}^{1/2} \hat{\sigma}_{kk}^{1/2} \hat{\rho}_{jk}$ (スケールは Mad で推定)
- Kendall's tau (Spearman's rho)
- Gaussian rank correlation (Boudt et al., 2012)

$$\hat{\rho}_{jk} = \frac{\sum_{i=1}^{n} \Phi^{-1}\left\{\frac{\mathrm{R}(X_{ij})}{n+1}\right\} \Phi^{-1}\left\{\frac{\mathrm{R}(X_{ik})}{n+1}\right\}}{\sum_{i=1}^{n} \left\{\Phi^{-1}\left(\frac{i}{n+1}\right)\right\}^2}$$

Remark

- Gaussian rank のみ $\hat{\boldsymbol{\Sigma}} \geq \boldsymbol{O}$ を保証
- Nonparanormal (Liu et al., 2009) モデルの下で最適
  - ある単調変換 $f_1, \ldots, f_p$ が存在して $(f_1(X_1), \ldots, f_p(X_p)) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - Nonparanormal SKEPTIC (Liu et al., 2012)

比較対象

共分散の変形に基づくアプローチ (Tarr et al., 2016)

- Gnanadesikan and Kettenring (1972)

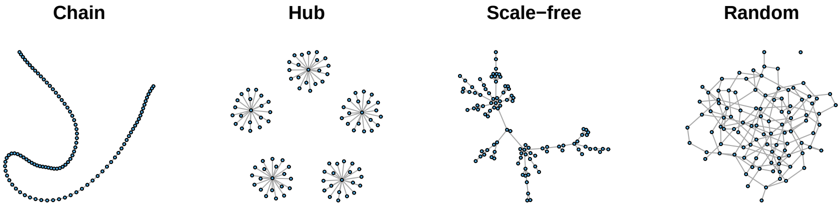
$$\mathbb{C}\mathrm{ov}(X, Y) = \frac{1}{4\alpha\beta}\{\mathbb{V}(\alpha X + \beta Y) - \mathbb{V}(\alpha X - \beta Y)\}$$
$$\alpha^{-1} = \sqrt{\mathbb{V}(X)}, \quad \beta^{-1} = \sqrt{\mathbb{V}(Y)}$$

- $\mathbb{V}(\cdot)$ をロバスト推定
  - $Q_n = 2.222\{|Z_i - Z_j| : i < j\}_{(k)}$, $k \simeq \binom{n}{2}/4$

Remark

- $\hat{\Sigma} \geq \boldsymbol{O}$ は保証されない
- 経験的にスケールの推定はあまり上手くいかない

いろいろなグラフ ($p = 100$)



	Chain	Hub	Scale-free	Random
Edges	99	95	99	156
Max. degree	2	19	15	4
Mean weight	0.57	0.43	0.28	0.25

その他の設定

データの生成

- $(n,p) = (200, 100)$, $\varepsilon_1 = \cdots = \varepsilon_p \in \{0.05, 0.15, 0.25\}$

セルワイズな外れ値

- Asymmetric と Symmetric

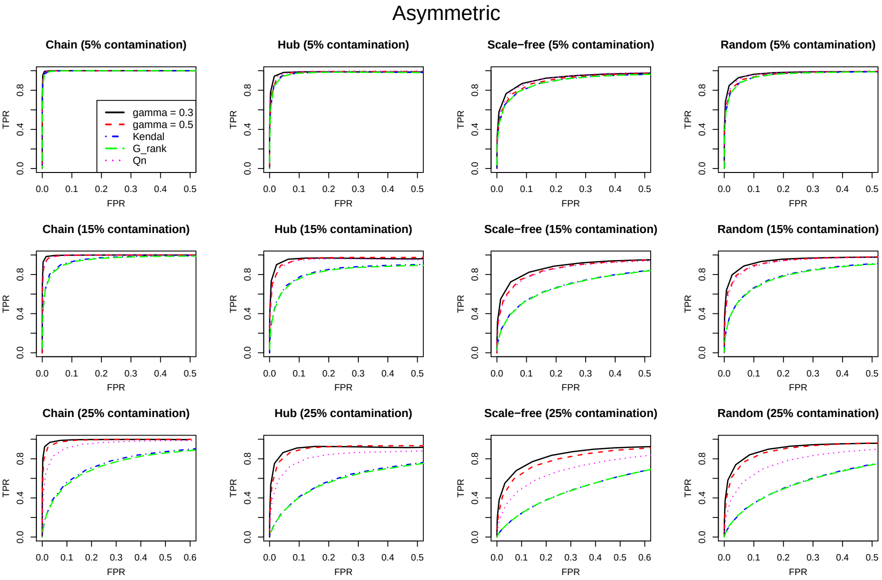
評価指標

- ROC curve $(\mathrm{FPR}(\lambda), \mathrm{TPR}(\lambda))$

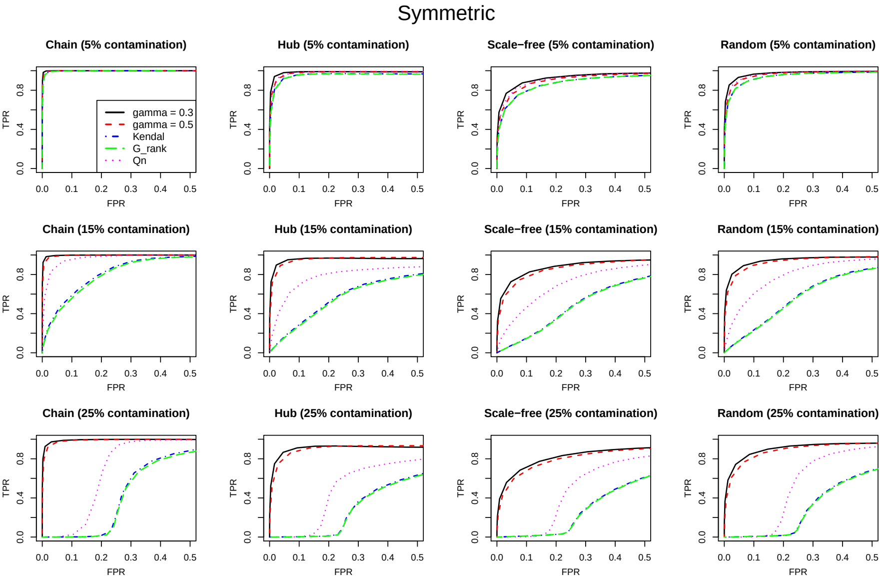
$$\mathrm{FPR}(\lambda) = \frac{|\hat{E} \cap E^c|}{|E^c|}, \quad \mathrm{TPR}(\lambda) = \frac{|\hat{E} \cap E|}{|E|}$$

- λ をチューニング $\Rightarrow$ TPR, FPR, MSE ($\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\|_{\mathrm{F}}/p$)

ROC curves for Glasso



ROC curves for Glasso

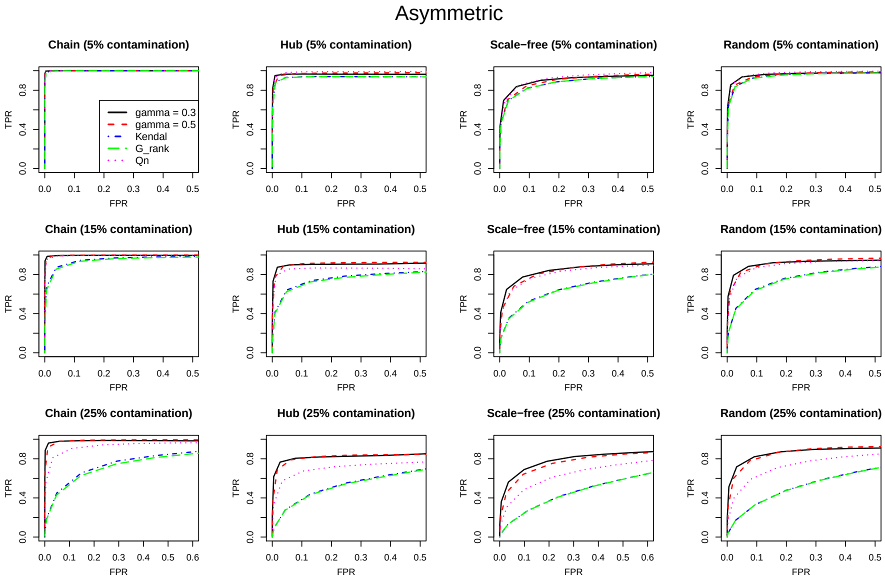


Tuned λ for Glasso

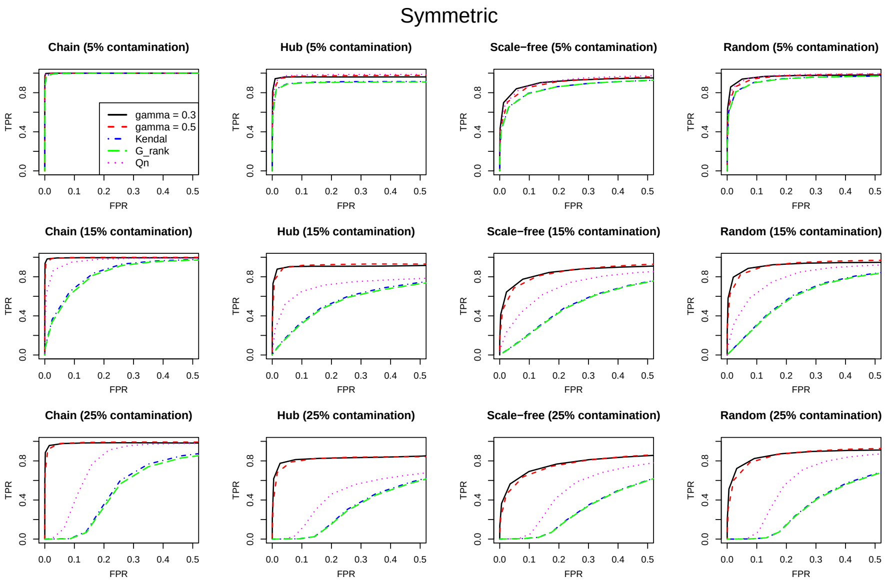
$\varepsilon = 0.25$

		Chain			Scale-free			Random		
		MSE	TPR	FPR	MSE	TPR	FPR	MSE	TPR	FPR
Asym.	$\gamma = 0.3$	**0.074**	**0.992**	0.106	**0.053**	**0.584**	0.044	**0.049**	**0.753**	0.063
		(0.004)	(0.011)	(0.025)	(0.002)	(0.103)	(0.020)	(0.005)	(0.165)	(0.031)
	$\gamma = 0.5$	0.085	0.972	**0.064**	0.055	0.458	**0.029**	0.051	0.692	**0.045**
		(0.003)	(0.026)	(0.023)	(0.002)	(0.114)	(0.017)	(0.002)	(0.079)	(0.021)
	Q_n	0.140	0.941	0.132	0.098	0.488	0.084	0.086	0.665	0.112
		(0.001)	(0.028)	(0.035)	(0.001)	(0.090)	(0.028)	(0.001)	(0.080)	(0.037)
Sym.	$\gamma = 0.3$	**0.046**	**0.821**	0.079	**0.053**	**0.595**	0.047	**0.046**	**0.825**	0.078
		(0.002)	(0.038)	(0.020)	(0.002)	(0.068)	(0.018)	(0.002)	(0.045)	(0.020)
	$\gamma = 0.5$	0.050	0.700	**0.049**	0.055	0.436	**0.025**	0.050	0.707	**0.049**
		(0.002)	(0.082)	(0.021)	(0.002)	(0.102)	(0.014)	(0.002)	(0.083)	(0.022)
	Q_n	0.082	0.744	0.292	0.094	0.505	0.248	0.082	0.742	0.288
		(0.002)	(0.073)	(0.028)	(0.001)	(0.096)	(0.025)	(0.002)	(0.065)	(0.027)

ROC curves for Node-wise regression



ROC curves for Node-wise regression



Tuned λ for Node-wise regression

$\varepsilon = 0.25$

		Chain		Scale-free		Random	
		TPR	FPR	TPR	FPR	TPR	FPR
Asym.	$\gamma = 0.3$	**0.978**	0.059	**0.614**	0.057	**0.776**	0.062
		(0.013)	(0.024)	(0.065)	(0.019)	(0.060)	(0.021)
	$\gamma = 0.5$	0.972	**0.046**	0.530	**0.041**	0.680	**0.042**
		(0.019)	(0.016)	(0.090)	(0.019)	(0.075)	(0.017)
	Q_n	0.881	0.062	0.429	0.058	0.547	0.059
		(0.043)	(0.022)	(0.069)	(0.019)	(0.069)	(0.020)
Sym.	$\gamma = 0.3$	**0.977**	0.063	**0.622**	0.059	**0.782**	0.066
		(0.016)	(0.027)	(0.062)	(0.021)	(0.047)	(0.020)
	$\gamma = 0.5$	0.975	**0.047**	0.526	**0.041**	0.678	**0.040**
		(0.018)	(0.016)	(0.084)	(0.016)	(0.067)	(0.017)
	Q_n	0.600	0.137	**0.137**	0.132	**0.183**	0.132
		(0.116)	(0.018)	(0.066)	(0.015)	(0.094)	(0.019)

Real data analysis

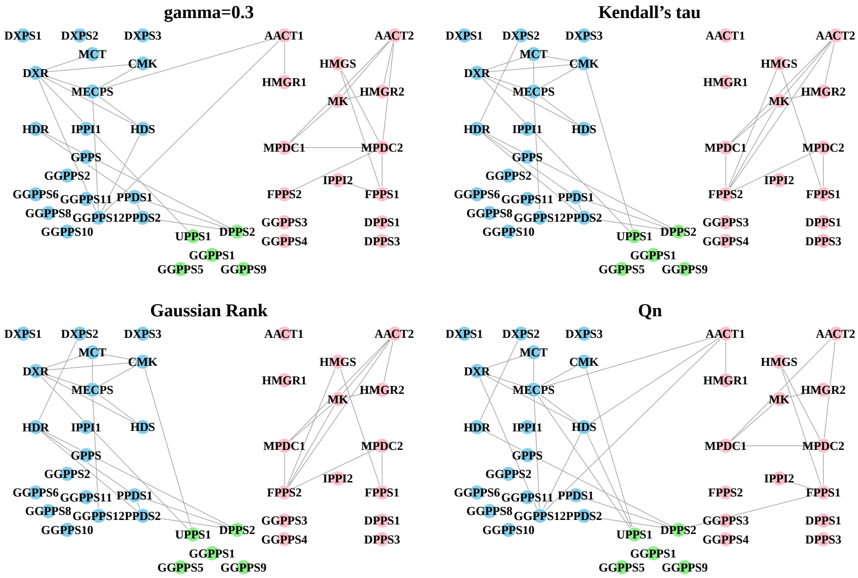
シロイヌナズナの遺伝子データ

- Wille et al. (2004)
- $n = 118$, $p = 39$ (small dimension)
- 変数 (遺伝子) は 3 つのグループに分けられる
  - MEP パスウェイ, MVA パスウェイ, その他

Daily S&P 500 stocks

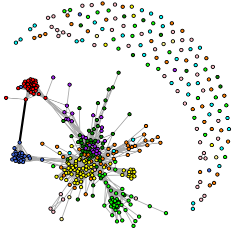
- $n = 1257$ (Jan. 2003 – Jan. 2008), $p = 452$ (large dimension)
- 変数は 10 のグループ (GICS) に分けられる
  - 金融, 公共事業, エネルギー, 素材, etc

シロイヌナズナの遺伝子データ (30 edges)

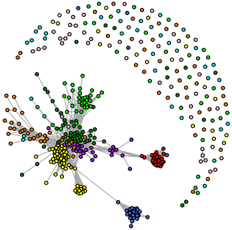


S&P 500 stocks (2,500 edges)

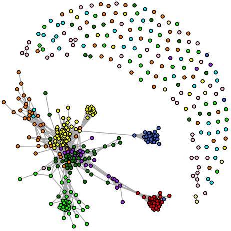
gamma=0.1



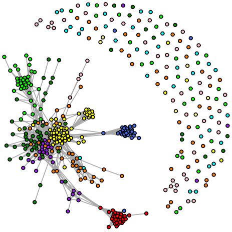
gamma=0.3



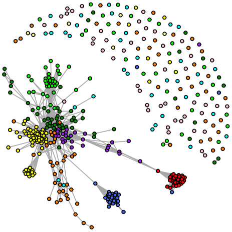
Kendall's tau



Gaussian rank



Qn



パフォーマンス比較の結果

Simulation

- $\gamma = 0.3, 0.5$ ともに混入率 ε からの影響をほとんど受けない
- Q_n の性能はグラフ推定法 (特にチューニング) に左右される

シロイヌナズナの遺伝子データ

- $\gamma = 0.3$ と Q_n はパスウェイ間の関連性を捉えることに成功

S&P 500

- どの方法でもグループ構造 (GICS) が観測される
- $\gamma = 0.1$ で公共事業 (青) と素材 (赤) 間にパス

Discussion

Theory?

- $\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty$ の集中不等式が導出できるか?
- $\hat{\boldsymbol{\Sigma}}$ が陽に書ける or 凸最適化問題の解 $\Rightarrow$ OK
  - Miao (2010), Catoni (2012), Loh and Tan (2018+)
- 今回は非凸最適化問題

Extension?

- ガウス以外のグラフィカルモデルに適用できるか?
  - 特にカウントデータへの適用が重要
- 周辺分布が扱い易い形で与えられることが必須条件
  - 例えば Poisson-log normal (Aitchison and Ho, 1989) など

ありがとうございました!!

今日のスライド：私のホームページに後で置いておきます

R code：https://github.com/shkatayama/robust_graphical_model

参考文献

- Aitchison, J., & Ho, H. (1989), The multivariate Poisson-log normal distribution, *Biometrika*, **76**, 643–653
- Anderson, W. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley-Interscience
- Boudt, K., Cornelissen, J., & Croux, C. (2012), The Gaussian rank correlation estimator: robustness properties, *Statistics and Computing*, **22**, 471–483
- Cai, T., Liu, W. & Luo, X. (2011), A constrained ℓ_1 minimization approach to sparse precision matrix estimation, *Journal of the American Statistical Association*, **106**, 594–607
- Catoni, O. (2012), Challenging the empirical mean and empirical variance: a deviation study, *Annales de l'Institut Henri Poincaré, Probabilitès et Statistiques*, **48**, 1148–1185
- Finegold, M. & Drton, M. (2011), Robust graphical modeling of gene networks using classical and alternative t-distributions, *The Annals of Applied Statistics*, **2A**, 1057–1080
- Friedman, H., Hastie, T. & Tibshirani, R. (2008), Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, **9**, 432–441

参考文献

- Fujisawa, H. & Eguchi, S. (2008), Robust parameter estimation with a small bias against heavy contamination, *Journal of Multivariate Analysis*, **99**, 2053–2081
- Hsieh, J., Sustik, A., Dhillon, S. & Ravikumar, P. (2011), Sparse inverse covariance matrix estimation using quadratic approximation, *In Advances in Neural Information Processing Systems*, 2330–2338
- Hirose, K., Fujisawa, H., & Sese, J. (2017), Robust sparse Gaussian graphical modeling, *Journal of Multivariate Analysis*, **161**, 172–190
- Katayama, S., Fujisawa, H. & Drton, M. (2018), Robust and sparse Gaussian graphical modeling under cell-wise contamination, *Stat*, **7**, e181
- Lauritzen, S. (1996), *Graphical models*, Clarendon Press
- Liu, H., Lafferty, J., & Wasserman, L. (2009), The nonparanormal: Semiparametric estimation of high dimensional undirected graphs, *Journal of Machine Learning Research*, **10**, 2295–2328
- Loh, P. & Tan, X. (2018+), High-dimensional robust precision matrix estimation: Cellwise corruption under epsilon-contamination, *Electronic Journal of Statistics*

参考文献

- Liu, H., Roeder, K. & Wasserman, L. (2010), Stability approach to regularization selection for high dimensional graphical models, *Advances in Neural Information Processing Systems*
- Liu, H., Han, F., Yuan, M., Lafferty, J., & Wasserman, L. (2012), High-dimensional semiparametric Gaussian copula graphical models, *The Annals of Statistics*, **40**, 2293–2326
- Gnanadesikan, R. & Kettenring, R. (1972), Robust estimates, residuals and outlier detection with multiresponse data, *Biometrics*, **28**, 81–124
- Meinshausen, N. & Bühlmann, P. (2006), High-dimensional graphs and variable selection with the lasso, *The Annals of Statistics*, **34**, 1436–1462
- Miao, O. (2010), Concentration inequality of maximum likelihood estimator, *Applied Mathematics Letters*, **23**, 1305–1309
- Tarr, G., Müller, S. & Weber C. (2016), Robust estimation of precision matrices under cellwise contamination, *Computational Statistics and Data Analysis*, **93**, 404–420