フィッシャー情報量行列が特異になる場合のパラメータ推定　ーSARモデルの場合ー

1

早稲田大学　力丸　佑紀（発表者）
慶應義塾大学　柴田　里程

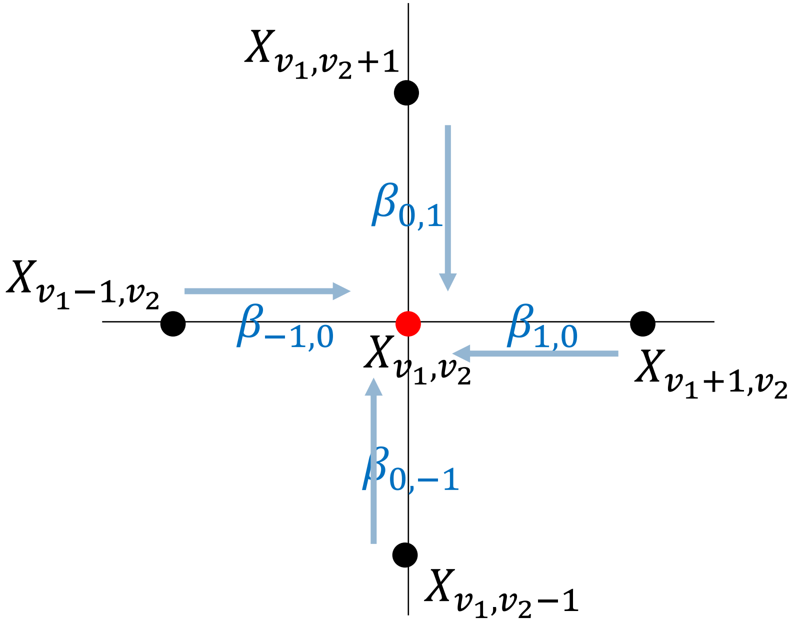
2016/9/7

Simultaneous autoregressive (SAR) model

- n 次元 SAR モデル $\sum_{\boldsymbol{k}\in K} \beta_{\boldsymbol{k}} X_{\boldsymbol{v}+\boldsymbol{k}} = \varepsilon_{\boldsymbol{v}}$, $\beta_{\boldsymbol{0}} = 1$, $\varepsilon_{\boldsymbol{v}} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 I_{\mathrm{n}})$
  - 添え字は n 次元ベクトル
  - K は 0 ベクトルを含む有限集合

    方向性がない形を許す
  - $\varepsilon_{\boldsymbol{v}}$ はイノベーションではない

    すべての $\boldsymbol{u}$ について、$E(X_{\boldsymbol{u}}\varepsilon_{\boldsymbol{v}}) \neq 0$



（例）２次元 SAR モデル
$K = \{(0,0), (1,0), (0,1), (-1,0), (0,-1)\}$

伝達関数を用いた表現

- SARモデル（伝達関数を用いた表現）　$P(T_1,\dots,T_n)X_v = \varepsilon_v$

  ただし、$P(T_1,\dots,T_n) = \sum_{\boldsymbol{k}\in K} \beta_{\boldsymbol{k}} T_1^{k_1}\cdots T_n^{k_n}$,　$T_i X_{\boldsymbol{v}} = X_{v_1,\dots,v_i+1,\dots,v_n}$

- これ以降，弱定常 SAR モデルを考える.

弱定常性の仮定

$P(z_1,\dots,z_n) = 0$は$|z_1| = \cdots = |z_n| = 1$ 上で根をもたない

パラメータ推定

- パラメータは，回帰係数 $\beta_{\boldsymbol{k}}$ $(\boldsymbol{k} \neq \boldsymbol{0} \in K)$ と $\varepsilon_{\boldsymbol{v}}$ の標準偏差 σ
  - パラメータベクトル　$\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)^T$
  - $\boldsymbol{\beta}$ は $\beta_{\boldsymbol{k}}$ $(\boldsymbol{k} \neq \boldsymbol{0} \in K)$ を添え字が辞書式になるように並べたベクトル

- 最尤法を用いたパラメータ推定
  - 対数尤度

$$L = -\frac{1}{2}\log\det(\Sigma) - \frac{N}{2}\log 2\pi - \frac{1}{2}x^T\Sigma^{-1}x$$

SAR モデルでは，L を計算することが難しい！

近似尤度を用いたパラメータ推定

- 近似尤度を考案　（2012年本学会で発表）

$$L_A = \frac{1}{2}\log\det(\mathrm{A}) - \frac{N}{2}\log 2\pi - \frac{1}{2}x^T\,\tilde{\mathrm{A}}x$$

ただし，A は Σ を修正した巡回行列，$\tilde{A}$ は A に重みをつけた巡回行列

（例）２次元SARモデルであれば

$$A = \sum_{\boldsymbol{k},\boldsymbol{k}'\in K} \beta_{\boldsymbol{k}}\beta_{\boldsymbol{k}'} W_{n_1}^{-k_1+k_1'} \otimes W_{n_2}^{-k_2+k_2'}$$

$$\tilde{A} = \sum_{\boldsymbol{k},\boldsymbol{k}'\in K} \beta_{\boldsymbol{k}}\beta_{\boldsymbol{k}'} \alpha_{n_1}^{|-k_1+k_1'|} \alpha_{n_2}^{|-k_2+k_2'|} W_{n_1}^{-k_1+k_1'} \otimes W_{n_2}^{-k_2+k_2'}$$

（参考文献）Yuuki Rikimaru and Ritei Shibata, (2016), A good approximation of the Gaussian likelihood of simultaneous autoregressive model which yields us an asymptotically efficient estimate of parameters, *Journal of Statistical Planning and Inference*, 173, 31-46.

パラメータ推定の問題

- 近似尤度 L_A の利点
  - 簡単に計算できる
  - 推定量は一致性をもつ
  - 推定量はある条件のもとで漸近有効性をもつ
  - 乱数を用いた実験で確認済

これらの利点をすべて満たすのは L_A のみ

（参考）Whitleの近似尤度(1954)
Guyonの近似尤度(1986)
Robinsonの近似尤度(2006)

SAR モデルではフィッシャー情報量行列（の極限）$I(\boldsymbol{\theta})$ が非正則になることがある

（本物の対数尤度や他の近似尤度でも同様に起きる問題）

$I(\boldsymbol{\theta})$ が非正則であるとき，何が起きるのか？

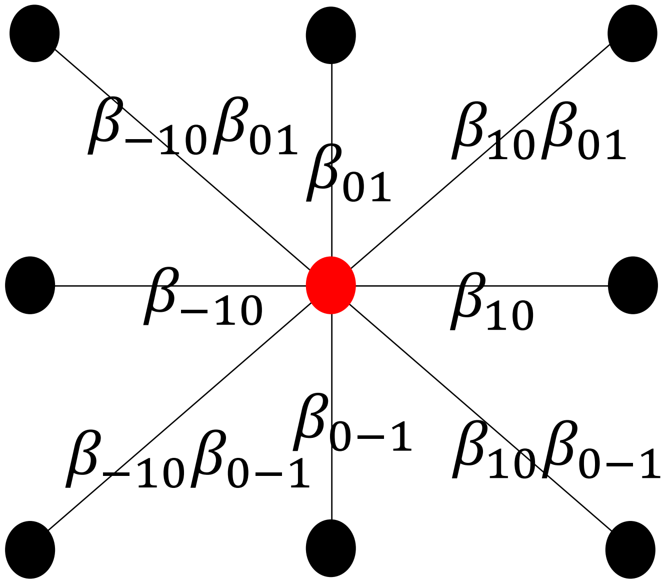
$\boldsymbol{I}(\boldsymbol{\theta})$が非正則のとき，何が起きるか？

- これ以降，２次元 SAR モデル $P_1(z_1)P_2(z_2)X_{v_1,v_2} = \varepsilon_{v_1,v_2}$ について考える．

  ただし，$P_1(z_1) = 1 + \beta_{10}z_1 + \beta_{-10}z_1^{-1}$，

  $$P_2(z_2) = 1 + \beta_{01}z_2 + \beta_{0-1}z_2^{-1}$$

  - パラメータは　$\boldsymbol{\theta} = (\beta_{-10}, \beta_{0-1}, \beta_{01}, \beta_{10}, \sigma)$



$I(\theta)$ が非正則になるとき，乱数実験の結果は？

- $I(\theta)$ が非正則になる場合の乱数を生成し，L_A によるパラメータ推定を行った.
  - $\beta_{-10}=\beta_{10}=0.4,\ \beta_{0-1}=0.2,\ \beta_{01}=0.7,\ \sigma=0.01$ として 20回の推定を行った.
  - 乱数の大きさは $40\times 40=1600$

	β_{10}	β_{-10}	β_{01}	β_{0-1}	σ
真の値	0.4	0.4	0.2	0.7	0.01
平均	$\boxed{0.31964}$	$\boxed{0.47359}$	0.20364	0.69619	0.0102
標準誤差	$\boxed{0.06707}$	$\boxed{0.06411}$	0.02132	0.02474	0.0003

非正則の原因になるパラメータの推定が明らかに悪い

$\boldsymbol{I}(\boldsymbol{\theta})$ はどのように非正則になっているか？

- SAR モデルにおけるフィッシャー情報量

$$\{I(\boldsymbol{\theta})\}_{pq} = \frac{1}{(2\pi)^n} \int \frac{\partial \log f(\boldsymbol{\omega})}{\partial \theta_p} \frac{\partial \log f(\boldsymbol{\omega})}{\partial \theta_q} \mathrm{d}\boldsymbol{\omega}$$

$$f(\boldsymbol{\omega}) = \frac{\sigma^2}{|P(z_1, \dots, z_n)|^2}, \qquad z_j = \exp(i\omega_1), j = 1, \dots, n$$

  - パラメータは $\boldsymbol{\theta} = (\beta_{-10}, \beta_{0-1}, \beta_{01}, \beta_{10}, \sigma)$

◆ 伝達関数の根の位置関係（パラメータの関係性）によって，$I(\theta)$ の形は異なる.

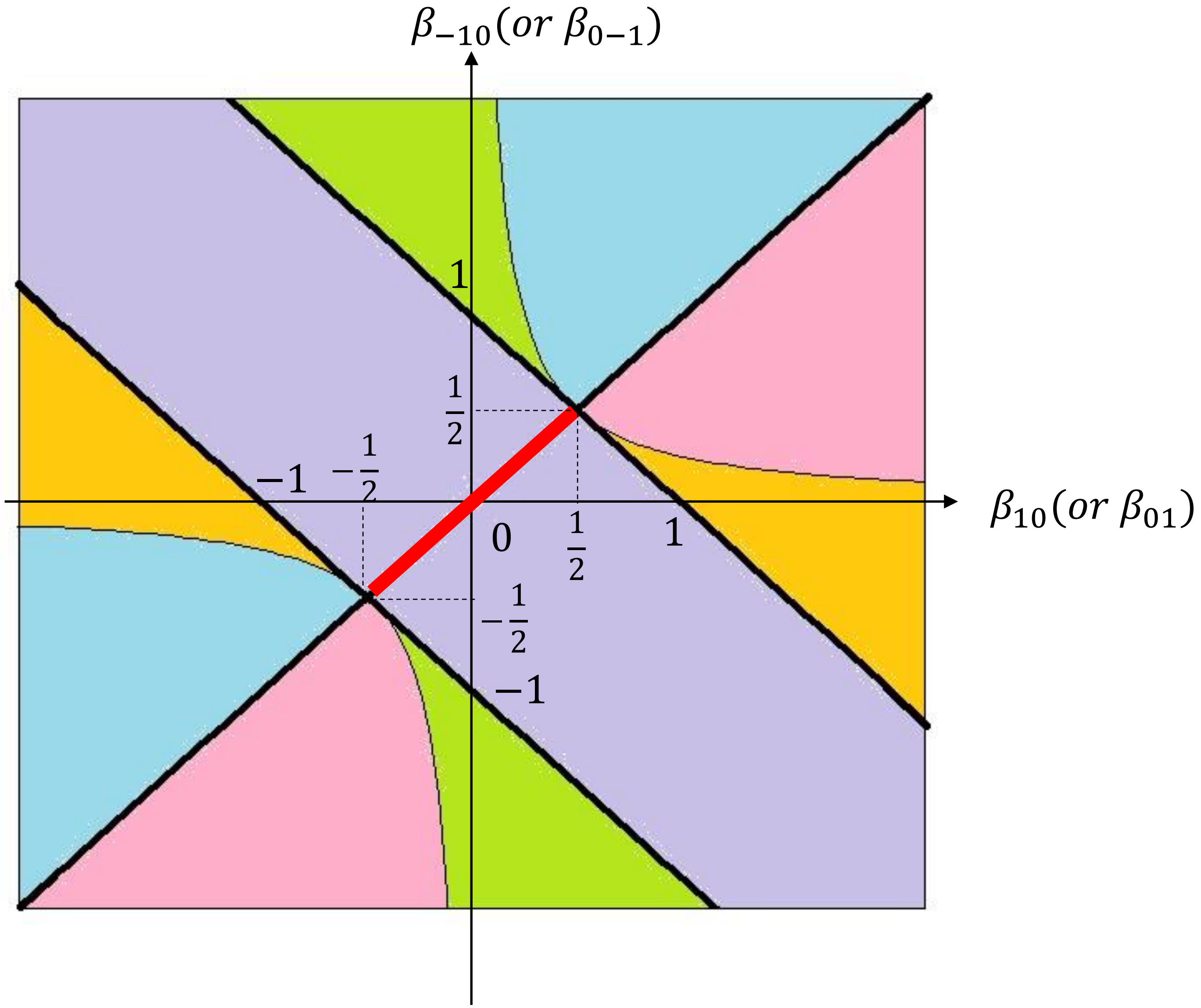
$$\{I(\boldsymbol{\theta})\}_{11}=\frac{1}{4\pi^2}\iint\left\{\frac{\partial\log f(\omega_1,\omega_2)}{\partial\beta_{-10}}\right\}^2 \mathrm{d}\omega_1\mathrm{d}\omega_2=\frac{1}{\pi}\int\left[\frac{z_1^{-2}}{\{P_1(z_1)\}^2}+\frac{1}{|P_1(z_1)|^2}\right]\mathrm{d}\omega_1$$

$$=\frac{1}{\beta_{10}^2\pi i}\int\frac{1}{z(z_1-\alpha_1)^2(z_1-\alpha_2)^2}\mathrm{d}z_1$$

$$+\frac{1}{\beta_{10}\beta_{-10}\,\pi i}\int\frac{z}{(z_1-\alpha_1)(z_1-\alpha_2)(z_1-\alpha_1^{-1})(z_1-\alpha_2^{-1})}\mathrm{d}z_1$$

根 α_1,α_2 が単位円の内側にあるかどうかで計算結果が変わる

- $P_1(z_1)=0$ の根を α_1,α_2 としている． $\alpha_1=\frac{-1+\sqrt{1-4\beta_{10}\beta_{-10}}}{2\beta_{10}},\alpha_2=\frac{-1-\sqrt{1-4\beta_{10}\beta_{-10}}}{2\beta_{10}}$



- 異なる色の領域で $I(\boldsymbol{\theta})$ は異なる表現をもつ.
- ━━ 上は非定常

（弱定常性の仮定より）

注目

非正則になるのは

紫の領域の ━━ 上のみ

紫の領域内で $I(\theta)$ は次のような形をしている.

$$A = 1 - 4\beta_{10}\beta_{-10}\ , \quad B = \frac{1}{2\sqrt{A}}\left\{\left(\frac{-1+\sqrt{A}}{\sqrt{A}}\right)^2 \left(1+2\sqrt{A}\right) + \frac{1}{P_1(1)P_1(-1)}\right\},$$

$$C = \left\{\frac{4}{\sqrt{A}(-1+\sqrt{A})}\right\}^2, \quad D = \frac{-2(1-\sqrt{A})}{\sigma\sqrt{A}}$$ とおく.

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \frac{B}{\beta_{-10}^2} & \beta_{-10}\beta_{0-1}C & \beta_{-10}\beta_{01}C & \frac{B}{\beta_{-10}\beta_{10}} & \frac{D}{\beta_{-10}} \\ \beta_{-10}\beta_{0-1}C & \frac{B}{\beta_{0-1}^2} & \frac{B}{\beta_{0-1}\beta_{01}} & \beta_{10}\beta_{0-1}C & \frac{D}{\beta_{0-1}} \\ \beta_{-10}\beta_{01}C & \frac{B}{\beta_{0-1}\beta_{01}} & \frac{B}{\beta_{01}^2} & \beta_{10}\beta_{01}C & \frac{D}{\beta_{01}} \\ \frac{B}{\beta_{-10}\beta_{10}} & \beta_{10}\beta_{0-1}C & \beta_{10}\beta_{01}C & \frac{B}{\beta_{10}^2} & \frac{D}{\beta_{10}} \\ \frac{D}{\beta_{-10}} & \frac{D}{\beta_{0-1}} & \frac{D}{\beta_{01}} & \frac{D}{\beta_{10}} & \frac{4}{\sigma^2} \end{pmatrix}$$

$\beta_{10} = \beta_{-10}$, $\beta_{01} = \beta_{0-1}$ とする.
（$I(\theta)$ を非正則にする）

$\beta_{10} = \beta_{-10} = \beta_1$，$\beta_{01} = \beta_{0-1} = \beta_2$（$I(\boldsymbol{\theta})$ が非正則になる）のとき，

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{a_1} & \boldsymbol{b} & \boldsymbol{b} & \boldsymbol{a_1} & \boldsymbol{c_1} \\ \boldsymbol{b} & \boldsymbol{a_2} & \boldsymbol{a_2} & \boldsymbol{b} & \boldsymbol{c_2} \\ \boldsymbol{b} & \boldsymbol{a_2} & \boldsymbol{a_2} & \boldsymbol{b} & \boldsymbol{c_2} \\ \boldsymbol{a_1} & \boldsymbol{b} & \boldsymbol{b} & \boldsymbol{a_1} & \boldsymbol{c_1} \\ \boldsymbol{c_1} & \boldsymbol{c_2} & \boldsymbol{c_2} & \boldsymbol{c_1} & \boldsymbol{d} \end{pmatrix}$$

このとき，MLE にどのような影響があるのか？

$\boldsymbol{I(\theta)}$ が非正則のとき，MLEで何が起きるか？

- m 個のパラメータに対し，$I(\boldsymbol{\theta})$ のランクが p 個落ちたとき，p 個のパラメータが漸近的に推定不可能．

$$0 = \frac{1}{\sqrt{N}} \frac{\partial \log p(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\hat{\theta}} \approx \frac{1}{\sqrt{N}} \frac{\partial \log p(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \boxed{\frac{1}{N} \frac{\partial^2 \log p(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}}$$

$$\frac{1}{N} \frac{\partial^2 \log p(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \longrightarrow -I(\boldsymbol{\theta})$$

$I(\boldsymbol{\theta})$ のランクが p 個落ちたら，方程式も p 個減る

$I(\theta)$が非正則のとき，何を推定するのか？

$$\begin{cases} a_1(\hat\beta_{-10}-\beta_{-10})+b(\hat\beta_{0-1}-\beta_{0-1})+b(\hat\beta_{01}-\beta_{01})+a_1(\hat\beta_{10}-\beta_{10})+c_1(\hat\sigma-\sigma)=0 \\ b(\hat\beta_{-10}-\beta_{-10})+a_2(\hat\beta_{0-1}-\beta_{0-1})+a_2(\hat\beta_{01}-\beta_{01})+b(\hat\beta_{10}-\beta_{10})+c_2(\hat\sigma-\sigma)=0 \\ b(\hat\beta_{-10}-\beta_{-10})+a_2(\hat\beta_{0-1}-\beta_{0-1})+a_2(\hat\beta_{01}-\beta_{01})+b(\hat\beta_{10}-\beta_{10})+c_2(\hat\sigma-\sigma)=0 \\ a_1(\hat\beta_{-10}-\beta_{-10})+b(\hat\beta_{0-1}-\beta_{0-1})+b(\hat\beta_{01}-\beta_{01})+a_1(\hat\beta_{10}-\beta_{10})+c_1(\hat\sigma-\sigma)=0 \\ c_1(\hat\beta_{-10}-\beta_{-10})+c_2(\hat\beta_{0-1}-\beta_{0-1})+c_2(\hat\beta_{01}-\beta_{01})+c_1(\hat\beta_{10}-\beta_{10})+d(\hat\sigma-\sigma)=0 \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} a_1\{(\hat\beta_{-10}+\hat\beta_{10})-(\beta_{-10}+\beta_{10})\}+b\{(\hat\beta_{0-1}+\hat\beta_{01})-(\beta_{0-1}+\beta_{01})\}+c_1(\hat\sigma-\sigma)=0 \\ b\{(\hat\beta_{-10}+\hat\beta_{10})-(\beta_{-10}+\beta_{10})\}+a_2\{(\hat\beta_{0-1}+\hat\beta_{01})-(\beta_{0-1}+\beta_{01})\}+c_2(\hat\sigma-\sigma)=0 \\ c_1\{(\hat\beta_{-10}+\hat\beta_{10})-(\beta_{-10}+\beta_{10})\}+c_2\{(\hat\beta_{0-1}+\hat\beta_{01})-(\beta_{0-1}+\beta_{01})\}+d(\hat\sigma-\sigma)=0 \end{cases}$$

$I(\theta)$ が非正則のとき，パラメータの和 $\beta_{-10}+\beta_{10}, \beta_{0-1}+\beta_{01}, \sigma$ を推定する

$\boldsymbol{I(\theta)}$ が非正則の場合の解決策

- m 個のパラメータに対し，$I(\boldsymbol{\theta})$ のランクが p 個落ちたとき，

  $m-p$ 本の尤度方程式と p 本の非正則条件を用いて推定する.

  - （例）$\beta_{10}+\beta_{-10}, \beta_{01}+\beta_{0-1}, \sigma$ を求める尤度方程式

$$
\begin{cases}
a_1\{(\widehat{\beta}_{-10}+\widehat{\beta}_{10})-(\beta_{-10}+\beta_{10})\}+b\{(\widehat{\beta}_{0-1}+\widehat{\beta}_{01})-(\beta_{0-1}+\beta_{01})\}+c_1(\hat{\sigma}-\sigma)=0 \\
b\{(\widehat{\beta}_{-10}+\widehat{\beta}_{10})-(\beta_{-10}+\beta_{10})\}+a_2\{(\widehat{\beta}_{0-1}+\widehat{\beta}_{01})-(\beta_{0-1}+\beta_{01})\}+c_2(\hat{\sigma}-\sigma)=0 \\
c_1\{(\widehat{\beta}_{-10}+\widehat{\beta}_{10})-(\beta_{-10}+\beta_{10})\}+c_2\{(\widehat{\beta}_{0-1}+\widehat{\beta}_{01})-(\beta_{0-1}+\beta_{01})\}+d(\hat{\sigma}-\sigma)=0
\end{cases}
$$

  と非正則条件 $\beta_{10}=\beta_{-10}, \beta_{01}=\beta_{0-1}$ から $\beta_{10}, \beta_{-10}, \beta_{01}, \beta_{0-1}, \sigma$ を推定する.

$\boldsymbol{I(\theta)}$が正則である場合に起きる問題

- $I(\boldsymbol{\theta})$ が正則になる場合でも，推定の過程で $\beta_{10} = \beta_{-10}$ や $\beta_{01} = \beta_{0-1}$ になると推定はその条件を保ったままの結果になってしまう．
  - $\beta_{10} = 0.3, \beta_{-10} = 0.5, \beta_{01} = 0.2, \beta_{0-1} = 0.7, \sigma = 0.01$ の乱数を生成
  - 20 回の試行のうち，6回が $\beta_{10} = \beta_{-10}$ になってしまった

試行番号	$\boldsymbol{\beta_{10}}$	$\boldsymbol{\beta_{-10}}$	$\boldsymbol{\beta_{01}}$	$\boldsymbol{\beta_{0-1}}$	$\boldsymbol{\sigma}$
7 回目	0.375912736	0.375912736	0.242644206	0.654915747	0.009873481
9 回目	0.393632223	0.393635391	0.134337373	0.756401097	0.010701736
10 回目	0.383414812	0.383424875	0.219344761	0.678277834	0.010160032
11 回目	0.396038230	0.396043109	0.192676695	0.694267185	0.009782115
17 回目	0.398392461	0.398397014	0.209448643	0.680307478	0.009892484
18 回目	0.405991267	0.405937688	0.241373211	0.658419457	0.009934828

推定の過程で $I(\theta)$ が非正則になると，推定がうまくいかない

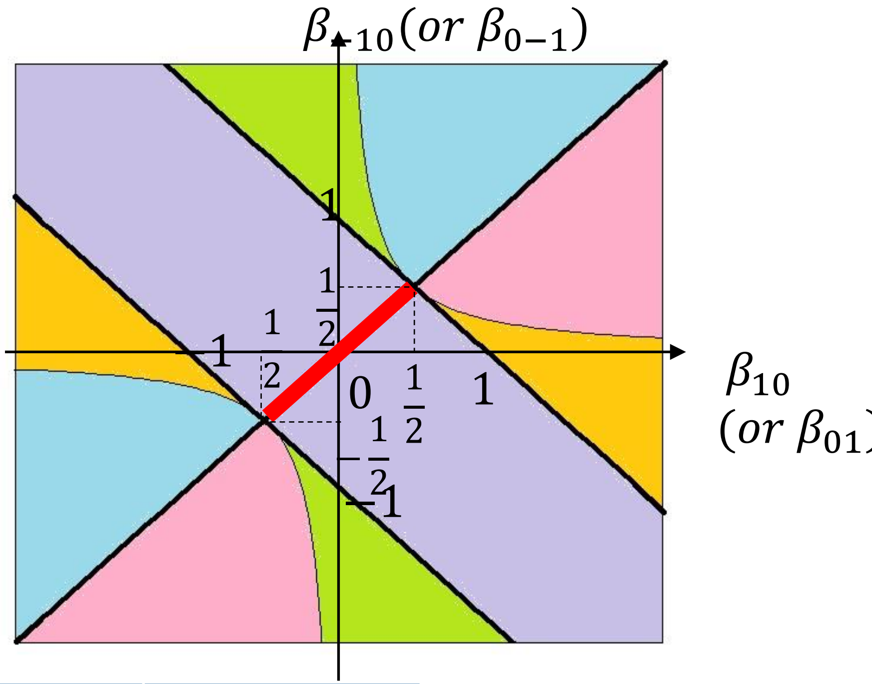
$I(\theta)$が正則な場合に非正則条件を避ける方法

- $|\beta_{-10}-\beta_{10}|<\varepsilon$ と $|\beta_{0-1}-\beta_{01}|<\varepsilon$ の範囲を避けて推定する.
  - R の constrOptim 関数で

（$\beta_{10}-\beta_{-10}\leq\varepsilon$ または $-\beta_{10}+\beta_{-10}\leq\varepsilon$）

かつ（$\beta_{01}-\beta_{0-1}\leq\varepsilon$ または $-\beta_{01}+\beta_{0-1}\leq\varepsilon$）

の 4 領域でそれぞれ推定し，尤度を比較し，推定値を決定する.



	L_A	β_{10}	β_{-10}	β_{01}	β_{0-1}	σ
真の値		0.3	0.5	0.2	0.7	0.01
領域1	7.50393	0.34033	0.48186	0.18647	0.71066	0.00952
領域2	0.77390	−0.34041	0.43109	0.64382	−0.81232	1.05585
領域3	−1.75965	0.32389	−0.41508	−0.70680	0.77031	3.75567
領域4	−5.02098	1.06396	0.35851	1.46577	−0.31399	19.18454

課題

- 初期値の選択
  - $I(\theta)$ の非正則条件が複数ある場合，分ける領域が増える
  - データによっては領域間で最大尤度が同じ値になることがある（対象な位置？）

	L_A	β_{10}	β_{-10}	β_{01}	β_{0-1}	σ
真の値		0.3	0.5	0.2	0.7	0.01
領域1	7.47957	0.29258	0.49819	0.19401	0.68762	0.01053
領域2	−5.37035	−0.34262	0.43037	0.64441	−0.80972	1.02145
領域3	7.47957	0.49830	0.29250	0.19402	0.68761	0.01053
領域4	0.83927	0.92898	−0.04678	1.44851	−0.52519	22.84632

- 与えられたデータについて，$I(\boldsymbol{\theta})$ が正則か非正則かを判定する方法