発表資料は「tanaken-basis(検索)」の「explasso」
からダウンロードできます。

（この案内は発表資料の中で何回も出てきます。）

実験計画法における計画行列の生成について

2019年3月7日（木）

成蹊大学　田中 研太郎

発表資料は「tanaken-basis(検索)」の「explasso」
からダウンロードできます。
（この案内は発表資料の中で何回も出てきます。）

今回の発表

- 前半：実験計画を生成するお話
- 後半：実験計画における代数的なお話

今回の発表資料は、色々なところで使ったものを
つなぎ合わせたりして作っているので、統一感が無いです。

発表資料は「tanaken-basis(検索)」の「explasso」
からダウンロードできます。

（この案内は発表資料の中で何回も出てきます。）

前半：実験計画を生成するお話

まず、

実験計画を生成することは、

スパースモデリングなんです！

（変数選択）

というお話から始めます。

実験計画法 の 目的 の１つ

少ない実験回数でなるべく多くの情報を獲得したい。

↓

スパースモデリング！！

スパースモデリング（英語: Sparse modeling、スパース sparse とは「すかすか」、「少ない」を意味する）または疎性モデリングとは、少ない情報から全体像を的確にあぶり出す科学的モデリング。

https://ja.wikipedia.org/wiki/スパースモデリング

以下、簡単のため、2水準(−1 or 1) の 因子が 3つ ある場合で説明。

- 2水準の因子が3つ $x_1, x_2, x_3 = -1 \text{ or } 1$
- y： a, b, c についての線形モデルで表されるとする

モデル
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$
未知パラメーター　誤差

回帰分析　未知のパラメーターを実験データから精度よく推定したい。

$2^3 = 8$ 通りの実験（候補）

	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

注：以降のページでも各列が1つの実験（観測）に対応！！

回帰分析：　線形モデルのいくつかの未知パラメーターを精度よく推定したい。

＋

実験計画法：　少ない実験回数でなるべく多くの情報を獲得したい。

実験データの候補（８通り）

	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

パラメーター推定のための
情報をなるべく失わずに、かつ、
なるべく実験回数を少なく
するためには、どの実験を
どれだけ選べばよいか？

この問題は、スパースモデリングの代表的な手法の
１つである Lasso の問題として定式化できる

モデル

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

未知パラメーター　　誤差

とりあえず
簡単のために

β_1, β_2 を推定したいと仮定。少ない実験回数 で 精度よく 推定したい。

不偏であり，分散を小さくするような
線形推定量 $\hat{\beta}_1, \hat{\beta}_2$ を考える。

つまり…

実験データの候補（8通り）

(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

線形推定量

$$\hat{\beta}_1 = \sum_{g=1}^{8} w_{1g}\, y_g$$

$$\hat{\beta}_2 = \sum_{g=1}^{8} w_{2g}\, y_g$$

係数 $\boldsymbol{w_{1g}, w_{2g}}$ を適切に選べば、不偏 にしたり 分散を小さく したりできる。

発表資料は「tanaken-basis(検索)」の「explasso」からダウンロードできます。
（この案内は発表資料の中で何回も出てきます。）

線形推定量

$$\hat{\beta}_1 = \sum_{g=1}^{8} w_{1g}\, y_g$$

$$\hat{\beta}_2 = \sum_{g=1}^{8} w_{2g}\, y_g$$

さらに…

$\hat{\beta}_1$と$\hat{\beta}_2$の両方で、例えば y_2 の係数が 0（$w_{12} = w_{22} = 0$）だとすると、2番目の実験結果をパラメーターの推定に使わないということなので、この実験自体やる必要がなくなる。

つまり…

$w_{12} = w_{22} = 0$ なら、この2番目の実験は不要になる。

(定数)	1	~~1~~	1	1	1	1	1	1
x_1	-1	~~-1~~	-1	-1	1	1	1	1
x_2	-1	~~-1~~	1	1	-1	-1	1	1
x_3	-1	~~1~~	-1	1	-1	1	-1	1
	y_1	~~y_2~~	y_3	y_4	y_5	y_6	y_7	y_8

回帰分析：　線形モデルのいくつかの未知パラメーターを精度よく推定したい。

線形推定量

$$\boxed{\hat{\beta}_1 = \sum_{g=1}^{8} w_{1g}\, y_g \quad , \quad \hat{\beta}_2 = \sum_{g=1}^{8} w_{2g}\, y_g}$$

係数 w_{1g}, w_{2g} を適切に選べば、
不偏にしたり分散を小さくしたりできる。

＋

実験計画法：　少ない実験回数でなるべく多くの情報を獲得したい。

（例えば $w_{1g} = w_{2g} = 0$ なら、g番目の実験は不要になる。）

線形制約付きの Group Lasso というもので定式化できる。

扱いたい問題

線形モデル $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ を考える。
未知パラメータのうち β_1, β_2 を推定したいと仮定。
なるべく少ない実験回数で 精度よく推定したい。
不偏であり、分散を小さくするような 線形推定量 を考えたい。

線形制約付きの Group Lasso

$$\underset{\boldsymbol{w}_1, \boldsymbol{w}_2 \in \boldsymbol{R}^8}{\text{minimize}} \underbrace{\{\|\boldsymbol{w}_1\|^2 + \|\boldsymbol{w}_2\|^2\}}_{\text{分散}} + \underbrace{\sum_{g=1}^{8} \lambda_g \left| \sqrt{w_{1g}^2 + w_{2g}^2} \right|}_{\text{グループごと（実験}g\text{ごと）のスパースネス（⇨ 少ない実験回数）}}$$

subject to

$$\underbrace{\begin{pmatrix} \boldsymbol{M}\boldsymbol{w}_1 \\ \boldsymbol{M}\boldsymbol{w}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{pmatrix}}_{\text{不偏性}}$$

詳しくは次のページから…

不偏性のための線形制約

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \qquad (\, x_1, x_2, x_3 = -1 \text{ or } 1 \,)$$

未知パラメーター　　誤差

(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

β_1 の線形不偏推定量の例： $\hat{\beta}_1 = \dfrac{1}{2}\{-y_1 + y_5\}$

たしかに

$$E[\hat{\beta}_1] = E\left[\frac{1}{2}\{-y_1 + y_5\}\right] = \frac{1}{2}\{-(\beta_0 - \beta_1 - \beta_2 - \beta_3) + (\beta_0 + \beta_1 - \beta_2 - \beta_3)\} = \beta_1$$

となって、不偏。

不偏性のための線形制約

$$
\begin{array}{c|cccccccc}
 & ① & ② & ③ & ④ & ⑤ & ⑥ & ⑦ & ⑧ \\
(\text{定数}) & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
x_1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
x_2 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
x_3 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8
\end{array}
= M
$$

行列

β_1 の線形不偏推定量の例をもう一度考えると…

$$E[\hat{\beta}_1] = E\left[\frac{1}{2}\{-y_1 + y_5\}\right] = \frac{1}{2}\{-(\beta_0 - \beta_1 - \beta_2 - \beta_3) + (\beta_0 + \beta_1 - \beta_2 - \beta_3)\}$$

$$= (\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3)\underbrace{\left\{-\frac{1}{2}\begin{pmatrix}1\\-1\\-1\\-1\end{pmatrix} + \frac{1}{2}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\right\}}_{=M\boldsymbol{w}_1} = (\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3)\underbrace{\begin{pmatrix}0\\1\\0\\0\end{pmatrix}}_{=\boldsymbol{e}_1} = \beta_1$$

つまり、 $M\boldsymbol{w}_1 = \boldsymbol{e}_1$ であることが、不偏であることと同値。 ⇨ 線形制約

推定量の分散

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \qquad (x_1, x_2, x_3 = -1 \text{ or } 1)$$

未知パラメーター　誤差

誤差 ε は、平均0, 分散σ^2 であり、各実験で独立であると仮定する。

線形推定量 $\hat{\beta}_1 = \sum_{g=1}^{8} w_{1g}\, y_g$ の分散：

$$\boxed{\mathrm{var}[\hat{\beta}_1] = \sum_{g=1}^{8} w_{1g}^2\, \mathrm{var}[y_g] = \sigma^2 \sum_{g=1}^{8} w_{1g}^2 = \sigma^2 \|\boldsymbol{w}_a\|^2} \Rightarrow \text{目的関数}$$

$\hat{\beta}_2$についても同様。

扱いたい問題

線形モデル $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ を考える。
未知パラメータのうち β_1, β_2 を推定したいと仮定。
なるべく少ない実験回数で 精度よく推定したい。
不偏であり、分散を小さくするような 線形推定量 を考えたい。

線形制約付きの Group Lasso

$$\underset{\boldsymbol{w}_1, \boldsymbol{w}_2 \in \boldsymbol{R}^8}{\text{minimize}} \underbrace{\{\|\boldsymbol{w}_1\|^2 + \|\boldsymbol{w}_2\|^2\}}_{\text{分散}} + \sum_{g=1}^{8} \lambda_g \left|\sqrt{w_{1g}^2 + w_{2g}^2}\right|$$

グループごと（実験gごと）のスパースネス
（⇨ 少ない実験回数）

subject to

$$\begin{pmatrix} \boldsymbol{M}\boldsymbol{w}_1 \\ \boldsymbol{M}\boldsymbol{w}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{pmatrix}$$

不偏性

実験計画法の問題 が 最適化問題 として定式化される。
（今回はLassoを使ったが、変数選択の他の手法を使ってもできそう。）

最適化問題を解くためのアルゴリズムはすでに色々あり、利用可能。

今回の説明では、不偏な推定量を考えたが、多少の交絡を許すような感じでの定式化も余裕でできる。
もっと因子数や水準数が多い場合でも余裕（定式化するだけなら）。

いずれにしても、2次錘計画を解くアルゴリズムを利用すれば解ける。
（線形制約があるので、Lasso 関連のパッケージがそのままでは使えなかったりする。）

状況に応じた実験計画を生成する Python の プログラムを作成してアップロードしている。
(https://github.com/tanaken-basis(検索)/explasso) ⇐「田中研太郎」で検索

たとえば…

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \qquad (\, x_1, x_2, x_3 = -1 \text{ or } 1 \,)$$

未知パラメーター

誤差

β_1, β_2 の２つを推定したいとして、Group lasso を(で？)解くと…

全候補

(定数)	1	1	1	1	1	1	1	1
x_1	−1	−1	−1	−1	1	1	1	1
x_2	−1	−1	1	1	−1	−1	1	1
x_3	−1	1	−1	1	−1	1	−1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

３つ選出

(定数)	1	1	1
x_1	−1	−1	1
x_2	−1	1	−1
x_3	−1	−1	−1
	y_1	y_3	y_5

ノートPCとかでも数秒！！

さらに…

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \qquad (\ x_1, x_2, x_3 = -1 \ \text{or} \ 1\ )$$

未知パラメーター　　誤差

$\beta_1, \beta_2, \beta_3$ の 3 つを推定したいとして、Group lasso すると…

全候補

(定数)	1	1	1	1	1	1	1	1
x_1	-1	1	-1	1	-1	1	-1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	-1	-1	-1	1	1	1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

4 つ選出

(定数)	1	1	1	1
x_1	1	1	-1	-1
x_2	1	-1	1	-1
x_3	1	-1	-1	1
	y_8	y_2	y_3	y_5

L_4 直交表！

やっぱり数秒！！

発表資料は「tanaken-basis(検索)」の「explasso」からダウンロードできます。
（この案内は発表資料の中で何回も出てきます。）

実験計画を自動生成する１つの手段として有効。

多少、因子数や水準数が増えても、ノートPCで計算できそう。

ただし、因子数が１０００個！ とかは現状では全然無理。

さらに、本当はパラメーターをいろいろといじっている。

じつは、このパラメーターの設定のところに罠が…。

対称性の問題

もし、罰則項の係数がすべて等しいと…　（$\lambda_1 = \cdots = \lambda_8$）

$$\underset{\boldsymbol{w}_1, \boldsymbol{w}_2 \in \boldsymbol{R}^8}{\text{minimize}} \ \{\|\boldsymbol{w}_1\|^2 + \|\boldsymbol{w}_2\|^2\} + \lambda \sum_{g=1}^{8} \left| \sqrt{w_{1g}^2 + w_{1g}^2} \right|$$

subject to

$$\begin{pmatrix} \boldsymbol{M}\boldsymbol{w}_1 \\ \boldsymbol{M}\boldsymbol{w}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{pmatrix}$$

- 上の問題では、変数について対称性がある。 $\left(\begin{array}{l} w_{i1} \leftrightarrow -w_{i8} \ , \ w_{i2} \leftrightarrow -w_{i7} \\ w_{i3} \leftrightarrow -w_{i6} \ , \ w_{i4} \leftrightarrow -w_{i5} \end{array} \right)$
- Group Lasso の問題は凸最適化なので、解は一つ。

対称性の問題

もし、罰則項の係数がすべて等しいと…

- Group Lasso の問題において, **変数についての対称性** が生じる。
- Group Lasso の問題は, **凸**最適化なので, **解は一つ**。

３因子の場合は全部で8通りの実験がある。

説明のために、一時的に少し設定を変えて、 $\beta_1, \beta_2, \beta_3$を推定したい場合を考える。
この場合、実験計画法の考え方では、以下のL4とかが計算で出てくれるとうれしい。
しかし、例えば、以下の **2 つは等価**で、どちらも直交表。

L4

定数	**1**	**1**	**1**	**1**
x1	**1**	**1**	**−1**	**−1**
x2	**1**	**−1**	**1**	**−1**
x3	**1**	**−1**	**−1**	**1**

L4’

定数	**1**	**1**	**1**	**1**
x1	**−1**	**−1**	**1**	**1**
x2	**−1**	**1**	**−1**	**1**
x3	**−1**	**1**	**1**	**−1**

よって, L4 は Group lasso の最適解として出てこない!!

つまり, Group Lasso の問題において, 変数についての
対称性を破るように, 罰則項などの係数を決めないと,
全部の実験を使うような解しか出てこない。
そもそも, 疎な解は非対称。

解のパターン

得られる最適解は、λ_gの値に依存する。

$$\underset{\boldsymbol{w}_a, \boldsymbol{w}_b \in \boldsymbol{R}^8}{\text{minimize}} \ \{\|\boldsymbol{w}_a\|^2 + \|\boldsymbol{w}_b\|^2\} + \sum_{g=1}^{8} \lambda_g \left| \sqrt{w_{1g}^2 + w_{2g}^2} \right|$$

$$\text{subject to} \ \begin{pmatrix} \boldsymbol{M}\boldsymbol{w}_a \\ \boldsymbol{M}\boldsymbol{w}_b \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_a \\ \boldsymbol{e}_b \end{pmatrix}$$

λ_g：g 番目の実験のやりたくなさ(？)を表す

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	−1	−1	−1	−1	1	1	1	1
x_2	−1	−1	1	1	−1	−1	1	1
x_3	−1	1	−1	1	−1	1	−1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

解のパターン

$$\boxed{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon}$$

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

例1） $\boxed{\beta_1, \beta_2}$ を推定したいとする。

⇨ $(\lambda_1, \cdots, \lambda_8) = (10, 100, 20, 200, 40, 400, 80, 800)$

⇨ Group lasso の問題を解く ⇨ 最適解：

	①	③	⑤
定数	1	1	1
x_1	−1	−1	1
x_2	−1	1	−1
x_3	−1	−1	−1

解のパターン

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

例2) $\boxed{\beta_1, \beta_2, \beta_3}$ を推定したいとする。

⇨ $(\lambda_1, \cdots, \lambda_8) = (100, 0, 0, 100, 0, 100, 100, 0)$

⇨ Group lasso の問題を解く ⇨ **L4**

	⑧	⑤	③	②
定数	1	1	1	1
x_1	1	1	−1	−1
x_2	1	−1	1	−1
x_3	1	−1	−1	1

解のパターン

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

例3) $\boxed{\beta_1, \beta_2}$ を推定したいとする。

⇨ $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8$ （変数に対称性がある！）

⇨ Group lasso ⇨

	①	②	③	④	⑤	⑥	⑦	⑧
定数	1	1	1	1	1	1	1	1
x_1	−1	−1	−1	−1	1	1	1	1
x_2	−1	−1	1	1	−1	−1	1	1
x_3	−1	1	−1	1	−1	1	−1	1

全部選ばれてしまう…。

解のパターン

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

例4) $\boxed{\beta_1, \beta_2}$ を推定したいとする。

$\Rightarrow$ $\lambda_1, \lambda_3, \lambda_5 < \lambda_2, \lambda_4, \lambda_6, \lambda_7, \lambda_8$

$\Rightarrow$ Group lasso $\Rightarrow$

	①	③	⑤
定数	1	1	1
x_1	-1	-1	1
x_2	-1	1	-1
x_3	-1	-1	-1

or

	①	③	⑤	⑦
定数	1	1	1	1
x_1	-1	-1	1	1
x_2	-1	1	-1	-1
x_3	-1	-1	-1	1

or…

解のパターン

例5) $\boxed{\beta_1, \beta_2}$ を推定したいとする。

⇨ $\lambda_1, \lambda_2, \lambda_5 < \lambda_4 < \lambda_3, \lambda_6, \lambda_7, \lambda_8$

⇨ Group lasso ⇨

	①	②	④	⑤
定数	1	1	1	1
x_1	−1	−1	−1	1
x_2	−1	−1	1	−1
x_3	−1	1	1	−1

or…

例6) $\boxed{\beta_1, \beta_2}$ を推定したいとする。

⇨ $\lambda_1, \lambda_2, \lambda_4 < \lambda_5 < \lambda_3, \lambda_6, \lambda_7, \lambda_8$

⇨ Group lasso ⇨

	①	②	④	⑤
定数	1	1	1	1
x_1	−1	−1	−1	1
x_2	−1	−1	1	−1
x_3	−1	1	1	−1

or…

異なった $\lambda_1, \ldots, \lambda_8$ の順番が同じ解のパターンを与える。

Question 解のパターンを決めているのは何か？ ➡ 代数(グレブナー基底)で分かる？

ちなみに、対称性の問題は、今回の実験計画の生成だけの問題ではなく、
普通の Lasso でも起こりうると考えられる。

例えば、、、

モデル

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

データ

(定数)	1	1
x_1	0	1
x_2	0	1
x_3	0	1
y	0	1

つまり、4つの変数に対して2つのデータがある場合に、Lassoでパラメーターの推定を
考えると、以下のような感じ。

$$\underset{\beta_0,\cdots,\beta_3 \in \boldsymbol{R}}{\text{minimize}} \ \left[\{0-\beta_0\}^2 + \{1-(\beta_0+\beta_1+\beta_2+\beta_3)\}^2\right] + \lambda(|\beta_0| + |\beta_1| + |\beta_2| + |\beta_3|)$$

$$\underset{\beta_0,\cdots,\beta_3 \in \boldsymbol{R}}{\text{minimize}} \quad [\{0-\beta_0\}^2 + \{1-(\beta_0+\beta_1+\beta_2+\beta_3)\}^2] + \lambda(|\beta_0|+|\beta_1|+|\beta_2|+|\beta_3|)$$

上の式には $\beta_1, \beta_2, \beta_3$ の交換に対する対称性がある。
一方で、凸な関数なので、最適解は1つ。

$\beta_1 > 0$ で、 $\beta_2 = \beta_3 = 0$ という解は最適解にならないはず。

ただし、Rとかで実際に計算してみると、
Coordinate Descent 法 (座標降下法) で最適化する場合が多いためか、
1番目のパラメーター β_1 のみが非ゼロで、あとはゼロになる感じの結果(スパースな結果)になったりする。
変数の順番に依存してしまうような感じで、それはそれで問題かもしれない。。。
スパースモデリングでは、ただ単にスパースにすればよいわけではなく、
対称性の破り方も考えないといけないかも。

発表資料は「tanaken-basis(検索)」の「explasso」
からダウンロードできます。

（この案内は発表資料の中で何回も出てきます。）

後半：実験計画における代数的なお話

発表資料は「tanaken-basis(検索)」の「explasso」
からダウンロードできます。
（この案内は発表資料の中で何回も出てきます。）

つぎに、

実験計画法における

計算代数の使われ方について

大雑把に説明します。

発表資料は「tanaken-basis(検索)」の「explasso」
からダウンロードできます。
（この案内は発表資料の中で何回も出てきます。）

「計算代数統計」は、Pistone先生 と Wynn先生が
「計算代数」を「実験計画法」に応用したことが始まり(おそらく)。

この辺の話は、「計算代数統計（青木 敏 (著), 共立出版）」で
とても分かりやすく説明されている。

今回は、計算代数がどんなものかというお話はあまりしない。
大体の感じだけ説明する。

まず、 Pistone-Wynn流の計算代数の実験計画への応用方法を
おおまかに説明してから、別の応用方法について説明する。

Pistone-Wynn流の計算代数の実験計画への応用では、

たとえば、以下のような **実験計画が与えられている** とする。

	①	②	③	④
(定数)	1	1	1	1
x_1	1	1	-1	-1
x_2	1	-1	1	-1
x_3	1	-1	-1	1
	y_1	y_2	y_3	y_4

これは、3つの変数 x_1, x_2, x_3 に対する4つの実験からなる計画行列。（L_4直行表）

このとき、たとえば、

- x_1の行を2乗したものは、定数の行と同じ ➡ この実験では x_1^2 と 定数 を区別できず、これらの効果を同時には推定できない。
- x_1の行とx_2の行をかけたものは、x_3の行と同じ ➡ この実験では x_1x_2と x_3を区別できず、これらの効果を同時には推定できない。

	①	②	③	④
(定数)	1	1	1	1
x_1	1	1	-1	-1
x_2	1	-1	1	-1
x_3	1	-1	-1	1
	y_1	y_2	y_3	y_4

つまり、計画行列における行同士の関係を見ていくと、

（注：今回の話では、慣例的な計画行列の書き方とは行と列の関係が逆になっている。）

- x_1の行を2乗したものは、定数の行と同じ ➡ $x_1^2 - 1 = 0$
- x_1の行とx_2の行をかけたものは、x_3の行と同じ ➡ $x_1 x_2 - x_3 = 0$

みたいな代数的な関係が見えてくる。

	①	②	③	④
(定数)	1	1	1	1
x_1	1	1	-1	-1
x_2	1	-1	1	-1
x_3	1	-1	-1	1
	y_1	y_2	y_3	y_4

計算代数では、計画行列を行ごとにみてグレブナー基底を作ることで、

（注：慣例的な計画行列の書き方では列ごと。）

$$x_1^2 - 1 = 0 \ , \quad x_2^2 - 1 = 0 \ , \quad x_3^2 - 1 = 0 \ ,$$

$$x_1 x_2 - x_3 = 0 \ , \quad x_2 x_3 - x_1 = 0 \ , \quad x_1 x_3 - x_2 = 0$$

という関係を導出し、さらにそこから、以下の標準単項式(上の多項式で割れないもの)

$$1, \quad x_1, \quad x_2, \quad x_3$$

が得られる。そして、これらの標準単項式の効果だけであれば推定できる。

つまり、Pistone-Wynn流の計算代数の実験計画への応用では、
以下のような計画行列が与えられたときに、

	①	②	③	④
(定数)	1	1	1	1
x_1	1	1	-1	-1
x_2	1	-1	1	-1
x_3	1	-1	-1	1
	y_1	y_2	y_3	y_4

この計画行列を行ごとに見て、計算代数的な手法で導出される
「標準単項式」からなるモデル

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

におけるパラメーター $\beta_0, \beta_1, \beta_2, \beta_3$ の推定が可能であることがわかる。

要するに

Pistone-Wynn流

計画行列 $\xrightarrow{\text{計算代数}}$ 推定可能なモデル
(行ごと)

という感じ。

(注：慣例的な計画行列の書き方では列ごと。)

発表資料は「tanaken-basis(検索)」の「explasso」
からダウンロードできます。
（この案内は発表資料の中で何回も出てきます。）

（注：慣例的な計画行列の書き方では行ごと。）

つぎに、計画行列を 列ごと に見た場合に、

計算代数の手法がどのように応用できそうか考えてみる。

今度は、まず、モデルが与えられている とする。

そして、そのモデルのパラメーターを推定するためには

どのような計画行列が必要なのかを考えていく。

まず、推定したいパラメーターを持つモデルが、以下のように与えられたとする。

$$y = \underbrace{\beta_0}_{} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \qquad (x_1, x_2, x_3 = -1 \text{ or } 1)$$

未知パラメーター（$\beta_0, \beta_1, \beta_2, \beta_3$）　誤差（$\varepsilon$）

そして、以下の全実験 **1** から **8** の、どれがあればパラメーターを推定できるのかを考えていく。

	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

	①	②	③	④	⑤	⑥	⑦	⑧
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

ここで、計画行列の列ごとの関係を考えてみる。

例えば、**1** と **4** の列を足し、 **2** と **3** の列を引くと、ゼロベクトルになる。

つまり、全ての効果が打ち消されて、 $y_1 + y_4 - y_2 - y_3$ は平均的にはゼロになるはず。

$y_1 + y_4 - y_2 - y_3$ は、パラメーターの推定に特に役には立たなそう。

つぎに、例えば、全実験の行列において、 x_1の行を消したものを考える。

$$
\begin{array}{c|cccccccc}
 & ① & ② & ③ & ④ & ⑤ & ⑥ & ⑦ & ⑧ \\
(\text{定数}) & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\cancel{x_1} & \cancel{-1} & \cancel{-1} & \cancel{-1} & \cancel{-1} & \cancel{1} & \cancel{1} & \cancel{1} & \cancel{1} \\
x_2 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
x_3 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8
\end{array}
$$

このとき、例えば、 **5** の列から **1** の列を引くとゼロベクトルになる。

ただし、x_1の行があるとすると、その成分はゼロにならない。

つまり、x_1以外の効果が打ち消されて、 $y_5 - y_1$ には x_1 の効果のみが現れる。

よって、$y_5 - y_1$ (の定数倍)で

$$
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon
$$

のパラメーター β_1 を推定できることがわかる。

(注: 前のページの $y_1 + y_4 - y_2 - y_3$ でもゼロベクトルになるが、 β_1 は推定できない。)

もっとちゃんと言うと…　（前半でも同じような説明をしている）

$$
\begin{array}{c|cccccccc}
 & ① & ② & ③ & ④ & ⑤ & ⑥ & ⑦ & ⑧ \\
(\text{定数}) & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
x_1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
x_2 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
x_3 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8
\end{array}
= M
$$

行列

β_1 は、$(y_5 - y_1)/2$ によって不偏推定できることが以下のようにわかる。

$$
E[\hat{\beta}_1] = E\left[\frac{1}{2}\{-y_1 + y_5\}\right] = \frac{1}{2}\{-(\beta_0 - \beta_1 - \beta_2 - \beta_3) + (\beta_0 + \beta_1 - \beta_2 - \beta_3)\}
$$

$$
= (\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3)\underbrace{\left\{-\frac{1}{2}\begin{pmatrix}1\\-1\\-1\\-1\end{pmatrix} + \frac{1}{2}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\right\}}_{= M\boldsymbol{w}_1} = (\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3)\underbrace{\begin{pmatrix}0\\1\\0\\0\end{pmatrix}}_{= \boldsymbol{e}_1} = \beta_1
$$

つまり、 $M\boldsymbol{w}_1 = \boldsymbol{e}_1$ 　であることが、不偏であることと同値。

モデルが以下のように与えられていて、

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \qquad (\ x_1, x_2, x_3 = -1 \ \text{or} \ 1\ )$$

全実験を以下の行列で表したときに、

$$\begin{array}{c|cccccccc}
 & ① & ② & ③ & ④ & ⑤ & ⑥ & ⑦ & ⑧ \\
(定数) & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
x_1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
x_2 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
x_3 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8
\end{array} = M$$

行列

以下のことがいえる。

T ：列ベクトル①,⋯,⑧　の部分集合

T の線形結合で
β_1 を不偏推定できる

同値

T の線形結合で $\begin{pmatrix} 0 \\ k \\ 0 \\ 0 \end{pmatrix}$ が作れる。

$(k \neq 0)$

さらに、x_1の行を消した時だけゼロベクトルになる列の組み合わせで不偏推定できるので…

$$\begin{array}{c|c}
 & \begin{array}{cccccccc} ① & ② & ③ & ④ & ⑤ & ⑥ & ⑦ & ⑧ \end{array} \\
\begin{array}{c} (\text{定数}) \\ x_1 \\ x_2 \\ x_3 \end{array} & \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}
\end{array} = \boldsymbol{M}$$

$$\begin{array}{c|c}
 & \begin{array}{cccccccc} ① & ② & ③ & ④ & ⑤ & ⑥ & ⑦ & ⑧ \end{array} \\
\begin{array}{c} (\text{定数}) \\ \cancel{x_1} \\ x_2 \\ x_3 \end{array} & \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \cancel{-1} & \cancel{-1} & \cancel{-1} & \cancel{-1} & \cancel{1} & \cancel{1} & \cancel{1} & \cancel{1} \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}
\end{array} = \boldsymbol{M_{-1}}$$

補題　（準同型定理の応用？）

T ：列ベクトル ①,⋯,⑧ の部分集合

T の線形結合で β_1 を不偏推定できる ⟺ T の線形結合で $\begin{pmatrix} 0 \\ k \\ 0 \\ 0 \end{pmatrix}$ が作れる。 $(k \neq 0)$ ⟺ $T \in \ker M_{-1} \,/\, \ker M$

$\ker M$ と $\ker M_{-1}$ の基底は、M とM_{-1} を「配置行列」としたトーリックイデアルのグレブナー基底を計算することで得られる。

$$
\begin{array}{c|c}
 & \begin{array}{cccccccc} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 & t_8 \end{array} \\
\begin{array}{c} (\text{定数}) \\ x_1 \\ x_2 \\ x_3 \end{array} &
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
-1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1
\end{pmatrix}
\end{array}
= M
$$

$$
\begin{array}{c|c}
 & \begin{array}{cccccccc} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 & t_8 \end{array} \\
\begin{array}{c} (\text{定数}) \\ \cancel{x_1} \\ x_2 \\ x_3 \end{array} &
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\cancel{-1} & \cancel{-1} & \cancel{-1} & \cancel{-1} & \cancel{1} & \cancel{1} & \cancel{1} & \cancel{1} \\
-1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1
\end{pmatrix}
\end{array}
= M_{-1}
$$

$$
\left\{
\begin{array}{ll}
\ker M: & \text{toric ideal} = \langle t_1t_4 - t_2t_3,\ t_1t_6 - t_2t_5,\ t_1t_8 - t_2t_7,\ \cdots \rangle \\
\ker M_{-1}: & \text{toric ideal} = \langle t_1 - t_5,\ t_2 - t_6,\ t_3 - t_7,\ t_4 - t_8,\ t_1t_4 - t_2t_3 \rangle \\
\ker M \,/\, \ker M_{-1}: & \langle t_1 - t_5,\ t_2 - t_6,\ t_3 - t_7,\ t_4 - t_8 \rangle
\end{array}
\right.
$$

Gröbner bases

β_1 の任意の線形不偏推定量は以下の線形結合で表される 。

$$\boxed{y_1 - y_5,\ y_2 - y_6,\ y_3 - y_7,\ y_4 - y_8}$$

ちなみに、β_2 を推定したい場合は…

$$
\begin{array}{c|c}
 & \begin{array}{cccccccc} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 & t_8 \end{array} \\
\begin{array}{c} (\text{定数}) \\ x_1 \\ x_2 \\ x_3 \end{array} &
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
-1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1
\end{pmatrix}
\end{array} = M
$$

$$
\begin{array}{c|c}
 & \begin{array}{cccccccc} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 & t_8 \end{array} \\
\begin{array}{c} (\text{定数}) \\ x_1 \\ \cancel{x_2} \\ x_3 \end{array} &
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
\cancel{-1} & \cancel{-1} & \cancel{1} & \cancel{1} & \cancel{-1} & \cancel{-1} & \cancel{1} & \cancel{1} \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1
\end{pmatrix}
\end{array} = M_{-2}
$$

ker M :　toric ideal = $< t_1t_4 - t_2t_3, t_1t_6 - t_2t_5, t_1t_8 - t_2t_7, \cdots >$

ker M_{-2} :　toric ideal = $< t_1 - t_3, t_2 - t_4, t_5 - t_7, t_6 - t_8, t_1t_6 - t_2t_5 >$

ker M / ker M_{-2} :　$< t_1 - t_3, t_2 - t_4, t_5 - t_7, t_6 - t_8 >$

Gröbner bases

β_2 の任意の線形不偏推定量は以下の線形結合で表される 。

$$\boxed{y_1 - y_3, y_2 - y_4, y_5 - y_7, y_6 - y_8}$$

つまり、今回紹介した計算代数の実験計画への応用では、
以下のようなモデルが与えられたときに、

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

全実験からなる行列を列ごとに見て、

$$\begin{array}{c|c}
 & \begin{array}{cccccccc} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 & t_8 \end{array} \\
\begin{array}{c} (\text{定数}) \\ x_1 \\ x_2 \\ x_3 \end{array} & \begin{pmatrix} \boxed{\begin{array}{c} 1 \\ -1 \\ -1 \\ -1 \end{array}} & \boxed{\begin{array}{c} 1 \\ -1 \\ -1 \\ 1 \end{array}} & \boxed{\begin{array}{c} 1 \\ -1 \\ 1 \\ -1 \end{array}} & \boxed{\begin{array}{c} 1 \\ -1 \\ 1 \\ 1 \end{array}} & \boxed{\begin{array}{c} 1 \\ 1 \\ -1 \\ -1 \end{array}} & \boxed{\begin{array}{c} 1 \\ 1 \\ -1 \\ 1 \end{array}} & \boxed{\begin{array}{c} 1 \\ 1 \\ 1 \\ -1 \end{array}} & \boxed{\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array}} \end{pmatrix}
\end{array} = \boldsymbol{M}$$

計算代数的な手法を使うことで、 $\beta_0, \beta_1, \beta_2, \beta_3$ の推定に必要な
実験の組み合わせの最小単位が導かれる。

要するに

今回の方法

$$\text{モデル，全実験の行列} \xrightarrow{\text{計算代数}} \text{計画行列}$$

(列ごと)

という感じ。

（注：慣例的な計画行列の書き方では行ごと。）

ちなみに、前半の話で、変数についての対称性が罰則項によってどう破られるのかが問題になったが、推定に必要な実験の組み合わせの最小単位を考えると、ある程度は分かる。

$$
\begin{array}{c|cccccccc}
 & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 & \lambda_7 & \lambda_8 \\
\begin{array}{c}\text{(定数)}\\ x_1 \\ x_2 \\ x_3\end{array} &
\multicolumn{8}{c}{\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
-1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1
\end{pmatrix}} \\
 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8
\end{array}
$$

$\boxed{\beta_1, \beta_2}$ を推定したいとする。

- β_1 の任意の線形不偏推定量は以下の線形結合で表される：

  $\boxed{y_1 - y_5, y_2 - y_6, y_3 - y_7, y_4 - y_8}$

- β_2 の任意の線形不偏推定量は以下の線形結合で表される：

  $\boxed{y_1 - y_3, y_2 - y_4, y_5 - y_7, y_6 - y_8}$

$\boxed{\lambda_1, \lambda_2, \lambda_5 < \lambda_4 < \lambda_3, \lambda_6, \lambda_7, \lambda_8}$ $\boxed{\lambda_1, \lambda_2, \lambda_4 < \lambda_5 < \lambda_3, \lambda_6, \lambda_7, \lambda_8}$

どちらの場合も $y_1 - y_5$ と $y_2 - y_4$ が選ばれる。

まとめ

発表資料は「tanaken-basis(検索)」の「explasso」からダウンロードできます。
（この案内は発表資料の中で何回も出てきました。）

- 実験計画の生成は、スパースモデリング的な方法でできる。
ただし、変数の対称性があるとスパースにならなかったり。対称性をどう破るかが重要。
- 実験計画法への計算代数の応用

Pistone-Wynn流

計画行列（**行**ごと） $\xrightarrow{\text{計算代数}}$ 推定可能なモデル

（注：慣例的な計画行列の書き方では<u>列ごと</u>。）

今回の方法

モデル，全実験の行列（**列**ごと） $\xrightarrow{\text{計算代数}}$ 計画行列

（注：慣例的な計画行列の書き方では<u>行ごと</u>。）

- 任意の線形不偏推定量は、「グレブナー基底で与えられるデータ列の線形結合」の線形結合で表される。つまり最小単位がある。例えば、最小二乗推定量も、「グレブナー基底で与えられるデータ列の線形結合」の線形結合の形に分解できる。