

グレブナー基底で MCMC

@kyusque

グレブナー基底

- 計算機代数と実際のコンピューターの計算をつなぐ
- 僕もあまりよく分かってません
 - 数学にしては具体的でイメージしやすい(らしい)
- 取っ掛かりとしては、
グレブナー基底大好きbotさんの本を読むとい
いと思います→
- 実は今回、グレブナー
基底の計算はほぼ出な
いです
 - (裏では計算しているみ
たいです)



妹がグレブナー基底に興味を持ち始めたのだが。①
グレブナー基底大好きbot、ぴーす

★★★★★☆ v 7

Kindle版 (電子書籍)

¥0 kindleunlimited

Kindle Unlimited会員は読み放題で読書
または、¥533で購入



妹がグレブナー基底に興味を持ち始めたのだが。②
グレブナー基底大好きbot、ぴーす

★★★★★☆ v 1

Kindle版 (電子書籍)

¥0 kindleunlimited

Kindle Unlimited会員は読み放題で読書
または、¥533で購入



妹がグレブナー基底に興味を持ち始めたのだが。③
グレブナー基底大好きbot

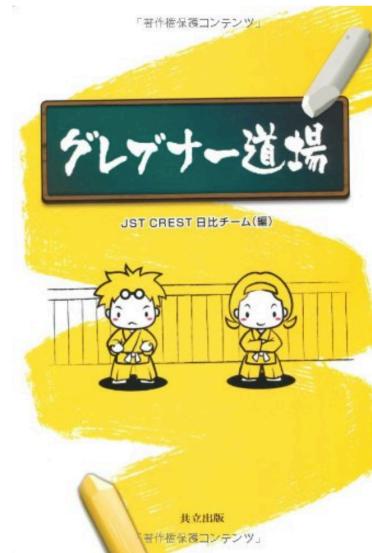
Kindle版 (電子書籍)

¥0 kindleunlimited

Kindle Unlimited会員は読み放題で読書
または、¥533で購入

さらに学ぶために(統計スタート)

- グレブナー道場への入門書
→
 - 数学科はCox, Little, O'Shea の *Ideals, Varieties, and Algorithms* の第一章を英語で読むがお勧めだそうです
 - <http://www.math.kobe-u.ac.jp/OpenXM/Math/dojo/do-jo-mult/dojo-howto-read.html>
- グレブナー道場へ入門するための入門書→
- 統計が分かる人はこちらの方が分かりやすい?→



実はRの先人が



Rで代数統計 TokyoR #42 LT

11,240 views

Share

Like

Download

...



Toru Imai, Assistant Professor at Kyoto University

+ Follow

- <https://www.slideshare.net/motivic/tokyor42algstat>

流れ

- ↓このPDFに沿ってやっていきます。
- マルコフ基底と実験計画（青木、竹村）
 - <http://www.math.kobe-u.ac.jp/crest-c/cs-2012/2012-02-22-aoki-takemura.pdf>
- あんまり自信がないので間違ってたらご指摘お願いします
- プログラムは以下に置いてます
 - <https://github.com/kyusque/japanr2019>

カイ二乗検定

胃ガンに対する喫煙のリスクを調べるため、胃ガン患者 20 人と健常者 100 人に対し、過去の喫煙習慣の有無を調査したところ、以下の結果を得た。喫煙習慣は胃ガンに関係あるといえるか？

	喫煙歴あり	喫煙歴なし	合計
胃ガン患者	14	6	20
健常者	56	44	100

(数値は架空)

- 行和、列和を固定して検定統計量のカイ二乗を計算
- カイ二乗分布に従うと仮定してp値を導出
 -  ここを代わりにMCMCで計算

カイ二乗検定

+ -
- +

の形の推移のみで十分（マルコフ基底をなす）

	喫煙歴あり	喫煙歴なし	合計
胃ガン患者	14 + -	6	20
健常者	56 - +	44	100

(数値は架空)

- ・グレブナー基底の計算で導出したマルコフ基底で行和、列和を固定して推移させていく

各種変換とマルコフ基底による推移

↓ここをMCMCで推移させる

$$\begin{array}{cc|cc} 14 & 6 & 20 & \\ 56 & 44 & 100 & \\ \hline 70 & 50 & 120 & \end{array} \rightarrow \left(\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array} \right) X = \left(\begin{array}{c} 20 \\ 100 \\ 70 \\ 50 \end{array} \right)$$

↑行和、列和は固定

A.mat

1	4	4	
2	1	1	0
3	0	0	1
4	1	0	1
5	0	1	0
6			

(4ti2)
\$ markov A

A.mar

1	1	4	
2	1	-1	-1
3			

👉 裏でグレブナー基底を計算しているらしい

+ - の形の推移 →
- +

計算の流れ

- 歸無仮説では、ある X が取り出される確率は超幾何分布に従う
- 観測された X_{obs} から得られる統計量 $\chi^2(X_{\text{obs}})$ より、MCMCで得られる X から得られる統計量 $\chi^2(X)$ が大きい状況がどれくらいの頻度で起きるか(p 値)
- 減多になければ棄却

MCMCのアルゴリズム

アルゴリズムの例 (アルゴリズム 4.2.4)

入力: 観測値 x^o , マルコフ基底 \mathcal{B} , 総ステップ数 N , 配置行列 A ,

帰無分布 $f(\cdot)$, 検定統計量 $T(\cdot)$, ($t = Ax^o$ とおく)

出力: p 値の推定値

変数: obs , count , sig , \mathbf{x} , \mathbf{x}_{next}

ステップ 1: $\text{obs} = T(\mathbf{x}^o)$, $\mathbf{x} = \mathbf{x}^o$, $\text{count} = 0$, $\text{sig} = 0$

ステップ 2: $\mathbf{z} \in \mathcal{B}$ をランダムに, $\varepsilon \in \{-1, +1\}$ を等確率で選ぶ.

ステップ 3: $\mathbf{x} + \varepsilon \mathbf{z} \notin \mathcal{F}_t$ であれば $\mathbf{x}_{next} = \mathbf{x}$ としてステップ 5 へ. $\mathbf{x} + \varepsilon \mathbf{z} \in \mathcal{F}_t$ であれば u を 0 と 1 の間の一様乱数とする.

ステップ 4: $u \leq \frac{f(\mathbf{x} + \varepsilon \mathbf{z})}{f(\mathbf{x})}$ であれば $\mathbf{x}_{next} = \mathbf{x} + \varepsilon \mathbf{z}$ としてステップ 5 へ. $u > \frac{f(\mathbf{x} + \varepsilon \mathbf{z})}{f(\mathbf{x})}$ であれば $\mathbf{x}_{next} = \mathbf{x}$ としてステップ 5 へ.

ステップ 5: $T(\mathbf{x}_{next}) \geq \text{obs}$ であれば $\text{sig} = \text{sig} + 1$

ステップ 6: $\mathbf{x} = \mathbf{x}_{next}$, $\text{count} = \text{count} + 1$

ステップ 7: $\text{count} < N$ であればステップ 2 へ

ステップ 8: p 値の推定値は sig/N

```
{r}  
mcmc(data_, move, 100000, chi_square, hypergeometric_dist)  
```
```

```
[1] 9 11 61 39
[1] 14 6 56 44
[1] 11 9 59 41
[1] 12 8 58 42
[1] 13 7 57 43
[1] 13 7 57 43
[1] 11 9 59 41
[1] 10 10 60 40
[1] 9 11 61 39
[1] 8 12 62 38
[1] 12 8 58 42
[1] 9 11 61 39
[1] 15 5 55 45
[1] 13 7 57 43
[1] 13 7 57 43
[1] 14 6 56 44
[1] 7 13 63 37
[1] 12 8 58 42
[1] 16 4 54 46
[1] 12 8 58 42
[1] 14 6 56 44
[1] 9 11 61 39
[1] 14 6 56 44
[1] 10 10 60 40
[1] 9 11 61 39
[1] 10 10 60 40
[1] 13 7 57 43
[1] 12 8 58 42
[1] 10 10 60 40
[1] 11 9 59 41
[1] 10 10 60 40
[1] 9 11 61 39
[1] 14 6 56 44
```

```
[1] 10 10 60 40
[1] 11 9 59 41
[1] 15 5 55 45
[1] 12 8 58 42
[1] 12 8 58 42
[1] 10 10 60 40
[1] 13 7 57 43
[1] 15 5 55 45
[1] 9 11 61 39
[1] 11 9 59 41
[1] 16 4 54 46
[1] 12 8 58 42
[1] 8 12 62 38
[1] 9 11 61 39
[1] 13 7 57 43
[1] 14 6 56 44
[1] 13 7 57 43
[1] 16 4 54 46
[1] 15 5 55 45
[1] 11 9 59 41
[1] 12 8 58 42
[1] 13 7 57 43
[1] 8 12 62 38
[1] 15 5 55 45
[1] 12 8 58 42
[1] 8 12 62 38
[1] 8 12 62 38
[1] 14 6 56 44
[1] 14 6 56 44
[1] 11 9 59 41
[1] 12 8 58 42
[1] 13 7 57 43
[1] 13 7 57 43
[1] 0.32363
```

```
[1] 10 10 60 40
[1] 11 9 59 41
[1] 15 5 55 45
[1] 12 8 58 42
[1] 12 8 58 42
[1] 10 10 60 40
[1] 13 7 57 43
[1] 15 5 55 45
[1] 9 11 61 39
[1] 11 9 59 41
[1] 16 4 54 46
[1] 12 8 58 42
[1] 8 12 62 38
[1] 9 11 61 39
[1] 13 7 57 43
[1] 14 6 56 44
[1] 13 7 57 43
[1] 16 4 54 46
[1] 15 5 55 45
[1] 11 9 59 41
[1] 12 8 58 42
[1] 13 7 57 43
[1] 8 12 62 38
[1] 15 5 55 45
[1] 12 8 58 42
[1] 8 12 62 38
[1] 8 12 62 38
[1] 14 6 56 44
[1] 14 6 56 44
[1] 11 9 59 41
[1] 12 8 58 42
[1] 13 7 57 43
[1] 13 7 57 43
[1] 0.32363
```

```
```{r}  
chisq.test(data)  
```
```

Pearson's Chi-squared test with Yates' continuity cor

data: data  
X-squared = 0.82971, df = 1, p-value = 0.3624

# まとめ

- グレブナー基底(マルコフ基底)から計算した遷移でカイ二乗検定のp値をMCMCで計算できた。
  - カイ二乗検定の妥当性についてはJulia言語界隈などで激論されている?みたいなので、今後も勉強していきたい

黒木玄 Gen Kuroki @genkuroki · 11月17日  
返信先: @genkurokiさん  
#統計 2×2の分割表では、独立性の帰無仮説と期待値のオッズ比が1になることは同値になります。  
仮説  $M(w)$  のもとでサンプルのP値が5%未満にならない  $w$  の範囲を  $w$  に関する95%信頼区間と定義すれば、色々すっきりするのに、#R言語の fisher.test では全然そうならないように見えます。続く

```
1 fisher.test(matrix(c(10, 6, 5, 15), ncol=2, byrow=T))
```

R言語の場合 Fisher's Exact Test for Count Data

```
data: matrix(c(10, 6, 5, 15), ncol = 2, byrow = T)
p-value = 0.04095
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.983679 26.768384
sample estimates:
odds ratio
4.756095
```

P値が5%未満なのに、オッズ比の95%信頼区間に1が含まれている！

- 今回は遷移が一つだったが、表がより複雑になると遷移の数が増える
  - 3x3x3分割表などでも遷移を導出できる
- 実験計画法にも使えるらしい