

RANDOM行列(仮)

@katoshoo

2019/03/02 Tokyo.R

自己紹介

Twitter	@katoshoo
職業	学部4年生（4月から院生）
卒論	ランダム行列のファイナンスへの応用的な
趣味	映画鑑賞、飲酒
最近	インターン頑張る
悩み	研究テーマが決まらない、彼女ができない

自己紹介

Twitt

職業

卒論

趣味

最近

悩み

R超初心者です
温かい目で見守ってください

的な

ない

今日の目標

- ・ 自分の卒業研究の内容を絡めて、ランダム行列の漸近固有値分布について伝える
- ・ 色んな人と仲良くなる

今日の目標

- ・ 自分の卒業研究の内容を絡めて、ランダム行列の漸近固有値

分布について伝える

- ・ 色んな人と仲良く

Rを有効活用する場面が全然無い
(ごめんなさい)

ランダム行列とは

ランダム行列とは

▶ 確率変数を要素にもつ行列

(例) ランダム行列 $\mathbf{X} \in \mathbb{R}^{N \times M}$ の各要素が独立に $N(0, 1^2)$

に従う場合

$$\mathbf{X}\mathbf{X}^T \sim W_N(M, \mathbf{I}_N)$$

N次元Wishart分布に従う

Wishart行列と呼びましょう



ランダム行列の歴史

- ▶ 1920年代に数理統計学の分野で導入された
- ▶ 1950年代にWignerが原子物理学へ応用し、ランダム行列の固有値分布についての統計的性質を明らかにした
- ▶ 現在は遺伝子工学、金融工学、無線工学、複雑ネットワークなどへ応用されている



- ▶ 1920年代に数理統計学の分野で導入された
- ▶ 1950年代にWignerが原子物理学へ応用し、ランダム行列の固有値分布についての統計的性質を明らかにした
- ▶ 現在は量子工学、金融工学、無線工学、複雑ネットワーク

今回はここに注目！



ランダム行列の漸近固有値分布

- ▶ Wishart行列の漸近固有値分布がすごく有名で有用

- ▶ Wishart行列の漸近固有値分布がすごく有名で有用

Marčenko & Pastur
distribution

データ行列が標準正規乱数により定まる
"雑音成分"のみで与えられるときの
標本分散共分散行列の固有値分布

Marčenko & Pastur
distribution

$Q = M/N$ とし、 Q 一定のもとで $N \rightarrow \infty$ としたとき、

Wishart行列の固有値 λ の分布は

$$P_{RM}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}$$

$$\lambda_{max,min} = \left[1 \pm \frac{1}{\sqrt{Q}} \right]^2$$

ランダム行列といえは

▶ psychパッケージのfa.parallel関数

分析するデータと同じサンプルサイズのランダム行列を用意し、その相関行列の固有値の推移と分析データのそれを比較してくれる

平行分析をやってくれる関数

ランダム行列といえば

▶ 人口データ

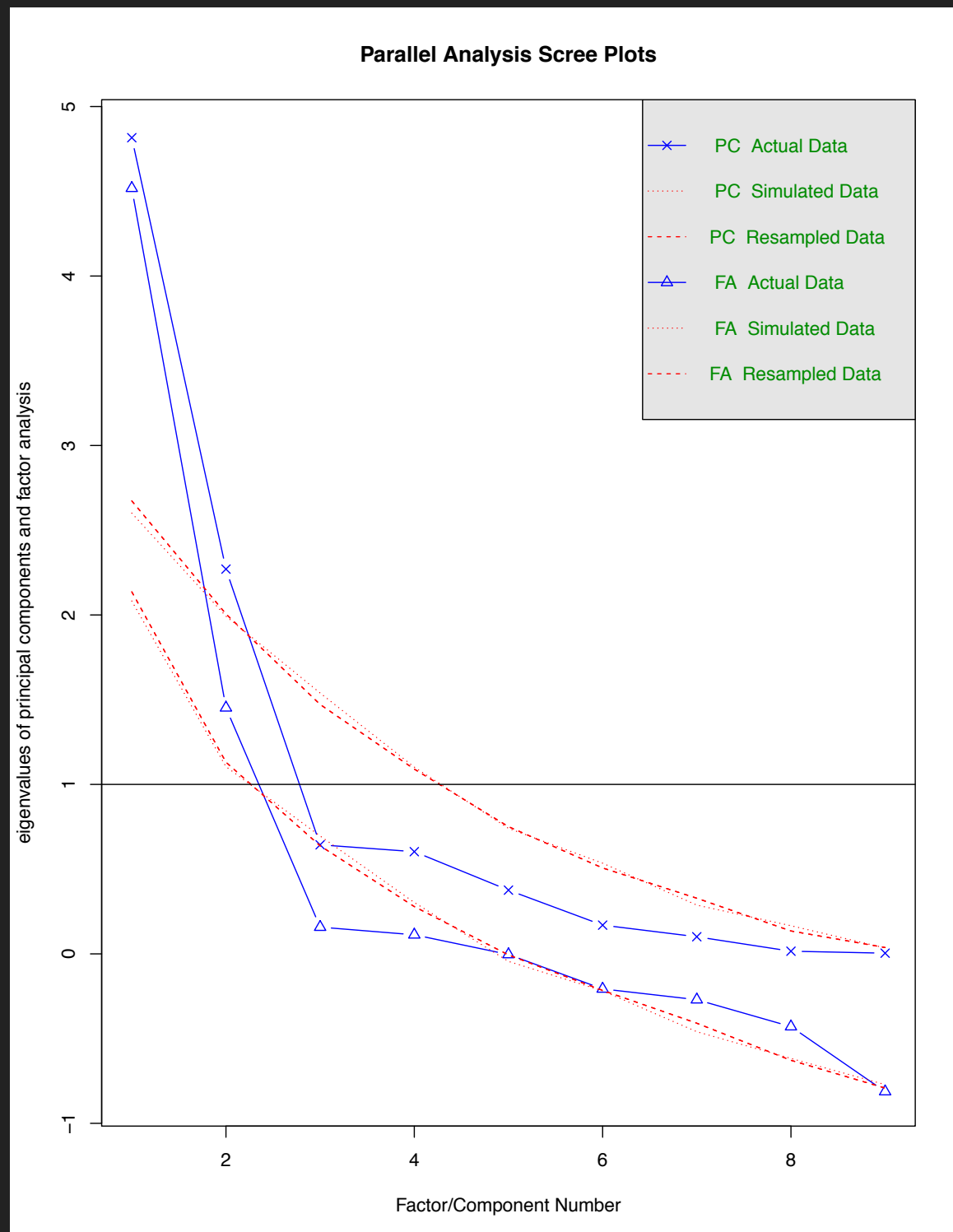
自分が吸ったことのあるタバコの銘柄を個人的な評価指標で並べてみました

smoking

	spicy	sweet	refreshed	heavy	savory	smell	Burning time	price	taste
SevenStars(7mm)	3	5	6	5	2	4	4	5	3
Mevius(6mm)	3	4	6	4	3	3	4	5	5
CABIN(8mm)	7	2	3	6	7	5	4	4	6
Peace(6mm)	2	7	5	5	3	2	5	7	3
Caster(5mm)	1	9	4	2	1	1	3	4	2
American Spirit(8mm)	6	2	5	3	6	2	9	9	7
HOPE(6mm)	6	2	3	6	6	5	3	1	6
LUCKY STRIKE(6mm)	8	1	2	8	8	6	2	7	5
LARK(6mm)	4	3	4	5	7	6	3	4	5
KENT(6mm)	3	4	5	5	5	4	2	4	5
Marlboro Gold(6mm)	5	3	5	4	4	5	5	8	4
CAMEL(6mm)	4	4	4	3	2	4	1	2	2

ランダム行列といえば

```
parallel <- fa.parallel(dat) #平行分析
```



因子数2つで説明できそう

ほんとは `vss()` でMAP基準
とかも見て因子数決めた方が
良いと思うけどとりあえず

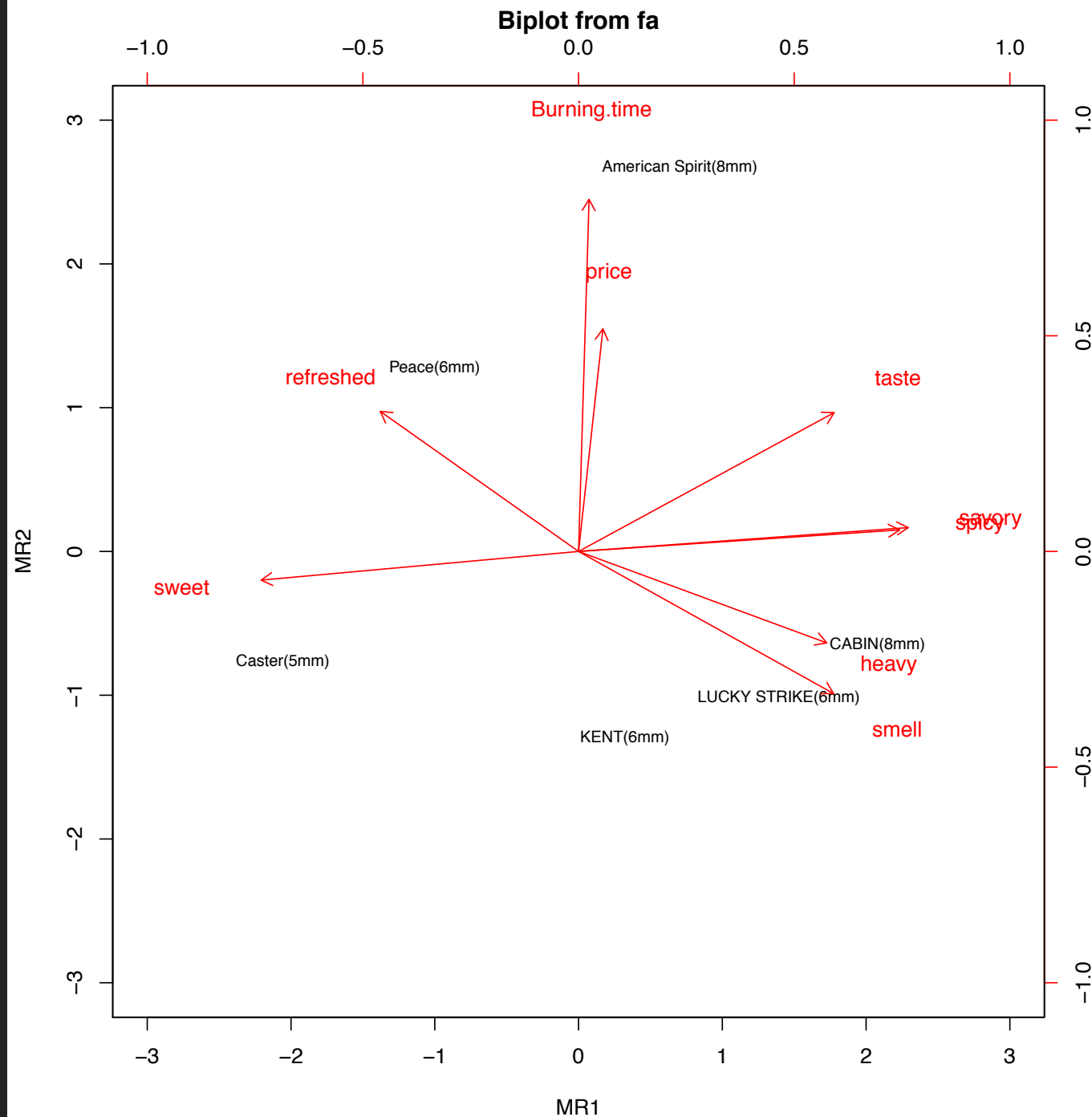
本当は実際のデータに合わせて因子の抽出法も
色々検討すべきですがとりあえず

```
res <- fa(dat, nfactors=2, fm="minres", rotate="oblimin")
```

```
print(res, digits=3)
```

```
biplot(res, labels=rownames(dat))
```

ランダム行列といえば



香ばしさと辛さ、重さや臭いが次元1に
燃焼時間と価格が次元2
に対応しているのが分かる

僕の好みは燃焼時間が
長くて香ばしくて辛い
値段はあまり気にしない

別にこんなのを
紹介したいわけじゃなくて

Marčenko & Pastur distribution が

色んな分野で活躍してる

イメージはfa.parallel()と同じ、ランダム行列の固有値よりも

大きい固有値に注目

▶ イメージはノイズ除去

ランダム行列の最大固有値よりも大きいところに分布している固有値に真の相関があるんじゃないか

ランダム行列の固有値分布に入っている部分は、ノイズとして除去してしまえ

ファイナンスへの応用

▶ 株式市場における相関行列のフィルタリング

標準化した対数収益率の相関行列を考える

$$\mathbf{C} = \mathbf{C}^{market} + \mathbf{C}^{group} + \mathbf{C}^{random} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \sum_{i=2}^R \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \sum_{i=R+1}^N \lambda_i \mathbf{u}_i \mathbf{u}_i^T$$

市場全体の要素、業種間の要素、ランダムな要素

本当に紹介したいこと

- ▶ 株式市場における相関行列を分解したいが

標準化した対数収

ここがランダム行列の固有値分布に従う部分
ここをノイズとして除去しよう

$$\mathbf{C} = \mathbf{C}^{market} + \mathbf{C}^{group} + \underline{\mathbf{C}^{random}} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \sum_{i=2}^R \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \sum_{i=R+1}^N \lambda_i \mathbf{u}_i \mathbf{u}_i^T$$

▶ 利用したデータ

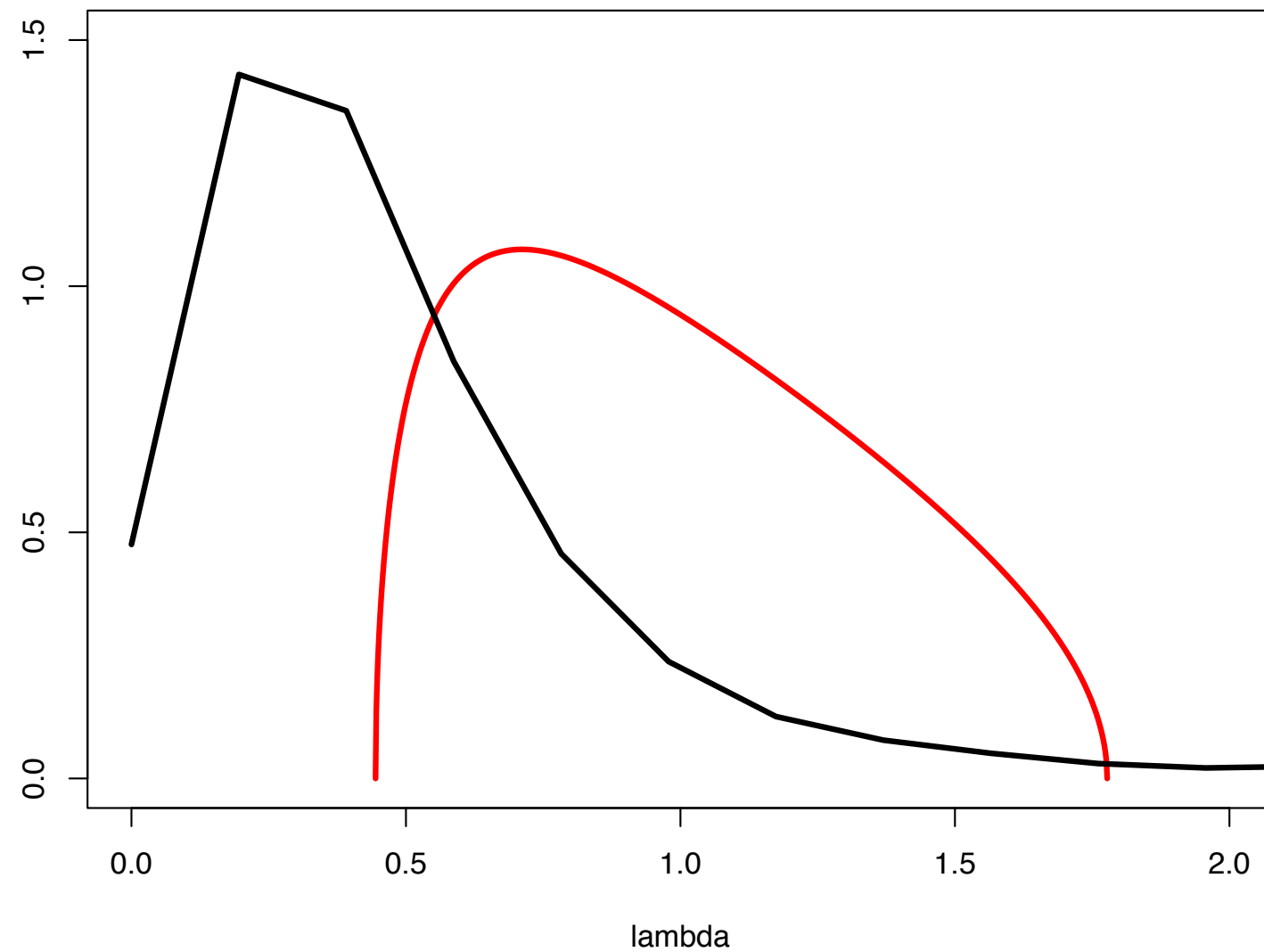
日経平均株価の算出に使われている225銘柄

ブルームバーグ様からデータをいただきました

Yahoo!ファイナンスなどからデータ取ってくるなら
quantmodパッケージのgetSymbols関数で取ってこれる
(これしか知らないだけ)

本当に紹介したいこと

Distribution of Eigenvalue

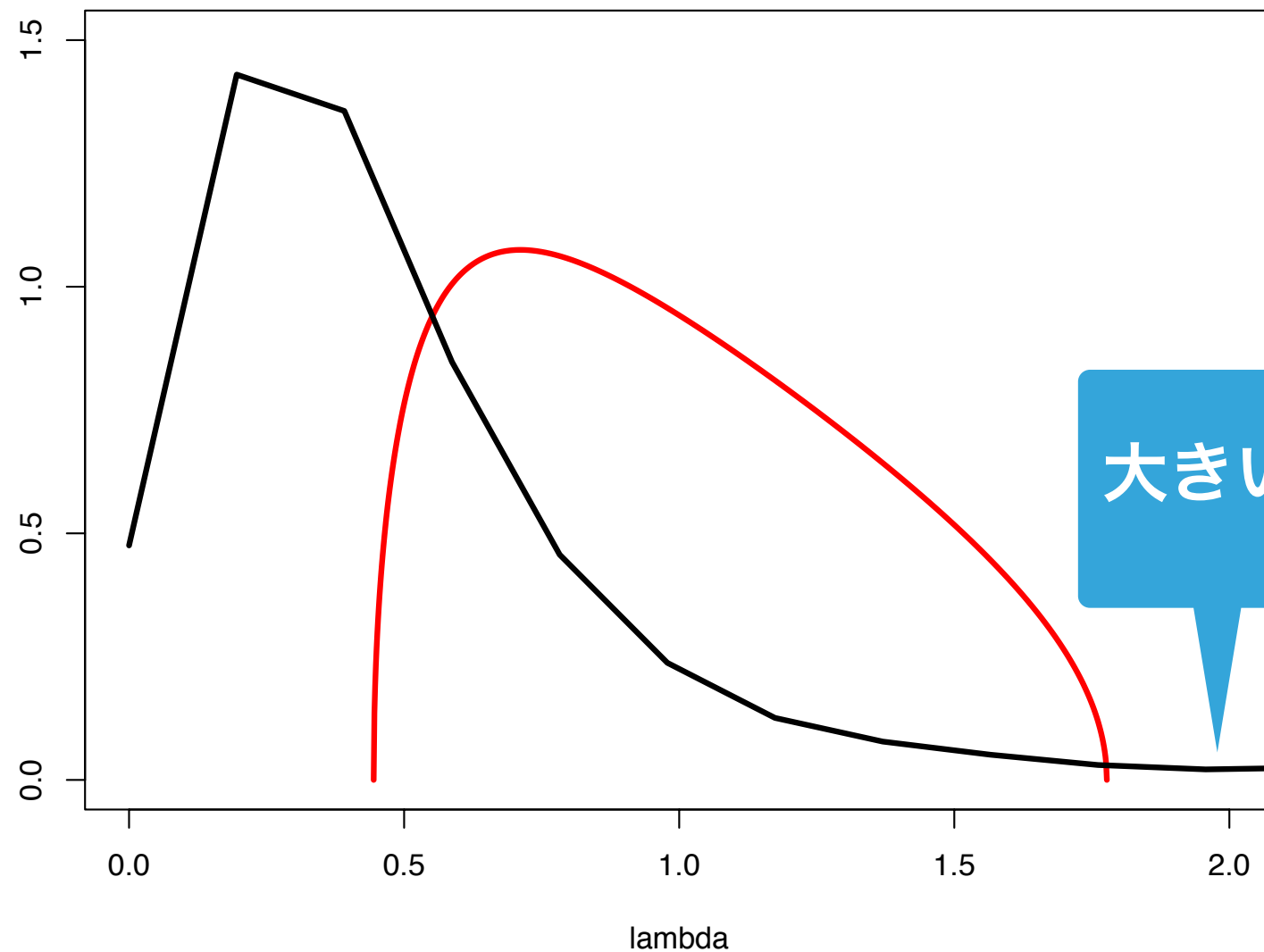


赤線：マルチェンコ=パスツール分布

黒線：実データの相関行列の固有値分布

本当に紹介したいこと

Distribution of Eigenvalue



赤線：マルチェンコ=パスツール分布

黒線：実データの相関行列の固有値分布

大きい固有値だけに注目

▶ フィルタリング後の相関行列

$$\mathbf{C}^{mg} = \mathbf{C}^{market} + \mathbf{C}^{group} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \sum_{i=2}^R \lambda_i \mathbf{u}_i \mathbf{u}_i^T$$

このデータだと、 $R=8$ だった

\mathbf{C}^{market} と \mathbf{C}^{group} のそれぞれで
ネットワーク分析したかった
igraphパッケージとか？

▶ ポートフォリオ最適化

フィルタリング後の相関行列を利用して、

Markowitzの平均分散モデルで最小分散ポートフォリオを組む

制約条件は、
投資比率の和が1であること
各投資比率は0以上1以下であること
に設定

▶ Rsolnpパッケージのsolnp関数

非線形計画問題を解いてくれる

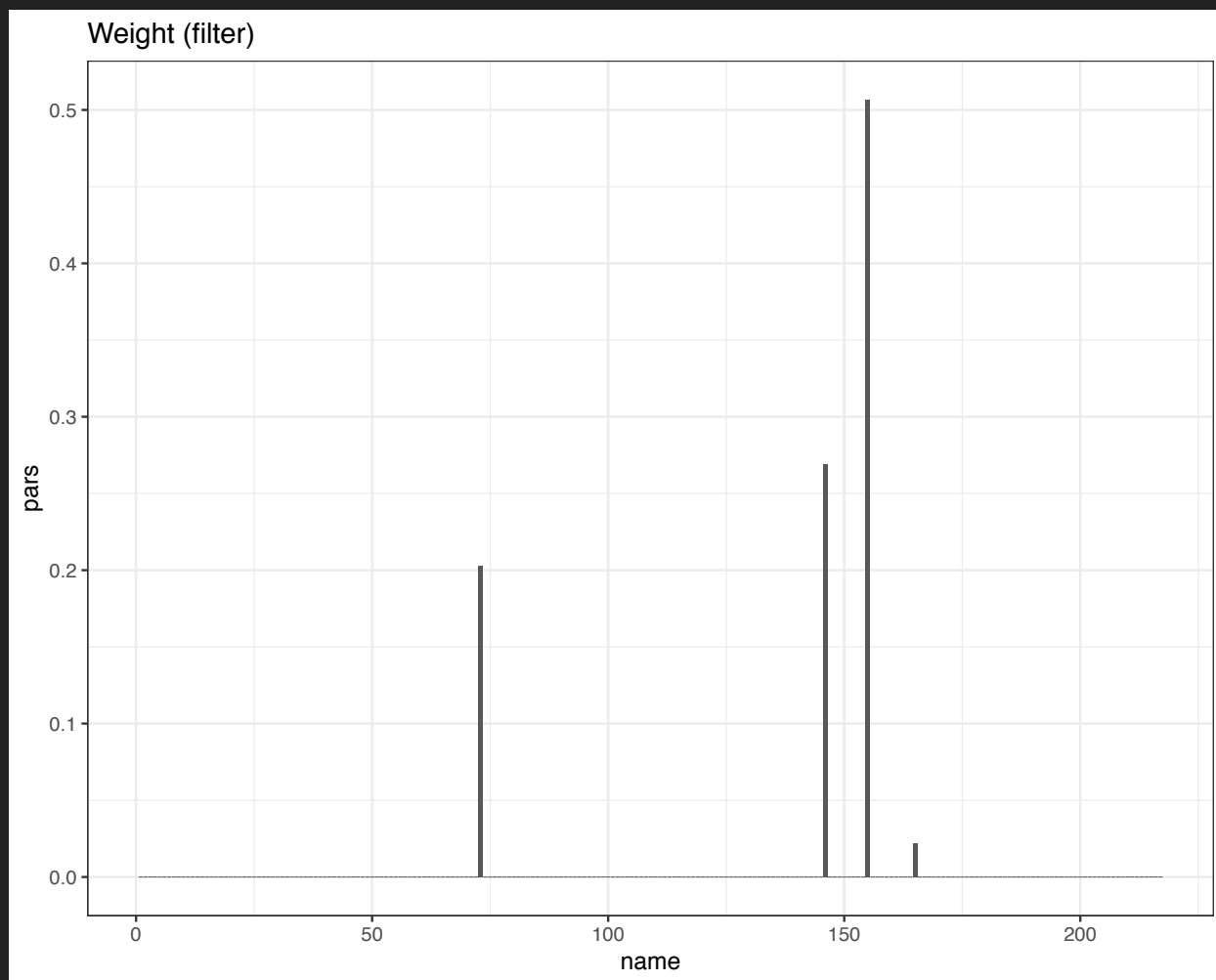
▶ Rsolnpパッケージのsolnp関数

最適形計画問題を解いてくれる

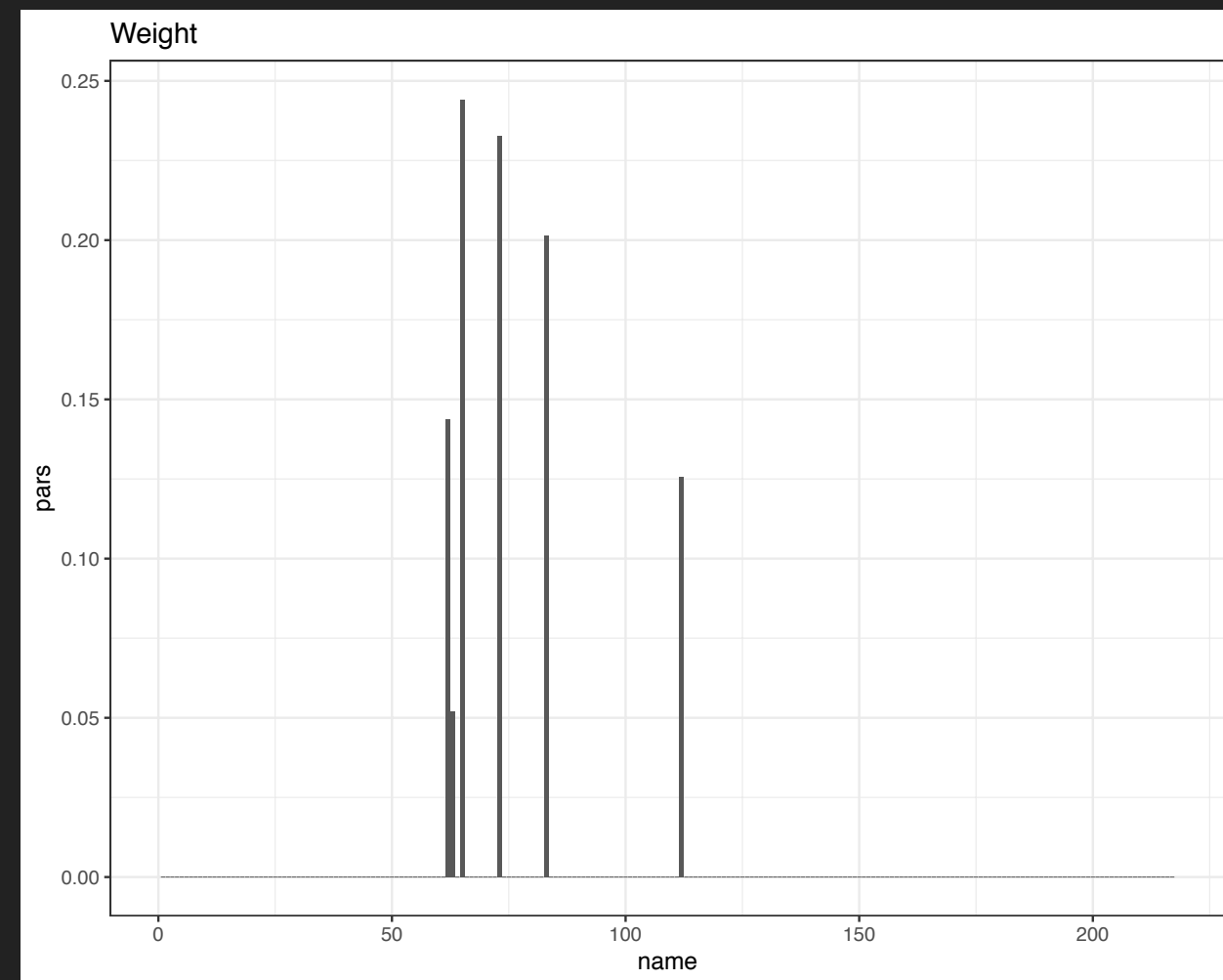
```
solution <- solnp(pars = weight,  
  fun = objectiveFunction,  
  eqfun = equalityConstraint,  
  eqB = eq.value,  
  ineqfun = inequalityConstraint,  
  ineqLB = ineq.lower,  
  ineqUB = ineq.upper)
```

目的関数
等式制約、不等式制約

本当に紹介したいこと



フィルタリングした方の投資比率



フィルタリングしてない方の投資比率

```
weightbar <- ggplot(result, aes(x=name, y=pars ))
```

```
weightbar <- weightbar + geom_bar(stat="identity") + theme_bw()
```

▶ 半年間の運用

smoking

	対数収益率	標準偏差
フィルタリングした方	0.2582	4.98%
フィルタリングしてない方	-0.0636	2.64%

なんだか良さそう(?)
標準偏差上がってるけどそれなりにリターン上がってるし

まとめ

- ▶ ランダム行列の漸近固有値分布をポートフォリオ最適化に応用してみた
- ▶ それなりに良い結果(?)が出た
- ▶ ノイズ除去というイメージでランダム行列理論は幅広い分野で使われている

参考文献をさらっと

渡辺 澄夫 『ランダム行列の数理と科学』 森北出版 2014

R.K. Pan and S. Sinha. Collective behavior of stock price movements in an emerging market. Phys. Rev.E76, 046116, 2007.

お粗末様でした

ご静聴ありがとうございました