ftanaka@stat.t.u-tokyo.ac.jp

時系列モデルの情報幾何とベイズ統計への応用(1)

２０１３年3月6日
初版

田中冬彦(*Tanaka Fuyuhiko*)

所属：東京大学 情報理工学系研究科　助教

1. 統計理論の位置づけ

研究対象は統計モデル！

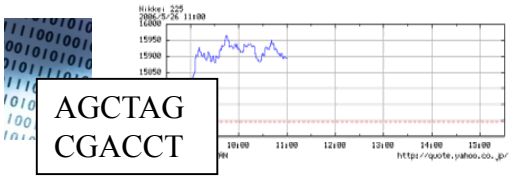
具体的

抽象的

実社会

遺伝子解析；情報理論；信号処理

脳計測；経済時系列　など



統計モデルの応用

個別の統計モデルを導入　$M = \{p(x \mid \theta); \theta \in \Theta \subset \mathrm{R}^k\}$

（脳波の時系列モデル、DNAの隠れマルコフモデルなど）

統計の基礎理論

統計モデルの全体 $\{M\}$ を扱う

ココ

発表の流れ

1. 統計理論の位置づけ
2. ベイズの公式
3. ベイズ統計入門
4. 無情報事前分布
5. 情報幾何の導入
6. 予測分布
7. 予測の決定理論的定式化
8. α平行事前分布
9. 優調和事前分布

2. ベイズの公式

ベイズの公式 (1/2)

ベイズの公式

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

$P(A)$	Aが成立している確率
$P(B \mid A)$	Aが成立している条件の下でのBが成立する確率（条件付き確率）
$P(B \cap A) = P(A \cap B)$	AとBが両方成立している確率

条件付き確率の定義

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B)$$

※ベイズの「公式」と呼ばれるが数学的には上の定義の書き換えにすぎない

ベイズの公式 (2/2)

ベイズの公式

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B)$$

例： 高校生のクラス30人のペット調査（確率＝割合）

A： 今、犬を飼っている

B： 昔、猫を飼っていた

$P(B \cap A) = P(A \cap B)$ 昔、猫を飼っていて、今、犬を飼っている割合

$P(A \mid B) = 4/9$ 昔、猫を飼っている人の中で、今、犬を飼っている人の割合

$P(A) = 1/6 \qquad P(B) = 3/10$

$$P(B \mid A) = \frac{4/9 \cdot 3/10}{1/6} = 4/5$$ 今、犬を飼っている人の中で、昔、猫を飼っていた人の割合

ベイズの公式 による推論

例： 高校生のクラス30人のペット調査をもとにした推理ゲーム
クラスの高校生1人と話してみる。

1. 手がかりがない場合

あなたは昔、猫を飼っていましたね？

当たる確率は3割　$P(B) = 3/10$

2. それまでの話から今、犬を飼っていることがわかった場合

あなたは昔、猫を飼っていましたね？

当たる確率は8割！　$P(B \mid A) = 4/5$

犬についての新たな手がかりと事前情報の両方を用いることで
猫に関する推論の精度があがる！

3. ベイズ統計入門

ベイズ統計の概観

1. 統計モデルの設定
2. 事前分布の設定
3. 事後分布に基いた推測

統計モデル (1/2)

統計モデル（確率モデル）

データを発生させている源に関して数学的に記述したもの

1. 試行的なモデル；現象論的なモデル（e.g., 駅からの距離と土地の値段の相関）
2. 物理法則など厳格なモデル（e.g., 物理定数の測定実験；真値＋観測誤差）

意識のギャップ

- 工学、経済　→　1を想定
- 物理（実験）などの科学　→　2を想定
- 統計研究者　→1，2を柔軟に考察

統計モデル (2/2)

記法:

データ x が確率分布（確率密度関数） $p(x)$ に従う時

$$x \sim p(x) \qquad \int_X p(x)\mathrm{d}x = 1,\ p(x) \geq 0$$

離散の時（以降はしばしば省略）

$$\sum_{x \in X} p(x) = 1,\ p(x) \geq 0$$

独立に同じ確率分布に従う時 (Independently identically distributed; i.i.d.)

$$y_1, \ldots, y_n \overset{\text{i.i.d.}}{\sim} p(y)$$

統計モデル

確率分布の有限次元パラメータ族を**統計モデル**と呼ぶ

$$p(x \mid \theta) \geq 0, \int_X p(x \mid \theta)\mathrm{d}x = 1 \qquad \forall\, \theta \in \mathrm{R}^k$$ ← θ の動く範囲は扱う問題による

未知パラメータ θ で決まる**真の確率分布**が存在し, データxはそこから発生している
とみなす（重要な仮定）

統計モデルの例(1/2)

・離散モデル　$X = \{a_1, a_2, \cdots, a_{k+1}\}$

$$p(x = a_i \mid \theta) = \theta_i \qquad \theta \in \mathrm{R}^k \quad \theta_i > 0$$

$$\theta_{k+1} = 1 - \theta_1 - \theta_2 \cdots - \theta_k$$

k=1: ベルヌーイ試行（１回のコイン投げ）

表の出る確率 $= \theta$　　$0 < \theta < 1$

・二項分布

n回投げて表がx回出る確率（独立性を仮定）

$$p_n(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad 0 < \theta < 1$$

統計モデルの例(2/2)

・平均未知の正規分布族(ガウスモデル)

$$p(x \mid \theta) = \frac{1}{(2\pi)^{k/2}} \exp\left\{ -\frac{\| x - \theta \|^2}{2} \right\} \quad \theta \in \mathrm{R}^k$$

事前分布

事前分布の導入

統計モデルに加えて，さらにパラメータ θ の確率分布も仮定(事前分布（prior distribution, prior ）と呼ぶ)

$$\theta \sim \pi(\theta) \qquad \pi(\theta) \geq 0, \int_{\Theta} \pi(\theta)\mathrm{d}\theta = 1$$

事前分布の例

k=1: ベルヌーイ試行（1回のコイン投げ）

表(H)の出る確率＝$p(x = H \mid \theta) = \theta$ $\qquad 0 < \theta < 1$

$\pi(\theta) = C\theta^{-1/2}(1-\theta)^{-1/2}$ Jeffreys事前分布(Jeffreys prior)

事後分布の構成

事後分布

未知パラメータも確率分布をもつとみなすことで条件付き分布がつくれる（ベイズの公式）

$$\pi(\theta \mid x) = \frac{p(x \mid \theta)\pi(\theta)}{p(x)} = \frac{p(x \mid \theta)\pi(\theta)}{\int p(x \mid \theta)\pi(\theta)\mathrm{d}\theta}$$

これをパラメータの**事後分布** (posterior distribution) と呼ぶ

ベイズの公式

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B)$$

ベイズ統計の考え方

統計モデルに基いた推論

（統計モデル所与の下）データから未知パラメータが推測できれば，
データを発生させる真の確率構造が把握できたことになる

ベイズ統計の考え方

未知パラメータの確率分布（事前分布）を導入し、

事後分布に基いて系統的に推測を行う

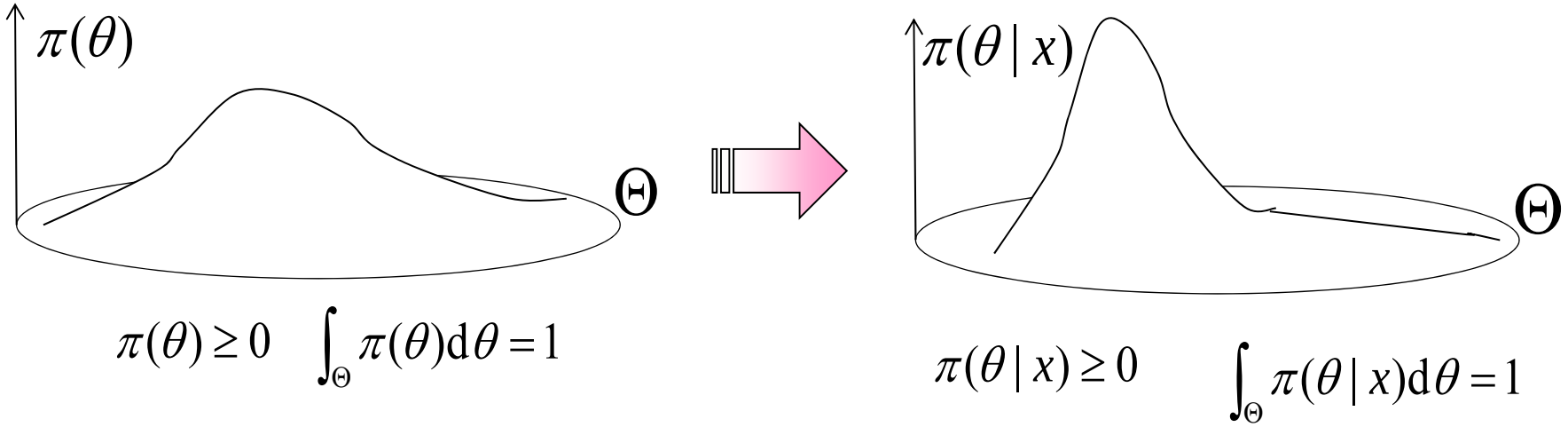
$\pi(\theta)$　事前情報に基づいたパラメータの分布

$\pi(\theta \mid x)$　事前情報＋実データに基づいたパラメータの分布

$\theta_B := \int \theta \pi(\theta \mid x) \mathrm{d}\theta$　パラメータのベイズ推定量（の例）

事後分布のイメージ

・事前分布＝パラメータに関する不確かさを表現
・事後分布＝データ x を得た後の不確かさを表現



$\pi(\theta) \geq 0 \quad \int_{\Theta} \pi(\theta) \mathrm{d}\theta = 1$

$\pi(\theta \mid x) \geq 0 \quad \int_{\Theta} \pi(\theta \mid x) \mathrm{d}\theta = 1$

(独立同一分布の下)データ数を増やしていくと
真値の周りに鋭いピークをもつ事後分布になる

ベイズ統計による推定の例 (1/2)

例：新商品の売上予測

問題

試験販売： $n = 100$ 人で30人購入

市場（10,000人）に出すとどれくらい売れるか？

統計モデル

潜在的な購入者の割合 θ $0 < \theta < 1$

n 人のうち実際にx人が購入する確率

$$p_n(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

ベイズ統計による推定の例 (1/2)

答え方(非ベイズ統計)

$$p_n(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad 0 < \theta < 1$$

最尤推定量 (MLE; Maximum Likelihood Estimator $\hat{\theta}(x) = \frac{x}{n}$)

$$\hat{\theta}(30) = \frac{30}{100} = 0.3$$

過去の似たような商品での情報は一切反映しない結果！

ベイズ統計による推定の例 (2/2)

答え方(ベイズ統計)

過去の似たような商品での購入者の割合の分布（事前分布）

$$\pi(\theta) = \frac{1}{2\sqrt{\theta}} \qquad 0 < \theta < 1$$

パラメータの事後分布

$$\pi(\theta \mid x) = \frac{1}{Be(30+1/2,71)} \theta^{30-1/2}(1-\theta)^{70}$$

事後分布に基いて，期待値やメディアン、MAP（事後確率が最大の値）などでパラメータを推定できる

ベイズ統計が有効なケース

実用上のニーズ

・厳格な法則よりも近似的なルールを与えたい（データマイニング；Amazonなどのおすすめ商品）

・過去のデータも踏まえて意思決定を行いたい（迷惑メールフィルタ；単語変換の予測；音声認識）

ベイズ統計発展の背景

・計算機の性能向上と事後分布計算のアルゴリズム開発（事後分布が計算可能）

・高次元の離散データの取り扱い(特にゲノムやWeb関係)

・人間行動のモデル化と予測に対するニーズ（ケータイ、カーナビ）

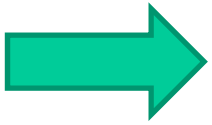
4. 無情報事前分布

無情報事前分布

無情報事前分布＝　パラメータに関する事前情報が全くない場合に使う事前分布

統計モデルのみから決めたい

$$M = \{ p(x \mid \theta) : \theta \in \Theta \subseteq R^k \}$$



$$\pi(\mathrm{d}\theta)?$$

注意点

1. いまだに議論があるが、望ましい条件などのコンセンサスもある(後述)

2.「無情報」は“noninformative”の訳語であり, objective prior, vague prior などとも言う

無情報事前分布の選択

歴史的には3段階

1. パラメータの動く範囲のみを見る（Laplace's criterion） $\Theta \subseteq \mathrm{R}^{k}$

2. 確率分布の空間への埋め込まれ方を見る

$$M = \{ p(x \mid \theta) : \theta \in \Theta \subseteq R^{k} \}$$

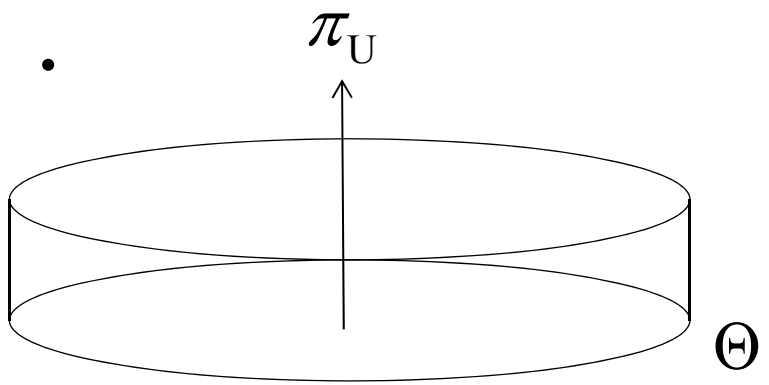
3. さらに、統計的決定理論の立場で眺める (統計理論の最先端)

Laplace's Criterion

素朴なアイディア

事前の情報が全くない ＝異なるパラメータに同じ重み
＝一様分布

$$\pi_{\mathrm{U}}(\theta) = const\ .$$



⇒ パラメータの取り方に依存した定義になっている

Ex. ゆがんだコイン

コイン投げのモデル (H=表, T＝裏)

$$M = \{(p(H \mid \theta), p(T \mid \theta)) = (\theta, 1-\theta) : \theta \in \Theta\} \qquad \Theta = (0,1)$$

$$= \{(p(H \mid \xi), p(T \mid \xi)) = (\tan \xi, 1 - \tan \xi) : \xi \in \Xi\}$$

$$\Xi = \left(0, \frac{\pi}{4}\right)$$

ある人の主張する「一様分布」 $\pi_U(\theta) = 1$ on Θ

別の人の主張する「一様分布」 $\pi_U(\xi) = 4 \cdot \pi^{-1}$ on Ξ

両者は同じ統計モデルとみなすため，「一様分布」という定義は望ましくない

無情報事前分布の指針

望ましい条件

- パラメータの取り方に依存しない定義（一様分布はダメ）
- 汎用性（対称性の低い複雑なモデルでも使える）
- 得られる結果が望ましい性質をもつ、なんらかの解釈が可能

Jeffreys 事前分布

Jeffreys 事前分布 $\pi_J(\theta)$

$$\pi_J(\theta) \propto \sqrt{\det(\, g_{ij}\,)}$$

g_{ij}　Fisher 情報行列

$$g_{ij} := \mathrm{E}^X\left[\frac{\partial \log\ p(x\mid\theta)}{\partial\theta^i}\frac{\partial \log\ p(x\mid\theta)}{\partial\theta^j}\right]$$

$$\mathrm{E}^X[\cdot] := \int \cdot\ p(x\mid\theta)\mathrm{d}x$$

- パラメータの取り方に依存しない

-汎用性（適当な正則条件のもとで常に存在）

→　他の事前分布との比較対象

- ただし, improper (積分発散)になるケースも多く

必ずしも良い選択肢とはいえない

例:ガウス分布族

ガウス分布族 (正規分布族)＝平均と分散で特徴づけられる確率分布族

$$M = \{p(x|\theta) : \theta \in \Theta \subset \mathrm{R}^2\} \quad p(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$\theta = (\theta_1, \theta_2) = (\mu, \sigma^2) \quad \Theta = \mathrm{R} \times (0, +\infty)$$

簡単のため分散＝1とおくと

$$\pi_{\mathrm{J}}(\mu) \propto 1 \qquad \mu \in \mathrm{R}$$

積分は発散 （improper とよばれる）

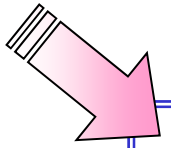
$$\int \mathrm{d}\mu \pi_{\mathrm{J}}(\mu) = \infty$$

無情報事前分布の提案

さまざまな無情報事前分布

- Jeffreys prior
- Maximum likelihood prior/Alpha parallel prior
- Superharmonic prior
- Reference prior/Latent information prior
- Haar measure (on a unimodular group)

今もなお多くの議論有り (cf. 量子系特有の例はFT 2012, Physical Review A)

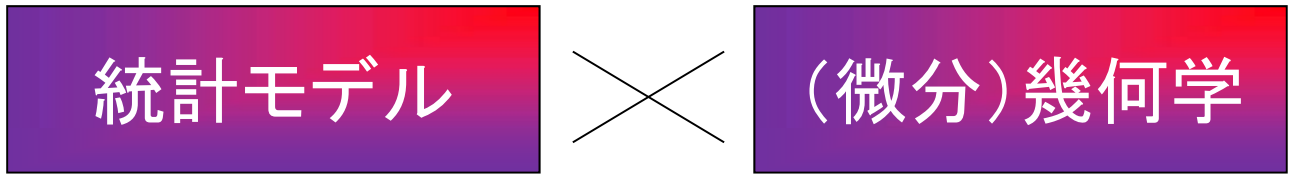


統計モデルに対し幾何学的に定義される事前分布（赤字）を紹介

5. 情報幾何の導入

よくある誤解

「統計モデル＋幾何学」だけだと（情報幾何の）他にも解釈がある



1．データが曲線・曲面上に分布　（cf. 方向統計学，スムージング）

2．確率密度関数がなめらかな曲線・曲面　（cf. 最大エントロピー法，ピアソン分布族）

以下、情報幾何のフォーマルな定義を紹介（1，2とは違う！）

統計モデルの幾何（＝情報幾何）

OVERVIEW

1. 確率分布のパラメータ族

$$p(x\,|\,\theta) \qquad \int p(x\,|\,\theta)\mathrm{d}x = 1,\ p(x\,|\,\theta) \geq 0 \qquad \theta = (\theta^1, \theta^2, \cdots, \theta^k)$$

2. 統計モデル多様体

$$\mathrm{M} = \{p(x\,|\,\theta) : \theta \in \Theta\}$$

3. Fisher計量

$$g_{ij} := \mathrm{E}^X\left[\frac{\partial \log\ p(x\,|\,\theta)}{\partial \theta^i}\frac{\partial \log\ p(x\,|\,\theta)}{\partial \theta^j}\right]$$

$$\mathrm{E}^X[\cdot] := \int \cdot\ p(x\,|\,\theta)\mathrm{d}x$$

4. Alpha 接続

$$\overset{(\alpha)}{\Gamma}_{ij;k} := \overset{(e)}{\Gamma}_{ij;k} + \frac{1-\alpha}{2}T_{ijk} \qquad \alpha \in \mathrm{R}$$

$$\overset{(e)}{\Gamma}_{ij;k} := \mathrm{E}^X\left[\frac{\partial^2 \log\ p(x\,|\,\theta)}{\partial \theta^i \partial \theta^j}\frac{\partial \log\ p(x\,|\,\theta)}{\partial \theta^k}\right]$$

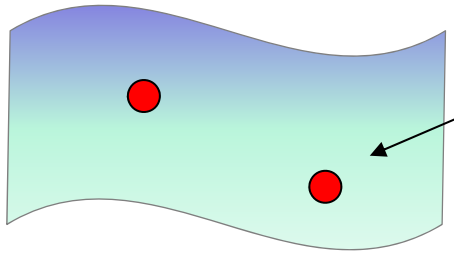
$$T_{ijk} := \mathrm{E}^X\left[\frac{\partial \log\ p(x\,|\,\theta)}{\partial \theta^i}\frac{\partial \log\ p(x\,|\,\theta)}{\partial \theta^j}\frac{\partial \log\ p(x\,|\,\theta)}{\partial \theta^k}\right]$$

統計モデル多様体

統計モデル多様体（確率分布のパラメータ族）

統計モデル

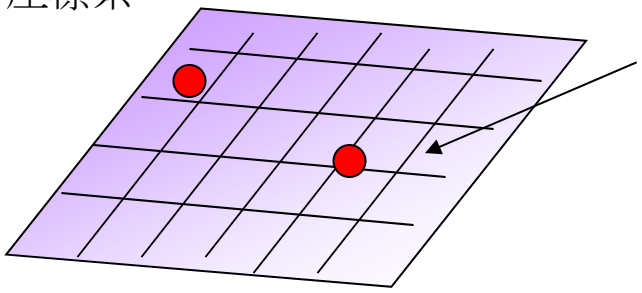
$$\mathrm{M} = \{p(x \mid \theta) : \theta \in \Theta\}$$



$$p(x \mid \theta_0) \in M$$

＊モデルのパラメータ＝多様体の座標系

$$\theta = (\theta^1, \theta^2, \cdots, \theta^d)$$



$$\theta = \theta_0 \in \Theta \subseteq R^k$$

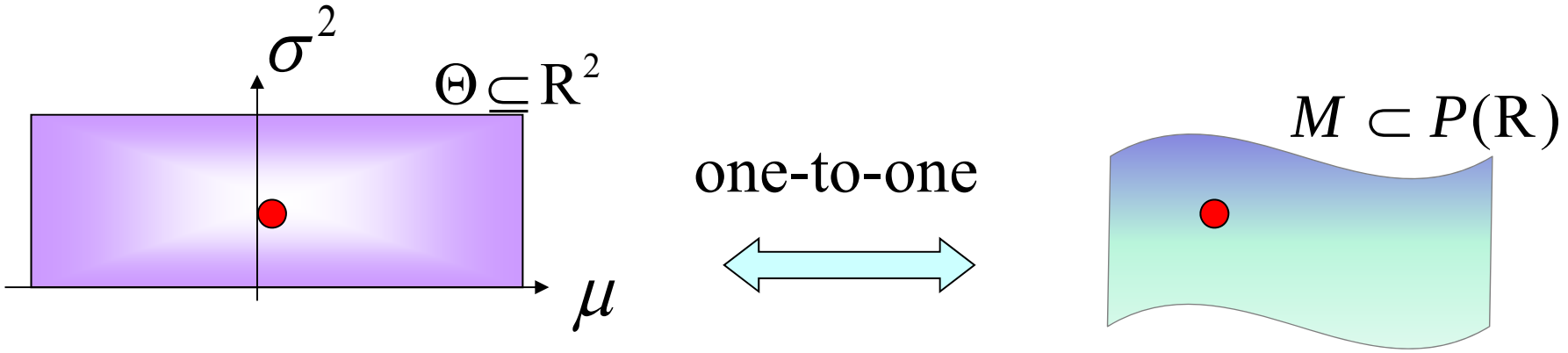
例:ガウス分布族 (1/2)

ガウス分布族 (正規分布族)＝平均と分散で特徴づけられる確率分布族

$$M = \{ p(x \mid \theta) : \theta \in \Theta \subset \mathrm{R}^2 \} \quad p(x \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

パラメータの動く範囲

$$\theta = (\theta_1, \theta_2) = (\underset{\text{平均}}{\mu}, \underset{\text{分散}}{\sigma^2}) \qquad \Theta = \mathrm{R} \times (0, +\infty)$$



例:ガウス分布族 (2/2)

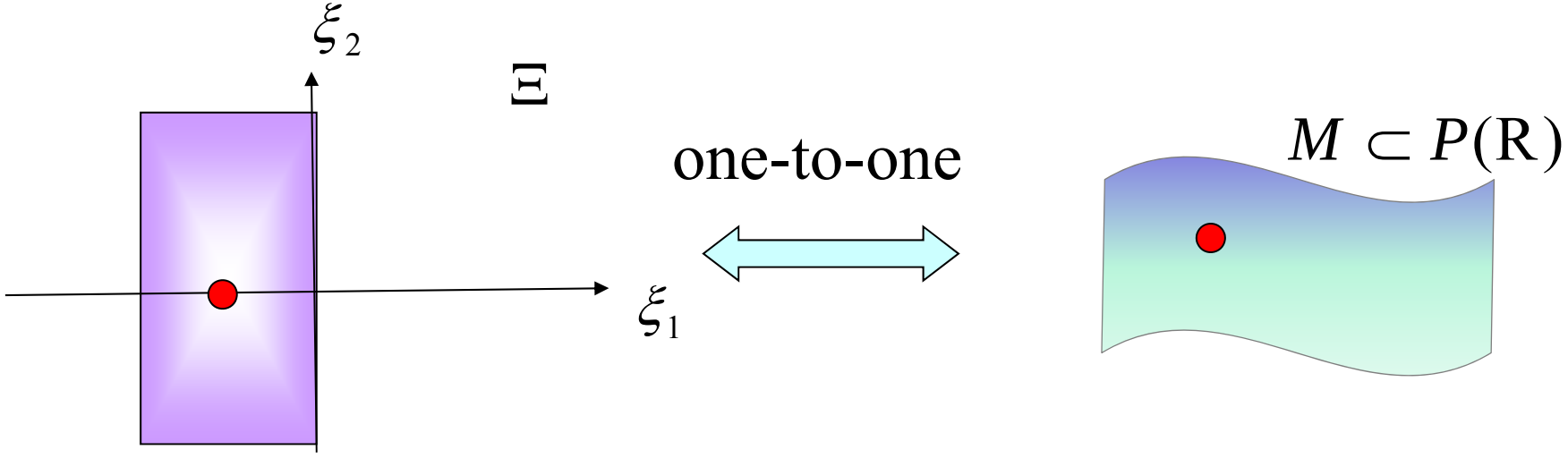
別のパラメータ(自然パラメータ)による表示

$$M = \{p(x \mid \xi) : \xi \in \Xi \subset \mathrm{R}^2\}$$

$$p(x \mid \xi) = \exp\left\{\xi_1 x^2 + \xi_2 x - \psi(\xi_1, \xi_2)\right\}$$

パラメータの動く範囲

$$\xi = (\xi_1, \xi_2) = (-1/(2\sigma^2), \mu/\sigma^2) \quad \Xi = (-\infty, 0) \times R$$

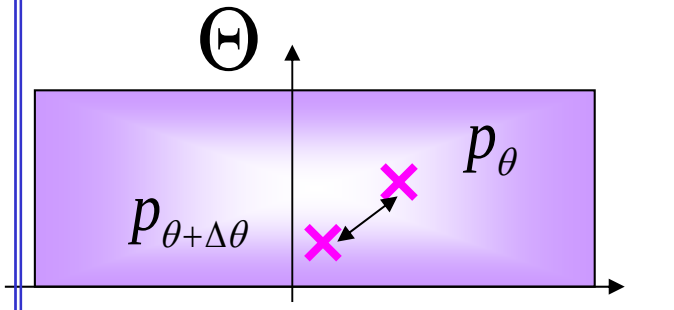


統計モデル上のリーマン計量

Fisher情報行列（Fisher計量）

$$g_{ij}(\theta) = \mathrm{E}_\theta\left[\frac{\partial}{\partial \theta^i}\log\ p_\theta\ \frac{\partial}{\partial \theta^j}\log\ p_\theta\right]$$

$$\mathrm{E}_\theta[\,\cdot\,] := \int \cdot\ p(x\,|\,\theta)\mathrm{d}x$$



$$M = \{p(x\,|\,\theta) : \theta \in \Theta \subset \mathrm{R}^k\}$$

性質

-実対称テンソル

-正定値（通常の統計モデル）

-様々な特徴づけ

統計モデル上の高次のテンソルと接続

対数尤度の高次微分

$$\partial_i l := \frac{\partial \log\ p(x|\theta)}{\partial \theta^i} \qquad D_{ij\cdots} l := \frac{\partial}{\partial \theta^i}\frac{\partial}{\partial \theta^j}\cdots \{\log\ p(x|\theta)\}$$

高次テンソル

$$L_{ijk\cdots}(\theta) := \mathrm{E}_\theta[D_i l\, D_j l\, D_k l \cdots] \qquad \mathrm{E}_\theta[\,\cdot\,] := \int \cdot\, p(x|\theta)\mathrm{d}x$$

Alpha 接続 (アファイン接続の１パラメータ族)

$$\overset{(\alpha)}{\Gamma}_{ij;k} := \overset{(e)}{\Gamma}_{ij;k} + \frac{1-\alpha}{2} T_{ijk} \qquad \alpha \in \mathrm{R}$$

$$\overset{(e)}{\Gamma}_{ij;k}(\theta) := \mathrm{E}_\theta\left[D_{ij} l\, D_k l\right] \qquad T_{ijk}(\theta) := \mathrm{E}_\theta[D_i l\, D_j l\, D_k l]$$

一意性

Chentsov's Theorem

データのあらゆる変換（stochastic map of random variables）に対する単調性から、リーマン計量は定数倍を除いて一意的に定まる（Fisher情報行列！）。

Remark:

1. See, e.g., Campbell (1986)
2. 量子情報幾何では一意でなくなる

統計モデル上の体積形式

統計モデル多様体上の体積形式

＊事前分布は統計モデル多様体上の体積要素とみなせる

$$\omega = \pi(\theta)\mathrm{d}\theta^1 \wedge \cdots \wedge \mathrm{d}\theta^d$$

＊Jeffreys 事前分布はモデル多様体上の不変体積要素

$$\omega = \sqrt{\det\ g_{ij}(\theta)}\mathrm{d}\theta^1 \wedge \cdots \wedge \mathrm{d}\theta^d \propto \pi_{\mathrm{J}}(\theta)\mathrm{d}\theta^1 \wedge \cdots \wedge \mathrm{d}\theta^d$$

事前分布の選択

パラメータの取り方に依存しない定義を考える

⇒ 微分幾何学 に基づいたアプローチが有効

例:ガウスモデル多様体

ガウスモデル

$$p(x\,|\,\theta)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad \theta=(\theta_1,\theta_2)=(\mu,\sigma^2)$$

$$\Theta = R\times(0,+\infty)$$

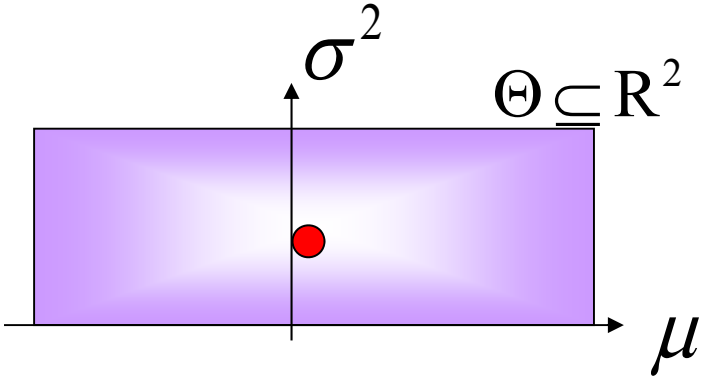
Fisher計量

$$g_{ij}(\theta)=\begin{pmatrix}\frac{1}{\theta_2} & 0\\ 0 & \frac{1}{2(\theta_2)^2}\end{pmatrix}$$

alpha 接続 (omitted)

体積要素

$$\omega=(\theta_2)^{-3/2}\,\mathrm{d}\theta_1\wedge\mathrm{d}\theta_2$$



無情報事前分布の選択（再掲）

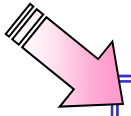
歴史的には3段階

1. パラメータの動く範囲のみを見る（Laplace's criterion） $\Theta \subseteq \mathrm{R}^k$

2. 確率分布の空間への埋め込まれ方を見る（情報幾何学的な記述）

$$M = \{ p(x \mid \theta) : \theta \in \Theta \subseteq R^k \}$$

3. さらに、統計的決定理論の立場で眺める



3番目のステップとして予測的な観点を導入

6. 予測分布

予測分布の基本的な考え(1/2)

例：新規開店コンビニの売上予測

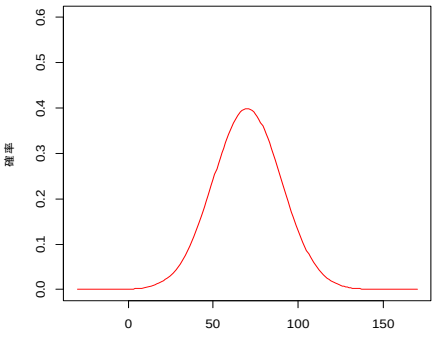
FuyuMart

4日間の売上データ(万円)

$$x = (x_1, x_2, x_3, x_4)$$
$$= (120, 70, 80, 130)$$

問：明日の売上　$y := x_5$ はどのくらいか？

他の店舗から経験的に平均　θ　分散20万円のガウス分布(毎日独立；平均は店で違う)



$$p(x \mid \theta) = \frac{1}{\sqrt{2\pi \cdot (20)^2}} \exp\left\{-\frac{(x-\theta)^2}{2 \cdot (20)^2}\right\}$$

明日の売上をどう予測するか？

予測分布の基本的な考え(2/2)

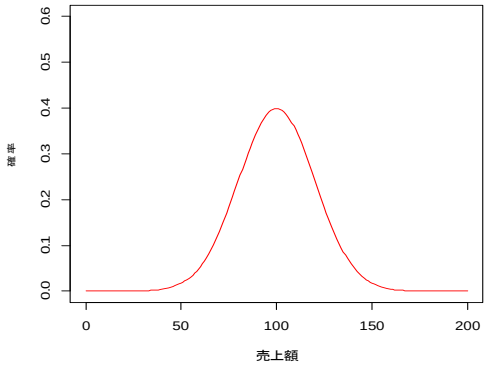
1．統計を知らない経営者

明日の売上の期待値（データ平均）は100万円
だいたい100万円前後という回答

でも、「前後」ってどのくらい？

2．パラメータ推定は知っている経営者

明日の売上も右のような分布で考えよう

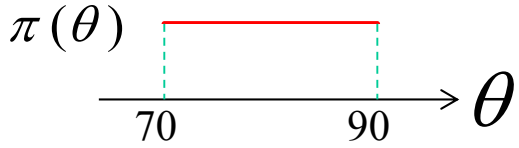


点で与えるより分布で与えるほうがよい

ベイズ予測分布の方法(1/2)

3．ベイズ予測分布を知っている経営者

同じ地区の別の店舗での平均は70万円～90万円くらい



パラメータに関する事前情報（事前分布）

パラメータに関する事前の情報と実際の４日間の売上データの両方をもとに2ステップで、明日の売上予測を行う

ステップ1．事後分布（データ所与の条件付き分布）

$$\pi(\theta \mid x_1, \cdots, x_n) = \frac{p(x_1, \cdots, x_n \mid \theta)\pi(\theta)}{\int p(x_1, \cdots, x_n \mid \theta)\pi(\theta)\mathrm{d}\theta}$$

$$\int_\Theta \pi(\theta \mid x_1, \cdots, x_n)\mathrm{d}\theta = 1 \qquad \pi(\theta \mid x_1, \cdots, x_n) \geq 0$$

ベイズ予測分布の方法(2/2)

3. ベイズ予測分布を知っている経営者

ステップ2. ベイズ予測分布(確率分布の平均化)

$$p_\pi(y) := \int_\Theta p(y \mid \theta)\pi(\theta \mid x_1, \cdots, x_n)\mathrm{d}\theta$$

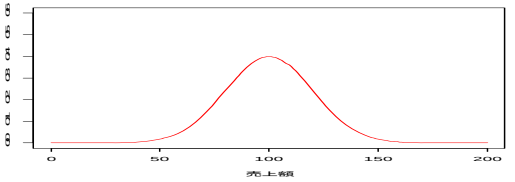
$$\pi(\theta \mid x_1, \cdots, x_n) = \frac{p(x_1, \cdots, x_n \mid \theta)\pi(\theta)}{\int p(x_1, \cdots, x_n \mid \theta)\pi(\theta)\mathrm{d}\theta}$$

ポイント(a)： 事前の情報と実データの両方をとりいれた予測

$$\pi(\theta) \qquad x = (120, 70, 80, 130)$$

ポイント(b)： もとのガウス分布より幅が広い

ポイント(c)： パラメータのベイズ推定とは違う



様々な分野でベイズ予測分布が利用されている

研究対象は統計モデル！（再掲）

具体的

抽象的

統計モデルの応用（マーケティング）

※コンビニ/マーケティングに特化した議論

コンビニのマーケティングを記述するモデル　$M = \{p(x \mid \theta); \theta \in \Theta \subset \mathrm{R}^k\}$

- 営業戦略を考える上での望ましいモデルとは何か？
- コンビニのPOSデータに基いた統計的解析
- モデルの妥当性の議論

統計の基礎理論

※抽象的な数学モデルを扱う

統計モデル全体（抽象的）　$\{M\}$

ココ

統計理論のスタンス

理論的な課題の例

ベイズ予測分布は何らかの意味で望ましい性質をもつのか？

$$p_\pi(y \mid x) := \int_\Theta p(y \mid \theta)\pi(\theta \mid x)\mathrm{d}\theta$$

$$\pi(\theta \mid x) := \frac{p(x \mid \theta)\pi(\theta)}{\int_\Theta p(x \mid \theta)\pi(\theta)\mathrm{d}\theta}$$

望ましい性質がある（統計的決定理論の枠組みで）

7. 予測の決定理論的定式化

統計モデルを使う場面

2種類の統計モデル

1. 試行的なモデル；現象論的なモデル

・将来のデータの確率分布（予測分布）の推定を重視

・現象のランダムネスを考慮しつつ、Decision Makingを行いたい

2. 物理法則など厳格なモデル

・興味あるパラメータの推定（点推定）を重視

・法則性を明らかにしようとする立場

実用上の区別はあいまいだが，定式化（理論）は明確に違う

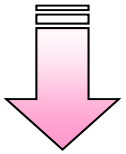
予測問題の定式化(1/3)

パラメータの推定

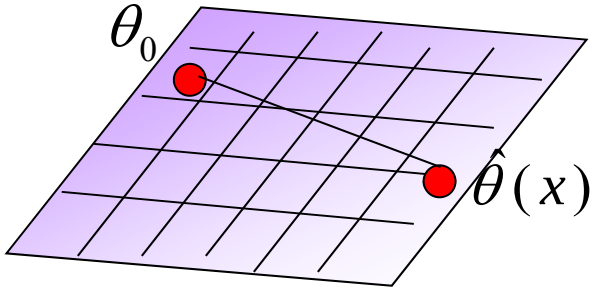
$$x=(x_1,x_2,\cdots,x_n)$$

$$\longrightarrow \quad \hat{\theta}(x)$$ ※統計では推定量にハット

（演算子ではない！）



パラメータの空間



$$d(\theta_0,\hat{\theta}(x)) = \left\|\theta_0 - \hat{\theta}(x)\right\|^2$$

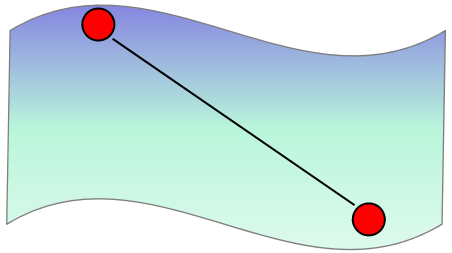
確率分布の推定（予測分布の構成）

$$x=(x_1,x_2,\cdots,x_n)$$

$$\longrightarrow \quad \hat{p}_x := \hat{p}(y\,|\,x)$$

確率分布の空間

$$p_0 = p(y\,|\,\theta_0)$$



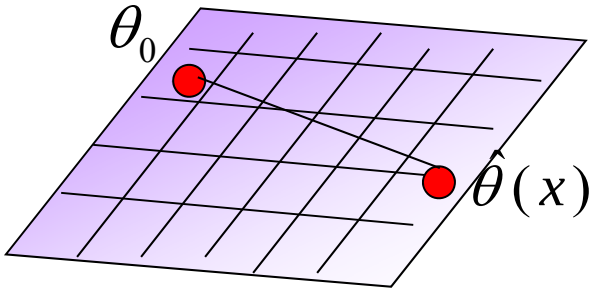
$$d(p_0,\hat{p}_x) \qquad \hat{p}_x = \hat{p}(y\,|\,x)$$

予測問題の定式化(2/3)

パラメータの最適推定

$d(\theta_0, \hat{\theta}(x))$ の期待値（平均二乗誤差）を最小にする

パラメータの空間



$$d(\theta_0, \hat{\theta}(x)) = \left\| \theta_0 - \hat{\theta}(x) \right\|^2$$

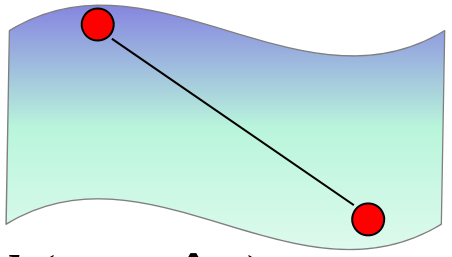
分布の最適推定（最適な予測分布の構成）

$d(p_0, \hat{p}_x)$ の期待値を最小にする

※「距離」に相当するこのような関数を損失関数とよぶ

確率分布の空間

$p_0 = p(y|\theta_0)$



$d(p_0, \hat{p}_x)$　　$\hat{p}_x = \hat{p}(y|x)$

$d(p_0, \hat{p}_x)$をどう与えるか？

予測問題の定式化(3/3)

確率分布の推定では以下の損失関数がよく使われる

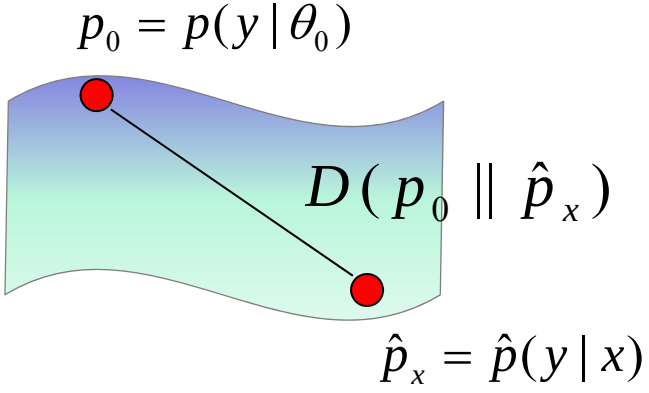
Kullback-Leibler divergence

$$D(p \,\|\, q) := \int \mathrm{d}x \{ p(x) \log(p(x) / q(x)) \}$$

性質

- 正値
- $D(p \,\|\, q) = 0 \leftrightarrow p = q$
- 対称ではない $D(p \,\|\, q) \neq D(q \,\|\, p)$
- 加法性

確率分布の空間



注意:ダイバージェンスの導入について

1.　ダイバージェンス = 異なる確率分布の(擬)距離

2.　導入する目的:

統計的決定理論の枠組みで推定方法の定量的な評価

&「よい推定方法」の理論的な検討

3.　「正しい」ダイバージェンスはないが、望ましい条件をもつダイバージェンスのクラスは知られている

※量子情報/物理の人には, ここの感覚があまり理解されていない;

i.e., 「正しい」距離があるというスタンス

cf) 無情報事前分布も「正しい」ものが存在するというスタンス

リスク関数の導入

リスク関数 ～推定誤差の平均

$$R_q(\theta) := \mathrm{E}^X[D(p_\theta \| q)]$$

$q = q(y \mid x)$ データに基いた予測分布

・異なる予測分布同士を比較するのに使う

・パラメータ（未知）に依存

・一様に小さくできる予測分布が望ましい（1点で0にできるため一般には無理）

ベイズ予測分布の特徴づけ(1/2)

問題設定

・n個の データの分布(i.i.d.)　$x_i \sim p(x_i \mid \theta)$

※i.i.d. = independently identical distributed , 「同じ分布から独立に」の意

・事前分布を仮定　$\theta \sim \pi(\theta)$

このとき、n個のデータ$x = (x_1, \ldots, x_n)$ から

次の　$n+1$個目の分布　$y \sim \hat{p}(y \mid x_1, \cdots, x_n)$　を推定する

Problem

平均リスク $\int R_{\hat{p}_x}(\theta)\pi(\mathrm{d}\theta)$ を最小化する予測分布 $\hat{p}(y \mid x)$ は？

ベイズ予測分布の特徴づけ(2/2)

KL損失のもとで最適なもの（Aitchison, 1975）

$$\hat{p}_{\pi}(y \mid x) := \int_{\Theta} p(y \mid \theta)\pi(\theta \mid x)\mathrm{d}\theta$$

$$\pi(\theta \mid x) := \frac{p(x \mid \theta)\pi(\theta)}{\int_{\Theta} p(x \mid \theta)\pi(\theta)\mathrm{d}\theta}$$

事前分布が与えられると最適な予測分布が定まる

無情報事前分布の選択（再掲）

歴史的には3段階

1. パラメータの動く範囲のみを見る（Laplace's criterion） $\Theta \subseteq \mathrm{R}^k$

2. 確率分布の空間への埋め込まれ方を見る（情報幾何学的な記述）

$$M = \{p(x \mid \theta) : \theta \in \Theta \subseteq R^k\}$$

3. さらに、統計的決定理論の立場で眺める （予測分布を念頭に）

※量子統計では特に3が見落とされている

予測分布の構成を主眼にした事前分布を導入

8. α平行事前分布

最尤推定型事前分布(MLE prior)

MLE prior (Hartigan1998)

事前分布 $\pi(\theta)$ が以下の条件を満たす時

最尤推定型事前分布 (*MLE prior, e-prior*) という: (π_M と書く)

$$\frac{\partial}{\partial \theta^j} \log \pi(\theta) = q_j(\theta) \qquad q_j := g^{ik} \Gamma^{(e)}_{ij;k} \qquad \Gamma^{(e)}_{ij;k} := \mathrm{E}^X\left[\partial_i \partial_j l \partial_k l\right]$$

g^{lk} Fisher計量の逆行列

Hartigan (1998)

最尤推定量を代入した予測分布はプラグイン予測分布の中で(漸近リスクで)許容的

⟺ π_M が存在

1. (可積分条件なので)1パラメータなら必ず存在
2. 漸近リスクと有限標本でのリスクの間にはギャップあり(Hartigan 自身が指摘)
3. 今回の講演は「統計的意味」は深入りしない

α 平行事前分布 (1/2)

MLE prior → 微分幾何学的な意味は？

α平行事前分布 (Takeuchi and Amari (2005))

d次元；向き付け可能な統計モデル多様体を考える

ω 正のd形式

$$\overset{(\alpha)}{\Gamma}_{ij;k} := \overset{(e)}{\Gamma}_{ij;k} + \frac{1-\alpha}{2} T_{ijk}$$

α 接続（多様体上の平行移動；共変微分を定義）

ω がα平行移動で不変（Statistical Equiaffine）

$$\overset{\text{def}}{\Longleftrightarrow} \quad \overset{(\alpha)}{\nabla}\omega = 0$$

α平行移動で不変なωが存在する時，
ω の成分を規格化したものを α 平行事前分布と呼ぶ

α 平行事前分布 (2/2)

事前分布の幾何学的意味

- MLE prior は e接続の下で平行な体積要素
- Jeffreys priorはLevi Civita 接続の下で平行な体積要素(常に存在)

存在条件(Takeuchi and Amari (2005))

$\alpha \neq 0$ の時

統計モデル上に

α 平行事前分布が存在

$\Leftrightarrow$

3階テンソルTが以下をみたす(αに依存しない)

$$\partial_i T_j - \partial_j T_i = 0$$

$$T_{ijk} := \mathrm{E}^X[\partial_i l \partial_j l \partial_k l] \quad \text{Cubic tensor}$$

$$T_i := T_{ijk} g^{jk} \quad \text{Tchebychev 形式}$$

参考: 曲指数型分布族への拡張 Matsuzoe *et al* (2006)

微分形式での定式化

α平行事前分布の存在条件は微分形式を用いた定式化も可能

WHY differential form?

- α平行事前分布が存在しない場合の統計モデルの分類を考える指標
- 量子情報幾何（一般にTorsionが残る）での定式化を意識（未着手）

接続形式・曲率形式

一般論(接続形式・曲率形式)

基底ベクトル場 $\left\{\frac{\partial}{\partial\theta^i}\right\}_{i=1,\dots,d}$ と接続係数 Γ_{ij}^{k} が与えられた時以下のように d^2 個の1形式を定める(**接続形式**(connection form)と呼ぶ)

$$\omega^{k}{}_{j} := \Gamma_{ij}^{k}\,\mathrm{d}\theta^{i}$$

さらに以下のように d^2 個の2形式を定める(**曲率形式**(curvature form)と呼ぶ)

$$\Omega^{k}{}_{j} := \mathrm{d}\omega^{k}{}_{j} + \omega^{k}{}_{l} \wedge \omega^{l}{}_{j}$$

※今回は座標基底を用いて簡単化した定義

(本来は, より一般の基底ベクトル場をとる)

不変微分形式

基底ベクトルの変換で曲率2形式は以下のように変換（形式的には相似変換）

$$\Omega^{a}{}_{b} = A^{a}{}_{c}\Omega^{c}{}_{d}(A^{-1})^{d}{}_{b}$$

2形式は可換なため不変多項式を用いて 不変 $2q$ 形式がつくれる

不変 $2q$ 形式

f ：次数q の不変多項式

$f(\Omega)$　f による不変 $2q$ 形式

f が不変多項式 $\Longleftrightarrow$ $f(X) = f(AXA^{-1})$

A　k 次の正則行列

X　k次の任意の行列

Ex: Tr X, det X　(固有値の対称多項式から生成される)

トレース2形式の導入 (1/2)

トレース2形式

統計モデル多様体について以下の不変2形式を定義

$$\Xi^{(\alpha)} = \mathrm{Tr}\Omega^{(\alpha)} := \Omega^{(\alpha)a}{}_{a}$$

α接続に関するトレース2形式と呼ぶ

※基底ベクトル場の取り方に依存しない定義

Lemma

α接続に関するトレース2形式はTchebychev形式（1形式）を用いて以下のようにかける（α=0はJeffreys事前分布に対応）

$$\Xi^{(\alpha)} = -\frac{\alpha}{2}\mathrm{d}T = -\frac{\alpha}{2}\sum_{i<j}(\partial_i T_j - \partial_j T_i)\mathrm{d}\theta^i \wedge \mathrm{d}\theta^j$$

トレース2形式の導入 (2/2)

必要十分条件

α 平行事前分布が存在 ⟺ α接続に関するトレース2形式が恒等的に0

$$\Xi^{(\alpha)} = 0$$

最尤推定が（漸近KLリスクで）許容的

⟺ MLE priorが存在

⟺ e 接続に関するトレース2形式が恒等的に0

$$\Xi^{(e)} = 0$$

平坦性との関係

モデル多様体は α 平坦 ⟺ α接続に関する曲率2形式が恒等的に0

$$\Omega^{(\alpha)} = 0$$

⟹ ⟸

$$\Xi^{(\alpha)} = \mathrm{Tr}\,\Omega^{(\alpha)} = 0$$

統計モデルの情報幾何学的な量による分類

α平行事前分布をもつ $\overset{(\alpha)}{\nabla}\omega = 0 \quad \forall \alpha \in \mathrm{R}$

$\Xi^{(\alpha)} = 0, \Omega^{(\alpha)} \neq 0$

・Lauritzenの例（人工的）

Conjugate symmetric $R^{(\alpha)}_{ijkl} = -R^{(\alpha)}_{ijlk} \quad \forall \alpha \in \mathrm{R}$

$\alpha(\neq 0)$ 平坦 $\exists \alpha \neq 0 \quad R^{(\alpha)}_{ijkl} = 0$

・指数型分布族

・混合分布族

＊ARモデル、MAモデルも含まれる

$\Xi^{(\alpha)} = 0, \Omega^{(\alpha)} = 0$

・ARMA(p,q)モデル $\Xi^{(\alpha)} \neq 0, \Omega^{(\alpha)} \neq 0$

9. 優調和事前分布

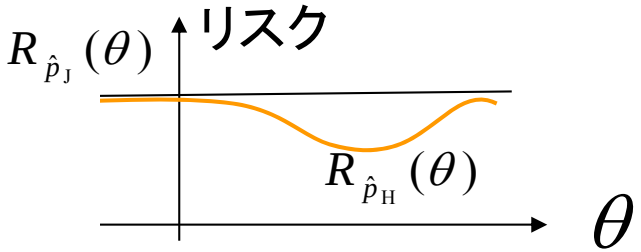
Jeffreys事前分布が望ましくない例

事実（Komaki, 2001）

分散既知、平均未知の正規分布モデル

$$p(x \mid \theta) = \frac{1}{(2\pi)^{k/2}} \exp\left\{ -\frac{\| x - \theta \|^2}{2} \right\} \quad \theta \in \mathrm{R}^k \quad k \geq 3$$

事前分布　$\pi_{\mathrm{H}}(\theta) \propto \dfrac{1}{\|\theta\|^{k-2}}$



ベイズ予測分布

$$\hat{p}_{\mathrm{H}}(y \mid x) := \int_{\Theta} p(y \mid \theta)\pi_{\mathrm{H}}(\theta \mid x)\mathrm{d}\theta$$

上の事前分布に基づくベイズ予測分布はJeffreys事前分布に基づくベイズ予測分布よりも性能が良い（優越する＝リスク関数を一様に改善）

上の結果を優調和事前分布の方法として展開

正の優調和関数

Def.

$h(\theta)$ 正の優調和関数

(positive superharmonic function) $\Leftrightarrow$

$$\begin{cases} h(\theta) > 0 \\ \Delta h(\theta) \le 0 \qquad {}^{\forall}\theta \in \Theta \end{cases}$$

ただし $\Delta h(\theta) := \frac{1}{\sqrt{g}} \frac{\partial}{\partial \theta^i} \left(\sqrt{g} g^{ij} \frac{\partial}{\partial \theta^j} h(\theta) \right)$

g_{ij} はFisher 情報行列(Fisher計量) $g_{ij} := \mathrm{E}^X \left[\frac{\partial \log p_\theta}{\partial \theta^i} \frac{\partial \log p_\theta}{\partial \theta^j} \right]$

注意

1. 正定数は自明であり, 正の非定数の優調和関数に興味がある
2. 一般のリーマン多様体上で定義される
3. 本発表では上の定義で十分

優調和事前分布

Def.

統計モデル上に正の非定数優調和関数 $h(\theta)$ が存在する時, 以下で定まる事前分布を優調和事前分布と呼ぶ

$$\pi_{\mathrm{H}}(\theta) := h(\theta)\pi_{\mathrm{J}}(\theta)$$

$$\pi_{\mathrm{J}}(\theta) \propto \sqrt{g} := \sqrt{\det(g_{ij})}$$ Jeffreys 事前分布

注意

1. 優調和事前分布も improper なケースだと規格化できない
2. Jeffreys 事前分布と違い, 存在しなかったり、無数に存在する場合もある

優調和事前分布に基いたベイズ予測分布

定理（Komaki, 2006）

一般の確率分布族について、予測分布の構成を考える。

もし、 優調和事前分布 $\pi_{\mathrm{H}}(\theta)$ が存在すれば、

$\hat{p}_{\pi_{\mathrm{H}}}(y)$ が漸近的に $\hat{p}_{\pi_{\mathrm{J}}}(y)$ を優越する（i.e. 性能がよい）

証明の概要（漸近リスクの比較）

$$n^2\left\{\mathrm{E}^X\left[D(p_\theta \,\|\, \hat{p}_{\pi_{\mathrm{J}}})\right]-\mathrm{E}^X\left[D(p_\theta \,\|\, \hat{p}_f)\right]\right\} \leftarrow \theta \text{ の関数}$$

$$=\frac{1}{2}g^{ij}(\theta)\left(\partial_i \log\frac{f(\theta)}{\pi_{\mathrm{J}}(\theta)}\right)\left(\partial_j \log\frac{f(\theta)}{\pi_{\mathrm{J}}(\theta)}\right)-\frac{\pi_{\mathrm{J}}(\theta)}{f(\theta)}\Delta\frac{f(\theta)}{\pi_{\mathrm{J}}(\theta)}+O(n^{-1/2})$$

$\hat{p}_f(y)$ 任意の事前分布 $f(\theta)$ に基づくベイズ予測分布

ここまでのまとめ

事前情報のない場合のベイズ予測分布の構成

⇒ 優調和事前分布 $\pi_{\mathrm{H}}(\theta)$ の利用が望ましい（＊）

$$\pi_{\mathrm{H}}(\theta) := h(\theta)\pi_{\mathrm{J}}(\theta)$$

$$h(\theta) > 0$$

$$\Delta h(\theta) = \frac{1}{\sqrt{g}} \frac{\partial}{\partial \theta^i} \left(\sqrt{g} g^{ij} \frac{\partial}{\partial \theta^j} h(\theta) \right) \leq 0$$

⇒ 以上の話を時系列モデル（主にARモデル）に拡張

＊先の定理は漸近論だが正規分布族の例など
有限標本でいえることが多い

ftanaka@stat.t.u-tokyo.ac.jp

時系列モデルの情報幾何とベイズ統計への応用(2)

２０１３年3月6日
初版

田中冬彦(*Tanaka Fuyuhiko*)

所属：東京大学 情報理工学系研究科　助教

発表の流れ

10. 時系列モデル
11. 定常ガウス過程の情報幾何
12. ARモデルの優調和事前分布
13. ARモデルの優調和事前分布のPAC表示
14. ARMAモデルのトレース2形式
15. まとめと今後の課題

10. 時系列モデル

時系列解析の基本的な考え方(1/4)

a. 時系列データの種類

- 1変量/多変量
- 等間隔（1日、1年; 1秒）/等間隔でない

※本講演では1変量の等間隔のデータを扱う

$$x_1, x_2, \cdots, x_n, \cdots$$

時系列データの例

- 売上の推移
- 経済指標（GDP,失業率, 雇用情勢）
- 天候データ（降水量、気温）

時系列解析の基本的な考え方(2/4)

b. 時系列データの分解

$$X_t = f_t + s_t + Y_t$$

トレンド項　季節項　定常雑音項

非確率的な部分

トレンド項（ゆるやかな変化）

季節項（周期的な変化）

⇒ ある程度の長さのデータから推定できる

確率的な部分

定常雑音項（ランダムな変化）

⇒ 統計的手法が必要

時系列解析の基本的な考え方(3/4)

$$X_t = f_t + s_t + Y_t$$

CO2濃度

X_t

トレンド項

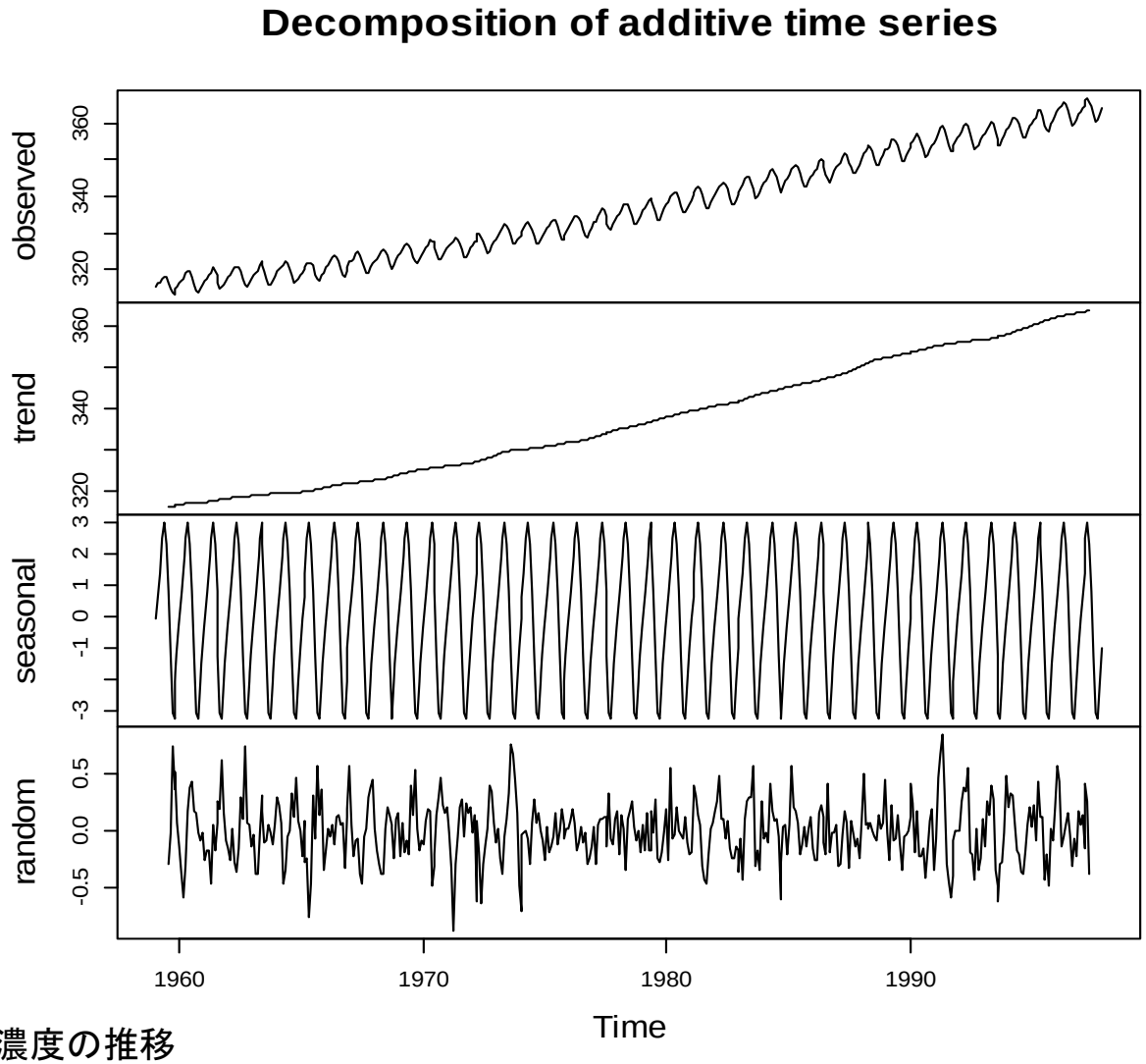
f_t

季節項

s_t

定常雑音項

Y_t



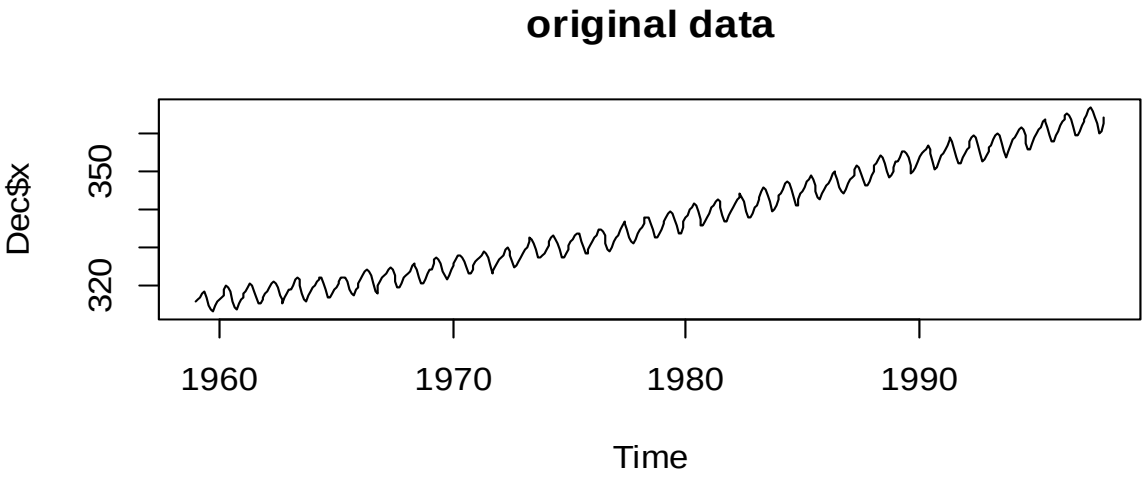
図：月次のCO2濃度の推移

時系列解析の基本的な考え方(4/4)

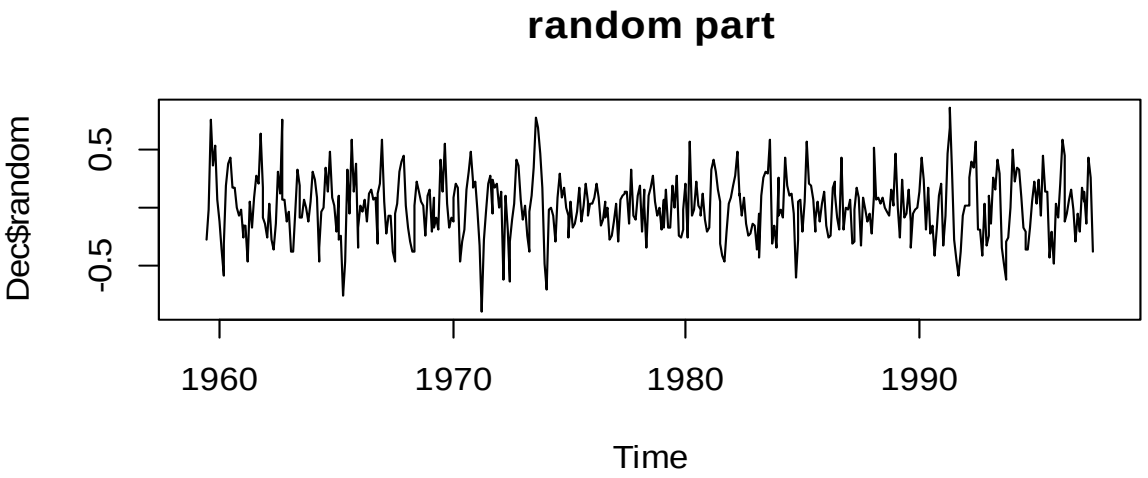
c. 時系列解析の基本的な考え方

$$X_t = f_t + s_t + Y_t$$

トレンド、季節成分を取り除いた部分＝ランダム項



以降、ランダム部分のみに注目（ X_t と表す）



時系列モデル (1/5)

a. ランダム項の取扱い

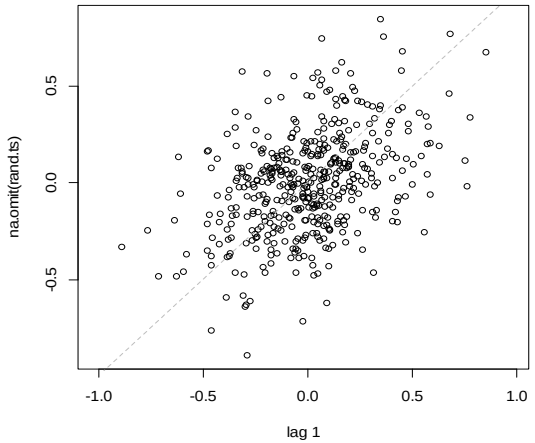
もっとも単純なモデル

異なる時点の値は独立　X_t, X_s

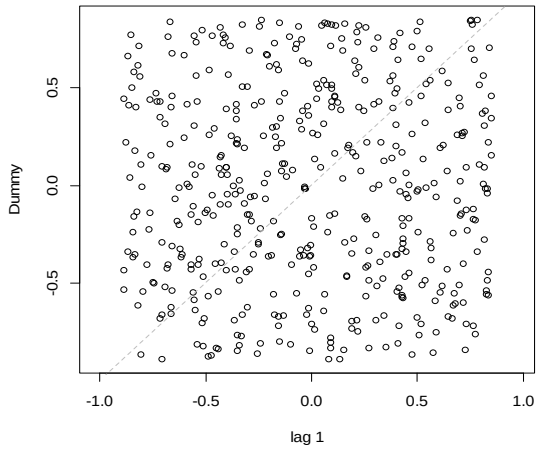
⇒ 実際には近い時点間では独立ではない！

例：CO2のランダム成分を一ヶ月ずらした値の散布図(lag plot)

(横軸, 縦軸)= (x_t, x_{t+1})



右図：独立に発生させたダミーデータ



時系列モデル (2/5)

b. 自己回帰モデル

1次の自己回帰モデル　　$-1 < a < 1$

$$x_t = ax_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$ 白色雑音（独立）

一般の自己回帰モデル（AutoRegressive model of p-th order）

$$x_t = \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$ 白色雑音

$a_1, \cdots, a_p$ ARパラメータ

※AR(p)モデル、p次のARモデルなどと呼ぶ

※ARパラメータの動く範囲は後で見る

時系列モデル (3/5)

c. 一般的な時系列モデル

（理論的な）時系列モデルの分類

- 定常時系列
- 非定常時系列（扱いにくい）

先に示した分解や差分、非線型変換などを用いてランダム部分を定常時系列モデルで解析

- 定常ガウス過程（平均と自己共分散で決まり扱いやすい）
  - 自己回帰モデル（ARモデル）
  - 移動平均モデル（MAモデル）
  - ARMAモデル
  - その他にも・・・

時系列モデル (4/5)

d. 定常ガウス過程

確率変数　$\{X_t : t = 0,1,2,\cdots\}$

期待値（時間によらない）

$$\mathrm{E}[X_t] = \mu$$

自己共分散（2時点の差のみで決まる）

$$\begin{aligned}\mathrm{Cov}(X_r, X_s) &= \mathrm{Cov}(X_{r+t}, X_{s+t}) \\ &= \mathrm{Cov}(X_{r-s}, X_0) \\ &= \gamma(r\text{-}s)\end{aligned}$$

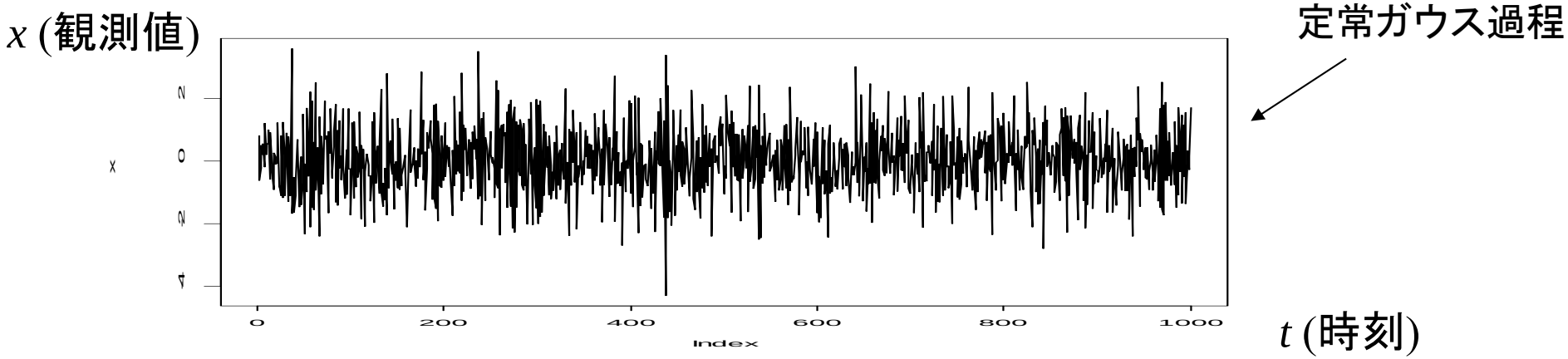
自己相関関数

$$\rho(t) := \frac{\gamma(t)}{\gamma(0)}$$

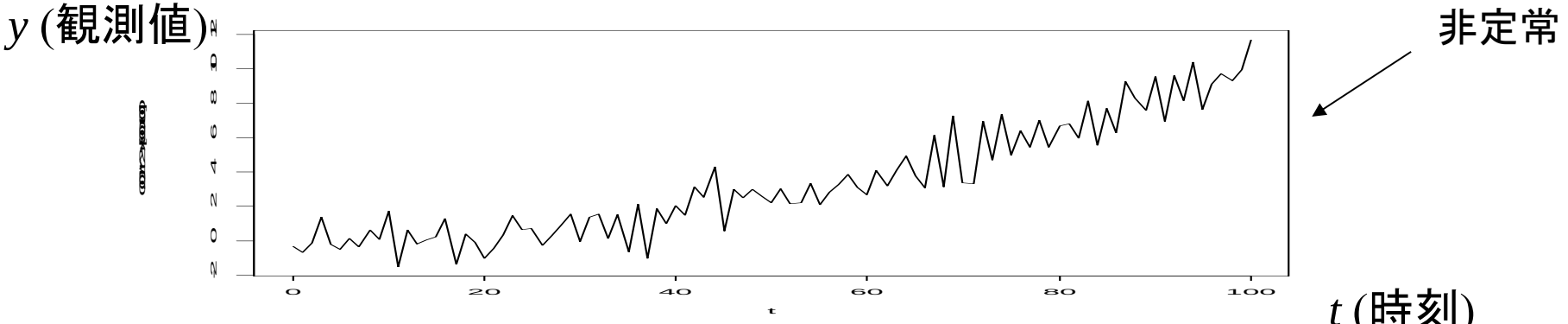
時系列モデル (5/5)

e. 時系列データのシミュレーション

AR(1)モデル　$x_t = -0.35\, x_{t-1} + \varepsilon_t$



AR(1)モデル＋トレンド項　$y_t = x_t + f_t$



11. 定常ガウス過程の情報幾何

定常ガウス過程の無情報事前分布

定常ガウス過程

- p 次のARモデル(AR(p))
- q 次のMAモデル(MA(q))
- (p,q) 次のARMAモデル（ARMA(p,q)）

無情報事前分布の議論

非i.i.d.のためAR（p）モデルでも難しい！！

- AR(1)での議論（Philips (1991), Berger and Yang (1994)）

情報幾何学的な観点から無情報事前分布を提案！！

定常ガウス過程とスペクトル密度

定常ガウス過程とスペクトル密度は1:1に対応

定常ガウス過程（ただし$\mathrm{E}[X_i] = \mu = 0$を仮定）

$$x_1, x_2, \cdots, x_n, \cdots \sim p(x_1, \cdots, x_n \mid \mu = 0, \Gamma)$$

1:1 （定常ガウス過程のとき）

自己共分散関数 $\gamma(n)$

$$\Gamma_{r,s} = \mathrm{Cov}(X_r, X_s) = \gamma(r\text{-}s) = \gamma(\mathrm{s}\text{-}r)$$

1:1（フーリエ変換）

スペクトル密度

$$S(\omega) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \gamma(n) e^{-in\omega}$$

例：AR過程のスペクトル密度 (1/2)

AR(p)過程

$a_1, \cdots, a_p$ ARパラメータ

$\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$ 白色雑音

$$x_t = -\sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t$$

AR(p) 過程のスペクトル密度

$$S(\omega \mid a_1, \cdots, a_p, \sigma^2) = \frac{\sigma^2}{2\pi} \frac{1}{|\mathrm{L}_a(e^{i\omega})|^2}$$

$$\mathrm{L}_a(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p}$$

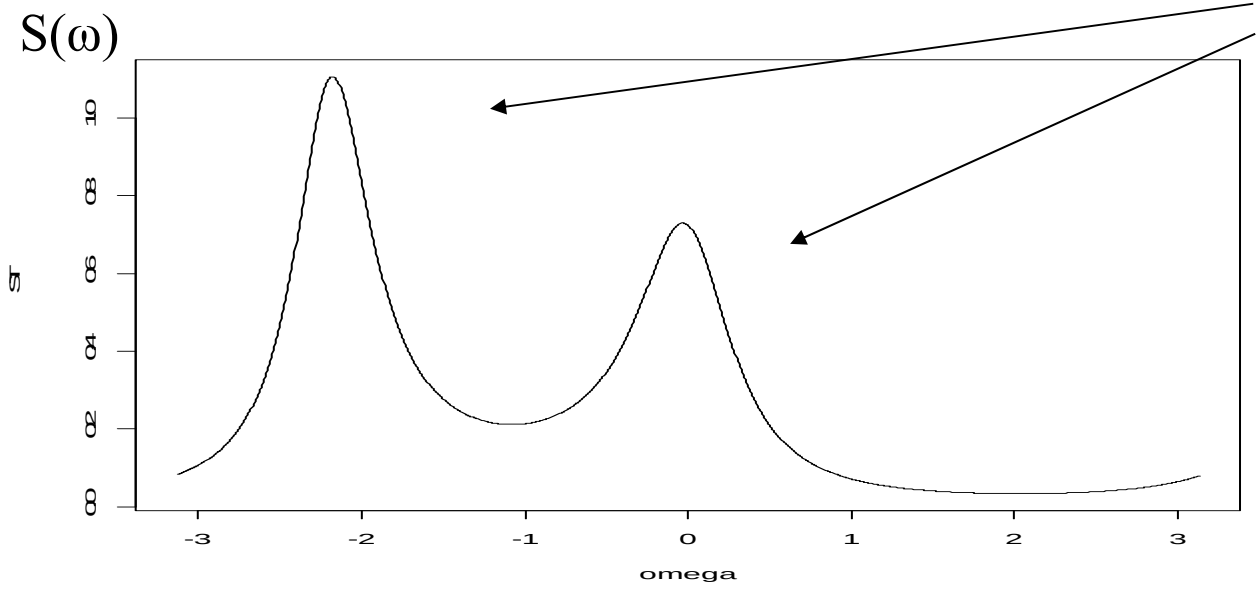
例：AR過程のスペクトル密度 (2/2)

AR(3)過程 $x_t = -0.18\, x_{t-1} + 0.05\, x_{t-2} + 0.39\, x_{t-3} + \varepsilon_t$

$$\varepsilon_t \sim \mathrm{N}(0, 1)$$

スペクトル密度

$$S(\omega \mid a_1, a_2, a_3) = \frac{1}{2\pi} \mid e^{3\omega i} + 0.18\, e^{2\omega i} - 0.05\, e^{\omega i} - 0.39 \mid^{-2}$$



ピークや裾幅は

ARパラメータで決まる

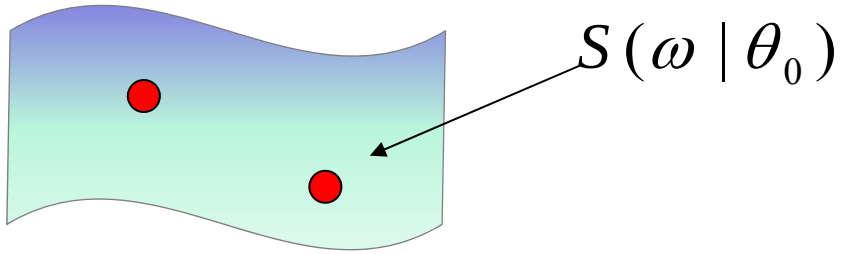
ω(周波数)

時系列モデルの幾何（情報幾何）(1/3)

時系列モデル多様体（スペクトル密度のパラメータ族）

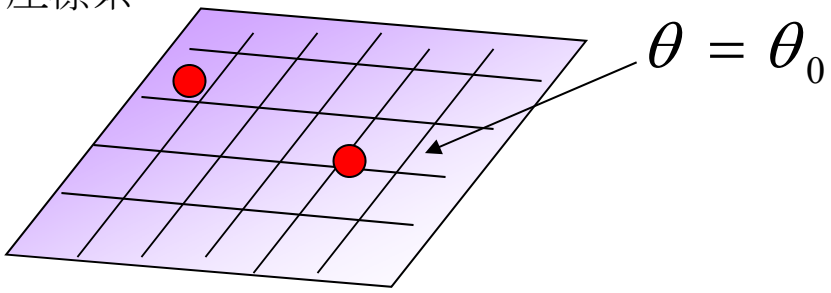
時系列モデル多様体

$$\mathrm{M} = \{S(\omega \mid \theta) : \theta \in \Theta\}$$



＊モデルのパラメータ＝多様体の座標系

$$\theta = (\theta^1, \theta^2, \cdots, \theta^d)$$



時系列モデルの幾何（情報幾何）(2/3)

独立同一分布の確率分布のdivergenceを拡張

Kullback-Leibler divergence

$$D(S \parallel S') := \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{d}\omega \left\{ \frac{S}{S'} - 1 - \log \frac{S}{S'} \right\}$$

※確率分布のダイバージェンスの適切な極限とも一致

例: 分散の異なる白色雑音

$$S(\omega) = \frac{\sigma^2}{2\pi} \qquad \gamma(n) = \delta_{0,n}\sigma^2$$

$$D(S \parallel S') = \frac{1}{2}\left\{ \frac{\sigma^2}{\sigma^{2\prime}} - 1 - \log \frac{\sigma^2}{\sigma^{2\prime}} \right\} = D(p \parallel p')$$

$$p(x \mid \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2} \right)$$

時系列モデルの幾何（情報幾何）(3/3)

OVERVIEW

1. スペクトル分布のパラメータ族

$$S(\omega \mid \theta) \qquad \theta = (\theta^1, \theta^2, \cdots, \theta^k)$$

2. モデル多様体

$$\mathrm{M} = \{S(\omega \mid \theta) : \theta \in \Theta\}$$

3. リーマン計量

$$g_{ij} := \frac{1}{4\pi} \int_{-\pi}^{\pi} \mathrm{d}\omega \, \frac{\partial \log S(\omega \mid \theta)}{\partial \theta^i} \frac{\partial \log S(\omega \mid \theta)}{\partial \theta^j}$$

4. Alpha 接続 $\alpha \in \mathrm{R}$

$$\overset{(\alpha)}{\Gamma}_{ij;k} := \overset{(e)}{\Gamma}_{ij;k} + \frac{1-\alpha}{2} T_{ijk}$$

$$T_{ijk} := \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{d}\omega \, \frac{\partial \log S(\omega \mid \theta)}{\partial \theta^i} \frac{\partial \log S(\omega \mid \theta)}{\partial \theta^j} \frac{\partial \log S(\omega \mid \theta)}{\partial \theta^k}$$

$$\overset{(e)}{\Gamma}_{ij;k} := \frac{1}{4\pi} \int_{-\pi}^{\pi} \mathrm{d}\omega \left\{ \frac{\partial^2 \log S(\omega \mid \theta)}{\partial \theta^i \partial \theta^j} - \frac{\partial \log S(\omega \mid \theta)}{\partial \theta^i} \frac{\partial \log S(\omega \mid \theta)}{\partial \theta^j} \right\} \frac{\partial \log S(\omega \mid \theta)}{\partial \theta^k}$$

例：ARモデルのパラメータ範囲 (1/3)

AR(p)過程

$a_1, \cdots, a_p$ ：ARパラメータ

$\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$ 白色雑音

$$x_t = -\sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t$$

AR(p)　モデルのパラメータの範囲

定常性条件（=特性方程式の根に対する条件）で決まる

$$|z_1|<1, \ldots, |z_p|<1$$

$z_1, \ldots, z_p$ は特性方程式 $z^p + a_1 z^{p-1} + \cdots + a_p = 0$ の根

※複素共役根と実数根の両方の場合がありえる

例：ARモデルのパラメータ範囲 (2/3)

AR(p)　モデルのパラメータの範囲

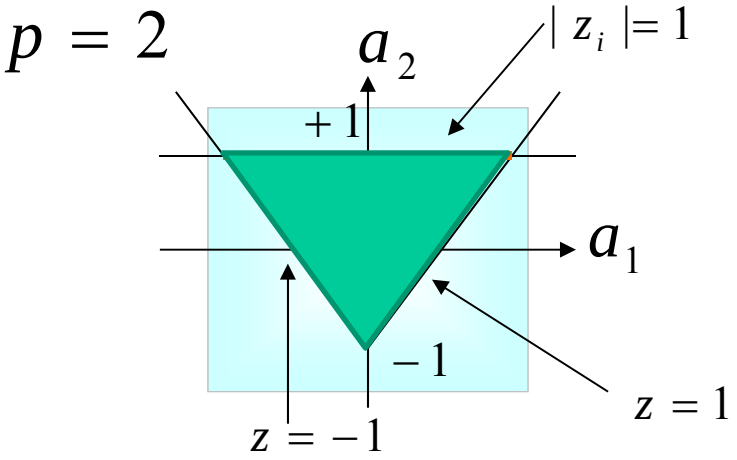
$$|z_1|<1,\ldots,|z_p|<1$$

$z_1,\ldots,z_p$ は特性方程式 $z^p + a_1 z^{p-1} + \cdots + a_p = 0$ の根

$p = 1$

$z + a_1 = 0 \quad |z_1| < 1$

$p = 2$



$z^2 + a_1 z + a_2 = 0$

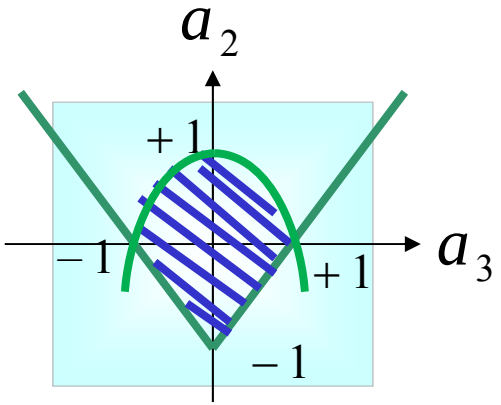
例：ARモデルのパラメータ範囲 (3/3)

3次以上は複雑！

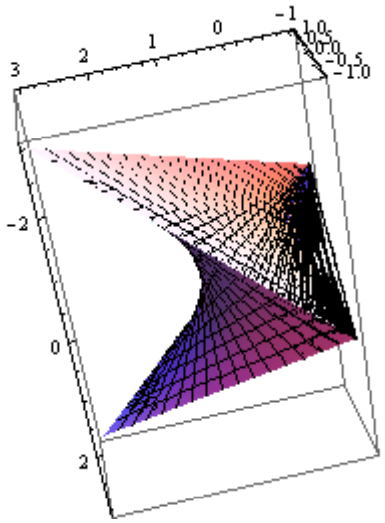
$p = 3$

$$a_1 + a_2 + a_3 > -1$$
$$a_1 - a_2 + a_3 < 1$$
$$a_3{}^2 - a_1 a_3 + a_2 < 1$$
$$|a_3| < 1$$

$a_1 = 0$ の断面



Mathematica で斜めから眺めた定常領域



例：ARモデルのFisher計量

根座標系

特性方程式

$z^p + a_1 z^{p-1} + \cdots + a_p = 0$

の根 $z_1, z_2, \cdots, z_p$

$\Longleftrightarrow$ 1 : 1

AR パラメータ

$a_1, a_2, \cdots, a_p$

AR(p) モデルのFisher計量

根を新しい局所座標と考えて計量を計算

$\theta^0 := \sigma^2 \qquad \theta^i := z_i \ (i = 1 \cdots p)$

$$g_{00} = \frac{1}{2\sigma^4} \qquad g_{ij} = \frac{1}{1 - \theta^i \theta^j}$$

$$g_{i0} = g_{0i} = 0$$

$$g_{ij} = \frac{1}{4\pi} \int_{-\pi}^{\pi} d\omega \, \frac{\partial \log S(\omega;\theta)}{\partial \theta^i} \frac{\partial \log S(\omega;\theta)}{\partial \theta^j}$$

$$S(\omega; a_1, \cdots, a_p, \sigma^2) = \frac{\sigma^2}{2\pi} \frac{1}{|L_a(e^{i\omega})|^2}$$

例：ARモデルのKL-divergence

p=1,2でも煩雑な式

AR(1)

$$D(S \,\|\, S') = \frac{1}{2}\frac{1}{1-|\alpha|^2}\left\{1+|\beta|^2-\alpha\overline{\beta}-\overline{\alpha}\beta\right\}-\frac{1}{2}$$

$$= \frac{1}{2}\frac{(\alpha-\beta)^2}{1-\alpha^2}$$

α, β は根座標

AR(2)

$$D(S \,\|\, S') = \frac{1}{2}\left\{\frac{\beta_1\beta_2}{\alpha_1\alpha_2}+\frac{(1-\alpha_1\overline{\beta_1})(1-\alpha_1\overline{\beta_2})(\alpha_1-\beta_1)(\alpha_1-\beta_2)}{\alpha_1(\alpha_1-\alpha_2)(1-\alpha_1\overline{\alpha_1})(1-\alpha_1\overline{\alpha_2})}\right.$$

$$\left.+\frac{(1-\alpha_2\overline{\beta_1})(1-\alpha_2\overline{\beta_2})(\alpha_2-\beta_1)(\alpha_2-\beta_2)}{\alpha_2(\alpha_2-\alpha_1)(1-\alpha_2\overline{\alpha_1})(1-\alpha_2\overline{\alpha_2})}\right\}-\frac{1}{2}$$

$\alpha_1, \alpha_2, \beta_1, \beta_2$ は根座標

スペクトル密度のベイズ推定(1/2)

問題設定(i.i.d. と類似)

・n個の データ ～定常ガウス過程　$S(\omega|\theta)$　　$x \sim p_n(x \mid \theta)$

・事前分布（仮定）　$\theta \sim \pi(\theta)$

・Kullback-Leibler 損失(推定誤差の評価)

$$D(S_\theta \| \hat{S}) := \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{d}\omega \left\{ \left(\frac{S(\omega \mid \theta)}{\hat{S}(\omega \mid x)} \right) - 1 - \log \left(\frac{S(\omega \mid \theta)}{\hat{S}(\omega \mid x)} \right) \right\}$$

・リスク関数　　$R_{\hat{S}}(\theta) := \mathrm{E}^X[D(S_\theta \| \hat{S})]$

このとき、n個のデータ　$x = (x_1, \ldots, x_n)$　から　スペクトル密度を推定

Problem

平均リスク　$\int R_{\hat{S}}(\theta)\pi(\mathrm{d}\theta)$　を最小化するスペクトル密度の推定

$\hat{S}(\omega \mid x)$　は？

スペクトル密度のベイズ推定(2/2)

KL損失のもとで最適な推定（i.i.d.と類似 ）

$$\hat{S}_{\pi}(\omega) := \int_{\Theta} S(\omega \mid \theta)\pi(\theta \mid x)\mathrm{d}\theta$$

$$\pi(\theta \mid x) := \frac{p_n(x \mid \theta)\pi(\theta)}{\int_{\Theta} p_n(x \mid \theta)\pi(\theta)\mathrm{d}\theta}$$

cf. 予測分布の場合

$$\hat{p}_{\pi}(y \mid x) := \int_{\Theta} p(y \mid \theta)\pi(\theta \mid x)\mathrm{d}\theta$$

MLE Plug-in との比較 (1/2)

AR(3)過程でベイズと最尤推定による推定を数値的に比較

真のスペクトル密度（任意にARパラメータ固定）

$$S_0(\omega) := S(\omega \mid \theta)$$

最尤推定にもとづくスペクトル密度 (MLE plug-in)

$$\hat{S}_{\mathrm{M}}(\omega) := S(\omega \mid \hat{\theta}_{\mathrm{MLE}}(x))$$

Jeffreys事前分布にもとづくベイズスペクトル密度

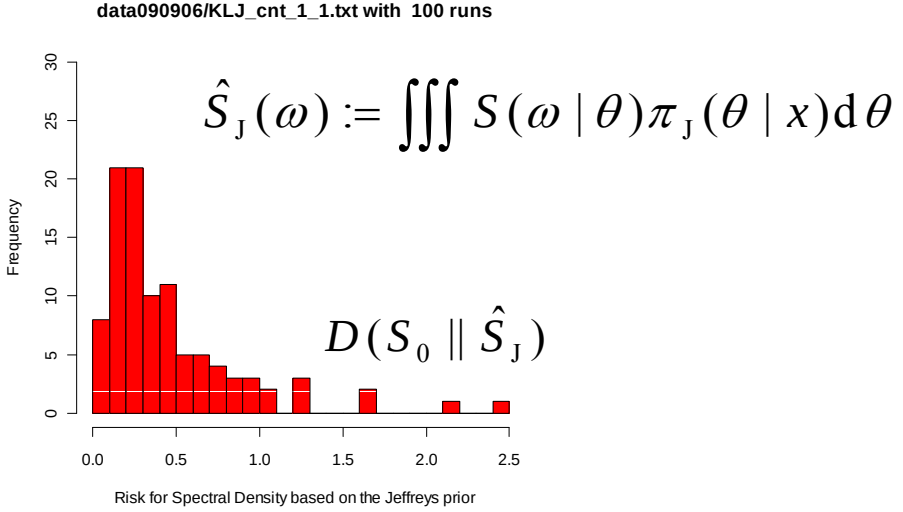
$$\hat{S}_{\mathrm{J}}(\omega) := \iiint S(\omega \mid \theta)\pi_{\mathrm{J}}(\theta \mid x)\mathrm{d}\theta$$

データの発生を繰り返して推定誤差 $D(S_0 \,||\, \hat{S})$ の分布を見る

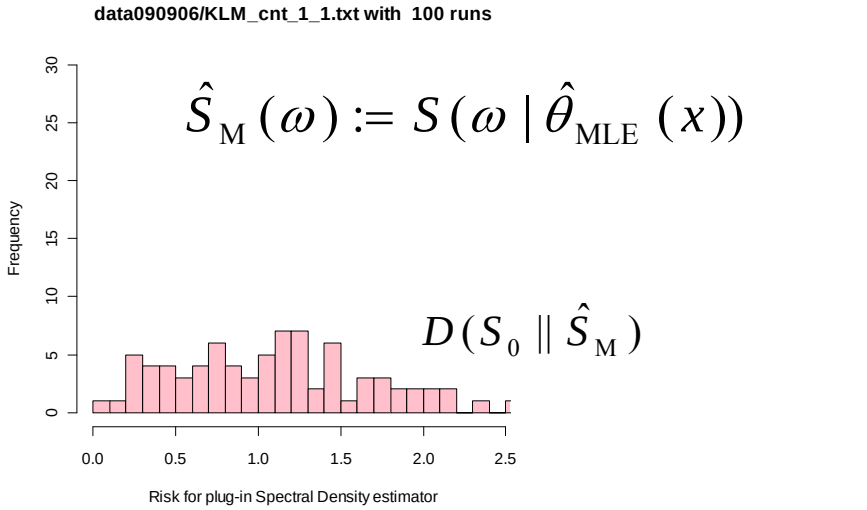
MLE Plug-in との比較　(2/2)

*推定誤差の分布（0に集中するほどよい推定）

Jeffreys事前分布



最尤推定のプラグイン



$(r_1, r_2, r_3) = (-0.6, \ -0.4, 0.2)$

ベイズの方がMLE Plug-inに比べて性能が良い

優調和事前分布の導入 (1/2)

パラメータの事前情報がない場合

先の例ではJeffreys 事前分布 $\pi_{\mathrm{J}}(\theta)$ に基づいて

$$\hat{S}_{\mathrm{J}}(\omega) := \int_{\Theta} S(\omega \mid \theta)\pi_{\mathrm{J}}(\theta \mid x)\mathrm{d}\theta$$

で推定

⇒ improperなら, さらに改善できる可能性がある

improper の例（ARモデル）

Jeffreys 事前分布は $(p \geq 2)$ で improper

$$\pi_{\mathrm{J}} \propto \left| \frac{\prod_{i<j}(z_i - z_j)^2}{\prod_{j=1}^{p}\prod_{i=1}^{p}(1 - z_i z_j)} \right|^{1/2} \qquad |z_i| < 1,\ z_1, \cdots, z_p$$

優調和事前分布の導入 (2/2)

定理（Tanaka and Komaki, 2011）

もし、優調和事前分布 $\pi_{\mathrm{H}}(\theta)$が存在すれば、

$\hat{S}_{\pi_{\mathrm{H}}}(\omega)$が漸近的に $\hat{S}_{\pi_{\mathrm{J}}}(\omega)$ を優越（i.e. 性能がよい）

リスクの比較（漸近的）

$$n^2\left\{\mathrm{E}^X\left[D(S_\theta \,\|\, \hat{S}_{\pi_{\mathrm{J}}})\right]-\mathrm{E}^X\left[D(S_\theta \,\|\, \hat{S}_f)\right]\right\}$$

← θ の関数

$$=\frac{1}{2}g^{ij}(\theta)\left(\partial_i \log\frac{f(\theta)}{\pi_{\mathrm{J}}(\theta)}\right)\left(\partial_j \log\frac{f(\theta)}{\pi_{\mathrm{J}}(\theta)}\right)-\frac{\pi_{\mathrm{J}}(\theta)}{f(\theta)}\Delta\frac{f(\theta)}{\pi_{\mathrm{J}}(\theta)}+O(n^{-1/2})$$

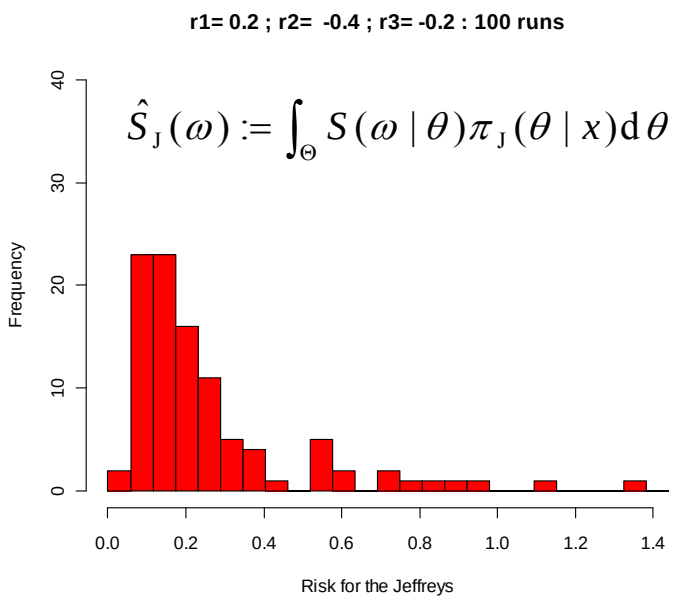
$\hat{S}_f(\omega)$ 任意の事前分布 $f(\theta)$ に基づくベイズ推定量

証明の際に原論文の標記に合わせているが、実は$\sqrt{h(\theta)}$ が正の非定数優調和関数なら十分であることが示される（$h(\theta)$ は優調和でなくとも良い）

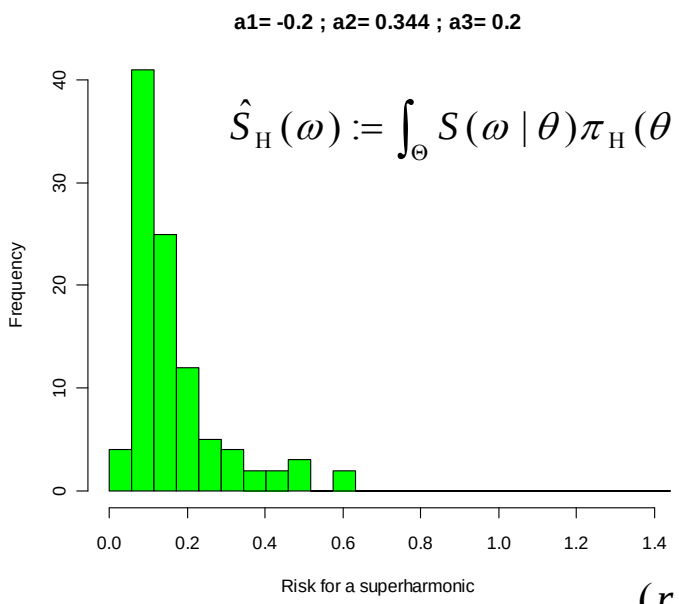
AR(3)過程での数値実験

*推定誤差の分布（0に集中するほどよい推定）

Jeffreys事前分布



優調和事前分布



$(r_1, r_2, r_3) = (0.2, -0.4, -0.2)$

ベイズ同士では優調和事前分布の方が良さそう

ここまでのまとめ

時系列モデルでの無情報事前分布

⇒ スペクトル密度のパラメータ族の幾何学的な量に注目！

（独立同一分布での結果が拡張できる）

幾何的に考察可能な事前分布

α 平行事前分布が存在 ⇔ α 接続に関するトレース2形式が0　$\Xi^{(\alpha)} = 0$

優調和事前分布が存在 ⇔ モデル多様体上に正の優調和関数が存在

$$h(\theta) > 0$$

$$\Delta h(\theta) = \frac{1}{\sqrt{g}} \frac{\partial}{\partial \theta^i} \left(\sqrt{g} g^{ij} \frac{\partial}{\partial \theta^j} h(\theta) \right) \leq 0$$

12. ARモデルの優調和事前分布

優調和事前分布を探す

Def. （再掲）

$\pi_{\mathrm{H}}(\theta)$ 優調和事前分布

$$\Longleftrightarrow \pi_{\mathrm{H}}(\theta) := h(\theta)\pi_{\mathrm{J}}(\theta)$$

$h(\theta)$　正の非定数優調和関数

$$\pi_{\mathrm{J}}(\theta) \propto \sqrt{g} := \sqrt{\det(g_{ij})}$$

Jeffreys 事前分布

統計的な課題

与えられた時系列モデルに対して優調和事前分布は存在するか？

同値な課題

与えられた時系列モデル多様体に対して正の非定数優調和関数は存在するか？

ARモデルの場合

正の非定数優調和関数

AR(p)モデルの場合, 根座標系を用いて以下のようにかける

$$h(z_1,\ldots,z_p)>0,$$

$$\Delta h=\frac{\partial}{\partial z_i}\left(g^{ij}\frac{\partial h}{\partial z_j}\right)+\frac{1}{2g}\frac{\partial g}{\partial z_i}\left(g^{ij}\frac{\partial h}{\partial z_j}\right)\le 0 \qquad |z_1|<1,\ldots,|z_p|<1$$

ただし計量の逆行列は以下のような煩雑な形のため
一般形の予想 and/or 実際に証明することが著しく難しい

$$g^{mh}=\frac{(1-z_mz_h)\prod_{l\neq h}(1-z_lz_m)\prod_{l\neq m}(1-z_lz_h)}{\prod_{l\neq h}(z_h-z_l)\prod_{l\neq m}(z_m-z_l)}$$

AR(p)モデルの優調和関数形の予想

AR(p)モデルのJeffreys事前分布

$$\pi_{\mathrm{J}} \propto \left| \frac{\prod_{i<j} (z_i - z_j)^2}{\prod_{j=1}^{p} \prod_{i=1}^{p} (1 - z_i z_j)} \right|^{1/2}$$

＊根座標系

$$\mathrm{L}_a(z) = z^p + a_1 z^{p-1} + \cdots + a_p = 0$$

AR(2)モデル多様体上の正の優調和関数 (比較的すぐ見つかる)

$$h(z_1, z_2) = 1 - z_1 z_2$$

以上から一般形をどう見つけたかheuristic ideaを説明

AR(2)での考察(1/2)

正の優調和関数の存在は体積増大度と関連

⇒ 不変体積要素($\pi_J(\theta)\mathrm{d}\theta$)の境界での発散に注目

$$\pi_J(\theta) \propto \sqrt{\det(g_{ij})} = \frac{|z_1 - z_2|}{\sqrt{(1-z_1^2)(1-z_2^2)(1-z_1 z_2)^2}}$$

境界でのJeffreys priorの発散は2種類

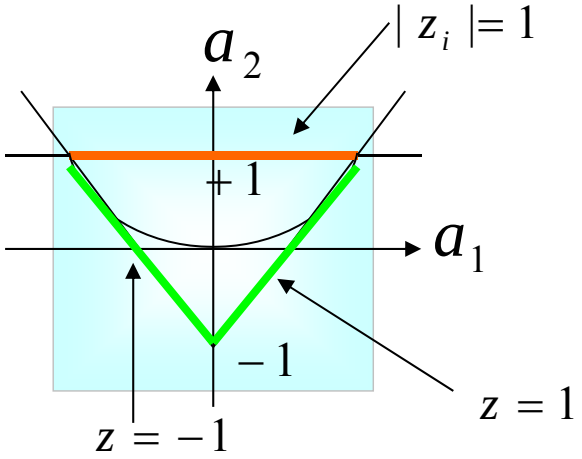
パラメータの範囲は定常条件 $|z_i| < 1$

実根(境界で積分有限)

$$\pi_J(\theta) \propto \frac{1}{\sqrt{1-z_i^2}} \approx \frac{1}{\sqrt{2(1-z_i)}}$$

複素共役根(境界で積分発散)

$$\pi_J(\theta) \propto \frac{1}{(1-z_1 z_2)^2} \approx \frac{1}{2(1-|z_1|)}$$



$$z^2 + a_1 z + a_2 = 0$$

AR(2)での考察(2/2)

正の優調和関数の境界での振舞いに注目

$$h(\theta) = 1 - z_1 z_2 (= 1 - a_2)$$

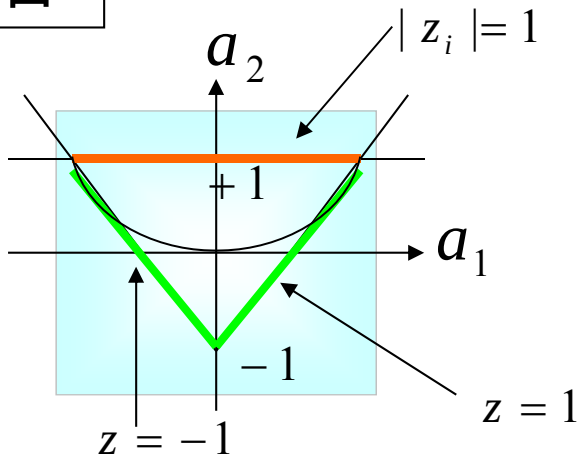
実根

$$h(\theta) = const\ .(\neq 0)$$

複素共役根

$$h(\theta) = 1 - |z_1|^2 \propto 2(1 - |z_1|)$$

$$\pi_H \propto \pi_J h$$ 境界で積分が有限



$$z^2 + a_1 z + a_2 = 0$$

予想

$\pi_J(\theta)$ の発散を抑える因子が優調和関数になるのでは？

予想を支持する補題

補題

$h(\theta)$ を正の優調和関数とする。このとき

$h_r(\theta) := (h(\theta))^r \quad 0 < r \leq 1$ も正の優調和関数

Proof.

$$-\frac{\Delta h_r}{h_r} = r(1-r)g^{ij}(\partial_i \log h)(\partial_j \log h) + r\left(-\frac{\Delta h}{h}\right) \geq 0$$

よりあきらか。

$$\Delta h = \frac{\partial}{\partial z_i}\left(g^{ij}\frac{\partial h}{\partial z_j}\right) + \frac{1}{2g}\frac{\partial g}{\partial z_i}\left(g^{ij}\frac{\partial h}{\partial z_j}\right)$$

$$g^{mh} = \frac{(1-z_m z_h)\prod_{l\neq h}(1-z_l z_m)\prod_{l\neq m}(1-z_l z_h)}{\prod_{l\neq h}(z_h - z_l)\prod_{l\neq m}(z_m - z_l)}$$

一般形の予想

正の優調和関数の形（予想）

$$h_r := \prod_{i>j} (1 - z_i z_j)^r \quad 0 < r \leq 1$$

$\pi_J(\theta)$ にかけるとどの境界での積分も有限！

AR（3）モデル

直接計算（Mathematica）で証明（補題によりr=1でのみ考える）

一般のAR(p)では？

$$\Delta h = \frac{\partial}{\partial z_i}\left(g^{ij} \frac{\partial h}{\partial z_j} \right) + \frac{1}{2g} \frac{\partial g}{\partial z_i}\left(g^{ij} \frac{\partial h}{\partial z_j} \right) \quad g^{mh} = \frac{(1 - z_m z_h) \prod_{l \neq h} (1 - z_l z_m) \prod_{l \neq m} (1 - z_l z_h)}{\prod_{l \neq h} (z_h - z_l) \prod_{l \neq m} (z_m - z_l)}$$

・・計算には何らかの工夫が必要

計算方法の工夫(1/3)

試行錯誤の結果、以下の三つの項に分けると良い

$$\frac{\Delta h}{h} = -\sum_{i=1}^{p} z_i \frac{\partial_i V_p}{V_p} - p$$

定義からすぐに評価できて

$$-\frac{p(p-1)}{2} - p = -\frac{p(p+1)}{2}$$

$$+\sum_{i=1}^{p} \frac{-z_i^2}{1-z_i^2} + \sum_{i=1}^{p}\sum_{j=1}^{p} g^{ij} \frac{z_i}{1-z_i^2}\frac{z_j}{1-z_j^2} \quad \text{(A)} \leftarrow \text{要工夫}$$

$$+\sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial_i V_p}{V_p} g^{ij} \frac{z_j}{1-z_j^2} + \sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial}{\partial z_i}\left\{ g^{ij} \frac{z_j}{1-z_j^2} \right\} \quad \text{(B)} \leftarrow \text{要工夫}$$

ただし

$$g^{mh} = \frac{(1-z_m z_h)\prod_{l\neq h}(1-z_l z_m)\prod_{l\neq m}(1-z_l z_h)}{\prod_{l\neq h}(z_h - z_l)\prod_{l\neq m}(z_m - z_l)} \qquad V_p := \prod_{i<j}(z_j - z_i)$$

計算方法の工夫(2/3)

(A),(B)は定数になる！！

$$(\mathrm{A}) := \sum_{i=1}^{p} \frac{-z_i^2}{1-z_i^2} + \sum_{i=1}^{p}\sum_{j=1}^{p} g^{ij} \frac{z_i}{1-z_i^2}\frac{z_j}{1-z_j^2} = \begin{cases} \dfrac{p}{2} & \text{p: 偶数} \\ \dfrac{p-1}{2} & \text{p: 奇数} \end{cases}$$

$$(\mathrm{B}) := \sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial_i V_p}{V_p} g^{ij} \frac{z_j}{1-z_j^2} + \sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial}{\partial z_i}\left\{ g^{ij} \frac{z_j}{1-z_j^2} \right\} = \begin{cases} \dfrac{p}{2} & \text{p: 偶数} \\ \dfrac{p+1}{2} & \text{p: 奇数} \end{cases}$$

$$g^{mh} = \frac{(1-z_m z_h)\prod_{l\neq h}(1-z_l z_m)\prod_{l\neq m}(1-z_l z_h)}{\prod_{l\neq h}(z_h-z_l)\prod_{l\neq m}(z_m-z_l)} \qquad V_p := \prod_{i<j}(z_j - z_i)$$

計算方法の工夫(3/3)

途中に出てくる計算の例:

$$\sum_{m=1}^{p}\left\{\frac{\prod_{l=1}^{p}(1-z_l z_m)}{(1-z_m^2)\prod_{l\neq m}(z_m-z_l)}\right\} = \begin{cases} 0 & \text{p: 偶数} \\ V_p & \text{p: 奇数} \end{cases}$$

$$V_p := \prod_{i<j}(z_j - z_i) \quad \text{Vandermonde行列式}$$

まともに展開すると大変だが、たとえば行列式の形に書いて評価できる。

$$V\left(f(X)\right) := (-1)^{p-1}\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{p-2} & z_2^{p-2} & \cdots & z_p^{p-2} \\ f(z_1) & f(z_2) & \cdots & f(z_p) \end{vmatrix} \qquad f(X) := \frac{\prod_{l=1}^{p}(1-z_l X)}{1-X^2}$$

＊効率のよい別解の探求は今後の課題

Schur多項式との関係

p=3のケース

$$\frac{V\left(X^{k+2}\right)}{V_3}=\frac{\begin{vmatrix}1 & 1 & 1\\ z_1 & z_2 & z_3\\ z_1^{k+2} & z_2^{k+2} & z_3^{k+2}\end{vmatrix}}{\begin{vmatrix}1 & 1 & 1\\ z_1 & z_2 & z_3\\ z_1^{2} & z_2^{2} & z_3^{2}\end{vmatrix}}=S_{(k,0,0)}(z_1,z_2,z_3)$$

＊3次のシューア多項式（の一部）

$$S_{(k,0,0)}(z_1,z_2,z_3)=\sum_{\substack{\nu_1+\nu_2+\nu_3=k\\ \nu_1,\nu_2,\nu_3\geq 0}} z_1^{\nu_1} z_2^{\nu_2} z_3^{\nu_3}$$

シューア多項式を用いて以下の公式を求めることも可能。
（逆に以下の公式の a の展開係数からシューア多項式が導出できる。）

$$\frac{V\left(\frac{1}{1-aX}\right)}{V_3}=a^2\,\frac{1}{1-az_1}\,\frac{1}{1-az_2}\,\frac{1}{1-az_3}$$

余談：Schur多項式とYoung図

シューア多項式の各単項式＝半標準盤

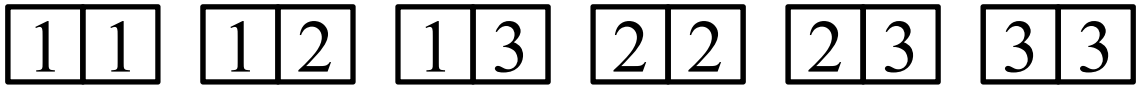
p=3のケース

$$S_{(k,0,0)}(z_1,z_2,z_3)=\sum_{\substack{\nu_1+\nu_2+\nu_3=k\\ \nu_1,\nu_2,\nu_3\geq 0}} {z_1}^{\nu_1}{z_2}^{\nu_2}{z_3}^{\nu_3}$$

k=1

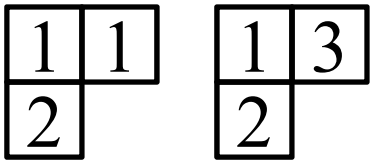
$$S_{(k=1,0,0)}(z_1,z_2,z_3)=z_1+z_2+z_3$$

k=2



$$S_{(k=2,0,0)}(z_1,z_2,z_3)={z_1}^2+z_1z_2+z_1z_3+{z_2}^2+z_2z_3+{z_3}^2$$

半標準盤



1~n を横は単調増加；縦は狭義単調増加で重複を許して書き入れる

AR(p)モデルの優調和事前分布

定理(Tanaka, 2009)

AR(p) $(p \geq 2)$ モデル多様体を考える。

$h := \prod_{i>j} (1 - z_i z_j)$ とおくと　$h > 0$　であって

$$\Delta h = -\frac{p(p-1)}{2} h \qquad \Delta h := \frac{1}{\sqrt{g}} \frac{\partial}{\partial \theta^i} \left(\sqrt{g} g^{ij} \frac{\partial h}{\partial \theta^j} \right)$$

つまり、h はAR(p)モデル多様体上の正の非定数優調和関数

さらにLaplacianの固有関数にもなっている。

系

AR(p)モデルの優調和事前分布（のひとつ）は以下で与えられる。

$$\pi_{\mathrm{H}} = \pi_{\mathrm{J}} h \propto \left| \frac{\prod_{i<j} (z_i - z_j)^2}{\prod_{i=1}^{p} (1 - z_i^2)} \right|^{1/2}$$

進捗状況

AR(p) モデル

・上でみたとおり

・AR(2)はLaplacianが簡単になるためもう少し多い

MA(q) モデル

・リーマン構造がARモデルと同じため、そのまま使える
（接続まで入れるAR, MA は違ってくる）

ARMA(p,q) モデル

・Open

（以上のモデルは時系列解析では初等的；

ARIMA, GARCH, VAR etc. など課題は多い）

13. ARモデルの優調和事前分布のPAC表示

実用上の問題点

ベイズスペクトル密度

$$\hat{S}_{\pi}(\omega) := \int_{\Theta} S(\omega \mid \theta)\pi(\theta \mid x)\mathrm{d}\theta$$

$\int_{\Theta} \cdot \mathrm{d}\theta$ はモンテカルロ積分を実行

$\Theta \subseteq \mathrm{R}^{p}$ はパラメータによっては複雑な領域

パラメータ表示

1. ARパラメータ ⇒ Θ は次数が上がると複雑な領域
2. 特性方程式の根（根パラメータ）
   ⇒ Θ は共役複素根が入ると煩雑

もっと実用的に扱いやすいパラメータ表示は？

PACパラメータの導入(1/2)

偏自己相関係数(PAC)パラメータ

p次ARモデルのパラメータを逐次的に定義(由来は省略)

$k = 1, \ldots, p$

$$a_k^{(k)} = -r_k$$
$$a_j^{(k)} = a_j^{(k-1)} - r_k a_{k-j}^{(k-1)} \quad j = 1, \ldots, k-1$$

$(r_1, \ldots, r_p)$ をPACパラメータと呼ぶ

Barndorff-Nielsen and Schou (1973, *J. Multivariate Anal.*)

Θ_p ARパラメータの定常性条件をみたす領域

$\Xi_p := (-1,1)^{\times p}$ p次元超立方体

$\varphi : (r_1, \ldots, r_p) \in \Xi_p \mapsto (a_1, \ldots, a_p) \in \Theta_p$ 上で定義された写像

⇒ $\Xi_p \cong \Theta_p$ 微分同相(1:1, φ, φ^{-1} C^∞ 写像)

PACパラメータの導入(2/2)

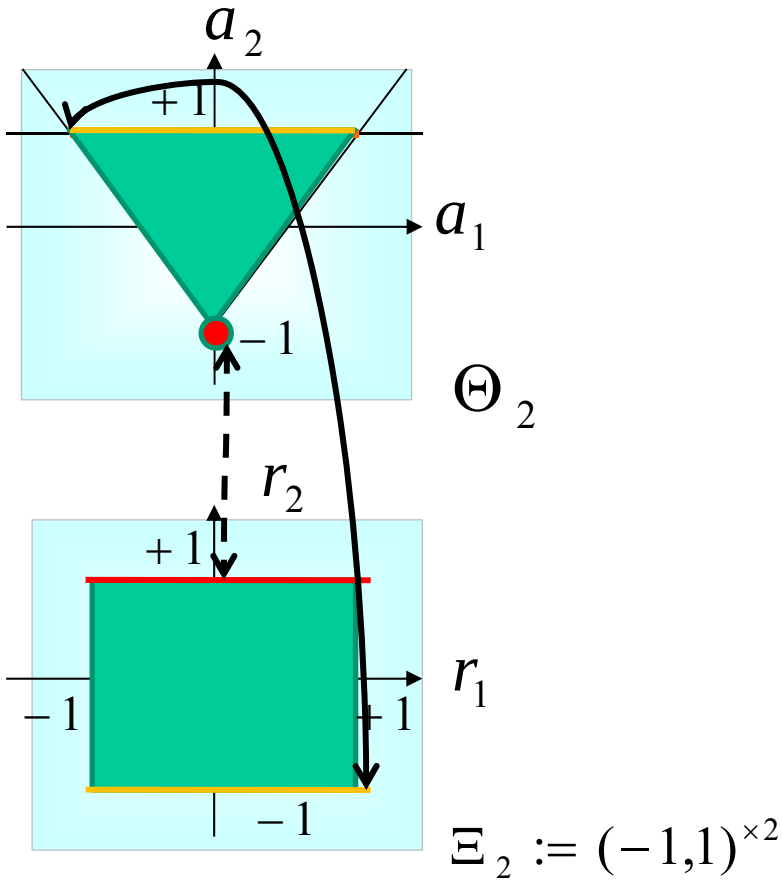
PACパラメータによる表示の例

$$a_1^{(1)} = -r_1$$

$$a_1^{(2)} = -r_1 + r_1 r_2$$
$$a_2^{(2)} = -r_2$$

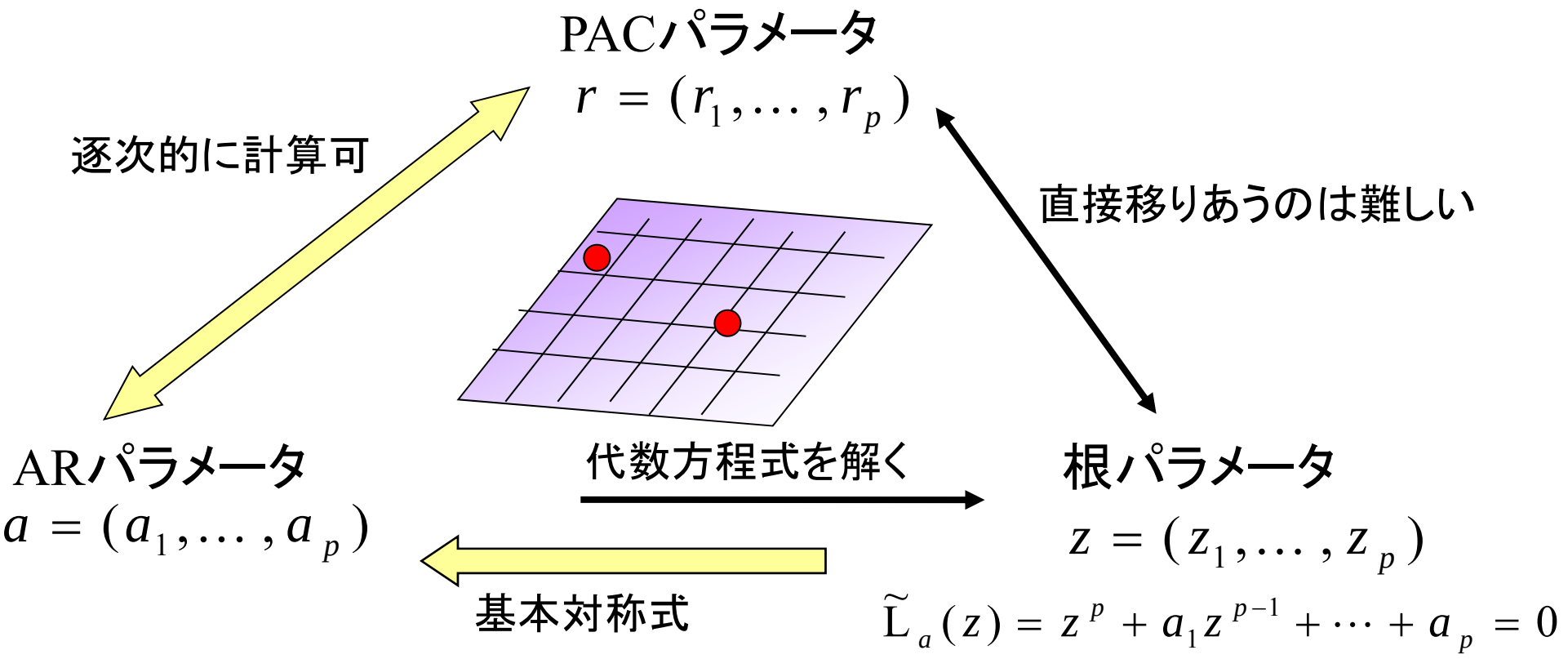
$$a_1^{(3)} = -r_1 + r_1 r_2 + r_2 r_3$$
$$a_2^{(3)} = -r_2 + r_1 r_3 - r_1 r_2 r_3$$
$$a_3^{(3)} = -r_3$$

2次のARパラメータとPACパラメータの対応図(座標変換)



ARモデル：3つのパラメータ表示

ARモデルのパラメータ表示

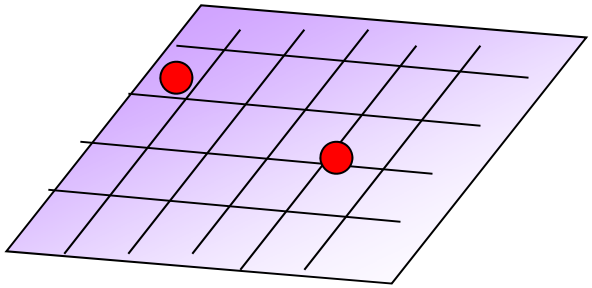


各パラメータは幾何学的には座標系に対応

ARモデル多様体:3つの座標系

PAC座標系　$r = (r_1, \ldots, r_p)$

・数値実験、数値積分用



AR座標系　$a = (a_1, \ldots, a_p)$

・統計的考察用

根座標系　$z = (z_1, \ldots, z_p)$

・幾何学的な量の計算;証明

必要に応じてうまい座標系を選ぶことがポイント

PACでのAR(p)での優調和事前分布

定理

AR(p) $(p \geq 2)$ 過程において 分散パラメータ固定で考える

優調和事前分布（のひとつ）はproperであり、PACパラメータを用いると独立なベータ分布の積でかける

$$\pi_{\mathrm{H}}(r)\mathrm{d}r = \prod_{i=1}^{p} \left\{ \pi_{\mathrm{H}}^{(i)}(r_i)\mathrm{d}r_i \right\}$$

$$r \sim \pi_{\mathrm{H}}^{(k)}(\mathrm{d}r) = \frac{(1-r)^{\frac{k}{2}-1}(1+r)^{\frac{k}{2}-1}}{2^{k-1}\mathrm{B}\left(\frac{k}{2},\frac{k}{2}\right)} \qquad k = 1, \cdots, p$$

系

上の優調和事前分布に基づいて構成したスペクトル密度の推定量は許容的

１４. ARMAモデルのトレース2形式

トレース2形式の計算からわかること

現状

・AR,　MAモデル

e(m) 平坦 ⇒ MLE prior (及びα平行事前分布)の存在はただちにいえる

・ARMAモデル

α平坦ではないことが知られている

⇒ MLE prior (及びα平行事前分布)の存在の有無は不明だった

⇒ ARMA(p,q) モデルで具体的な計算(＊)でα平行事前分布をもたないことが示された

⇒ ARMA(p,q) でMLEのplug-inはplug-inの中でも(漸近リスクの意味で)非許容的;

＊あとで計算例を紹介

ARMAモデル

ARMA (p,q) モデル

$$x_t = -\sum_{i=1}^{p} a_i x_{t-i} + \eta_t + \sum_{j=1}^{q} b_j \eta_{t-j}$$

$a_1, \cdots, a_p$: ARパラメータ

$b_1, \cdots, b_q$: MA パラメータ

$\eta_t \sim \mathrm{N}(0, \sigma^2)$ 白色雑音

スペクトル密度

$$S(\omega \mid \{a_i\}, \{b_j\}, \sigma^2) = \frac{\sigma^2}{2\pi} \frac{|\mathrm{M}_b(e^{i\omega})|^2}{|\mathrm{L}_a(e^{i\omega})|^2}$$

$$L_a(z) := 1 + \sum_{i=1}^{p} a_i z^{-i}, M_b(z) = 1 + \sum_{j=1}^{q} b_j z^{-j}$$

* 以降，分散は固定 $\sigma^2 = 1$

ARMAモデルのFisher計量

根座標系

$$\tilde{\mathrm{L}}_a(z) = z^p + a_1 z^{p-1} + \cdots + a_p = 0$$

$$z_1, \cdots, z_p \iff a_1, \cdots, a_p \quad \text{one-to-one}$$

$$\tilde{\mathrm{M}}_b(z) = z^q + b_1 z^{q-1} + \cdots + b_q = 0$$

$$w_1, \cdots, w_q \iff b_1, \cdots, b_q$$

根座標系での幾何学的な量

$$g_{ij} = \frac{\varepsilon_i \varepsilon_j}{1 - \theta^i \theta^j} \qquad |\theta^i| < 1$$

$$\theta^i := z_i \qquad \varepsilon_i := +1 \qquad (i = 1, \cdots, p)$$

$$\theta^i := w_i \qquad \varepsilon_i := -1 \qquad (i = p+1, \cdots, p+q)$$

$$T_{ijk} = \varepsilon_i \varepsilon_j \varepsilon_k \left\{ \frac{2\theta^i}{(1-\theta^i\theta^j)(1-\theta^i\theta^k)} + \frac{2\theta^j}{(1-\theta^j\theta^k)(1-\theta^j\theta^i)} + \frac{2\theta^k}{(1-\theta^k\theta^i)(1-\theta^k\theta^j)} \right\}$$

Ex: ARMA(1,1) モデル (1/3)

根座標系でのFisher計量

$$g_{ij} = \begin{pmatrix} \dfrac{1}{1-z^2} & -\dfrac{1}{1-zw} \\ -\dfrac{1}{1-wz} & \dfrac{1}{1-w^2} \end{pmatrix} \qquad |z|<1, |w|<1 \qquad \theta^1 := z \quad \theta^2 := w$$

Formula

$$g_{ij} = \frac{\varepsilon_i \varepsilon_j}{1-\theta^i \theta^j} \quad |\theta^i| < 1$$

ARMA(1,1) モデル

$$x_t = -a_1 x_{t-1} + \eta_t + b_1 \eta_{t-1}$$

$$\eta_t \sim \mathrm{N}(0, \sigma^2)$$

Ex: ARMA(1,1) モデル (2/3)

根座標系でのCubic tensor

$$T_{111} = \frac{6z}{(1-z^2)^2} \qquad T_{112} = -\left\{\frac{2w}{(1-zw)^2} + \frac{4z}{(1-zw)(1-z^2)}\right\}$$

$$T_{222} = -\frac{6w}{(1-w^2)^2} \qquad T_{122} = \frac{2z}{(1-zw)^2} + \frac{4w}{(1-zw)(1-w^2)}$$

$$|z| < 1, |w| < 1$$

Formula

$$T_{ijm} = \varepsilon_i \varepsilon_j \varepsilon_m \left\{\frac{2\theta^i}{(1-\theta^i\theta^j)(1-\theta^i\theta^m)} + \frac{2\theta^j}{(1-\theta^j\theta^m)(1-\theta^j\theta^i)} + \frac{2\theta^m}{(1-\theta^m\theta^i)(1-\theta^m\theta^j)}\right\} \quad |\theta^i| < 1$$

Ex: ARMA(1,1) モデル (3/3)

ARMA (1,1) の e-接続1形式

$$\omega^1{}_1 = \frac{-2z}{1-z^2}\mathrm{d}z + \frac{-z^2+2wz-1}{(1-wz)(w-z)}\mathrm{d}w$$

$$\omega^1{}_2 = \frac{-z^2+2wz-1}{(1-wz)(w-z)}\mathrm{d}z + \frac{-2(1-z^2)}{(1-w^2)(z-w)}\mathrm{d}w$$

$$\omega^2{}_1 = \frac{-w^2+2wz-1}{(1-wz)(w-z)}\mathrm{d}w$$

$$\omega^2{}_2 = \frac{-w^2+2wz-1}{(1-wz)(w-z)}\mathrm{d}z + \frac{2\{3w^2z^2-z^2-2w^3z-2wz+w^2+1\}}{(1-w^2)(1-wz)(w-z)}\mathrm{d}w$$

上を計算するとトレース2形式が非零であることがわかる

$$\Xi = f(w,z)\mathrm{d}w \wedge \mathrm{d}z \neq 0$$

以上の結果は, より簡潔で一般的な形で記述できる

ARMAモデル多様体のトレース2形式

定理 (FT2012)

ARMAモデル多様体の　α接続に関するトレース2形式は非零であり根座標系で表示すれば以下で与えられる

$$\Xi^{(\alpha)} = -4\alpha \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{1}{(1 - z_i w_j)^2} \mathrm{d}w_j \wedge \mathrm{d}z_i$$

$\{w_j\}, \{z_i\}$ はARMAモデルのAR,MA部分の特性多項式の根であり,
$|z_i| < 1, |w_j| < 1$

系

ARMAモデルでのパラメータの最尤推定量をプラグインしたスペクトル密度の推定量はプラグインのクラスの中でも（漸近リスクのもとで）非許容的

15. まとめと今後の課題

まとめ

★ベイズ統計での事前分布の選択は難しい問題。特に非i.i.d. モデルは理論的な研究が少ない

★幾何学的な量の計算や解析から統計モデル上のあるクラスの事前分布の存在の有無がわかる

★高次元の複雑なモデルの場合、幾何学的な量の計算で、最尤推定量や予測分布の性能に対する理論評価できるのは大きなメリット

今後の課題

★ 漸近的な結果と有限標本での結果のギャップの詳細な解析

★ 統計的な課題（実応用）・微分幾何学的な課題（優調和関数の存在と構成）の両方が山積み（個人研究の限界！）

文献 (1/2)

情報幾何のテキスト:

S. Amari:*Differential geometrical methods in statistics*. Springer-Verlag, 1985.
S. Amari and H. Nagaoka,: *Methods of Information Geometry*. AMS, Oxford, 2000.

スライド中で引用した文献 （見やすさを踏まえ省略された文献は原論文を参照）

J. Aitchison: Goodness of prediction fit. *Biometrika*, **62** (1975), 547-554.
O.L.E. Barndorff-Nielsen and G. Schou: On the parametrization of autoregressive models by partial autocorrelations. *J. Multivariate Anal.,* **3** (1973), 408-419.
J. Berger and R. Y. Yang: Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory*, **10** (1994), 461-482.
L. L. Campbell: An extended Cencov characterization of the information metric. *Proc. Amer. Math. Soc.* **vol 98** (1986), 135-141.
J. A. Hartigan: The Maximum Likelihood Prior. *Ann. Statist.*, **26 no.6** (1998), 2083-2103.
F. Komaki: Shrinkage priors for Bayesian prediction. *Ann. Statist.*, **34** (2006), 808-819.
F. Komaki: A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, **88** (2001), 859-864.
S. Lauritzen: Statistical manifolds. In *Differential Geometry in Statistical Inference*, IMS Lecture Notes: Monograph Series 10:
Institute of Mathematical Statistics, Hayward, California, (1987) 163-216.
H. Matsuzoe, J. Takeuchi, and S. Amari, Equiaffine structures on statistical manifolds and Bayesian statistics. *Differential Geom. Appl.*, **24** (2006), 567-578.

文献 (2/2)

P.C.B. Phillips: To criticize the critics: an objective Bayesian analysis of stochastic trends. *J. Appl. Econ*. **6** (1991), 333-364.

J. Takeuchi and S. Amari, α-parallel prior and its properties. *IEEE. Trans. Info. Theory*, **51, no.3** (2005), 1011-1023.

F. Tanaka, Superharmonic priors for autoregressive models. *Mathematical Engineering Technical Reports*, **2009-18**, (2009) 1-20.

F. Tanaka and F. Komaki:
Asymptotic expansion of the risk difference of the Bayesian spectral density in the autoregressive moving average model,
Sankhya Series A, Indian Statistical Institute, **Vol.73-A** (2011), pp. 162－184.

F. Tanaka: Curvature form on statistical model manifolds and its application to Bayesian analysis, *Journal of Statistics Applications and Probability*, Natural Sciences Publishing, **Vol.1** (2012), 35-43.

F. Tanaka: Noninformative prior in the quantum statistical model of pure states. *Phys. Rev. A*, **85** (2012): 062305.