非滑らかな確率密度推定の統計理論的解析

2018/11/6　IBIS2018

今泉允聡

(統計数理研究所 / 理研AIP / JSTさきがけ)

概要

- **やること**
  - データを生成する密度関数 f^* を推定する
- **知りたいこと**
  - 推定量 $\hat{f}$ の性能（誤差）

$$\|f^* - \hat{f}\| = ?$$

- **めざすもの**
  - f^* が非滑らかな場合の性能評価を可能に

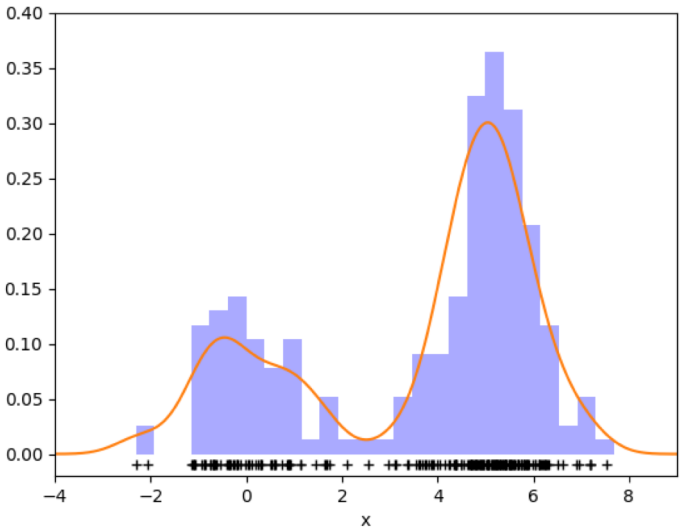
目次

確率密度関数の推定

- **設定**
  - 観測データ（i.i.d.）　$X_1, X_2, \ldots, X_n \sim F^*$
    - X_i は D-次元ベクトル

  - F^* の密度関数 f^* を推定

- **推定の例**
  - **青棒: ヒストグラム**
  - **赤線: カーネル密度推定量**



確率密度関数の推定

- **何に使うの？**
  - 密度関数による量 / 手法

$$\int \left(\log \frac{f(x)}{f'(x)} \right) f(x) dx$$

カルバック=ライブラ情報量

$$-\int f(x) \log f(x) dx$$

シャノン・エントロピ

$$\prod_{i=1}^{n} \frac{f(X_i)}{f'(X_i)}$$

密度比

  - 回帰（教師あり学習）

$$y = m(x) + \epsilon \quad \Leftrightarrow \quad y = \int y' f(y'|x) dy' + \epsilon$$

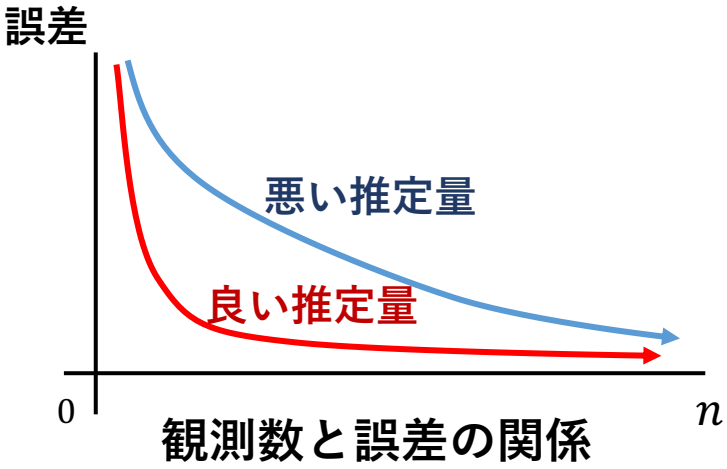
- **統計分析の基盤技術**
  - 多くの密度を推定する手法が開発されている

統計理論による解析

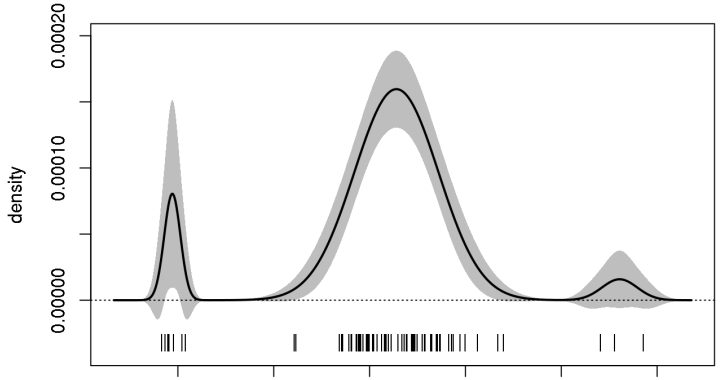
- **推定の良さを測るには**
  - データ数 n が増えた時の誤差の収束レート
  - 例: $\|f^* - \hat{f}\| = O(n^{-\alpha})$

収束レート$(-\alpha)$がわかると

- **統計的推論**
  - 信頼区間
  - 統計的検定
  - 良い正則化　　など



観測数と誤差の関係



密度推定の信頼区間の例

https://www.r-bloggers.com/

研究の出発点

- 密度関数が**滑らか** (≒ 微分可能) なら…
  - 収束レートはよく知られている

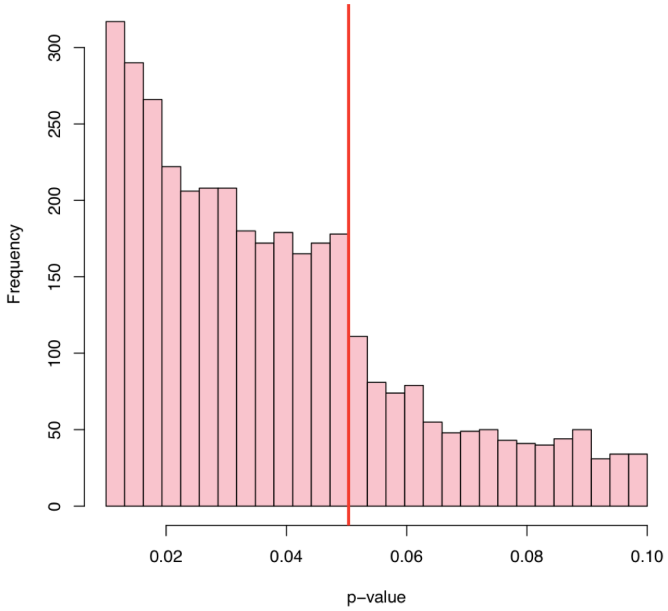
<u>カーネル推定量の例</u>
密度関数 f^* が β回連続微分可能なとき、カーネル法による推定量 $\hat{f}$ は以下を満たす：

$$\left\|\hat{f} - f^*\right\|_2^2 = O_P(n^{-2\beta/(2\beta+D)})$$

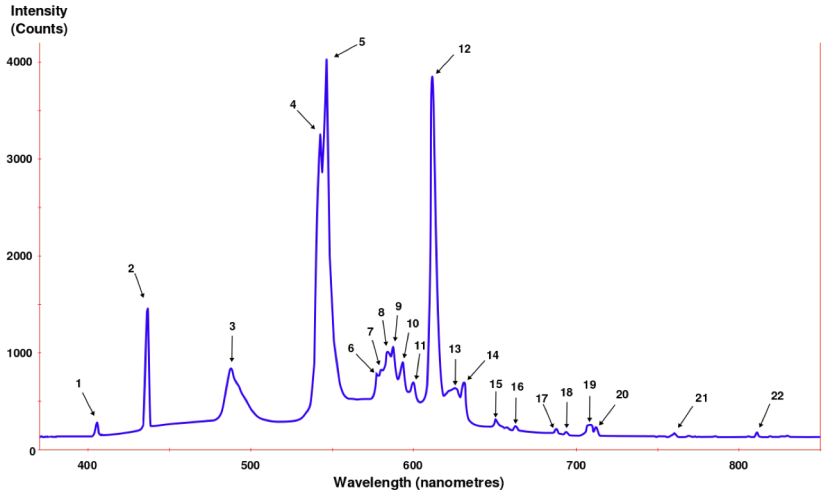
  - 同様の収束レートを与える方法
    - ヒストグラム法 / カーネル法 / ガウス過程 / シリーズ法 etc

研究の出発点

- 現実の密度関数はそんなに**滑らかではない**



論文で報告されるp値の分布
（赤線は0.05） Masicampo+ (2012)



蛍光灯のスペクトル
©Wikipedia

研究の出発点

- 微分を使わない収束評価は数学的に困難

収束レート

密度関数の性質	微分可能	連続	非連続
ヒストグラム	○	△	△
カーネル法	○		
シリーズ法	○		
ガウス過程法	○		

○：収束レートが明らか
△：一致性のみ（収束レートは不明）

本研究の目的：
収束レートが分かる、非滑らかな密度の推定量を作る

目次

Szemeredi の分割

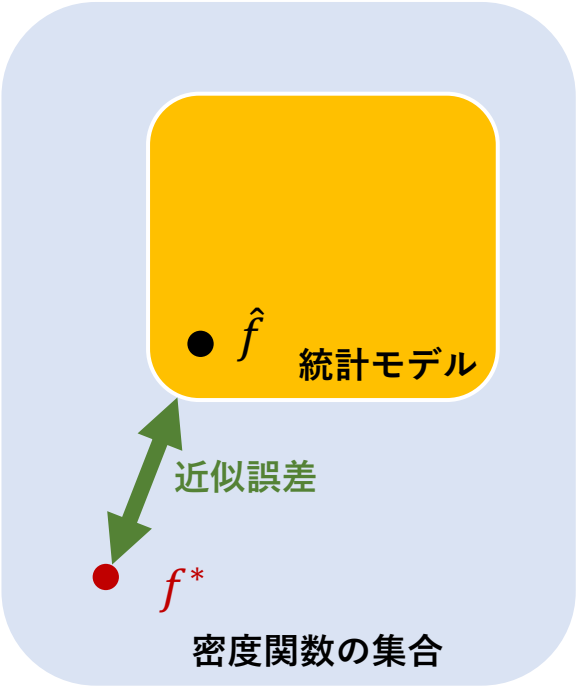
全体図

- **誤差の評価方法**

推定量の誤差

＝ 近似誤差 ＋ モデル複雑性

  - 近似誤差
    - 統計モデルが f^* を表現できるか
  - モデル複雑性
    - モデルの大きさ（過適合しやすさ）



全体図

- **誤差の評価方法**

推定量の誤差

＝ 近似誤差 ＋ モデル複雑性

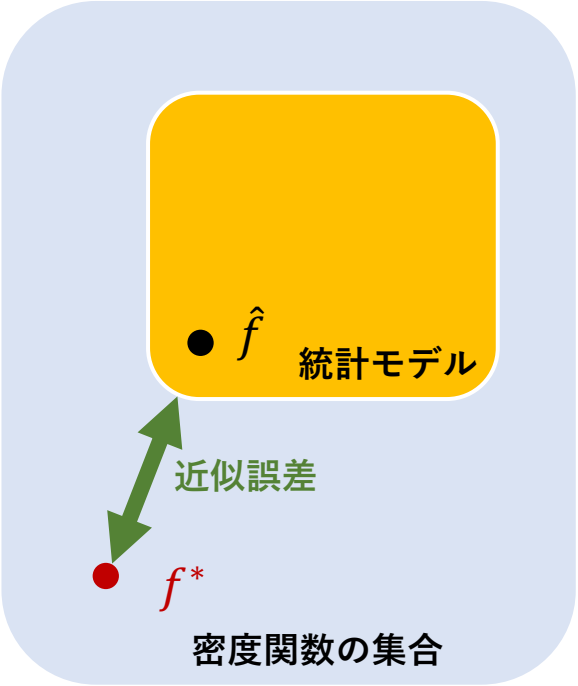
- **近似誤差**の評価が難しい

f^*が微分可能なら:

→ テイラー展開などで評価

f^*が微分できない:

→ **本研究：Szemeredi の分割理論を導入**



準備(1/3)

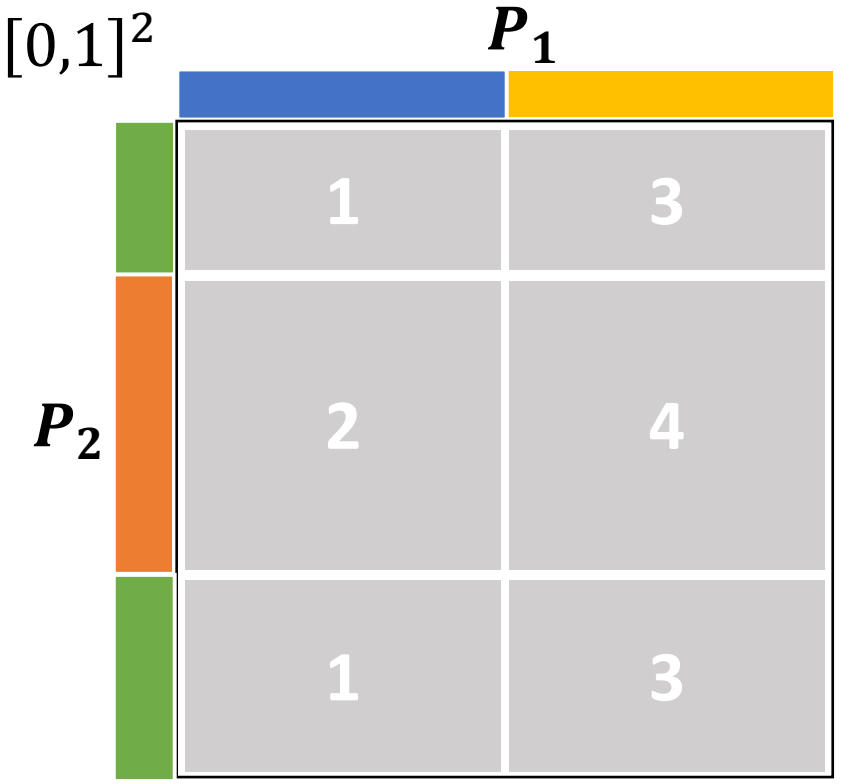
- **等分割（Equipartition）**
  - $I := [0,1]$ を3-等分割したい

  - (連結)等分割

  - (一般の)等分割
    - P: $I := [0,1]$のK-等分割

準備(2/3)

- I^D内の**セル**
  - P_d: d次元目のIの等分割
  - $\mathcal{S} = P_1 \times \cdots \times P_D$
    セルの集合
  - I^DをK^D個のセルに分割

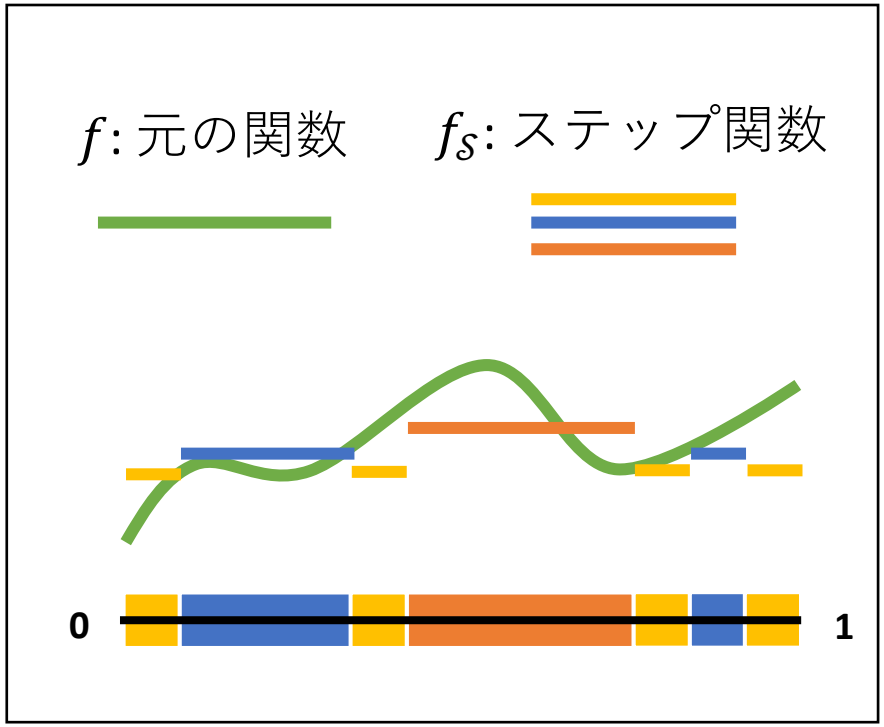


$K = 2, D = 2$の場合のセル（4個）

準備(3/3)

- **ステップ関数**
  - 関数 $f: I^D \to \mathbb{R}_+$
  - I^D内セルの集合 $\mathcal{S}$
  - f_S ： fによるステップ関数
    - $f_S(x) = \sum_{S \in \mathcal{S}} \frac{1_{\{x \in S\}} \int_S f d\lambda}{\lambda(S)}$
    - f を $\mathcal{S}$ 上で離散化

λ：ルベーグ測度



$K = 3, D = 1$の場合のf_S

Szemerediの分割理論

- 以下の補題を用いる

Weak Regularity Lemma (Szemeredi (1975))

どんな有界な $f: I^D \to \mathbb{R}_+$ にも、ある$\mathcal{S}^*$があり以下が成立：

$$\|f - f_{\mathcal{S}^*}\|_{\square} := \sup_{\{T_d \subset [0,1]\}_{d=1}^{D}} \left| \int_{T_1 \times \cdots \times T_D} (f - f_{\mathcal{S}^*}) d\lambda \right| \leq \frac{4}{\sqrt{D \log K}}$$

- どんな有界なfでも、fと$f_{\mathcal{S}^*}$の距離は$O(1/\sqrt{D \log K})$
  - D: データの次元、 K: 等分割の数（ハイパラ）
  - 誤差の下限も与えられる（別定理）

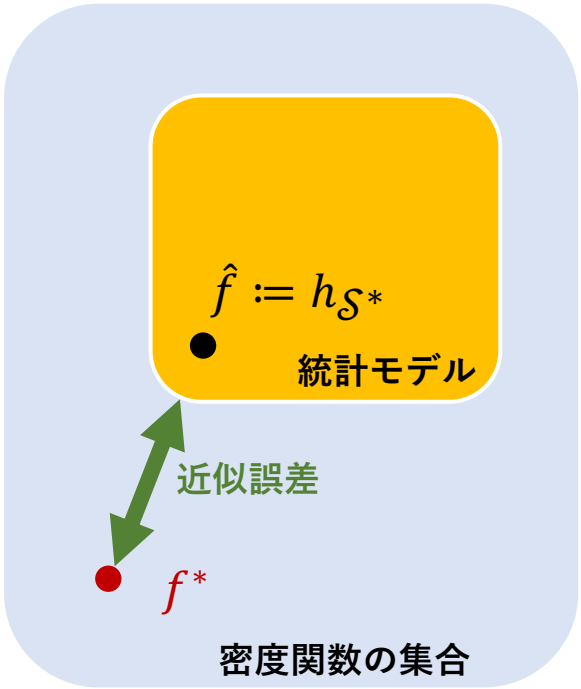
※巨大グラフ(グラフォン)の近似に用いられる理論

本研究のアイディア

- **アイディア**
  - 統計モデル：$\mathcal{S}^*$上のヒストグラム

$$h_{\mathcal{S}}(x) := \frac{\sum_{S\in\mathcal{S}} 1_S(x)\, n^{-1} \sum_{i\in[n]} 1_S(X_i)}{\sum_{S\in\mathcal{S}} 1_S(x)\lambda(S)}$$

  - 近似誤差（f^*と$h_{\mathcal{S}^*}$の距離）は先のLemmaより評価できる



- **するべきこと**
  - $\mathcal{S}^*$を得る

目次

イントロ

アイディア

方法1 (M-SDE)

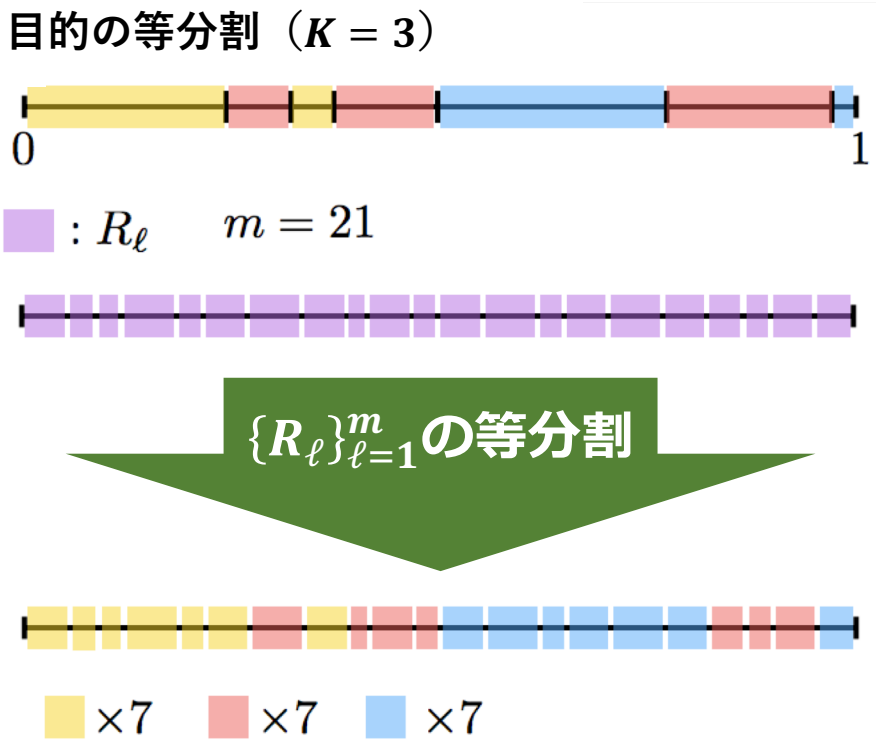
方法2 (V-SDE)

評価

ランダム分割法

提案法1 (M-SDE)

- **方針**
  - ランダムな有限分割で S^* を近似
- 概略
  - Iをm個にランダムに分割
    - $\{R_\ell\}_{\ell=1}^m$とする
  - $\{R_\ell\}_{\ell=1}^m$ の等分割で Iの等分割を近似
    - mの増加による近似レートは既知



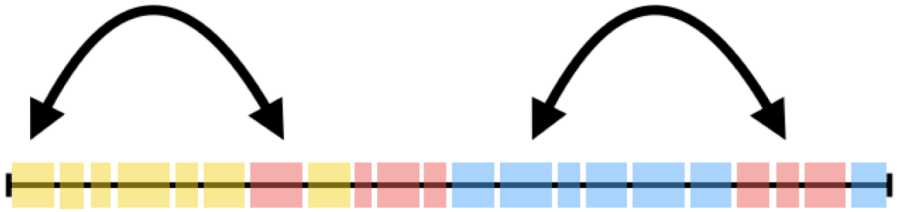
提案法1 (M-SDE)

- 任意の等分割を（漸近的に）計算できる

⇩

- **有限分割の選択**
  - CV損失を最小化
  - 探索は貪欲なスワッピング

各分割のラベルをスワップ



提案法1 (M-SDE)

- **M-SDE**
  - **M**inimized-**S**zemeredi **D**ensity **E**stimator
  - $\mathcal{S}^M$: 有限分割の近似で得られたセル

$$\hat{f}^M(x) := h_{\mathcal{S}^M}(x)$$

  - ◎ 実装がシンプル
  - ◎ 理想的には良い収束レートを達成（後述）
  - × アルゴリズム的に$\mathcal{S}^M$が$\mathcal{S}^*$に近い保証が無い
  - × 計算時間が分割数 K について指数的に増加

目次

イントロ

アイディア

方法1 (M-SDE)

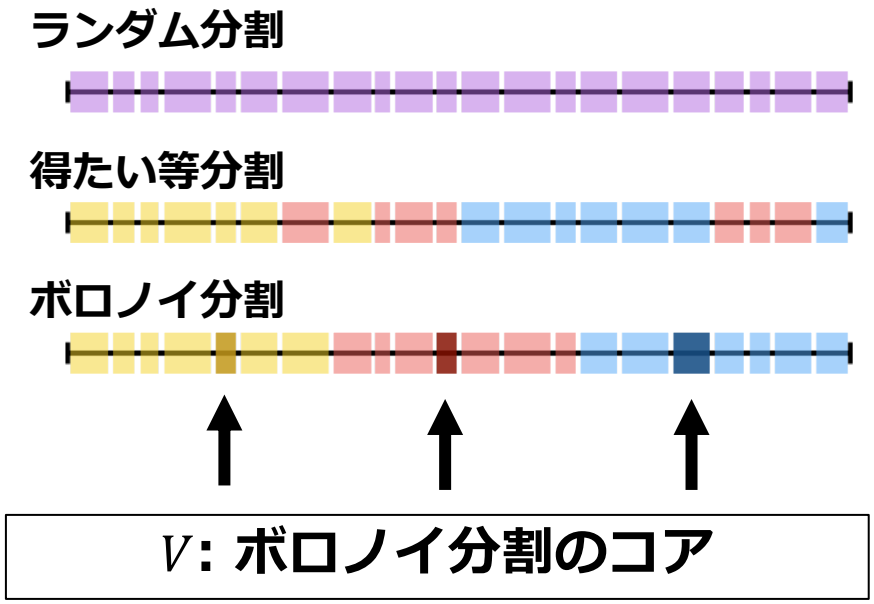
方法2 (V-SDE)

ボロノイ分割法

評価

提案法2 (V-SDE)

- **方針**
  - Iの等分割を$\{R_\ell\}_{\ell=1}^m$のボロノイ分割で近似
    - $V \subset \{R_\ell\}_{\ell=1}^m$：分割のコア
- 概略
  - $\{R_\ell\}_{\ell=1}^m$からランダムにR_ℓを選択してVに加える
  - 一定の条件を満たすまで繰り返す



提案法2 (V-SDE)

- **V-SDE**
  - **V**oronoi-**S**zemeredi **D**ensity **E**stimator
  - $\mathcal{S}^V$: ボロノイ分割で得たセル

$$\hat{f}^V(x) \coloneqq h_{\mathcal{S}^V}(x)$$

  - ◎ アルゴリズムの結果が理論的に保証される
  - ◎ 計算が早い（Kについて多項式オーダー）
  - × 全体の収束レートは悪くなる（後述）
  - × 理論は$D = 2$の場合のみ

目次

イントロ

アイディア

方法1 (M-SDE)

方法1 (V-SDE)

評価

収束レートの導出

収束レート (M-SDE)

定理 1

f^*が**有界**で$\mathcal{S}^M$が weak regularity lemma の不等式を満たすとする。この時、大きなnのもと十分大きな確率で以下が成立：

$$\left\| \hat{f}^M - f^* \right\|_1 = O\left(\frac{1}{\sqrt{D \log K}} + \frac{K^D \log n}{\sqrt{n}} \right)$$

K: 分割数

--：近似誤差　（weak regularity lemma 由来）

--：モデル複雑性　（ヒストグラム推定由来）

f^*には微分可能性を仮定していない（有界性のみ）

収束レート (M-SDE)

- 分割数 K の調整
  - **近似誤差**(Kで減少) **&** **モデル複雑性**(Kで増加)

トレードオフを調整する分割数 K

$$\frac{(K_n^*)^D \log n}{\sqrt{n}} = \Theta\left(\frac{1}{\sqrt{D \log K_n^*}}\right)$$

系 1

定理1の設定のもとで、$K = K_n^*$ とすると以下が成立：

$$\left\|\hat{f}^M - f^*\right\|_1 = \tilde{O}\left(\frac{1}{\sqrt{\log n}}\right)$$

$\tilde{O}(\cdot)$は$\log \log n$を無視した記法

収束レート (M-SDE)

- 実践的なKの選択方法（Lepski法）

$$\hat{K}_n = \min\left\{K \middle| \forall \ell > K, \left\|\hat{f}^{M,(K)} - \hat{f}^{M,(\ell)}\right\|_1 \leq \tau n^{-\frac{1}{2}} \ell^D \log n\right\}$$

$\tau \geq 4$: 定数, $\hat{f}^{M,(\ell)}$: 分割数ℓのM-SDE

定理 2

定理1の設定のもとで、$K = \hat{K}_n$ とすると以下が成立：

$$\left\|\hat{f}^M - f^*\right\|_1 = O\left(\frac{1}{\sqrt{\log n}}\right)$$

$\tilde{O}(\cdot)$は$\log\log n$を無視した記法

収束レート (V-SDE)

- V-SDEの収束レート
  - 準備：ボロノイ分割によるS^*の近似誤差

補題 （Lovasz (2012)）

$D = 2$かつ$|V| = \left(\frac{1}{\epsilon^4}\right)^{\Theta\left(\frac{1}{\epsilon^4}\right)} \log\frac{1}{\delta}$とする。この時、確率$1 - \delta$で以下が成立：

$$\left\| f_{S^V} - f_{S^*} \right\|_1 \leq 6\sqrt{\epsilon}$$

  - ϵ を調整して収束レートを調整

収束レート (V-SDE)

定理 3
f^*が**有界**とする。この時、大きなnのもと十分大きな確率で以下が成立：

$$\|\hat{f}^V - f^*\|_1 = O\left(\frac{1}{(\log n)^{1/8}}\right)$$

- ボロノイ分割の影響で、V-SDEの収束レートはM-SDEより少し悪い

理論的結果のまとめ

- **最適性**
  - M-SDEのレートは最適（ミニマックス）
    - 論文中では具体例を構成
    - メトリック・エントロピーによる証明も可

- **結果の概観**

方法	収束レート	計算コスト
M-SDE	◎：最適	▲：大きい (指数 in n)
V-SDE	▲：非最適	◎：小さい (多項式 in n)

理論的結果のまとめ

- **M/V-SDE**
  - 広い関数クラスについて、
    これまで未知だった収束レートを導出

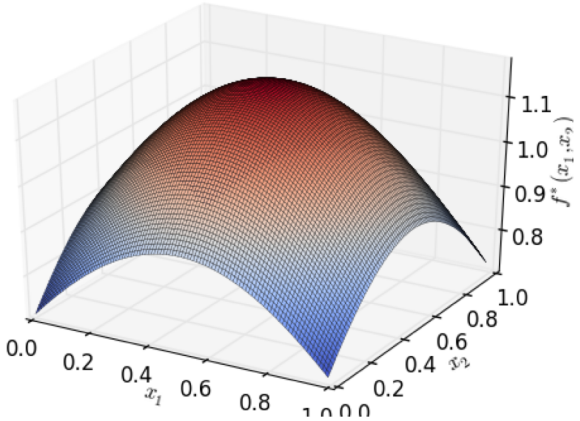
収束レート

密度関数の性質	微分可能	連続	非連続
ヒストグラム	○	△	△
カーネル法	○		
シリーズ法	○		
ガウス過程法	○		
M/V-SDE	○	○	○

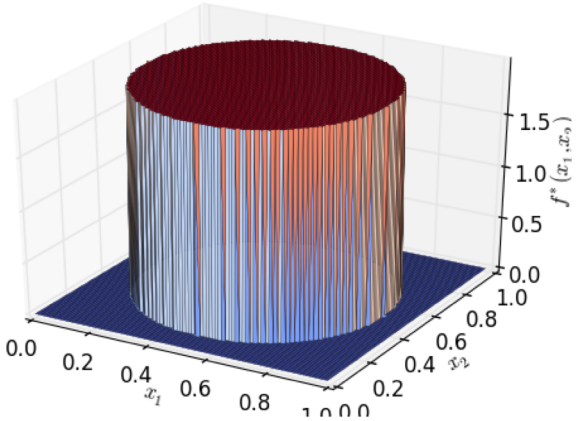
○：収束レートが明らか
△：一致性のみ（収束レートは不明）

実験（設定）

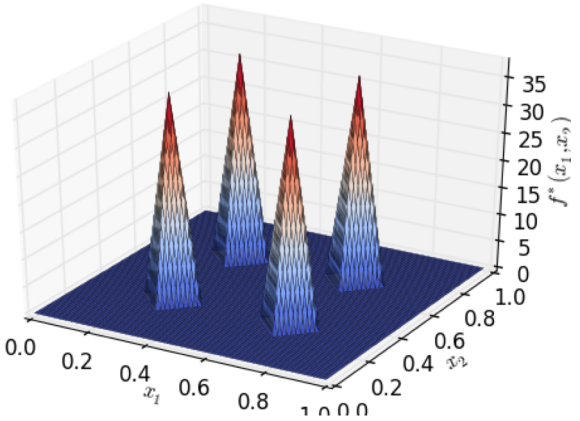
- シミュレーションによる実験
  - データを生成する密度関数 f^*



1. ガウス型



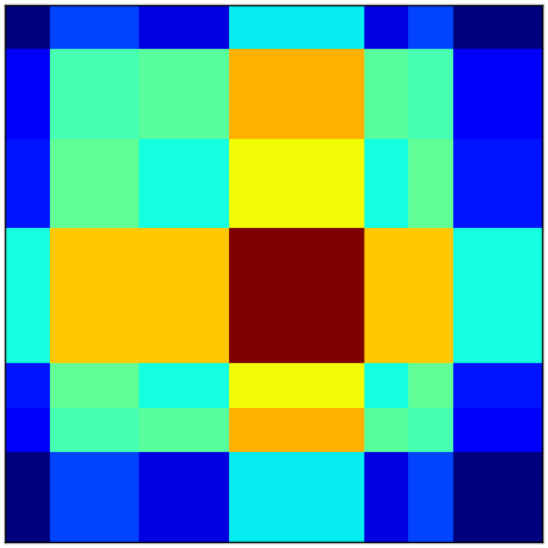
2. シリンダー型



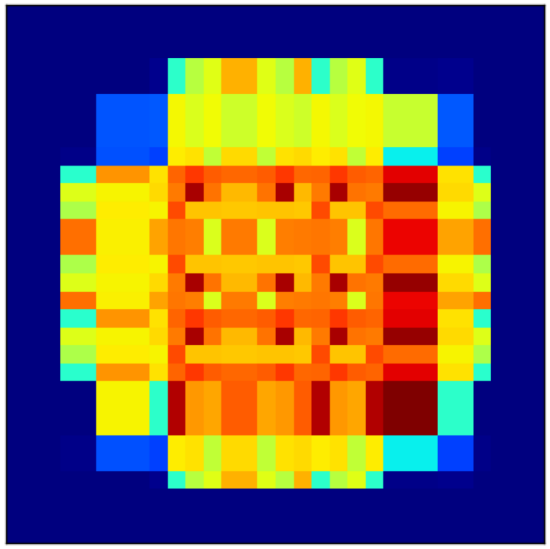
3. ピラミッド型

実験（分割）

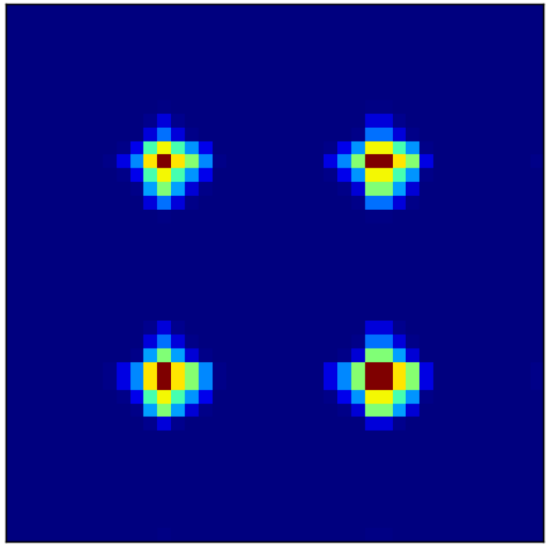
- M-SDEで得られたセル（分割）$\mathcal{S}^M$



1. ガウス型
$(K = 4)$

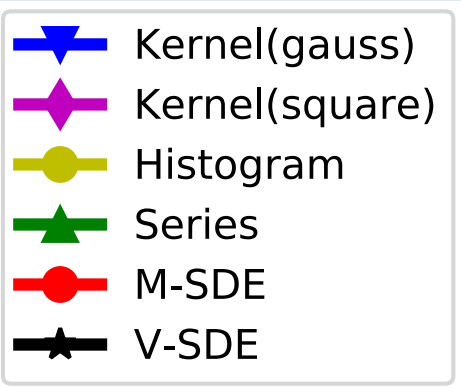


2. シリンダー型
$(K = 10)$

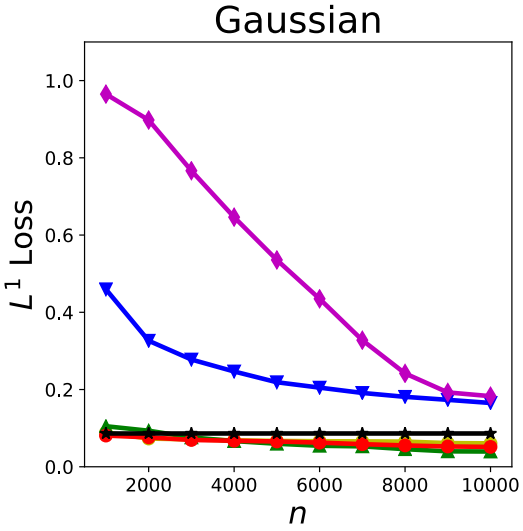


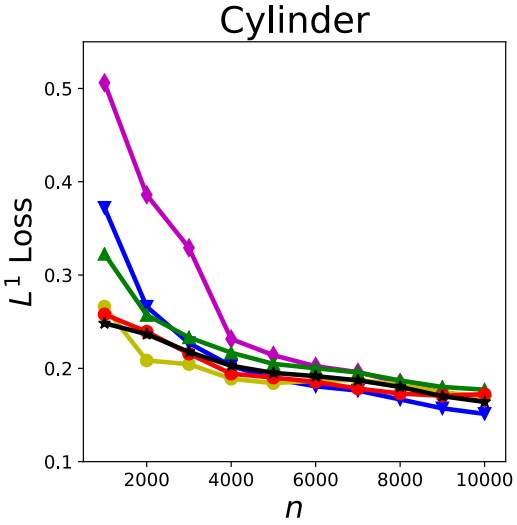
3. ピラミッド型
$(K = 13)$

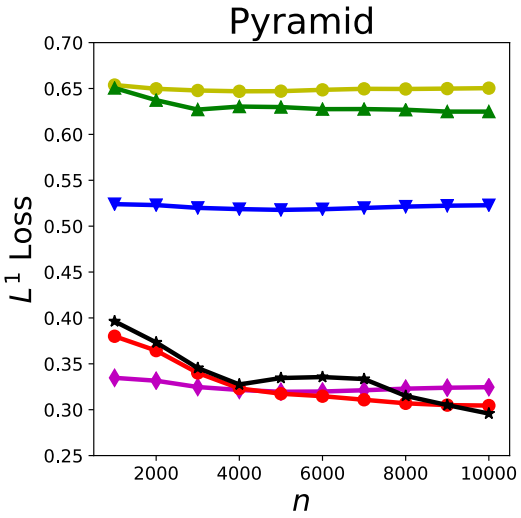
実験（誤差）



- 推定量の誤差 $\|\hat{f} - f^*\|_1$
  - 横軸: n, 縦軸: 誤差







1. ガウス型　　**2. シリンダー型**　　**3. ピラミッド型**

提案法はどの設定でもほどよく推定できる

まとめ

- **めざしたもの**
  - 微分できない密度関数の推定問題で
    推定量の良さ（収束レート）を導出する
- **やったこと**
  - 収束レートが分かる推定量（V-SDE / M-SDE）を提案
- **これから**
  - 非滑らか関数のための統計的推論の開発

ご静聴ありがとうございました

引用

- Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below. 05. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2271-2279.
- Szemerédi, Endre (1975), "On sets of integers containing no k elements in arithmetic progression", Polska Akademia Nauk. Instytut Matematyczny. Acta Arithmetica, 27: 199–245,
- Laszlo Lovasz, *Large networks and graph limits*, Vol. 60, American Mathematical Society Providence, 2012.