

博士論文

方向統計学における
ノンパラメトリック統計解析

金沢大学人間社会環境研究科
人間社会環境学専攻

学 籍 番 号	1521082009
氏 名	鶴田 靖人
主任指導 教員	寒河江 雅彦

目次

1	序論	3
1.1	背景と研究目的	3
1.2	標準的なカーネル密度推定量	4
1.3	標準的なノンパラメトリック回帰	11
1.4	方向統計学	14
1.5	概要	18
2	Di Marzio 型カーネル密度推定量 [Tsuruta and Sagae (2017a)]	20
2.1	カーネル密度推定量の定義とその理論的性質	20
2.2	フォン・ミーゼスカーネル	21
2.3	巻き込みコーシーカーネル	23
2.4	シミュレーション	24
2.5	Di Marzio 型カーネル密度推定量の課題	26
3	Hall 型カーネル密度推定量 [Tsuruta and Sagae (2017b)]	28
3.1	Hall 型カーネル族の定義	28
3.2	Hall 型カーネル密度推定量の理論的性質	29
3.3	高次オーダーカーネル密度推定量の構成法	31
3.4	シミュレーション	33
3.5	先行研究との比較	36
4	Hall 型カーネル密度推定量の平滑化パラメータ推定法 [Tsuruta and Sagae (2017c)]	37
4.1	はじめに	37
4.2	最小二乗クロスバリデーション法	37
4.3	ダイレクト・プラグイン法	39
4.4	シミュレーション	41
4.5	先行研究との比較	44
5	Di Marzio 型局所多項式回帰 [Tsuruta and Sagae (2017d)]	45
5.1	Di Marzio 型局所多項式回帰の定義とその理論的性質	45
5.2	フォン・ミーゼスカーネルと巻き込みコーシーカーネル	47
5.3	巻き込みコーシーカーネル	47
5.4	シミュレーション	48
5.5	先行研究との比較	51
6	Hall 型局所多項式回帰	52
6.1	はじめに	52
6.2	Sine 級数展開と Hall 型カーネル関数の理論的性質	52

6.3	Hall 型局所多項式回帰の理論的性質	54
6.4	シミュレーション	55
6.5	先行研究との比較	56
7	方向統計学における多変量ノンパラメトリック回帰	57
7.1	先行研究	57
7.2	実数・トーラス上のノンパラメトリック回帰とその理論的性質	60
7.3	実データ分析への応用	62
7.4	先行研究との比較	65
8	結論	66
8.1	はじめに	66
8.2	方向統計学におけるカーネル密度推定量	66
8.3	方向統計学におけるノンパラメトリック回帰	67
8.4	今後の研究の展望と課題	68
A	付録	71
A1	Appendix A	71
A2	Appendix B	72
A3	Appendix C	74
A4	Appendix D	75
A5	Appendix E	75
A6	Appendix F	76
A7	Appendix G	77
A8	Appendix H	78
A9	Appendix I	79
A10	Appendix J	81
A11	Appendix K	82
A12	Appendix L	83
A13	Appendix M	84
A14	Appendix N	84
A15	Appendix O	85
A16	Appendix P	86
A17	Appendix Q	86
A18	Appendix R	91
A19	Appendix S	92

1 序論

1.1 背景と研究目的

情報化が進んだ現在ではデータの取得や蓄積が容易になったので、大規模かつ多種多様なデータが日々生まれている。このため、社会の様々な分野でデータ分析を通して得た有益な知識を社会に活用したいというニーズが高まっている。すべてのデータは何らかのメカニズム (構造) を通して発生したものであり、データを分析する目的は、データの背後にある真の構造を解明することである。統計学では観測されたデータとすでに分かっている知識を組み合わせる統計モデルを構築する。統計学における最も適切な統計モデルとは真の構造に最も近い近似解となるモデルである。つまり、統計学とはデータを用いて最も適切な統計モデルを推定する学問であると言える。

統計手法は、統計モデルへの立場の違いからパラメトリック統計解析とノンパラメトリック統計解析の2つに分類できる。パラメトリック統計解析とは、事前に持つ知識から真の構造を含むと予想される関数形の統計モデル (パラメトリックモデル) を仮定し、そのモデルのパラメータをデータから推定する手法のことを言う。パラメトリック統計解析の長所は、そのモデルから真の構造を数学的なモデルとして説明できる点である。しかし、パラメトリック統計解析は、真の構造とは大きく異なるモデルを仮定したとき、真の構造の説明には役立たない誤った推定結果を与えてしまう。真の構造は未知であるから、間違った仮定を立ててしまう危険は常に存在する。

ノンパラメトリック統計解析は、この誤りを避けるためにパラメトリックなモデルを用いずにデータに応じて柔軟に構造が変化する統計モデル (ノンパラメトリックモデル) を用いることで真の構造が持つ特徴を推定する手法である。ノンパラメトリック統計解析の定義として異なるものがあるが、本稿ではノンパラメトリック統計解析とは統計モデルにカーネル関数と呼ばれる重み関数を用いて推定値を与えるものを指す。ノンパラメトリック統計解析は真の構造が複雑であるときや特定のパラメトリックモデルを想定できないときに有効な手法である。ただし、同じサンプルサイズの下でノンパラメトリック統計解析の理論的推定精度は、パラメトリック統計解析に比べて劣る。現在はサンプルサイズが大きなデータを容易に取得できるためにノンパラメトリック統計解析の推定精度も十分に高い。そのために、柔軟な推定量を与えるノンパラメトリック統計解析は社会の様々な分野で利用されている。

例えば、近年の計量経済学では経済現象の解析にノンパラメトリック統計解析手法がよく用いられている。その理由は、ノンパラメトリック統計解析手法は柔軟なモデリングを可能にするので、この統計解析手法を用いることで経済現象が持つ複雑な確率構造を推定できるからである。

従来の標準的な統計学は、データが実数直線 \mathbb{R} 上で定義できる確率分布に従うという前提の下で、回帰分析や検定など様々な統計手法を提案してきた。しかし、多くの研究によって現実にはすべてのデータが実数直線 \mathbb{R} 上で定義できるわけではないことが分かってきた。

その例として、風向・動物が移動する方向・時間の周期的変動などの周期性を持つデータが挙げられる。一般に周期性を持つ観測値を角度データと呼ぶ。経済分野でも財の需要量 (供給量) の変動は、1年周期・1週間周期・24時間周期など周期的な変動を持つことが多い。周期的変動をとまなうこのような経済データも角度データとみなすことができる。

角度データに関する統計学上の問題に関して、加藤 (2017) は、「角度の観測を含むデータには、統計解析をする上で大きな問題がある。それは、このようなデータを解析する上では、統計学が主に対象としている実数

値データのための解析手法をそのまま使うことができないという問題である。この問題は、角度（または円周）には周期性があり、その位相が実数の位相と異なっていることに起因している。例えば、平均や分散などの要約統計量は、角度のデータにそのまま適用すると不自然な定義となってしまう。また、実数値データのための確率分布、回帰モデル、時系列モデルなども、角度データにそのまま応用することには無理がある。」と述べている。

角度データはその周期性のために、実数直線 \mathbb{R} 上の確率分布ではなく単位円周上で定義された確率分布に従うことが知られている。加藤 (2017) が述べたような角度データに関する統計学上の問題を克服するために、円周上の確率分布に基づいた新しい統計手法がいくつか提案されてきた。このような研究成果は、Maridia (1972) という一冊の本にまとめられた。これ以来、角度データに関する統計手法を研究する学問のことを方向統計学と呼ぶようになり、方向統計学は統計学の新しい分野と見なされるようになった。

方向統計学におけるパラメトリック統計解析の理論研究は目覚ましい進歩を遂げているが、ノンパラメトリック統計解析の理論研究はほとんど進んでいない。本稿の研究目的は、方向統計学におけるノンパラメトリック統計解析の理論研究を行い、その理論的性質を明らかにすることである。ノンパラメトリック統計解析が角度データの分析においても有効であることを理論的に示すことで、角度データの分析を行う人々に新しい選択肢を提示できる。また、角度データにノンパラメトリック統計解析を適用することで、データが待つ複雑な構造を解明する手掛かりを与えるであろう。計量経済学の分野でも経済現象の周期的変動を明らかにするために、このようなノンパラメトリック統計解析手法は貢献できる。

方向統計学の発展は、従来の統計手法を角度データに適用したときに実数空間上のデータに適用した場合と比較してどのような理論的な違いがあるのかという疑問を解き、次に理論的な問題があるならばどのような工夫を用いてそれを克服するのかを考えるという形で進んできた。本稿におけるノンパラメトリック統計解析の研究も同じアプローチを採用する。

したがって、方向統計学におけるノンパラメトリック統計解析を議論する前に、1.2 節と 1.3 節で標準的な統計学におけるノンパラメトリック統計解析手法の理論的性質について説明する。次に 1.4 節で方向統計学の中で角度データがどのように扱われているかを見るために、円周上の確率分布の定義・性質について簡単に述べる。1.4 節の内容は方向統計学と標準的な統計学の違いを理解するのに役立つだろう。最後に 1.5 節では本稿の流れを説明する。

1.2 標準的なカーネル密度推定量

標本 X_1, X_2, \dots, X_n は、独立同一分布 $f(x)$ $x \in \mathbb{R}$ に従うとする。このとき、 f のカーネル密度推定量 (KDE: kernel density estimation) $\hat{f}_h(x)$ は

$$\hat{f}_h(x) := \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \quad (1.1)$$

と定義される。ただし、 h はバンド幅と呼ばれる平滑化パラメータであり、 $K_h(x - X_i) := h^{-1}K((x - X_i)/h)$ は対称なカーネル関数 (以降、カーネルと略す) である。

カーネル $K(x)$ の t 次モーメントを $\alpha_t(K) := \int_{\mathbb{R}} x^t K(x) dx$ とおく。カーネル K はその対称性のために、常に奇数次のモーメントが 0 となることに注意されたい。2 次までのモーメントの存在条件：

$$\alpha_0(K) = 1, \quad \alpha_1(K) = 0, \quad \alpha_2(K) < \infty$$

を満たすカーネル族 $K(x)$ を通常のカーネル族と呼ぶ。通常のカーネル族の代表的なものは正規密度関数 (正

規カーネル) と対称ベータ密度関数族である。正規カーネルは

$$K(x) := (2\pi)^{-1/2} e^{-x^2/2}$$

で定義され、対称ベータ密度関数族の定義は次式で与えられる。

$$K_{(m)}(x) := \frac{(2m+1)!!}{2^{m+1}m!} (1-x^2)^m I_{\{|x|\leq 1\}}(x),$$

ただし、 $I_A(x)$ は x が条件 A を満たすとき 1 の値を取り、そうでなければ 0 の値を取る定義関数、2 重階乗 $(2m+1)!!$ は $(2m+1)!! = (2m+1)(2m-1)\cdots 3\cdot 1$ である。 $m \rightarrow \infty$ のとき、 $K_{(m)}(x)$ は正規カーネルとなることが知られている。

KDE では対称なカーネルを採用することが多い。その理由の一つは、真の密度 f は未知であるのでデータからの距離だけに依じて (左右対称に) 重みを与えることは合理的な選択であるからである。むしろ、 f への仮定をできるだけ緩めて推定するというノンパラメトリック統計解析の立場から考えると、左右非対称なカーネルを採用するためには f に対して新しい仮定 (事前知識) が必要であると考えられる。 f の構造によっては非対称カーネル \tilde{K} を用いた方が良い場合がある。非対称カーネル \tilde{K} の議論は対称なカーネル K の議論の後に行うことにしたい。

KDE の推定値を評価する際の誤差基準として、平均積分二乗誤差 (MISE: mean integrated squared error) が最もよく使われる。MISE の定義は次式で与えられる。

$$\text{MISE}_f[\hat{f}_h(\cdot)] := E_f \left[\int_{\mathbb{R}} \{\hat{f}_h(x) - f(x)\}^2 dx \right] \quad (1.2)$$

$$= \int_{\mathbb{R}} \text{MSE}_f[\hat{f}_h(x)] dx, \quad (1.3)$$

ただし、 $\text{MSE}_f[\hat{f}_h(x)] := E_f[\{\hat{f}_h(x) - E[\hat{f}_h(x)]\}^2]$ は KDE の平均二乗誤差 (MSE: mean squared error) を表す。KDE のバイアスの定義を $\text{Bias}_f[\hat{f}_h(x)] := E_f[\hat{f}_h(x)] - f(x)$ とし、KDE の分散の定義を $\text{Var}_f[\hat{f}_h(x)] := E_f[\{\hat{f}_h(x) - E_f[\hat{f}_h(x)]\}^2]$ とすると、 $\text{MSE}_f[\hat{f}_h(x)]$ はバイアスの 2 乗と分散の和となる。

$$\text{MSE}_f[\hat{f}_h(x)] = \text{Bias}_f[\hat{f}_h(x)]^2 + \text{Var}_f[\hat{f}_h(x)]. \quad (1.4)$$

(1.3) と (1.4) から $\text{MISE}_f[\hat{f}_h(\cdot)]$ は次の 2 つの項に分解できる。

$$\text{MISE}_f[\hat{f}_h(\cdot)] = \int_{\mathbb{R}} \text{Bias}_f[\hat{f}_h(x)]^2 dx + \int_{\mathbb{R}} \text{Var}_f[\hat{f}_h(x)] dx \quad (1.5)$$

$$= \text{ISB} + \text{IV}, \quad (1.6)$$

ただし、 $\text{ISB} := \int_{\mathbb{R}} \text{Bias}_f[\hat{f}_h(x)]^2 dx$ 、 $\text{IV} := \int_{\mathbb{R}} \text{Var}_f[\hat{f}_h(x)] dx$ とおいた。関数 g の二乗積分を $R(g) := \int_{\mathbb{R}} g(x)^2 dx$ とおく。 f は十分に滑らかであるという仮定の下でバイアスは

$$\text{Bias}_f[\hat{f}_h(x)] = \frac{h^2}{2!} \alpha_2(K) f^{(2)}(x) + \frac{h^4}{4!} \alpha_4(K) f^{(4)}(x) + \cdots + \frac{h^p}{p!} \alpha_p(K) f^{(p)}(x) + \cdots \quad (1.7)$$

$$= \frac{h^2}{2!} \alpha_2(K) f^{(2)}(x) + o(h^2) \quad (1.8)$$

となり、分散は次式となる。

$$\text{Var}_f[\hat{f}_h(x)] = \frac{f(x)R(K)}{nh} + o((nh)^{-1}). \quad (1.9)$$

(1.8) と (1.9) より KDE の MISE は以下の定理として与えられる (Wand and Jones, 1995, pp.20–22)。

定理 1.1. 次の 2 つの仮定 :

1. f は十分に滑らかな関数である.
2. $n \rightarrow \infty$ のとき, $h(n) \rightarrow 0$ かつ $nh(n) \rightarrow \infty$ である.

の下で, MISE は

$$\text{MISE}_f[\hat{f}_h(\cdot)] := \text{AMISE}_f[\hat{f}_h(\cdot)] + o((nh)^{-1} + h^4)$$

で与えられる. ただし, 漸近的な MISE は次式に等しい.

$$\text{AMISE}_f[\hat{f}_h(\cdot)] = \frac{h^4}{4} \alpha_2(K)^2 R(f^{(2)}) + \frac{R(K)}{nh}. \quad (1.10)$$

(1.10) の右辺の第 1 項は ISB を第 2 項は IV を表す. これ以後, 漸近的な MISE の式に関して ISB と IV をこの順序で並べることを暗黙の約束とする.

ISB はバンド幅 h に対して単調に増加し, IV はバンド幅 h に対して単調に減少する. つまり, $\text{AMISE}_f[\hat{f}_h(\cdot)]$ はバンド幅 h に対して凸性を満たし, 必ず $\text{AMISE}_f[\hat{f}_h(\cdot)]$ を最小にする最適なバンド幅 h_* が存在する. そのため, 最適なバンド幅 h_* は

$$h_* = \left[\frac{R(K)}{\alpha_2(K)^2 R(f^{(2)})} \right]^{1/5} n^{-1/5} \quad (1.11)$$

となる. このとき, 最小化された MISE は $\text{MISE}_f[\hat{f}_{h_*}(\cdot)] = O(n^{-4/5})$ で与えられる. \square

KDE の定義と定理 1.1 から KDE の長所と短所を以下のように示す.

KDE の長所として以下の 2 点がある.

- KDE は, カーネルの算術平均という単純な形で定義できるのでプログラムの実装が容易である.
- KDE は, 十分に滑らかな密度関数 f への一致性が成り立つ.

KDE の短所として以下の 2 点がある.

- 最尤推定量 (MLE: maximum likelihood estimator) などのパラメトリックな推定量の収束レートは $O(n^{-1})$ であるが, KDE の MISE の収束レートは $O(n^{-4/5})$ にすぎない.
- (1.11) が示すように最適なバンド幅 h_* は密度関数 f の導関数 $f^{(2)}$ に依存しているので, h_* の推定の問題が残る.

長所として挙げた一致性の性質は f が滑らかであるという仮定さえ満たせば, f の構造がどのようなものであっても KDE は f の構造を推定可能であることを示している. f の構造が複雑なときや f に関して特定の分布を仮定できないとき, KDE は魅力的な推定量である. 言い換えれば, データの背後にある f の構造を探索することが目的であるとき, KDE は有用な推定量であると言える. 以降では上記の短所の解決策として, MISE の収束レートの改善方法とバンド幅の推定方法を議論する.

1.2.1 高次オーダーカーネル

(1.7) から, 4 次以上の高次モーメントが 0 になるようなカーネルがあれば, バイアスを改良可能な事が分かる. 高次モーメントが 0 になるようなカーネルは, 高次オーダーカーネルと呼ばれ, 次のように定義される.

定義 1.1. p 次オーダーカーネル $K_{[p]}$ は

$$\alpha_0(K_{[p]}) = 1, \quad \alpha_t(K_{[p]}) = 0, \quad 0 < t < p, \quad \alpha_p(K_{[p]}) \neq 0$$

を満たす.

p 次オーダーカーネル $K_{[p]}$ を用いれば, (1.7) よりバイアスは $\text{Bias}_f[\hat{f}_h(x)] = h^p(p!)^{-1}\alpha_p(K_{[p]})f^{(p)}(x) + o(h^p)$ となる. このとき, 分散は (1.9) で与えられる. しかし, 定義 1.1 から分かるように, 4 次以上の p 次オーダーカーネルは, 必然的に非負性を犠牲にし, 密度関数の定義を満たさないことに注意が必要である.

高次オーダーカーネルの MISE を次の定理で与える.

定理 1.2. 定理 1.1 の仮定の下で, p 次オーダーカーネルを用いると, その漸近的な MISE は

$$\text{AMISE}_f[\hat{f}_h(x)] = \frac{h^{2p}\alpha_p^2(K_{[p]})R(f^{(p)})}{(p!)^2} + \frac{R(K_{[p]})}{nh} \quad (1.12)$$

となる. (1.12) を最小化する最適なバンド幅 h_* は

$$h_* = \left[\frac{(p!)^2 R(K_{[p]})}{(2p)\alpha_p(K_{[p]})R(f^{(p)})} \right] n^{-1/(2p+1)}$$

で与えられる. このとき, 最小化された MISE は

$$\text{MISE}_f[\hat{f}_{h_*}(\cdot)] = O(n^{-2p/(2p+1)}) \quad (1.13)$$

となる. (1.13) を満たす KDE を p 次オーダー KDE と呼ぶことにする. 特に $p \geq 4$ 次オーダー KDE のことを高次オーダー KDE とする. \square

2 次オーダーカーネルから 4 次オーダーカーネルを構成する方法として, Jones and Foster (1993) が提案した加法型構成法がある.

命題 1.1. 2 次オーダーカーネル $K_{[2]}(x)$ は微分可能とする. このとき, 4 次オーダーカーネルは

$$K_{[4]}(x) = \frac{3}{2}K_{[2]}(x) + \frac{1}{2}xK'_{[2]}(x)$$

で与えられる. 加法型構成法は, 同様な方法で p 次オーダーカーネル $K_{[p]}$ から $p+2$ 次オーダーカーネル $K_{[p+2]}$ を構成可能である (詳細は省略する). \square

定理 1.2 より, 定義 1.1 によって与えられた $K_{[4]}$ からなる KDE は 4 次オーダー KDE である. ただし, 定義 1.1 から分かるように, 加法型構成法によって得られたカーネルは積分の値は 1 となるが, 非負性を犠牲にする欠点がある. 非負性を満たし, かつ, MISE を改良できる高次オーダー KDE の構成法として, Terrell and Scott (1980) が提案した乗法型構成法がある.

命題 1.2. $\hat{f}_h(x)$ は正の値を持つ KDE とする. 異なるバンド幅を持つ 2 つの KDE: $\hat{f}_h(x)$ を用いて, KDE:

$$\tilde{f}_h(x) := \hat{f}_h(x)^{4/3}\hat{f}_{2h}(x)^{-1/3}$$

を定義する. $\tilde{f}_h(x)$ のバイアスは $\text{Bias}_f[\tilde{f}_h(x)] = O(h^4)$ となり, その分散は $\text{Var}_f[\tilde{f}_h(x)] = O((nh)^{-1})$ となる. したがって, 最小化された MISE は $\text{MISE}_f[\tilde{f}_h(\cdot)] = O(n^{-8/9})$ となるので $\tilde{f}_h(x)$ は 4 次オーダー KDE である. \square

しかし、 $\hat{f}_h(x)$ は非負性を満たすが積分の値は 1 とはならず、密度関数の定義を満たさない。非負性を満たし、かつ、積分の値も 1 となるような高次オーダー KDE は存在しない。実データの分析に高次オーダー KDE を用いる際はその特徴をよく理解して使用する必要があるだろう。

1.2.2 バンド幅推定法

2 次オーダー KDE の最適なバンド幅 h_* の推定方法について議論する。 h_* の推定方法の提案はいくつかなされているが、最もよく使われる推定方法である最小二乗クロスバリデーション法 (LSCV: least squares cross-validation) とダイレクト・プラグイン法 (DPI: direct plug-in rule) を取り上げることにする。

Rudemo (1982) や Bowman (1984) によって提案された LSCV は

$$\text{MISE}_f[\hat{f}_h(x)] - \int_{\mathbb{R}} f(x)^2 dx = E_f \left[\int_{\mathbb{R}} \hat{f}_h(x)^2 dx - 2 \int_{\mathbb{R}} \hat{f}_h(x) f(x) dx \right] \quad (1.14)$$

を最小にする h を推定量として選択すればよいというアイデアから考案された。 $\hat{f}_{-i}(x) := (n-1)^{-1} \sum_{j \neq i}^n K_h(x - X_j)$ とおく。(1.14) の不偏推定量は

$$\text{CV}(h) = \int_{\mathbb{R}} \hat{f}_h(x)^2 dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{-i}(X_i)$$

として与えられる。LSCV 推定量は

$$\hat{h}_{\text{CV}} = \arg \min_{h>0} \text{CV}(h)$$

として定義される。LSCV はその定義の単純さと解釈の分かりやすさから実データ解析の際に広く用いられている。

DPI は Sheather and Jones (1991) によって提案された。DPI の基本的なアイデアは、最適なバンド幅 h_* は未知な量 $R(f^{(2)})$ に依存しているので、 $R(f^{(2)})$ を推定することで h_* の推定値を与えるというものである。 f についての汎関数を $\psi_r := \int_{\mathbb{R}} f^{(r)}(x) f(x) dx$ とおく。このとき、 $R(f^{(s)}) = (-1)^s \psi_{2s}$ という関係が成り立つ。 ψ_r に関する KDE $\hat{\psi}_r(g)$ は次式で定義される。

$$\hat{\psi}_r(g) := n^{-1} \sum_{i=1}^n \hat{f}_g^{(r)}(X_i) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n D_g^{(r)}(X_i - X_j),$$

ただし、 g と D はそれぞれバンド幅とカーネルである。 $\hat{\psi}_r(g)$ を用いて DPI 推定量 \hat{h}_{DPI} を次式のように定義する。

$$\hat{h}_{\text{DPI}} = \left[\frac{R(K)}{a_2(K)^2 \hat{\psi}_4(g)} \right]^{1/5} n^{-1/5}.$$

$\hat{\psi}_r(g)$ の MSE は次の定理として与えることができる (Wand and Jones, 1995, pp.67–70)。

定理 1.3. 次の 3 つの仮定：

1. $n \rightarrow \infty$ のとき、 $g(n) = 0$ かつ $ng^{2r+1} = \infty$ が成り立つ。
2. p 次オーダーカーネル D は r 回微分可能であり、かつ、 $(-1)^{(r+p)/2+1} D^{(r)}(0) \alpha_p(D) > 0$ を満たす。
3. 密度関数 f は十分に滑らかである。

を満たすならば、そのとき、バイアスは

$$\text{Bias}_f[\hat{\psi}_r(g)] = \text{Abias}_f[\hat{\psi}_r(g)] + O(g^{p+2})$$

となる。ただし、 $\text{Abias}_f[\hat{\psi}_r(g)]$ は

$$\text{Abias}_f[\hat{\psi}_r(g)] = n^{-1}g^{-r-1}D^r(0) + \frac{1}{p!}g^p\alpha_p(D)\psi_{r+p} \quad (1.15)$$

で与えられる。また、分散は

$$\text{Var}_f[\hat{\psi}_r(g)] = 2n^{-1}g^{-2r-1}\psi_0R(D^{(r)}) + 4n^{-1}\text{Var}_f[f^{(r)}(X)] + o(n^{-2}g^{-2r-1} + n^{-1}) \quad (1.16)$$

となる。 $\text{Abias}_f[\hat{\psi}_r(g)]\Big|_{g=g_*} = 0$ を満たす g_* は

$$g_* = \left[\frac{p!D^{(r)}(0)}{-\alpha_p(D)\psi_{r+p}} \right]^{1/(r+p+1)} n^{-1/(r+p+1)} \quad (1.17)$$

で与えられる。このとき、MSE に与えるバイアスの影響は無視することで、 $\text{MSE}_f[\hat{\psi}_r(g^*)] \simeq \text{Var}_f[\hat{\psi}_r(g^*)]$ となるので、次の等式が成り立つ。

$$\inf_{g>0} \text{MSE}_f[\hat{\psi}_r(g^*)] = \begin{cases} O(n^{-2(r+1)/(r+p+1)}), & p < r, \\ O(n^{-1}), & p \geq r. \end{cases}$$

□

(1.17) を見れば分かるように最適なバンド幅 g_* は未知な値 ψ_{r+p} を含んでおり、バンド幅 g_* を推定するために今度は ψ_{r+p} を推定する必要がある。通常は ψ_{r+p} の推定量として正規分布を仮定して導出した推定量 $\tilde{\psi}_{r+p}$ を用いて g_* を推定する。

バンド幅推定量 \hat{h} と最適なバンド幅 h_* との間の誤差基準として $\hat{h}/h_* - 1$ がよく用いられる。 \hat{h} の理論的な性能は $\hat{h}/h_* - 1$ の漸近正規分布:

$$n^\alpha(\hat{h}/h_* - 1) \xrightarrow{d} N(0, \sigma^2), \quad 0 < \alpha \leq 1/2$$

で評価可能である。ただし、分散 σ^2 はサンプルサイズ n に依存しない定数である。LSCV 推定量 \hat{h}_{CV} の漸近正規分布は次の定理として与えられる (Hall et al., 1987, Scott and Terrell, 1987)。

定理 1.4. 次の 2 つの仮定:

1. 対称なカーネル関数 K はコンパクトな台を持ち、導関数 K' はヘルダー連続である。
2. f は十分に滑らかである。

の下で、LSCV 推定量 \hat{h}_{CV} は

$$n^{1/10}(\hat{h}_{\text{CV}}/h_* - 1) \xrightarrow{d} N(0, \sigma_{\text{CV}}^2)$$

を満たす。

□

また、DPI 推定量 \hat{h}_{DPI} の漸近正規分布は次の定理として与えられる (Sheather and Jones, 1991)。

定理 1.5. 次の 3 つの仮定:

1. $n \rightarrow \infty$ のとき, $g(n) = 0$ かつ $ng^{2r+1} = \infty$ が成り立つ.
2. 2次オーダーカーネル $D \geq 0$ は4回微分可能である.
3. 密度関数 f は十分に滑らかである.

の下で, DPI 推定量 \hat{h}_{DPI} は

$$n^{5/14}(\hat{h}_{\text{DPI}}/h_* - 1) \xrightarrow{d} N(0, \sigma_{\text{DPI}}^2)$$

を満たす. □

LSCV 推定量 \hat{h}_{CV} の収束レートが $O(n^{-1/10})$ であるのに比べて, DPI 推定量 \hat{h}_{DPI} の収束レートは $O(n^{-5/14})$ であるので, \hat{h}_{DPI} は \hat{h}_{CV} よりも理論的に優れている.

1.2.3 非対称カーネル

今までの KDE の議論では, f の定義域は実数直線全体 \mathbb{R} であることを仮定していた. 実データの中には f の定義域が有限区間 $[0, c]$ (例: センター試験の得点データ) である場合がある. 対称カーネル K を採用した KDE は, f の定義域が有限区間であるとき (または f の定義域が半無限区間 $[0, \infty)$ であるとき) には, 定義域の境界付近で不一致性を持つという問題があり, これを境界問題と呼ぶ.

次に, 境界問題を説明する. 簡便化するために, f の定義域を $[0, 1]$ とする. 通常のカーネル族 K の定義域を $[-1, 1]$ とおく. ここで, $\alpha_t(K; \delta) := \int_{-1}^{\delta} x^t K(x) dx$, $R(K; \delta) := \int_{-1}^{\delta} K(x)^2 dx$ とすると, f の定義域の境界付近: $x = \delta h$, ($\delta \in [0, 1)$) では, バイアスと分散はそれぞれ,

$$\begin{aligned} \text{Bias}_f[\hat{f}_h(x)] &= f(x)\{\alpha_0(K, \delta) - 1\} + o(1) = O(1), \\ \text{Var}_f[\hat{f}_h(x)] &= \frac{1}{nh} f(x)R(K; \delta) + o(n^{-1}h^{-1}) \end{aligned}$$

となる. したがって, $\hat{f}_h(x)$ は境界付近では一貫性を持たない. $\alpha_0(K, \delta)$ の定義から分かるように, 境界問題が生じる原因は, 対称なカーネル K の裾が境界点で切れてしまうからである.

境界問題を解決するための方法として境界点で裾が切れなような非対称カーネル \check{K} を採用する方法がある. 例えば, Chen (1999) は, f の定義域が $[0, 1]$ のとき, ベータ密度関数族 (非対称のものを含む) を採用したベータ KDE :

$$\check{f}_b(x) := n^{-1} \sum_{i=1}^n \check{K}_{x/b+1, (1-x)/b+1}(X_i)$$

を提案した. ただし, ベータ密度関数族 $\check{K}_{p,q}$ は次式のように定義される.

$$\check{K}_{p,q}(t) := t^{p-1}(1-t)^{q-1}/B(p, q), \quad (1.18)$$

ここで, $B(p, q)$ はベータ関数であり, $b > 0$ はバンド幅で, $b \rightarrow 0$ かつ $nb \rightarrow \infty$ ($n \rightarrow \infty$) を満たす. 定義から明らかにベータ KDE は非負性を満たす. さらに, ベータ KDE のバイアスは

$$\text{Bias}_f[\check{f}_b(x)] = \{(1-2x)f'(x) + x(1-x)f''(x)/2\}b + o(b)$$

となる. また, 分散は次式に等しい.

$$\text{Var}_f[\check{f}_b(x)] = \begin{cases} \frac{n^{-1}b^{-1/2}}{2\sqrt{\pi x(1-x)}} \{f(x) + o(1)\}, & x/b \rightarrow \infty \text{ かつ } (1-x)/b \rightarrow \infty, \\ \frac{n^{-1}b^{-1}\Gamma(2k+1)}{2^{1+2k}\Gamma(k+1)^2} \{f(x) + o(1)\}, & x/b \rightarrow k \text{ もしくは } (1-x)/b \rightarrow k, \end{cases}$$

ただし、 $k \in [0, \infty)$ は定数である。したがって、ベータ KDE の最小化された MSE と MISE の収束レートは $O(n^{-4/5})$ となる。

また、 f の定義域が半無限区間 $[0, \infty)$ (例：所得データ) のときは、境界問題を解決するために、非対称カーネル \tilde{K} の一つであるガンマカーネルを採用したガンマ KDE が提案されている (Chen, 2000, Igarashi and Kakizawa, 2014)。

1.3 標準的なノンパラメトリック回帰

独立同一分布から得られた 2 変数標本 $(X_1, Y_1), \dots, (X_n, Y_n)$ を得たとする。目的変数 Y_i と説明変数 X_i の間の関係を推定するためによく用いられるのが回帰分析である。目的変数 Y_i は以下の仮定を満たすとする。

$$Y_i = m(X_i) + v(X_i)^{1/2}\varepsilon_i, \quad i = 1, \dots, n, \quad (1.19)$$

ただし、 $\varepsilon_1, \dots, \varepsilon_n$ は平均が 0、分散が 1 である互いに独立な確率変数、 $m(x) = E_Y[Y|X = x]$ は $x \in [0, 1]$ で台をもつ回帰関数、 $v(x) = \text{Var}[Y|X = x]$ は Y の条件付分散である。

本稿では局所多項式回帰 (LPR: local polynomial regression) と呼ばれるカーネル型ノンパラメトリック回帰を扱う。テイラー展開から局所多項式 $\beta_0 + \beta_1(\cdot - x) + \dots + \beta_p(\cdot - x)^p$ は x の近傍で $m(x)$ の近似値を取ることが分かる。これが LPR の基本となるアイデアである。LPR $\tilde{m}(x; p, h)$ は重みつき二乗誤差:

$$\sum_{i=1}^n \{Y_i - \beta_0 - \beta_1(X_i - x) - \dots - \beta_p(X_i - x)^p\}^2 K_h(X_i - x) \quad (1.20)$$

を最小にする $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T$ の切片 $\hat{\beta}_0$ として定義される。カーネル $K_h(X_i - x)$ はデータ X_i と点 x の距離 $|X_i - x|$ が長くなるほど値が小さくなる重み関数である。点 x から近い位置になる観測値 X_i は推定量 $\tilde{m}(x; p, h)$ に強い影響を与えるが、点 x から遠い観測値 X_i が $\tilde{m}(x; p, h)$ に与える影響は小さいか無視される。つまり、カーネル $K_h(X_i - x)$ は x からの距離に応じて観測値 X_i が $\tilde{m}(x; p, h)$ に与える影響の大きさを決める役割を果たす。このため、LPR はカーネルを重みとすることで柔軟な推定量を与えることができる。(1.20) と標準的な重み付き最小二乗定理から LPR $\tilde{m}(x; p, h)$ は次式で与えられる。

$$\tilde{m}(x; p, h) := \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{Y}, \quad (1.21)$$

ただし、 $\mathbf{e}_1 := (1, 0, \dots, 0)^T$ は $(p+1) \times 1$ ベクトル、

$$\mathbf{X}_x = \begin{bmatrix} 1 & (X_1 - x) & \dots & (X_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x) & \dots & (X_n - x)^p \end{bmatrix}$$

は $n \times (p+1)$ デザイン行列であり、 $\mathbf{W}_x := \text{diag}\{K_h(X_1 - x), \dots, K_h(X_n - x)\}$ は $n \times n$ 対角重み付き行列である。

例えば、 $p = 0$ のとき、 $\tilde{m}(x; 0, h)$ はナダラヤ・ワトソン推定量と呼ばれ、

$$\tilde{m}(x; 0, h) = \sum_i K_h(X_i - x) Y_i / \sum_i K_h(X_i - x) \quad (1.22)$$

として定義される。また、 $p = 1$ のときの LPR を局所線形回帰 (LLR: local linear regression) $\tilde{m}(x; 1, h)$ と呼び、これを次式のように定義する。

$$\tilde{m}(x; 1, h) := n^{-1} \sum_i \frac{\{\tilde{s}_2(x; h) - \tilde{s}_1(x; h)(X_i - x)\} K_h(X_i - x) Y_i}{\tilde{s}_2(x; h) \tilde{s}_0(x; h) - \tilde{s}_1(x; h)^2}, \quad (1.23)$$

ただし, $\tilde{s}_l(x; h) := n^{-1} \sum_i (X_i - x)^l K_h(X_i - x)$ である.

記法を簡便化するために新しい p 次オーダーカーネルを定義する. \mathbf{N}_p は (i, j) 成分が $a_{i+j+2}(K)$ である $(p+1) \times (p+1)$ 行列とする. $\mathbf{M}_p(u)$ は \mathbf{N}_p とそれぞれの成分が同じであるが第 1 列は $(1, u, \dots, u^p)^T$ で置き換えた $(p+1) \times (p+1)$ 行列とする. そのとき, カーネル関数

$$K_{(p)}(u) = \{|\mathbf{M}_p(u)|/|\mathbf{N}_p|\}K(u)$$

は, p が奇数のとき $(p+1)$ 次オーダーカーネルであり, p が偶数のとき, $(p+2)$ 次オーダーカーネルである. $K_{(p)}(u)$ の例として,

$$K_{(2)}(u) = K_{(3)}(u) = \frac{\alpha_4(K) - \alpha_2(K)u^2}{\alpha_4(K) - \alpha_2(K)^2}K(u)$$

となる.

ここで, $(X_1, \dots, X_n)^T = \mathbf{X}_n$ とおく. LPR の条件付バイアス $\text{Bias}_Y[\tilde{m}(x; p, h)|\mathbf{X}_n]$ と条件付き分散 $\text{Var}_Y[\tilde{m}(x; p, h)|\mathbf{X}_n]$ を次の定理で与える (Wand and Jones, 1995, pp.120–125).

定理 1.6. 次の仮定:

1. $h = h(n)$ とする. $h(n) \rightarrow \infty$ かつ $nh(n) \rightarrow \infty$ である.
2. 導関数 $m^{(p+2)}(x)$ と条件付分散 $v(x)$ はそれぞれ定義域 $[0, 1]$ 上で連続である.
3. 周辺密度 $f(x) > 0$ は定義域が $x \in [0, 1]$ であり, $f(x)$ 連続微分可能である.
4. カーネル K は定義域が $[-1, 1]$ であり, 点 0 で対称である.
5. n_0 は任意の定数とする. 推定する点 x は, すべての $n \geq n_0$ に対して $h < x < 1 - h$ を満たす.

の下で条件付バイアスは, 次数 p が奇数のとき

$$\text{Bias}_Y[\tilde{m}(x; p, h)|\mathbf{X}_n] = \frac{1}{(p+1)!}h^{p+1}m^{(p+1)}(x)\alpha_{p+1}(K_{(p)}) + o_p(h^{p+1})$$

となり, 次数 p が偶数のとき

$$\begin{aligned} \text{Bias}_Y[\tilde{m}(x; p, h)|\mathbf{X}_n] &= h^{p+2} \left\{ \frac{1}{(p+1)!}m^{(p+1)}(x)\alpha_{p+1}(K_{(p)})\frac{f'(x)}{f(x)} \right. \\ &\quad \left. + \frac{1}{(p+2)!}m^{(p+2)}(x)\alpha_{p+2}(K_{(p)}) \right\} + o_p(h^{p+2}) \end{aligned}$$

となる. また, 条件付分散は

$$\text{Var}_Y[\tilde{m}(x; p, h)|\mathbf{X}_n] = (nh)^{-1}\frac{R(K_{(p)})}{f(x)}v(x) + o_p((nh)^{-1})$$

となる. □

重み付条件付 MISE は, $\text{MISE}_Y[\tilde{m}(x; p, h)|\mathbf{X}_n] = \text{E}_Y \left[\int_{\mathbb{R}} \{\tilde{m}(x; p, h) - m(x)\}^2 f(x) | \mathbf{X}_n \right]$ で定義される. 本稿では, 重み付条件付 MISE はノンパラメトリック回帰でのみ使うリスクであるので, 文脈から誤解を与えないと判断したときは重み付条件付 MISE を単に MISE と表記することにする.

p が奇数のとき LPR の $\text{MISE}_Y[\tilde{m}(x; p, h)|\mathbf{X}_n]$ と最適なバンド幅は次の定理で与えられる.

系 1.1. 定理 1.6 の仮定の下で漸近的な MISE は, p が奇数ならば,

$$\text{AMISE}_Y[\tilde{m}(x; p, h) | \mathbf{X}_n] = h^{2p+2} \frac{\alpha_{p+1}(K)^2 \int_{\mathbb{R}} m^{(p+1)}(x)^2 f(x) dx}{\{(p+1)!\}^2} + (nh)^{-1} R(K_{(p)}) \int_{\mathbb{R}} v(x) dx \quad (1.24)$$

となる. (1.24) を最小にするバンド幅 h_* は

$$h_* = \left[\frac{(p+1)(p!)^2 R(K_{(p)}) \int_{\mathbb{R}} v(x) dx}{2\alpha_{p+1}(K)^2 \int_{\mathbb{R}} m^{(p+1)}(x)^2 f(x) dx} \right]^{1/(2p+3)} n^{-1/(2p+3)}$$

で与えられるので, 最小化された MISE は $\text{MISE}_Y[\tilde{m}(x; p, h) | \mathbf{X}_n] = O_p(n^{-(2p+2)/(2p+3)})$ である. p が奇数の場合と同様に考えれば, p が偶数の場合の最小化された MISE は $\text{MISE}_Y[\tilde{m}(x; p, h_*) | \mathbf{X}_n] = O_p(n^{-(2p+4)/(2p+5)})$ となるのが簡単に分かる. \square

多変量局所線形回帰について説明する. d 次元説明変数ベクトルを $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^T$ とおく. デザイン行列を

$$\mathbf{X}_x = \begin{bmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \end{bmatrix}$$

とする. バンド幅行列を $\mathbf{H} := \text{diag}\{h_1^2, h_2^2, \dots, h_d^2\}$ とおく. また, 2 次オーダーカーネル K の積からなるプロダクトカーネル

$$K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) = \frac{1}{|\mathbf{H}|^{1/2}} \prod_{j=1}^d K((X_{ij} - x_j)/h_d)$$

を重み関数として用いる. (1.21) と同様にして標準的な重み付き最小二乗定理から多変量 LLR は次式で与えられる.

$$\tilde{m}(\mathbf{x}; \mathbf{H}) = \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{Y}.$$

関数 $g(\mathbf{x})$ のヘッセ行列を $\mathcal{H}_g(\mathbf{x})$ とする. $\mathcal{H}_g(\mathbf{x})$ の (i, j) 成分を $\frac{\partial^2}{\partial x_i \partial x_j} g(\mathbf{x})$ とおく. 重み付条件付 MISE を $\text{MISE}_Y[\tilde{m}(\mathbf{x}; \mathbf{H}) | \mathbf{X}] := E_Y[\int \{\tilde{m}(\mathbf{x}; \mathbf{H}) - m(\mathbf{x})\}^2 f(\mathbf{x}) d\mathbf{x} | \mathbf{X}_1, \dots, \mathbf{X}_n]$ とする.

多変量 LLR の条件付バイアス・条件付分散・重み付条件付 MISE を次の定理で示す.

定理 1.7. 次の仮定:

1. $n \rightarrow \infty$ のとき, \mathbf{H} の各成分と $n^{-1}|\mathbf{H}|^{-1/2}$ は 0 に収束する.
2. ヘッセ行列 $\mathcal{H}_m(\mathbf{x})$ は, 連続とする.
3. 周辺密度 $f(\mathbf{x}) \geq 0$ は連続微分可能であり, 条件付分散 $v(\mathbf{x})$ は連続とする.

を満たすならば, そのとき, 条件付バイアスは

$$\text{Bias}_Y[\tilde{m}(\mathbf{x}; \mathbf{H})] = \frac{1}{2} \alpha_2(K) \text{Tr}\{\mathbf{H} \mathcal{H}_m(\mathbf{x})\} + o_p(\mathbf{H}) \quad (1.25)$$

となり, 条件付分散は

$$\text{Var}_Y[\tilde{m}(\mathbf{x}; \mathbf{H})] = n^{-1} \mathbf{H}^{-1/2} R(K) \frac{v(\mathbf{x})}{f(\mathbf{x})} + o_p(n^{-1} \mathbf{H}^{-1/2}) \quad (1.26)$$

となる。 \mathbf{I}_d を $d \times d$ 単位行列とし、 $\mathbf{H} = h\mathbf{I}_d$ とおくと、 (1.25) と (1.26) より、漸近的な MISE は次式となる。

$$\text{AMISE}_Y [\hat{m}(\mathbf{x}; \mathbf{H}) | \mathbf{X}] = \frac{h^4 \alpha_2(K)^2 \int \{\nabla^2 m(\mathbf{x})\}^2 f(\mathbf{x}) d\mathbf{x}}{4} + \frac{R(K)^d \int v(\mathbf{x}) d\mathbf{x}}{nh^d}, \quad (1.27)$$

ただし、 $\nabla^2 m(\mathbf{x}) := \sum_{j=1}^d (\partial^2 / \partial x_j^2) m(\mathbf{x})$ である。 (1.27) を最小化する最適な h_* は

$$h_* = \left[\frac{dR(K)^d \int v(\mathbf{x}) d\mathbf{x}}{\alpha_2(K)^2 \int \{\nabla^2 m(\mathbf{x})\}^2 f(\mathbf{x}) d\mathbf{x}} \right]^{1/(d+4)} n^{-1/(d+4)} \quad (1.28)$$

となる。したがって、 (1.27) と (1.28) より最小化された MISE は $\text{MISE}_Y [\hat{m}(\mathbf{x}; \mathbf{H}) | \mathbf{X}] = O_p(n^{-4/(d+4)})$ となる。 \square

定理 1.7 は、多変量 LLR の MISE は次元の数 d が上がるにつれて収束レートが遅くなることを示している。つまり、多変量 LLR は次元の呪いと呼ばれる、次元の数 d が大きくなると推定精度が落ちる性質を持つことが分かる。

1.4 方向統計学

方向や時間など周期性の性質 ($\theta = \theta + 2\pi$) を持つ観測値は角度データと呼ばれる。角度データは単位円周上に分布する点として表すことができる。方向統計学は角度データを扱う統計学の新しい分野である。角度データの例として、図 1 は同一地点で観測された風向に関するデータ (wind データ) をプロットしたものである。Wind データは統計ソフト R の circular パッケージから入手できる。図 1 から wind データは、北方向付近にモードを持つ単峰形の分布であることが推測できる。

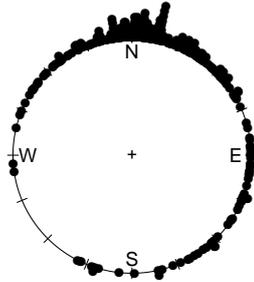


図 1: Wind データ ($n=310$) を円周上にプロットしたものの。

角度観測値を単位円周上の点とみなす理由を説明するために、wind データを通常の実数直線上で与えた 2 つのヒストグラムとして表すことにする。図 2 は wind データを北方向を原点とし時計回りに 0 から 2π までの角度で表したときのヒストグラムである。図 2 を見たとき、wind データは 0 と 2π 付近で 2 つのモードを持つ U 字型の分布であると考えよう。wind データは周期性を持つから $[2\pi - 5/8\pi, 2\pi)$ の範

図2の観測値を $[-5/8\pi, 0)$ の範囲の観測値に置き換えることができる。図3はこの置き換えをして作成した $[-5/8\pi, 2\pi - 5/8\pi)$ の範囲のヒストグラムである。図3からは wind データは 0 付近 (北方向付近) でモードを持つ単峰型の分布だと推測できるだろう。

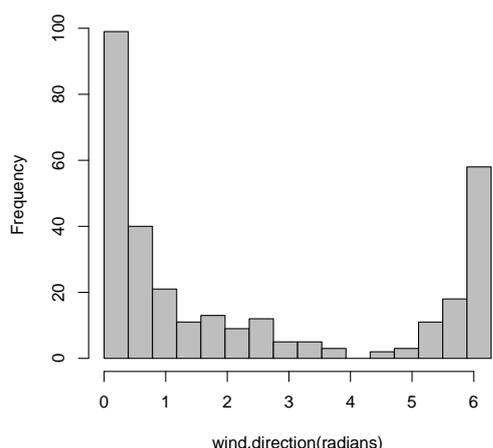


図 2: Wind データのヒストグラム。原点が北方向を意味し、観測値の方向を北から時計回りに正の値で 0 から 2π までの角度として表している。

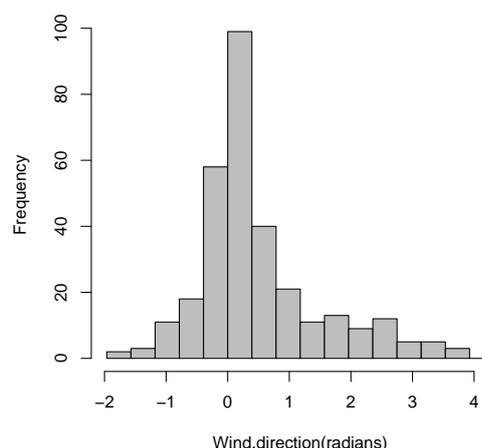


図 3: 図 2 のヒストグラムの $[2\pi - 5/8\pi, 2\pi)$ の範囲のデータを $[-5/8\pi, 0)$ の範囲の観測値に置き換えたときのヒストグラム。

図 2 と図 3 は角度データを実数直線上のデータとみなすと座標の取り方によってデータの特徴が違って見えることを示している。角度データを図 1 のように円周上にプロットすればこのような問題を避けることができるので、統計学者は、角度データを円周上のデータとして扱い、角度データを分析するための統計理論を発展させてきた。

方向統計学は角度データを単位円周上で定義される角度確率変数 (角度変数) とみなす。方向統計学の理論的基礎となる角度変数 Θ が従う密度関数 $f(\theta)$ の定義を次のように与える。

定義 1.2. 密度関数 $f(\theta)$ は次の 3 つの条件:

- 非負性: $f(\theta) \geq 0, \quad -\infty < \theta < \infty.$
- 周期性: $f(\theta + 2\pi) = f(\theta), \quad -\infty < \theta < \infty.$
- 積分の性質: $\int_0^{2\pi} f(\theta)d\theta = \int_{-\pi}^{\pi} f(\theta)d\theta = 1.$

を満たすとする。

今後、通常の密度関数と区別したいときは、定義 1.2 で与えた $f(\theta)$ を円周上の密度関数と呼ぶことにする。円周上の密度関数 $f(\theta)$ の定義は通常の密度関数に周期性の性質を加えたものになっている。便宜上、 $f(\theta)$ の定義域 \mathbb{T} を $\mathbb{T}_1 := [0, 2\pi)$ もしくは $\mathbb{T}_2 := [-\pi, \pi)$ とおくことが多い。どちらの \mathbb{T} の定義も $f(\theta)$ の性質に影響を与えないのでどちらの定義を用いても良い。本稿では \mathbb{T} の範囲は先行研究に準じて用いる。そのため、どちらの \mathbb{T} の定義を用いるかは各章ごとに異なるが、どちらの定義域も \mathbb{T} が円周上で定義されていることを意味していることに注意されたい。

角度変数 $\Theta \sim f(\theta)$ の特性関数 ϕ_p は

$$\phi_p := \mathbf{E}[e^{ip\theta}] = \int_{-\pi}^{\pi} e^{ip\theta} f(\theta) d\theta, \quad p = 0, \pm 1, \pm 2, \dots$$

として定義される.

$$\gamma_p := \mathbf{E}[\cos(p\theta)] = \int_{-\pi}^{\pi} \cos(p\theta) f(\theta) d\theta$$

と

$$\beta_p := \mathbf{E}[\sin(p\theta)] = \int_{-\pi}^{\pi} \sin(p\theta) f(\theta) d\theta$$

を用いれば, ϕ_p は

$$\phi_p = \gamma_p + i\beta_p$$

となることを示せる. 逆転公式から $f(\theta)$ は次式のようなフーリエ級数で与えることができる.

$$\begin{aligned} f(\theta) &= \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \phi_p e^{ip\theta} \\ &= \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \left(\gamma_p \cos(p\theta) + \beta_p \sin(p\theta) \right) \right\}. \end{aligned} \quad (1.29)$$

方向統計学では γ_p と β_p を三角モーメントと呼んでいる. 方向統計学では通常の p 次モーメント $\mathbf{E}[\Theta^p]$ の代わりに三角モーメントをモーメントとして用いる. 例えば, 角度データの散らばりの尺度としてよく用いられる円周上の分散 ν は平均方向が μ のとき,

$$\nu := 1 - \mathbf{E}[\cos(\theta - \mu)], \quad 0 \leq \nu < 1$$

である.

角度データでよく用いられるフォン・ミーゼス (VM: von Mises) 分布と巻き込みコーシー (WC: Wrapped Cauchy) 分布の密度関数を紹介する.

VM 分布の密度関数の定義は

$$f_{\text{VM}}(\theta; \mu, \kappa) := \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad 0 < \kappa < \infty \quad (1.30)$$

である. ただし, μ は平均方向を κ は分布の集中度を調節する集中度パラメータを表し, $I_p(\kappa)$ は p 次第 1 種修正ベッセル関数である. VM 分布の三角モーメントは $\gamma_p = I_p(\kappa)/I_0(\kappa)$ となる.

WC 分布の密度関数の定義は

$$f_{\text{WC}}(\theta; \mu, \rho) := \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}, \quad 0 < \rho < 1 \quad (1.31)$$

である. ただし, μ は平均方向を, ρ は集中度パラメータである. また, WC 分布の三角モーメントは $\gamma_p = \rho^p$ で与えられる.

VM 分布と WC 分布の密度関数を図 4 で示す. 図 4 から VM 分布は緩やかな形をしていて正規分布に似ているが, WC 分布は裾が厚い形状となっていてコーシー分布に似た特徴を持つことが分かる.

VM 分布と WC 分布の特徴を述べる. まず, VM 分布の長所と短所を示す.

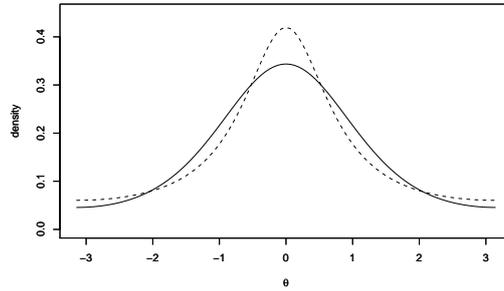


図 4: フォン・ミーゼス (VM) 分布と巻き込みコーシー (WC) 分布の密度関数. 実線は VM 分布の密度関数を表し, 破線は WC 分布の密度関数を表す. 2 つの分布の円周上の分散はともに $\nu = 0.55$ である.

- 長所 :
 1. VM 分布は $\kappa \rightarrow \infty$ のとき, 正規分布に分布収束する.
 2. VM 分布は指数型分布族に属し, MLE は明示的な解を持つ.
- 短所 :
 1. VM 分布の基準化定数は特殊関数である ($I_0(\kappa)$).
 2. VM 分布は再生性を満たさない.

VM 分布は角度データの解析の際に中心的な役割を果たし, 円周上の正規分布とも呼ばれている. しかし, 正規分布のような再生性を持たず, 正規分布と異なる特徴を持つことに注意が必要である.

次に, WC 分布の長所と短所を示す.

- 長所 :
 1. WC 分布の基準化定数は 2π である (特殊関数ではない).
 2. WC 分布は再生性を満たす.
- 短所 :
 1. 一般的に, WC 分布の MLE は明示的な解を持たない.

WC 分布は再生性を満たすために最近になって WC 分布に関する統計理論の研究が盛んに行われている.

実データ分析の際に角度変数 Θ_i と実数直線上の変数 Y_i からなるランダム 2 変量標本 $\{(Y_1, \Theta_1), \dots, (Y_n, \Theta_n)\}$ をしばしば目にする可能性がある. このようなデータの例として, 図 5 の風向 (Θ) と風速 (Y) に関するデータや図 6 の 1 月から 12 月までの月 (Θ) と月間の肺疾患による死亡者数 (Y) に関するデータを挙げる. このような 2 変量標本から明らかにしたいことは, 目的変数 Y と説明変数 Θ の関係性であろう.

Y と Θ の関係性を推定するための最も一般的な推定方法は回帰分析である. 方向統計学における代表的な回帰モデルは Johnson and Wehrly (1978) の回帰モデルである. 彼らの回帰モデルは, Y の条件付分布として正規分布を仮定した次式で与えられる.

$$Y_i = \beta_0 + \beta_1 \cos(\Theta_i) + \beta_2 \sin(\Theta_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2). \quad (1.32)$$

彼らの研究の直感的な説明を与える. $\cos(\cdot)$ と $\sin(\cdot)$ は直交系なので, $\cos(\cdot)$ と $\sin(\cdot)$ を別々な 2 つの実数直線上の変数と見なせば, (1.33) は通常の線形回帰モデルと同じ形であることに気づく. したがって, (1.33)

の回帰係数は通常の最小二乗法によって推定できる。同様の理由により、(1.33) の回帰モデルは、説明変数に $X_i(x_i \in \mathbb{R})$ を加えた

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \cos(\Theta_i) + \beta_3 \sin(\Theta_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad (1.33)$$

に拡張できる。

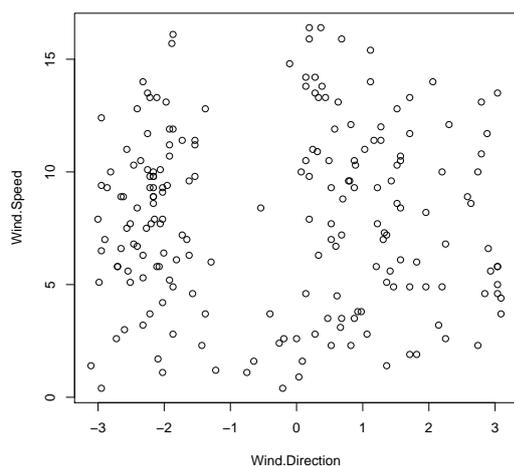


図 5: 風向と風速の散布図。X 軸は風向を表し、Y 軸は風速を表している。

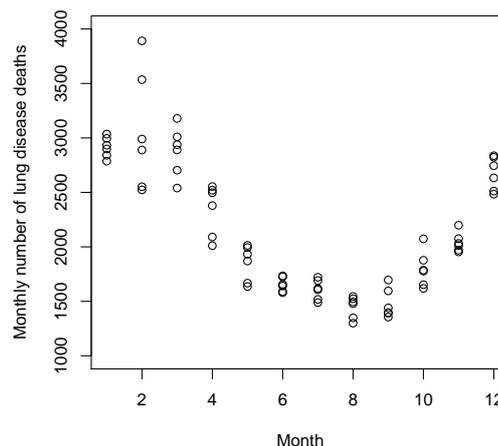


図 6: 月毎の肺疾患の死亡者数の散布図。X 軸は 1 月から 12 月までの月を表し、Y 軸は各月の肺疾患の死亡者数を表している。各観測値は 1974 年から 1979 年までの間に観測された。

1.5 概要

2 章から 4 章までで方向統計学における KDE の理論的性質を議論する。2 章では Di Marzio et al.(2011) が与えた Di Marzio 型 KDE に、VM カーネルと WC カーネルを適用したときの両者の理論的性質を比較する。また、彼らの KDE と標準的な KDE を比較することで彼らの KDE の課題を明らかにする。

3 章では Hall et al.(1987) が与えた Hall 型カーネル族を KDE に適用した Hall 型 KDE を提案し、その理論的性質を明らかにすることを目指す。特に MISE の改良方法について議論する。また、4 章では Hall 型 KDE に関する平滑化パラメータ推定量の理論的性質について議論をする。3 章と 4 章の最後の節で、これらの結果をもとに Hall 型 KDE・Di Marzio 型 KDE・標準的な KDE の理論的特徴を比較する。

5 章から 7 章までは方向統計学における LPR の理論的性質を議論することが目的である。5 章では Di Marzio et al.(2009) が与えた Di Marzio 型 LPR の理論的性質を議論するために、VM カーネルと WC カーネルを Di Marzio 型 LPR に適用した場合の理論的性質を明らかにする。最後に Di Marzio 型 LPR と標準的な LRR の理論的特徴を比較している。

6 章では Hall 型カーネル族を適用した Hall 型 LPR を提案する。Di Marzio 型 LPR は多項式の次数が $p > 1$ の場合の理論的性質はよくわかっていないが、Hall 型 LPR は多項式の次数を上げることで MISE を改良できることを示す。そして、Hall 型 LPR・Di Marzio 型 LPR・標準的な LRR の理論的性質を比較する。

7章では方向統計学における多変量 LLR の理論的性質について議論する. 説明変数が実数・トーラス $\mathbb{R}^d \times \mathbb{T}^q$ 上のベクトルであるときの Hall 型 LLR を提案する. 実数・トーラス $\mathbb{R}^d \times \mathbb{T}^q$ は一般的な表現であるので, 先行研究で提案された多変量 LLR の定義を含むという特徴がある. この提案する多変量 Hall 型 LLR に関する理論的性質を導出し, 方向統計学における多変量 LLR に関する一般的な性質について議論する.

8章では2章から7章までの研究成果をまとめることで, 本稿が方向統計学におけるノンパラメトリック統計解析の理論構築に果たした役割を議論したい. そして, 最後に今後の研究の展望と課題について簡単に触れる. なお, 各章の定理や補題の証明は付録を参照されたい.

2 Di Marzio 型カーネル密度推定量 [Tsuruta and Sagae (2017a)]

2章では Di Marzio et al. (2011) が与えた Di Marzio 型カーネル密度推定量の漸近的性質について説明する。また, Tsuruta and Sagae (2017a) が議論した VM カーネルと WC カーネルの理論的性質の違いについて述べることにする。

2.1 カーネル密度推定量の定義とその理論的性質

Di Marzio et al. (2011) に従って, この節では Di Marzio 型カーネル関数の定義と Di Marzio 型 KDE の漸近的な MISE について述べる。

定義 2.1. 関数 $K_\kappa(\theta)$ を円周上のカーネル関数とする。ただし, 集中度パラメータ κ は通常のバンド幅の逆数: $\kappa = h^{-2}$ を満たす平滑化パラメータである。カーネル $K_\kappa(\theta) : [0, 2\pi) \rightarrow \mathbb{R}$ は次の定義を満たす。

(m.1) $K_\kappa(\theta)$ はフーリエ級数展開可能である。つまり,

$$K_\kappa(\theta) = (2\pi)^{-1} \left\{ 1 + \sum_{j=1}^{\infty} \gamma_j(\kappa) \cos(j\theta) \right\}, \quad \theta \in [0, 2\pi)$$

を満たす。ただし, 三角モーメント (フーリエ係数) $\gamma_j(\kappa) := E_f[\cos(j\theta)]$ は κ の単調関数である。

(m.2) $\int_0^{2\pi} K_\kappa(\theta) d\theta = 1$ を満たす。また, もし, $K_\kappa(\theta)$ が負の値をとれば, そのとき任意の $\kappa > 0$ について $\int_0^{2\pi} |K_\kappa(\theta)| d\theta \leq M$ を満たす有界な値 $0 < M < \infty$ が存在する。

(m.3) $0 < \delta < \pi$ について $\lim_{\kappa \rightarrow \infty} \int_\delta^{2\pi-\delta} |K_\kappa(\theta)| d\theta = 0$ が成り立つ。

□

定義 2.1 で与えた Di Marzio 型カーネルは VM 分布や WC 分布といった多くの対称な分布を含む広いクラスであるという特徴を持つ。Sine 型 j 次モーメントを $\eta_j(K_\kappa) := \int_0^{2\pi} \sin^j(\theta) K_\kappa(\theta) d\theta$ とおく。そのとき, $j = 2s (s = 1, 2, \dots)$ について, $\eta_j(K_\kappa)$ は次式のような三角モーメント $\gamma_j(\kappa)$ の線形結合で表せる:

$$\eta_j(K_\kappa) = \frac{1}{2^{2s-1}} \left[\binom{j-1}{j/2} + \sum_{s=1}^{j/2} (-1)^{j+s} \binom{j}{j/2+s} \gamma_{2s}(\kappa) \right].$$

特に, $j = 2$ のとき

$$\eta_2(K_\kappa) = \{1 - \gamma_2(\kappa)\}/2 \tag{2.1}$$

となる。

定義 2.2. カーネル K_κ が sine p 次オーダーカーネルであるとは, K_κ が

$$\eta_0(K_\kappa) = 1, \quad \eta_j(K_\kappa) = 0, \quad 0 < j < p, \quad \eta_p(K_\kappa) \neq 0$$

であることを言う。

□

定義 2.2 から, 通常の高次オーダーカーネルと同様に sine 2 次オーダーカーネルは円周上の密度関数であり, $p > 2$ 以上の sine p 次オーダーカーネルは非負性を犠牲にしなければならないことが分かる。

定義 2.3. 標本 $\Theta_1, \dots, \Theta_n$ は未知な独立同一分布 $f(\theta)$ に従うとする。このとき、円周上のカーネル密度推定量は

$$\hat{f}(\theta; \kappa) := \frac{1}{n} \sum_{i=1}^n K_\kappa(\theta - \Theta_i)$$

として定義される (ただし, 3 章以降では $\hat{f}(\theta; \kappa)$ を $\hat{f}_\kappa(\theta)$ と略記する)。 □

円周上のカーネル関数はその周期性から境界点を持たないので, 定義 2.3 で与えられる円周上の KDE は, 有限区間の KDE に生じる境界問題が起きない長所を持っている。

パーセヴァルの等式から $R(K_\kappa) = (1 + 2 \sum_{j=1}^{\infty} \gamma_j^2(\kappa)) / (2\pi)$ が成り立つ。Di marzio et al. (2009) は sine 2 次オーダーカーネルの MISE を次の定理として与えた。

定理 2.1. 次の 3 つの仮定：

1. $n \rightarrow \infty$ のとき, κ は単調に増加し, かつ, それぞれの $j \in \mathbb{Z}^+$ について $\lim_{n \rightarrow \infty} \gamma_j(\kappa) = 1$ となる。
2. $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{\infty} \gamma_j^2(\kappa) = 0$ となる。
3. 導関数 f'' は連続でかつ 2 乗積分可能である。

を満たすならば, バイアスは

$$\text{Bias}[\hat{f}(\theta; \kappa)] = \eta_2(K_\kappa) f^{(2)}(\theta) / 2$$

となり, 分散は

$$\text{Var}[\hat{f}(\theta; \kappa)] = R(K_\kappa) f(\theta) / n$$

となる。sine 2 次オーダーカーネルの漸近的な MISE は次式で与えられる。

$$\text{AMISE}[\hat{f}(\theta; \kappa)] = \frac{\eta_2^2(K_\kappa)}{4} R(f'') + \frac{R(K_\kappa)}{n}. \quad (2.2)$$

(2.2) の右辺の第一項はバイアスの 2 乗積分であり, 第 2 項は分散の積分である。 □

(2.2) を最小にする最適な κ の値と最小化された MISE の収束レートは, 「 $\eta_2^2(K_\kappa)$ 」と「 $R(K_\kappa)$ 」をそれぞれ, 「 κ^α と任意の定数 $C(K)$ 」の 2 つの形に分離する必要がある。しかし, この分離を可能とするカーネルの条件は分かっていない。そのため, 特定のカーネルに関して「 $\eta_2^2(K_\kappa)$ 」と「 $R(K_\kappa)$ 」を計算する必要がある。そこで, 2 章では VM カーネルと WC カーネルの 2 つを適用した場合の KDE の理論的性質を議論する。

2.2 フォン・ミーゼスカーネル

Di Marzio et al. (2011) が与えた VM カーネルに関する理論的性質について述べる。VM カーネルは (1.30) で与えられる。

$$K_\kappa(\theta) := \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos \theta\}, \quad 0 < \kappa < \infty.$$

VM カーネルの三角モーメントは

$$\gamma_j(\kappa) = I_j(\kappa) / I_0(\kappa)$$

である。VM カーネルの sine 型 2 次モーメント $\eta_2(K_\kappa)$ と $R(K_\kappa)$ を次の補題として与える。

補題 2.1. Mardia and Jupp (1972) の (3.5.37) と (2.1) から VM カーネルの sine 型 2 次モーメントは次式となる.

$$\eta_2(K_\kappa) = \frac{I_1(\kappa)}{\kappa I_0(\kappa)}. \quad (2.3)$$

もし, κ が十分大きければ, そのとき, Mardia and Jupp (1972) の (3.5.34) と (2.3) より, sine 型 2 次モーメント $\eta_2(K_\kappa)$ は次式に等しい.

$$\eta_2(K_\kappa) = \frac{1}{\kappa} \{1 + o_p(1)\}.$$

Mardia and Jupp (1972) の (3.5.27) より, $R(K_\kappa)$ は

$$R(K_\kappa) = \frac{I_0(2\kappa)}{2\pi I_0(\kappa)^2}, \quad (2.4)$$

で与えられる. もし, κ が十分大きければ, そのとき, Mardia and Jupp (1972, 3.5.33) と (2.4) より $R(K_\kappa)$ は次式で近似できる.

$$R(K_\kappa) \simeq \kappa^{1/2}/(2\pi^{1/2}).$$

□

Di marzio et al. (2011) は定理 2.1 と補題 2.1 から次の定理を導いた.

定理 2.2. $n \rightarrow \infty$ のとき, $\kappa \rightarrow \infty$ かつ $n^{-1}\kappa^{1/2} \rightarrow 0$ が成り立つならば, そのとき, 定理 2.1 の仮定の下でバイアスは

$$\text{Bias}[\hat{f}(\theta; \kappa)] = f''(\theta)/(2\kappa)$$

となり, 分散は

$$\text{Var}[\hat{f}(\theta; \kappa)] = \kappa^{1/2}f(\theta)/(2n\pi^{1/2})$$

となる. また, 漸近的な MISE は

$$\text{AMISE}_{\text{VM}}[\hat{f}(\theta; \kappa)] = \frac{1}{4\kappa^2} R(f'') + \frac{\kappa^{1/2}}{2n\pi^{1/2}}. \quad (2.5)$$

で与えられる. (2.5) を最小化する最適な κ_* は

$$\kappa_* = \left(2\pi^{1/2}R(f'')n\right)^{2/5} \quad (2.6)$$

である. このとき, 最小化された MISE の収束レートは $\text{MISE}_{\text{VM}}[\hat{f}(\theta; \kappa_*)] = O(n^{-4/5})$ となる. □

VM カーネルの MISE の収束レートは $O(n^{-4/5})$ となるが, これは通常の 2 次オーダーカーネルの MISE の収束レートに等しい.

2.3 巻き込みコーシーカーネル

Tsuruta and Sagae (2017a) が与えた WC カーネルの理論的性質について議論する。WC カーネルは (1.31) で与えられる。

$$K_\rho(\theta) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta)}, \quad 0 < \rho < 1.$$

WC カーネルの三角モーメントは

$$\gamma_j(\rho) = \rho^j$$

である。したがって、WC カーネルの特性関数は

$$\phi_p = \rho^{|p|}$$

となる。WC カーネルの sine 型 2 次モーメント $\eta_2(K_\rho)$ と $R(K_\rho)$ を次の補題として与える。

補題 2.2. (2.1) から WC カーネルの sine 型 2 次モーメントは

$$\eta_2(K_\rho) = (1 - \rho^2)/2,$$

となる。パーセヴァルの公式： $R(K_\rho) = (2\pi)^{-1} \{1 + 2 \sum_{j=1}^{\infty} \gamma_j(\rho)^2\}$ から、 $R(K_\rho)$ は

$$\begin{aligned} R(K_\rho) &= \frac{1}{\pi(1 - \rho^2)} - \frac{1}{2\pi} \\ &= \frac{1}{\pi(1 - \rho^2)} \{1 + o_p(1)\} \end{aligned}$$

で近似できる。 □

ここで、 $h = (1 - \rho^2)$ とおく。ただし、 $0 < h < 1$ である。そのとき、定理 2.1 と補題 2.2 より次の定理が導かれる。

定理 2.3. $n \rightarrow \infty$ のとき、 $h \rightarrow 0$ かつ $nh \rightarrow \infty$ が成り立つならば、そのとき、定理 2.1 の仮定の下でバイアスは

$$\text{Bias}[\hat{f}(\theta; h)] = h^2 f''(\theta)/4$$

となり、分散は

$$\text{Var}[\hat{f}(\theta; h)] = f(\theta)/(nh\pi)$$

となる。また、漸近的な MISE は

$$\text{AMISE}_{\text{WC}}[\hat{f}(\cdot; h)] = \frac{h^2 R(f'')}{16} + \frac{1}{n\pi h}, \quad (2.7)$$

与えられる。(2.7) を最小にする最適な h_* は

$$h^* = \left(\frac{8}{\pi R(f'') n} \right)^{1/3}, \quad n > 8(\pi R(f''))^{-1} \quad (2.8)$$

となる。このとき、最小化された MISE は $\text{MISE}_{\text{WC}}[\hat{f}(\cdot; h_*)] = O(n^{-2/3})$ である。 □

数値実験の中で h_* はしばしば制約条件 $h < 1$ を超えて $h_* \geq 1$ の値を取ったので、実データ分析の際は次のような最適な ρ_* を選ぶべきである。

$$\rho_* := \arg \min_{0 < \rho < 1} \left\{ \text{AMISE}_{\text{WC}}[\hat{f}(\cdot; \rho)] \right\}, \quad (2.9)$$

ただし、 $\text{AMISE}_{\text{WC}}[\hat{f}(\cdot; \rho)]$ は (2.7) の h を $1 - \rho^2$ で置き換えた

$$\text{AMISE}_{\text{WC}}[\hat{f}(\theta; \rho)] = \frac{\{1 - \rho^2\}^2 R(f'')}{16} + \frac{1}{n\pi(1 - \rho^2)} \quad (2.10)$$

である。定理 2.3 と標準的な中心極限定理から WC カーネルを適用した KDE は漸近正規性を満たすことを示せる。

定理 2.4. c を任意の定数とし、 $h = cn^{-\alpha}$ とおく。ただし、 $0 < h < 1$ とする。もし、 $\alpha > 1/3$ かつ $n \rightarrow \infty$ を満たすならば、そのとき、

$$\sqrt{nh}[\hat{f}_h(\theta) - f(\theta)] \xrightarrow{d} N(0, f(\theta)/\pi), \quad n \rightarrow \infty \quad (2.11)$$

が成り立つ。 □

定理 2.4 の証明の詳細に関しては Appendix A を参照すること。

WC カーネルと VM カーネルはともに 2 次 sine オーダーカーネルであるが、両者の MISE の収束レートは異なる。例えば、WC カーネルは MISE の収束レートが $O(n^{-2/3})$ である一方、VM カーネルは MISE の収束レートが $O(n^{-4/5})$ である。この収束レートの不一致は標準的な KDE とは異なる特徴である。

WC カーネルの収束レートが $O(n^{-2/3})$ となる理由を考えたい。Davis (1975) は標準的な KDE に関してコーシーカーネルは MISE の収束レートが $O(n^{-2/3})$ となることを明らかにした。コーシーカーネルは分散を持たないために 0 次オーダーカーネルである。WC カーネルとコーシーカーネルの MISE の収束レートが一致する理由は $\rho = e^{-a}$ とおけば WC カーネルの特性関数 $\phi_p = e^{-a|p|}$ とコーシーカーネルの特性関数 $\phi(p) = e^{-a|p|}$ が一致するからである (Mardia and Jupp, 1999, (3,5.59)).

2.4 シミュレーション

最適な集中度パラメータ (κ_*, ρ_*) は未知な汎関数 $R(f'')$ に依存しているので、何らかの方法で $R(f'')$ を推定する必要がある。密度 f は VM 分布 $f_{\text{VM}}(\theta; \tau)$ であるという仮定の下で $R(f'')$ の推定量を与えることで最適な集中度パラメータを推定する方法をプラグイン法と呼ぶことにする。 $\hat{\tau}$ は VM 分布の集中度パラメータの MLE とする。このとき、 $R(f'')$ の推定量 $\hat{R}(f''_{\text{VM}}(\cdot; \hat{\tau}))$ は次式で与えられる。

$$\hat{R}(f''_{\text{VM}}(\cdot; \hat{\tau})) = \frac{\hat{\tau}[3\hat{\tau}I_2(2\hat{\tau}) + 2I_1(2\hat{\tau})]}{8\pi I_0^2(\hat{\tau})}. \quad (2.12)$$

(2.9), (2.10) と (2.12) より WC カーネルのプラグイン推定量 $\hat{\rho}$ は

$$\hat{\rho} = \arg \min_{0 < \rho < 1} \left\{ \frac{\{1 - \rho^2\}^2 \hat{R}(f''_{\text{VM}}(\cdot; \hat{\tau}))}{16} + \frac{1}{n\pi(1 - \rho^2)} \right\}$$

で与えられる。(2.6) の $R(f'')$ を (2.12) で推定することで VM カーネルのプラグイン推定量 $\hat{\kappa}$ は

$$\hat{\kappa} = \left(2\pi^{1/2} \hat{R}(f''_{\text{VM}}(\cdot; \hat{\tau})) n \right)^{2/5}$$

で与えられる.

前節で議論したように, VM カーネルと WC カーネルの理論上の性能を比較すると VM カーネルの性能はより優れている. しかし, 有限標本下でも同様な結果を得られると限らないので, 数値実験によって両者の性能を比較する. 推定量 $\hat{\rho}$ と $\hat{\kappa}$ は VM 分布を参照分布として仮定しているために, VN 分布の下で VM カーネルの性能の方が良いことは理論的性質から自明であろうから, 数値実験では VM 分布以外の分布 (混合 VM 分布と WC 分布) を真の分布として採用ことで, VM 分布と WC 分布の性能の安定性を調べる.

実施した実験 2.1 と実験 2.2 の手順を示す.

実験 2.1. (混合 VM 分布)

1. WC カーネルの場合:

(a) 真の混合 VM 分布を

$$f_{\text{MVM}}(\theta; \mu_1, \tau_1, \mu_2, \tau_2, p) = pf_{\text{VM}}(\theta; \mu_1, \tau_1) + (1-p)f_{\text{VM}}(\theta; \mu_2, \tau_2),$$

とする. $f_{\text{MVM}}(\theta; \mu_1 = \pi/2, \tau_1 = \tau, \mu_2 = 3\pi/2, \tau_2 = \tau, p = 0.5)$ からサンプルサイズ n の標本を生成する.

(b) 推定量 $\hat{\rho}$ を与える.

(c) 積分二乗誤差 (ISE:integrated squared error) を $\text{ISE} = \int_0^{2\pi} \{\hat{f}(\theta; \hat{\rho}) - f(\theta)\}^2 d\theta$ とおく. ISE の数値積分: $\overline{\text{ISE}}(\hat{f}(\cdot; \rho))$ を計算する.

(d) (a)–(c) を 1000 回繰り返し, $\overline{\text{MISE}}(\hat{f}(\cdot; \rho)) = \sum_{i=1}^{1000} \overline{\text{ISE}}_i(\hat{f}(\cdot; \rho))/1000$ を求める.

2. VM カーネルの場合: WC カーネルと同様に (a)–(d) の手順を行う. そして, $\overline{\text{MISE}}(\hat{f}(\cdot; \hat{\kappa}))$ を計算する.

3. $\overline{\text{MISE}}(\hat{f}(\cdot; \hat{\kappa})) - \overline{\text{MISE}}(\hat{f}(\cdot; \hat{\rho}))$ を計算する. ただし, 小数点第 4 位を四捨五入する.

実験 2.2. (WC 分布)

1. 真の WC 分布を

$$f_{\text{wc}}(\theta) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \pi)}$$

とする.

2. 実験 2.1 と同じ手順を行い, $\overline{\text{MISE}}(\hat{f}(\cdot; \hat{\kappa})) - \overline{\text{MISE}}(\hat{f}(\cdot; \hat{\rho}))$ を計算する.

表 1 は, 真の分布が混合 VM 分布であるとき, $\tau \leq 1$ かつ $n \leq 100$ であるならば, WC カーネルの性能は VM カーネルよりも優れていることを示している. n が増加するにつれて両者の差は小さくなることが分かる. 表 2 から真の分布が WC 分布であるとき, $\rho \geq 0.8$ であれば WC カーネルは VM カーネルよりも優れた性能を持つことが分かる. つまり, 真の分布が多峰性を持つかまたは真の分布の裾が厚い (山の頂上付近のカーブが急) ならば, WC カーネルは VM カーネルよりも優れた性質を持つ.

表 1: 実験 2.1 の結果. 各セルの値は $\overline{\text{MISE}}(\hat{f}(\cdot; \hat{\kappa})) - \overline{\text{MISE}}(\hat{f}(\cdot; \hat{\rho}))$ である. n はサンプルサイズを, τ はサンプリングした混合 VM 分布の集中度パラメータを表す.

	$n = 10$	50	100	200	500	1000
$\tau = 0.3$	0.011	0.002	0.001	0	0	0
0.5	0.012	0.002	0.001	0	0	0
0.7	0.011	0.002	0.001	0	0	0
1	0.012	0.001	0	0	0	0
2	0.007	-0.004	-0.005	-0.005	-0.005	-0.005
5	-0.017	-0.023	-0.025	-0.024	-0.024	-0.023
10	-0.032	-0.036	-0.037	-0.037	-0.037	-0.037
15	-0.037	-0.044	-0.043	-0.042	-0.042	-0.042
20	-0.039	-0.045	-0.045	-0.045	-0.048	-0.046

表 2: 実験 2.2 の結果. 各セルの値は $\overline{\text{MISE}}(\hat{f}(\cdot; \hat{\kappa})) - \overline{\text{MISE}}(\hat{f}(\cdot; \hat{\rho}))$ である. n はサンプルサイズを, ρ はサンプリングした巻き込みコーシー分布の集中度パラメータを表す.

	$n = 50$	100	200	500	1000
$\rho = 0.5$	-0.015	-0.009	-0.006	-0.003	-0.002
0.7	-0.011	-0.006	-0.003	-0.001	-0.001
0.8	0.007	0.012	0.013	0.011	0.009
0.9	0.133	0.141	0.134	0.112	0.093

2.5 Di Marzio 型カーネル密度推定量の課題

表 3: 標準的な KDE と Di Marzio 型 KDE の性質の比較. \circ : 当てはまる, \times : 当てはまらない, \triangle : VM カーネルのみ当てはまる.

KDE	標準的な KDE	Di Marzio 型
標本空間	実数直線: \mathbb{R}	円周: \mathbb{T}
p 次オーダーカーネル \Rightarrow $\text{MISE} = O(n^{-2p/(2p+1)})$	\circ	\times
2 次オーダーカーネルから高次オーダー KDE を構成	\circ	\triangle
平滑化パラメータ推定量の収束レートを導出済み	\circ	\times

5 節では表 3 でまとめた標準的な KDE と Di Marzio 型 KDE の特徴の違いに関する議論を行うことで Di Marzio 型 KDE の課題を整理する. Di Marzio 型 KDE の最大の問題点は, sine p 次オーダーカーネルの次数 p は MISE の収束レートに対応しておらず, その MISE の収束レートが $O(n^{-2p/(2p+1)})$ を満たさないことである. 例えば, 本章で示したように, sine 2 次オーダーカーネルに関して VM カーネルの MISE の収束レートは $O(n^{-4/5})$ であるが, WC カーネルの MISE の収束レートは $O(n^{-2/3})$ に過ぎない. つまり, sine 2 次オー

ダーカーネルに関して一般的な MISE の収束レートを与えることはできない。また、Di Marzio et al.(2011) は「 $\text{sine } p$ 次オーダーカーネルはバイアスを改良できるとは限らない。」と説明している。つまり、高次の $\text{sine } p$ 次オーダーカーネルの中に MISE を改良できないものが含まれている。

Di Marzio et al.(2011) は VM カーネルに Twiecing という方法を適用することで 4 次以上の高次オーダー KDE を構築できることを示しているが、その他の $\text{sine } 2$ 次オーダーカーネルから高次オーダー KDE を構築できるかどうかについては明らかにしていない。また、Di Marzio 型 KDE は集中度パラメータの推定量に関する収束レートの導出にも成功していない。筆者は Di Marzio 型 KDE がこれらの問題点を持つのは、 sine 型モーメントがバイアスの改良には無関係な量であるからではないかと考える。

3 Hall 型カーネル密度推定量 [Tsuruta and Sagae (2017b)]

2.5 節で Di Marzio 型 KDE の課題を議論したが、VM カーネルは、「MISE の収束レートが $O(n^{-4/5})$ である」や「2 次オーダーカーネルから高次オーダーカーネルを構築できる」といった標準的な KDE が持つ良い性質を持つ。そこで、3 章では VM カーネルを含む別のカーネル族として、Hall et al. (1987) が提案した Hall 型カーネル族を採用する。

Hall et al. (1987) は超球面上の KDE に Hall 型カーネル族を適用することで MISE の収束レートの導出に成功している。彼らの研究は非負カーネル (2 次オーダーカーネル) を採用したときの MISE に関して議論したが、高次オーダーカーネルを用いて MISE を改良することには成功していない。これは、彼らは超球面上の多変量テイラー展開に関して 2 次までの近似式を与えているが、 $t > 2$ 次の近似式を導出できていないためである。また、超球面上の KDE における集中度パラメータ推定法の理論的性能について明らかにした研究は、筆者の知る限り存在しない。

ここでは、この高次のテイラー展開に関する問題を回避するために、1 次元球面である円周上の KDE を対象を絞ったうえで高次オーダーカーネルの議論を行う。Tsuruta and Sagae (2017b) が与えた Hall 型カーネル族に関する新しい p 次オーダーカーネルを導入することで、Di Marzio 型 KDE の欠点を克服した Hall 型 KDE を提案する。また、Tsuruta and Sagae (2017b) の結果に基づいて Hall 型 KDE の MISE の改良方法について議論する。Hall 型 KDE に関する平滑化パラメータ推定量の収束レートについての議論は次章で行う。

3.1 Hall 型カーネル族の定義

Tsuruta and Sagae (2017b) が与えた Hall 型カーネル族と新しいモーメントおよび、それらを用いた p 次オーダーカーネルの定義について説明する。

Hall 型カーネル族 $K_\kappa(\theta) : [-\pi, \pi] \rightarrow \mathbb{R}$ は次の定義に従う。

定義 3.1. Hall 型カーネル族 $K_\kappa(\theta)$ を $K_\kappa(\theta) := C_\kappa^{-1}(L)L_\kappa(\theta)$ とする。ただし、関数 $L_\kappa(\theta) := L(\kappa\{1 - \cos(\theta)\})$ 、基準化定数 $C_\kappa(L) := \int_{-\pi}^{\pi} L_\kappa(\theta)d\theta$ である。

$r := \kappa\{1 - \cos(\theta)\}$ とおくと、関数 $L(r)$ は次の 2 つの条件を満たす。

(h.1) 関数 $L(r)$ は連続的微分可能である。

(h.2) r が十分大きいとき、任意の正の偶数 $v \geq 0$ について $L(r) = o(r^{-(v+1)})$ が成り立つ。

□

Hall 型カーネル族は VM カーネルを含むが WC カーネルを含まないカーネル族であり、Di Marzio 型カーネル族よりも狭いクラスである。

次に Hall 型カーネル族が持つ性質を述べる。正の偶数 $l \geq 0$ について、Hall 型カーネル族の l 次モーメントを

$$\mu_l(L) := \int_0^\infty L(r)r^{(l-1)/2}dr$$

と定義する。また、 $\mu_{\kappa,l}(L) := \int_0^\kappa L(r)r^{(l-1)/2}dr$ とおく。条件 (h.1) と (h.2) より次の性質が与えられる。

性質 3.1. 関数 L に関する次の性質が成り立つ。

- (i) $0 \leq l \leq v$ について、モーメント $\mu_l(L)$ は有界であり、かつ、 $\mu_l(L) = \mu_{\kappa,l}(L) + o(\kappa^{-(v+3)/2})$ となる。
- (ii) KDE の分散の定数項 $\delta_{2t}(L) := \int_{-\infty}^{\infty} L^2(z^2/2)z^{2t} dz$ に関して、 $t = 0, 1$ のとき、 $\delta_{2t}(L)$ は有界である。
- (iii) 部分積分 $\int_0^{\infty} L'(r)r^{(v+1)/2} dr = -(j+1)\mu_v(L)/2$ が成り立つ。

□

Hall 型カーネル族に関する高次オーダーカーネルを次のように定義する。

定義 3.2. $p \geq 2$ を正の偶数とおく。このとき、もし、 $K_{\kappa}(\theta)$ が $v \geq p+2$ 、かつ、

$$\mu_0(L) \neq 0, \quad \mu_l(L) = 0, \quad l = 2, 4, \dots, p-2, \quad \mu_l(L) \neq 0, \quad l = p$$

を満たすならば、 $K_{\kappa}(\theta)$ を p 次オーダーカーネルという。

□

定義 3.2 から明らかに 4 次以上の高次オーダーカーネルは非負性を犠牲にしていることが分かる。

3.2 Hall 型カーネル密度推定量の理論的性質

定義 2.3 で与えた KDE に関して Hall 型カーネル族を適用したものを Hall 型 KDE と呼ぶことにする。前節で定義した p 次オーダーカーネルを用いると、Hall 型 KDE に関する次のような漸近的性質が導かれる。

定理 3.1. 次のような緩やかな仮定：

1. $\kappa = \kappa(n)$ とおくと、 $\lim_{n \rightarrow \infty} \kappa(n) = \infty$ である。
2. $\lim_{n \rightarrow \infty} n^{-1}\kappa^{1/2}(n) = 0$ である。
3. 密度 f が $p+2$ 回微分可能であり、かつ、 $s = 1, 2, \dots, p$ に関して、 $f^{(s)}$ が二乗積分可能である。
4. カーネル K_{κ} は p 次オーダーカーネルである。

を満たすならば、そのとき、Hall 型 KDE の MISE は次式で与えられる。

$$\text{MISE}_f[\hat{f}_{\kappa}] = \frac{\mu_p^2(L)}{\mu_0^2(L)} R \left(\sum_{t=1}^{p/2} \frac{b_{p,2t} f^{(2t)}}{2t!} \right) \kappa^{-p} + n^{-1} \kappa^{1/2} d(L) + o(\kappa^{-p} + n^{-1} \kappa^{1/2}), \quad (3.1)$$

ただし、定数項 $d(L)$ と $b_{p,2t}$ はそれぞれ次のように与えられる。

$$d(L) = 2^{-1} \mu_0^{-2}(L) \delta_0(L),$$

かつ、

$$b_{p,2t} := 2^{1/2} \sum_{q=t}^{p/2} A_q(p, t) (\{p/2 - q\}!)^{-1} g_{2q}^{(p/2-q)}(0),$$

ただし、

$$g_j(r/\kappa) := \{2 - r/\kappa\}^{(j-1)/2}, \quad j = 0, 1, 2, \dots, \\ a_s := (2s - 2)!! / \{(2s - 1)!!\} s, \quad s = 0, 1, 2, \dots,$$

と

$$A_q(z, t) := \sum_{\sum_{s=1}^{z/2} t_s = t, \sum_{s=1}^{z/2} s t_s = q} \frac{t!}{t_1! t_2! \cdots t_{z/2}!} \prod_{l=1}^{z/2} a_l^{t_l}$$

である。(3.1)の右辺にある最初の2つの項を $\text{AMISE}[\hat{f}_\kappa]$ とする。 $\text{AMISE}[\hat{f}_\kappa]$ を最小にする κ^* は

$$\kappa^* = \left[\frac{2p\mu_p^2(L)R(\sum_{t=1}^{p/2} [b_{p,2t}f^{(2t)}/(2t!)]n)}{\mu_0^2(L)d(L)} \right]^{2/(2p+1)} \quad (3.2)$$

となる。(3.1)と(3.2)から、 p 次オーダーカーネルを用いたHall型KDEの最適なMISEの収束レートは $O(n^{-2p/(2p+1)})$ である。

また、(3.2)と $b_{2,2} = 2$ より2次オーダーカーネルを用いたときの漸近的なMISEは

$$\text{AMISE}_f[\hat{f}_\kappa] = \frac{\mu_2^2(L)}{\mu_0^2(L)} R(f'')\kappa^{-2} + n^{-1}\kappa^{1/2}d(L) \quad (3.3)$$

となる。(3.3)を最小にする集中度パラメータ κ_* は次式である。

$$\kappa_* = \beta(L)R(f^{(2)})^{2/5}n^{2/5} \quad (3.4)$$

ただし、 $\beta(L) := [4\mu_2^2(L)/\{\mu_0^2(L)d(L)\}]^{2/5}$ である。□

定理3.1の証明の概略を与える(定理3.1の証明に関する便利な補題をAppendix Bに置いた)。

定理3.1の証明. 最初にHall型KDEのバイアスを導出する。まず、 $\alpha_j(K_\kappa) := \int_{-\pi}^{\pi} K_\kappa(\theta)\theta^j d\theta$ とおく。補題B.2から正の偶数 $z \geq 0$ が $2t \leq z \leq v$ の範囲であるとき、 $\alpha_{2t}(K_\kappa)$ は

$$\alpha_{2t}(K_\kappa) = 2C_\kappa^{-1}(L)\kappa^{-1/2} \sum_{q=t}^{z/2} \sum_{m=0}^{z/2-q} \kappa^{-(q+m)} A_q(z,t)(m!)^{-1} g_{2q}^{(m)}(0)\mu_{2(q+m)}(L) + O(\kappa^{-(z+2)/2}) \quad (3.5)$$

となる。(3.5)を見ると、 p 次オーダーカーネルを用いれば低次の項が消えて $\alpha_{2t}(K_\kappa)$ のオーダーを上げることができることが分かる。 $0 \leq 2t \leq p+2$ について p 次オーダーカーネルの $\alpha_{2t}(K_\kappa)$ は

$$\alpha_{2t}(K_\kappa) = b_{p,2t}\mu_0^{-1}(L)\mu_p(L)\kappa^{-p/2} + O(\kappa^{-(p+2)/2}), \quad 0 \leq 2t \leq p, \quad (3.6)$$

かつ、

$$\alpha_{p+2}(K_\kappa) = O(\kappa^{-(p+2)/2}) \quad (3.7)$$

で与えられる。

(3.6)と(3.7)から期待値 $E_f[\hat{f}_\kappa(\theta)]$ は次式に等しい。

$$\begin{aligned} E_f[\hat{f}_\kappa(\theta)] &= \int_{-\pi}^{\pi} K_\kappa(\theta-y)f(y)dy \\ &= \int_{-\pi}^{\pi} K_\kappa(u) \left[\sum_{j=0}^{p+1} \frac{f^{(j)}(\theta)}{j!} u^j + O(u^{p+2}) \right] du \\ &= \sum_{t=0}^{p/2} \frac{f^{(2t)}(\theta)}{2t!} \alpha_{2t}(K_\kappa) + O(\alpha_{p+2}(K_\kappa)) \\ &= f(\theta) + \kappa^{-p/2} \frac{\mu_p(L)}{\mu_0(L)} \sum_{t=1}^{p/2} \frac{b_{p,2t}f^{(2t)}(\theta)}{2t!} + O(\kappa^{-(p+2)/2}). \end{aligned} \quad (3.8)$$

次に Hall 型 KDE の分散を求める。分散は次式で与えられる。

$$\begin{aligned} \mathbb{E}_f[K_\kappa^2(\theta - y)] &= \int_{-\pi}^{\pi} K_\kappa^2(u)[f(\theta) + f'(\theta)u + O(u^2)] \\ &= R(K_\kappa)f(\theta) + O(R(K_\kappa(u)u)), \end{aligned} \quad (3.9)$$

ただし、 $R(K_\kappa(\theta)\theta^t) = \int_{-\pi}^{\pi} K_\kappa^2(\theta)\theta^{2t}d\theta$ である。補題 B.3 から定数項 $R(K_\kappa)$ と $R(K_\kappa(u)u)$ はそれぞれ

$$R(K_\kappa) = \kappa^{1/2}[d(L) + o(1)], \quad (3.10)$$

と

$$R(K_\kappa(u)u) = O(\kappa^{-1/2}) \quad (3.11)$$

となる。(3.9), (3.10) と (3.11) によって分散は

$$\begin{aligned} \text{Var}_f[\hat{f}_\kappa(\theta)] &= n^{-1}\mathbb{E}_f[K_\kappa^2(\theta - Y)] - n^{-1}\mathbb{E}_f[K_\kappa(\theta - Y)]^2 \\ &= n^{-1}[\kappa^{1/2}\{f(\theta)d(L) + o(1)\} - n^{-1}\{f(\theta) + o(1)\}^2] \\ &= n^{-1}\kappa^{1/2}f(\theta)d(L) + o(n^{-1}\kappa^{1/2}) \end{aligned} \quad (3.12)$$

となる。したがって、(3.8) と (3.12) から Hall 型 KDE の MISE を得る。 \square

定理 3.1 から Hall 型 KDE は p 次オーダーカーネルの次数 p をあげることで MISE の収束レートを改善可能であることが分かる。Tsuruta and Sagae (2017b) が与えた p 次オーダーカーネルは MISE を改良可能であるという意味で魅力的な高次オーダーカーネルである。

VM カーネルの関数 $L_{\kappa, \text{VM}}(\theta)$ を

$$L_{\kappa, \text{VM}}(\theta) := e^{-\kappa\{1-\cos(\theta)\}} \quad (3.13)$$

とする。このとき、そのモーメントは $\mu_l(L_{\text{VM}}) = \Gamma(l+1/2)$ となるので VM カーネルを用いたときの MISE は 2 章で与えた (2.5) に一致することを示せる。

通常を中心局限定理によって Hall 型 KDE の漸近正規分布を次のように与えることができる。

定理 3.2. $\kappa = cn^\alpha$ とおく。定理 3.1 の条件を満たすとき、もし、 $\alpha > (2p+1)/2$ かつ $n \rightarrow \infty$ であるならば、

$$\sqrt{n\kappa^{-1/2}}[\hat{f}_\kappa(\theta) - f(\theta)] \xrightarrow{d} N(0, f(\theta)d(L))$$

が成り立つ。 \square

定理 3.2 の証明の詳細に関しては Appendix C を参照すること。

定理 3.2 は、Hall 型 KDE に関する信頼区間を構築可能であることを示している。この信頼区間は、密度 f の構造を探索するときに有益な情報を与えるであろう。

3.3 高次オーダーカーネル密度推定量の構成法

前節の定理 3.1 は高次オーダーカーネルを用いることで MISE を改良可能であることを示している。標準的な KDE の議論と同様に Hall 型 KDE でも Jones and Foster (1993) が与えた加法型構成法を用いれば、2 次オーダーカーネルから 4 次以上の高次オーダーカーネルを構成できる。

円周上のカーネルの加法型構成法を以下のように与える。

定義 3.3. p 次オーダーカーネルの関数 L を $L_{\kappa,[p]}(\theta) := L_{[p+2]}(\kappa\{1 - \cos\theta\})$ とおく. $r = \kappa\{1 - \cos(\theta)\}$ とおいて,

$$L_{[p+2]}^{\text{JF}}(r) := \frac{p+1}{p}L_{[p]}(r) + \frac{2}{p}rL'_{[p]}(r) \quad (3.14)$$

を与える. 関数 $L_{[p+2]}^{\text{JF}}(r)$ で与えられるカーネルを JF カーネル $K_{\kappa,[p+2]}^{\text{JF}}(\theta) := C_{\kappa}^{-1}(L_{[p+2]}^{\text{JF}})L_{\kappa,[p+2]}^{\text{JF}}(\theta)$ とする. また, $\text{JFK}_{\kappa,[p+2]}^{\text{JF}}(\theta)$ カーネルを用いた KDE を Jones and Foster 型 KDE(JFKDE) と呼ぶことにする. \square

Tsuruta and Sagae (2017b) は JF カーネルに関する次の命題を与えた.

命題 3.1. $K_{\kappa,[p+2]}^{\text{JF}}(\theta)$ は $p+2$ 次オーダーカーネルである. \square

この証明の詳細に関しては Appendix D を参照すること. 命題 3.1 から, JFKDE の MISE と最適な集中度パラメータは定理 3.1 で与えられるので, JFKDE は高次オーダー KDE であることが分かる.

加法型構成法の (3.14) を (3.13) に適用することで VM カーネルから次のような 4 次オーダーカーネルの関数:

$$\tilde{L}_{\kappa,[4]}^{\text{JF}}(\theta) := \{3/2 - \kappa\{1 - \cos(\theta)\}\} \exp[-\kappa\{1 - \cos(\theta)\}] \quad (3.15)$$

を構成できる.

(3.15) から構成されるカーネルに関して次の命題が成り立つ.

命題 3.2. (3.15) から構成されるカーネルを用いると JFKDE の最適な集中度パラメータは

$$\kappa_{\text{JF}} = \left[\frac{16\sqrt{\pi}}{3} R \left(\frac{5f^{(2)} + 2f^{(4)}}{12} \right) n \right]^{2/9}$$

となり, このとき, JFKDE の最適な MISE の収束レートとして $\text{MISE} = O(n^{-8/9})$ を得る. \square

JF カーネルの性質から, JFKDE は, 積分した値は 1 となるが非負性を犠牲にしてしまう欠点を持つ. Terrell and Scott (1980) が定義した加法型構成法を Hall 型 KDE に適用することで, この欠点を克服した 4 次オーダー KDE を構成できる. 円周上の KDE に関する乗法型構成法を次のように定義する.

定義 3.4. 異なる κ を持つ 2 つの 2 次オーダー KDE: $\hat{f}_{\kappa}(\theta)$ と $\hat{f}_{\kappa/4}(\theta)$ の比から, Terrell and Scott 型 KDE(TSKDE) を

$$\hat{f}_{\kappa}^{\text{TS}}(\theta) := \hat{f}_{\kappa}^{4/3}(\theta) \hat{f}_{\kappa/4}^{-1/3}(\theta)$$

として定義する. \square

TSKDE の漸近的な MISE, 最適な集中度パラメータと MISE の収束レートを次の定理として与える.

定理 3.3. TSKDE の漸近的な MISE は

$$\text{AMISE}_f[\hat{f}_{\kappa}^{\text{TS}}] = R(G)\kappa^{-4} + n^{-1}\kappa^{1/2}D(L) \quad (3.16)$$

である. (3.16) を最小にする最適な集中度パラメータは

$$\kappa_{\text{TS}} = \left[\frac{8R(G)n}{D(L)} \right]^{2/9} \quad (3.17)$$

となる. (3.16) と (3.17) から MISE の収束レート $\text{MISE} = O(n^{-8/9})$ が導かれる. 特に, 2つの VMKDE から構成される $\tilde{f}_\kappa^{\text{TS}}(\theta)$ の最適な集中度パラメータは

$$\kappa_{\text{TS}} = \left[\frac{288}{33 - 16\sqrt{2}/\sqrt{5}} R \left(\frac{\{f^{(2)}\}^2}{2f} - \frac{5f^{(2)} + 2f^{(4)}}{4} \right) n \right]^{2/9}$$

となる. □

定理 3.3 の証明の詳細に関しては Appendix E を参照すること.

定理 3.3 から TSKDE は 4 次オーダー KDE であることが分かる. TSKDE は非負性を満たすという長所を持つ. TSKDE の短所は積分の値が 1 とならない点である.

3.4 シミュレーション

最適な集中度パラメータは真の密度 f に依存しているので, ここではこのパラメータを推定するためにプラグイン法を採用する. プラグイン法とは, f が VM 分布 $f_{\text{VM}}(\theta; \mu, \tau)$ であると仮定して最適な集中度パラメータを推定する方法である. プラグイン法では初期値として τ の MLE $\hat{\tau}$ を用いる. 私たちは VMKDE · JFKDE · TSKDE の最適な集中度パラメータ $\kappa_{\text{VM}} \cdot \kappa_{\text{JF}} \cdot \kappa_{\text{TS}}$ のプラグイン法による推定量を, それぞれ $\hat{\kappa}_{\text{VM}} \cdot \hat{\kappa}_{\text{JF}} \cdot \hat{\kappa}_{\text{TS}}$ とする. ただし, それぞれの推定量は

$$\begin{aligned} \hat{\kappa}_{\text{VM}} &= [2\sqrt{\pi} \hat{R}_{\hat{\tau}}(f_{\text{VM}}^{(2)})n]^{2/5}, \\ \hat{\kappa}_{\text{JF}} &= \left[\frac{16\sqrt{\pi}}{3} \hat{R}_{\hat{\tau}} \left(\frac{5f_{\text{VM}}^{(2)} + 2f_{\text{VM}}^{(4)}}{12} \right) n \right]^{2/9}, \end{aligned}$$

と

$$\hat{\kappa}_{\text{TS}} = \left[\frac{288}{33 - 16\sqrt{2}/\sqrt{5}} \hat{R}_{\hat{\tau}}(m_{\text{vm}})n \right]^{2/9}$$

である. ただし, $m_{\text{vm}}(\theta) := [2\{f_{\text{VM}}^{(2)}\}^2/f_{\text{VM}} - 5f_{\text{VM}}^{(2)} + 2f_{\text{VM}}^{(4)}](\theta)/4$ である.

プラグイン法を用いたとき, VMKDE · JFKDE · TSKDE はどのような性能を持つのかを確かめるために, 実験 3.1 と実験 3.2 という 2 つの数値実験を行う. 実験 3.1 は真の分布として VM 分布を採用したプラグイン法にとって最も理想的な条件の下での数値実験である. プラグイン法は VM 分布以外の場合でも安定した性質を持つかどうかを確かめるために, 実験 3.2 では混合 VM 分布を採用する.

実験 3.1. 実験 3.1 は次の 3 つの手順に従う.

1. VMKDE $\tilde{f}_\kappa^{\text{VM}}(\theta)$:

- (a) VM 分布 $f_{\text{VM}}(\theta; \mu = 0, \tau)$ に従うサンプルサイズ n のランダム標本を生成する.
- (b) プラグイン法を用いて κ_{VM} を推定する.
- (c) 積分二乗誤差 $\text{ISE} = \int_{-\pi}^{\pi} \{\hat{f}_\kappa(\theta) - f(\theta)\}^2 d\theta$ の数値積分 $\overline{\text{ISE}}(\tilde{f}_\kappa^{\text{VM}})$ を計算する.
- (d) (a)–(c) を 1000 繰り返し, 算術平均 $\overline{\text{MISE}}(\tilde{f}_\kappa^{\text{VM}}) = \sum_{i=1}^{1000} \overline{\text{ISE}}_i(\tilde{f}_\kappa^{\text{VM}})/1000$ を求める.

2. JFKDE $\tilde{f}_\kappa^{\text{JF}}(\theta)$:

VMKDE の手順 (a)–(d) と同様の方法で $\overline{\text{MISE}}(\tilde{f}_\kappa^{\text{JF}})$ を求める.

3. TSKDE $\tilde{f}_\kappa^{\text{TS}}(\theta)$:

VMKDE の手順 (a)–(d) と同様の方法で $\overline{\text{MISE}}(\tilde{f}_\kappa^{\text{TS}})$ を求める.

実験 3.2. 実験 3.2 は次の 2 つの手順に従う.

1. 混合 VM 分布 $f_{\text{MVM}}(\theta; \mu_1 = \pi/4, \tau_1 = 6, \mu_2 = 5\pi/4, \tau_2 = 5, p = 0.5)$ に従うサンプルサイズ n のランダム標本を生成する.
2. 実験 3.1 と同様の方法に従って, $\overline{\text{MISE}}(\tilde{f}_\kappa^{\text{JF}}) \cdot \overline{\text{MISE}}(\tilde{f}_\kappa^{\text{TS}}) \cdot \overline{\text{MISE}}(\tilde{f}_\kappa^{\text{VM}})$ をそれぞれ計算する.

表 4: 実験 3.1 の結果. 各セルの値は VM 分布からサンプルサイズ n の標本を 1000 回生成して求めた $\text{MISE} \times 1000$ の値である. ただし, $n = 50, 100, 200, 500, 1000$ である. τ は VM 分布の集中度パラメータである. 太字は (n, τ) のとき 3 つの KDE の中で最適なものを表す.

τ	KDE	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
0.5	JFKDE	6.90	4.33	2.28	1.07	0.61
	TSKDE	6.67	4.15	2.18	1.00	0.53
	VMKDE	6.26	4.40	2.41	1.21	0.70
1	JFKDE	10.00	5.65	3.20	1.51	0.85
	TSKDE	9.33	5.20	2.88	1.35	0.77
	VMKDE	9.25	5.52	3.32	1.68	1.00
2	JFKDE	13.00	6.92	3.91	1.81	0.97
	TSKDE	13.45	7.20	4.07	1.87	1.01
	VMKDE	13.10	7.70	4.72	2.36	1.39
5	JFKDE	21.11	11.40	6.39	2.88	1.59
	TSKDE	22.82	12.22	6.84	3.08	1.70
	VMKDE	21.73	12.72	7.76	3.91	2.28

表 5: 実験 3.2 の結果. 各セルの値は, 混合 VM 分布からサンプルサイズ n の標本を 1000 回生成して求めた $\text{MISE} \times 1000$ の値である. ただし, $n = 50, 100, 200, 500, 1000$ である. 太字は同じ n の下で最も良い KDE を表す.

n	50	100	200	500	1000
JFKDE	5.35	03.73	2.78	2.29	2.19
TSKDE	5.55	3.56	2.46	1.83	1.67
VMKDE	4.84	3.73	3.09	2.75	2.68

表 4 から, $n \geq 100$ のとき, JFKDE $\tilde{f}_\kappa^{\text{JF}}$ と TSKDE $\tilde{f}_\kappa^{\text{TS}}$ の性能は VMKDE $\tilde{f}_\kappa^{\text{VM}}$ よりも優れていることが分かる. この結果は, JFKDE と TSKDE はともに 4 次オーダー KDE であるのに対して, VMKDE は 2 次オーダー KDE であるという理論的性質から生じたものである. 加えて, JFKDE は, 真の分布の集中度パラメータ τ が $\tau \geq 2$ のとき 3 つの中で一番性能が良い. TSKDE は, $n \geq 100$ かつ $\tau \leq 1$ のとき一番性能が良い. しかし, $n = 50$ のとき, VMKDE が一番良い性能を持つ.

表 5 は, $n \geq 100$ のとき, 一番良い性能を持つのは TSKDE であり, 2 番目は JFKDE であることを示している. また, $n = 50$ のとき, VMKDE が一番良い.

これらの実験結果をまとめると、4次オーダー KDE である JFKDE と TSKDE はサンプルサイズが $n \geq 100$ のとき優れた性能を示す。また、真の分布が混合 VM 分布であるときでも JFKDE と TSKDE は優れた性質を示している事実から、プラグイン法は真の分布が VM 分布でないときも安定した性能を持つことが分かる。

3.5 先行研究との比較

表 6 は Hall 型 KDE と Di Marzio 型 KDE の性質をまとめたものである。Di Marzio 型 KDE は、 $\text{sine } p$ 次オーダーカーネルの次数 p を挙げても MISE の収束レートが改善しない可能性を持つ。2 章で示したように Di Marzio 型 KDE は、2 次オーダーカーネルの中でも VM カーネルと WC カーネルのように MISE の収束レートが異なるものが存在する。

Hall 型 KDE に関する筆者のアイディアは、Hall 型カーネル族と新しく定義したモーメント $\mu_l(L)$ を採用して p 次オーダーカーネルを定義し直したことである。Hall 型カーネル族は、バイアスの中に含まれる項 $\alpha_t(K_\kappa)$ を、集中度パラメータ κ とモーメント $\mu_l(L)$ に分離することができる良い性質を持つ。そのため、低次のモーメント $\mu_l(L)$ が消えるような p 次オーダーカーネルを用いることでバイアスの収束レートを改良可能である。

つまり、Hall 型 KDE は、次数 p に応じて MISE の収束レートを改善可能である。そして、すべての p 次オーダーカーネルは MISE に関して同じ収束レートを持つ。また、2 章では Di Marzio 型 KDE に関して VM カーネルを用いた場合のみ高次オーダー KDE を構築できることが示されていたが、Hall 型 KDE はすべての 2 次オーダーカーネルに関して高次オーダー KDE を構築できる一般性を持っている。

表 6: Di Marzio 型 KDE と Hall 型 KDE の性質の比較。○: 当てはまる, ×: 当てはまらない, △: VM カーネルのみ当てはまる。

KDE	Hall 型	Di Marzio 型
p 次オーダーカーネル \Rightarrow MISE $= O(n^{-2p/(2p+1)})$	○	×
2 次オーダーカーネルから高次オーダー KDE を構成	○	△

4 Hall 型カーネル密度推定量の平滑化パラメータ推定法 [Tsuruta and Sagae (2017c)]

4.1 はじめに

4 章では, Hall 型 KDE の平滑化パラメータ推定法について議論する. 実データ分析の際に 2 次オーダー KDE が最も用いられると考えられるので, 2 次オーダー KDE の最適な集中度パラメータ:

$$\kappa_* = \beta(L)R(f^{(2)})^{2/5}n^{2/5}$$

の推定方法について議論する. ここでは, Tsuruta and Sagae (2017c) が導出した Hall et al.(1987) が提案した角度データの LSCV 推定量 (Hall et al., 1987) と角度データの DPI 推定量 (Di Marzio et al., 2011) の理論的性質 (収束レートなど) について議論する.

ただし, これら 2 つの推定量の理論的性質を導出するために Hall 型カーネル族の関数 $L(r)$ は, 定義 3.1 の条件 (h.1) と (h.2) に加えて次の条件を満たす必要がある.

定義 4.1. 関数 $L(r)$ は次の条件を満たす.

(h.3) $\eta(z) := \int_{-\infty}^{\infty} L(t^2/2)L((t+z)^2/2)dt$ について, $\lim_{|z| \rightarrow \infty} \eta(z)|z|^{3/2} = o(1)$ である.

(h.4) $\lambda(z) := \int_{-\infty}^{\infty} L'(t^2/2)L((t+z)^2/2)t^2/2dt$ について, $\lim_{|z| \rightarrow \infty} \lambda(z)|z|^{3/2} = o(1)$ である.

(h.5) 4 次導関数 $L^{(4)}(r)$ は連続である.

(h.6) $L_4(z^2/2) := 3L^{(2)}(z^2/2) - 6z^2L^{(3)}(z^2/2) + z^4L^{(4)}(z^2/2)$ とおく. このとき, 定数項 $\delta_t(L_4^m) := \int_{-\infty}^{\infty} L_4^{2m}(z^2/2)z^{2t}dz$ は, $t = 1, 2$ かつ $m = 1, 2$ について有界である.

□

VM カーネルは定義 4.1 のすべての条件を満たす. (h.3) と (h.4) は LSCV 推定量の漸近正規性を導出するために必要な条件であり, (h.5) と (h.6) は DPI 推定量の漸近正規性を導出するために必要な条件である.

4.2 最小二乗クロスバリデーション法

Hall 型カーネル $K_\kappa(\theta)$ は 2 次オーダーカーネルとする. $\hat{f}_{-i}(\Theta_i) = (n-1)^{-1} \sum_{i \neq j} K_\kappa(\theta - \Theta_j)$ とおくと, Hall 型 KDE の LSCV 推定量は次式で定義される.

$$\hat{\kappa}_{CV} = \arg \min_{h>0} CV(\kappa),$$

ただし, 関数 $CV(\kappa)$ は,

$$CV(\kappa) := R(\hat{f}) - \frac{2}{n} \sum_{i=1}^n \hat{f}_{-i}(\Theta_i) \quad (4.1)$$

である. $y_{ij} := \Theta_i - \Theta_j$, $\gamma(y) = \int_{-\pi}^{\pi} K_\kappa(w)K_\kappa(w+y)dy - 2K_\kappa(y)$ とおく. n が十分大きいとき, (4.1) は次式で近似できる.

$$CV(\kappa) := \frac{R(K_\kappa)}{n} + \frac{2}{n^2} \sum_{i<j} \gamma(y_{ij}). \quad (4.2)$$

通常は簡便化のために (4.1) の代わりに近似式 (4.2) を $\text{CV}(\kappa)$ として用いる. LSCV 推定量 $\hat{\kappa}_{\text{CV}}$ の理論的性質を導出するために, $\text{CV}(\kappa)$ に対して κ に依存しない補正項 $2n^{-1} \sum_i f(\Theta_i) - R(f)$ を加えた修正 CV 関数 $\overline{\text{CV}}(\kappa)$:

$$\overline{\text{CV}}(\kappa) := \text{CV}(\kappa) + \frac{2}{n} \sum_i f(\Theta_i) - R(f) \quad (4.3)$$

を用いることにする. (4.3) から $\overline{\text{CV}}(\kappa)$ を最小にする κ は $\hat{\kappa}_{\text{CV}}$ に一致することが分かる. この補正を加えることで $\hat{\kappa}_{\text{CV}}$ の期待値と分散は, 次の定理で示すように単純な形で与えることができる.

定理 4.1. 定理 3.1 の 3 つの条件に加えて, $R(f^{(4)} f^{1/2}) < \infty$ かつ $R((f^{(4)})^{1/2} f) < \infty$ を仮定する. このとき, 期待値は

$$\mathbb{E}_f[\overline{\text{CV}}(\kappa)] = \text{AMISE}_f[\hat{f}_\kappa] + o(\kappa^{-2} + n^{-1}\kappa^{1/2}) \quad (4.4)$$

となり, $Q(L) := \int_{-\infty}^{\infty} \left\{ 2^{-1}\mu_0^{-2}(L)\eta(z) - 2^{1/2}\mu_0^{-1}(L)L(z^2/2) \right\}^2 dz$ とおくと, 分散は次式となる.

$$\text{Var}_f[\overline{\text{CV}}(\kappa)] = \frac{2}{n^2}\kappa^{1/2}Q(L)R(f) + o(n^{-2}\kappa^{1/2} + n^{-1}\kappa^{-2}). \quad (4.5)$$

□

定理 4.1 の証明は Appendix F を参照すること.

背理法を用いることで定理 4.1 から LSCV 推定量 $\hat{\kappa}_{\text{CV}}$ に関する一致性が示される.

系 4.1. $0 < a < 1$ かつ $1 < b$ について, $\hat{\kappa}_{\text{CV}} := \arg \min_{\kappa \in (a\kappa_*, b\kappa_*)} \text{CV}(\kappa)$ とする. そのとき, 定理 3.1 と定理 4.1 のすべての条件が成り立つという仮定の下で, $n \rightarrow \infty$ であれば,

$$\hat{\kappa}_{\text{CV}}/\kappa_* \xrightarrow{p} 1$$

が成り立つ. □

系 4.1 の証明は Appendix G を参照すること.

系 4.1 より, 集中度パラメータ κ に関する探索範囲に真の最適な集中度パラメータ κ_* が含まれていれば, $\hat{\kappa}_{\text{CV}}$ に関する一致性が成り立つことが分かる.

LCSV 推定量 $\hat{\kappa}_{\text{CV}}$ の漸近正規分布を導出するために, 退化 U 統計量 (degenerate U-statistic) の漸近正規性の性質を使う. まず, 関数 $H_{ij} := H(\Theta_i, \Theta_j)$ は対称であり, かつ, $\mathbb{E}_f[H_{ij}] = 0$ を満たすとする. このとき, U 統計量は $U_n := \sum_{i < j} H_{ij}$ で定義される. 退化 U 統計量を $\mathbb{E}_f[H_{ij}|\Theta_i] = 0$ を満たす U 統計量 U_n と定義する. Hall (1984) が導出した退化 U 統計量の漸近正規性を次の補題で与える.

補題 4.1. 対称な関数 H_{ij} に関して, $\mathbb{E}_f[H_{ij}|\Theta_i] = 0$, かつ, ほとんどどこでもそれぞれの n について $\mathbb{E}_f[H_{ij}^2|\Theta_i] < \infty$ が成り立つと仮定する. また, $G_{ij} := \mathbb{E}[H_{ii}H_{ij}]$ とおく. もし, $n \rightarrow \infty$ のとき,

$$\frac{\mathbb{E}_f[G_{ij}^2] + n^{-1}\mathbb{E}_f[H_{ij}^4]}{\mathbb{E}_f[H_{ij}^2]^2} \rightarrow 0 \quad (4.6)$$

を満たすならば, そのとき,

$$\sum_{1 \leq i < j \leq n} H_{ij} \xrightarrow{d} N(0, n^2 \mathbb{E}[H_{ij}^2]/2)$$

が成り立つ. □

定理 4.2. 定理 3.1 と定理 4.1 のすべての条件が成り立つと仮定する. そのとき,

$$n^{1/10}(\hat{\kappa}_{\text{CV}}/\kappa_* - 1) \xrightarrow{d} N(0, \sigma_{\text{CV}}^2), \quad n \rightarrow \infty \quad (4.7)$$

が成り立つ. ただし,

$$m(L) := 2^{-1}\mu_0^{-2}(L)\{\eta(z) + \lambda(z) + \lambda(-z)\} - 2^{-1/2}\mu_0^{-1}(L)\{L(z^2/2) + L(z^2/2)z^2\}$$

を用いて $M_{m,t}(L) := \int_{-\infty}^{\infty} m(L)^{2m} z^{2t} dz$ とおくと, 分散 σ_{CV}^2 は

$$\sigma_{\text{CV}}^2 := 50d^{-2}(L)M_{1,0}(L)R(f)\beta^{-1/2}(L)R(f'')^{-1/5}$$

で表せる. □

定理 4.2 の証明の詳細に関しては Appendix H を参照すること.

4.3 ダイレクト・プラグイン法

実数直線の場合と同様に汎関数 $\psi_r := \int_{-\pi}^{\pi} f^{(r)}(\theta)f(\theta)d\theta$ を用いれば, $R(f^{(r)})$ は $R(f^{(r)}) = (-1)^r \psi_{2r}$ と表すことが可能である.

集中度パラメータを g とし, カーネルを $T_g^{(4)}(\theta) := C_{\kappa}^{-1}(L)S_g^{(4)}(\theta)$ とおく. また, 関数 $S_g^{(4)}(\theta)$ を

$$\begin{aligned} S_g^{(4)}(\theta) &:= -g \cos(\theta)S_g^{(1)}(\theta) + g^2\{-4 \sin^2(\theta) + 3 \cos^2(\theta)\}S_g^{(2)}(\theta) \\ &\quad + 6g^3 \cos(\theta) \sin^2(\theta)S_g^{(3)}(\theta) + g^4 \sin^4(\theta)S_g^{(4)}(\theta) \end{aligned} \quad (4.8)$$

とする. ただし, カーネル $T_g^{(4)}$ と集中度パラメータ g に関して, それぞれカーネル K_{κ} と集中度パラメータ κ と異なるものを選んで良い.

汎関数 ψ_r の KDE を次式のように定義する.

$$\hat{\psi}_4(g) := n^{-1} \sum_{i=1}^n \hat{f}_g^{(4)}(\Theta_i) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n T_g^{(4)}(\Theta_i - \Theta_j) \quad (4.9)$$

(4.9) を用いて, DPI 推定量を

$$\hat{\kappa}_{\text{PI}} = \beta(L)\hat{\psi}_4(g)^{2/5}n^{2/5}$$

と定義する.

KDE $\hat{\psi}_4$ の MSE の漸近的性質は, DPI 推定量 $\hat{\kappa}_{\text{PI}}$ の理論的性質に重要な役割を果たす. 私たちは $\hat{\psi}_4(g)$ の MSE を次の定理で示す.

定理 4.3. 次の 2 つの仮定:

1. 集中度パラメータ g を $g := g(n)$ とおく. このとき, $\lim_{n \rightarrow \infty} g(n) = \infty$ かつ $\lim_{n \rightarrow \infty} n^{-2}g^{9/2}(n) = 0$ である.
2. 密度 f は $4+p$ 回微分可能であり, $t = 1, 2, \dots, p/2$ について ψ_{4+2t} は有界である.

を満たし, かつ, p オーダーカーネルを用いるならば, そのとき, バイアスは次式となる.

$$\text{Bias}_f[\hat{\psi}_4(g)] = \text{Abias}_f[\hat{\psi}_4(g)] + O(n^{-1}g^{3/2} + g^{-(p+2)/2}), \quad (4.10)$$

ただし,

$$\text{Abias}_f[\hat{\psi}_4(g)] = \frac{3g^{5/2}S_g^{(2)}(0)}{2^{1/2}\mu_0(S)n} + \frac{\mu_p(S)}{\mu_0(S)} \sum_{t=1}^{p/2} \frac{b_{p,2t}\psi_{4+2t}}{(2t)!} g^{-p/2}$$

である. また, $G_{m,t}(S_4) := 2^{-m}\mu_0^{-2m}(S)\delta_t(S_4^m)$ とおくと, 分散は次式で与えられる.

$$\text{Var}_f[\hat{\psi}_4(g)] = \frac{4}{n}\text{Var}_f[f^{(4)}(\Theta_i)] + \frac{2G_{1,0}(S_4)\psi_0g^{9/2}}{n^2} + o(n^{-1} + n^{-2}g^{9/2}). \quad (4.11)$$

□

定理 4.3 の証明の詳細に関しては Appendix I を参照すること.

定理 4.3 から簡単に最適な集中度パラメータ g_* と最小化された MSE を次の系として与えることができる.

系 4.2. $\text{Abias}_f[\hat{\psi}(g)] = 0$ を満たすように最適な集中度パラメータ $g_* > 0$ を選択すると, g_* は次式となる.

$$g_* = W(S)n^{2/(p+5)}, \quad (4.12)$$

ただし, $W(S) = \left[-\{2^{1/2}\mu_p(S) \sum_{t=1}^{p/2} [\psi_{4+2t}b_{p,2t}/(2t)!]\} / \{3S_g^{(2)}(0)\}\right]^{2/(p+5)}$ である. 最適な集中度パラメータ g_* の選択によって (4.10) の残余項のみが残るので, バイアスは $\text{Bias}_f^2[\hat{\psi}_4(g_*)] = O(n^{-(2p+4)/(p+5)})$ となる.

つまり, バイアスは無視されるので 最小化された MSE は分散 $\text{Var}_f[\hat{\psi}_4(g)]$ に依存する. もし, カーネルの次数 p が $p < 4$ ならば, そのとき, $\inf_{g>0} \text{MSE}[\hat{\psi}_r(g)]$ は (4.11) の右辺の第 2 項となり, もし $p = 4$ であれば, そのとき, $\inf_{g>0} \text{MSE}[\hat{\psi}_r(g)]$ は (4.11) の右辺に等しくなる. また, $p > 4$ であれば, そのとき, $\inf_{g>0} \text{MSE}[\hat{\psi}_r(g)]$ は (4.11) の右辺の第 1 項に等しくなる. 以上の議論を整理すると, $\inf_{g>0} \text{MSE}[\hat{\psi}_4(g)]$ は

$$\inf_{g>0} \text{MSE}_f[\hat{\psi}_4(g)] = \begin{cases} O(n^{-(2p+1)/(p+5)}), & p < 4, \\ O(n^{-1}), & p \geq 4 \end{cases}$$

となる. □

系 4.2 は高次オーダーカーネルを採用すれば, MSE の収束レートはパラメトリックな収束レートである $O(n^{-1})$ を達成する. しかし, $\hat{\psi}_4(g)$ のカーネルとして高次オーダーカーネルを用いたとき, 最適な集中度パラメータ g_* が非負条件を満たすことを示すのは難しい. なぜならば, 高次オーダーカーネルの最適な集中度パラメータ g_* の符号は複数の未知の汎関数 ψ_r の和とカーネル T_g に依存してからである. 集中度パラメータ g_* の非負性を証明するためには真の密度 f を知っている必要があると思われるが, これは非現実的な仮定である.

カーネル T_g が 2 次オーダーカーネルであれば, そのとき, g_* の符号は T_g のみに依存する. それゆえに, $\psi_6 = -R(f^{(3)})$ に注意すると, $\mu_2(S)/S(0)$ が正となる適切なカーネル T_g を選択すれば, g_* の非負性は保証されることが分かる. 適切なカーネルの例として VM カーネルが挙げられる.

しかし, $c = -2^{-1/2}\mu_2(S)/(6S_g^{(2)}(0))$ とおくと, 適切な 2 次オーダーカーネルの集中度パラメータは $g_* = [c\psi_6n]^{2/7}$ となり, 集中度パラメータ g_* を推定するために未知な汎関数 ψ_6 を推定する必要がある. ψ_6 に関する単純な推定方法として, 真の密度関数が VM 分布 $f_{\text{VM}}(\theta; \mu, \tau)$ であるという仮定を用いる方法 (プラグイン法) を提案する. この仮定の下で ψ_6 の推定量を

$$\hat{\psi}_6^{\text{VM}} := -[4\hat{\tau}I_1(2\hat{\tau}) + 30\hat{\tau}^2I_2(2\hat{\tau}) + 15\hat{\tau}^3I_3(2\hat{\tau})]/\{16\pi I_0^2(\hat{\tau})\}$$

として定義する。ただし、 $\hat{\tau}$ は τ の MLE である。1 ステップ・プラグイン法というダイレクト・プラグイン法の実用的なアルゴリズムを提案する。

アルゴリズム 4.1. 1 ステップ・プラグイン法は次の手順に従って実行される。

Step.1 $\hat{\tau}$ と $\hat{\psi}_6^{\text{VM}}$ を推定する。

Step.2 推定量 $\hat{g} := [c\hat{\psi}_6^{\text{VM}}n]^{2/7}$ を求める。

Step.3 DPI 推定量 $\hat{\kappa}_{\text{PI}} = \beta(L)\hat{\psi}_4(\hat{g})^{2/5}n^{2/5}$ を与える。

DPI 推定量 $\hat{\kappa}_{\text{PI}}$ の漸近正規性を次の定理として与える。

定理 4.4. 定理 4.3 の 2 つの条件が成り立つと仮定し、かつ、適切な 2 次オーダーカーネルを用いることにする。そのとき、 $n \rightarrow \infty$ ならば、

$$n^{5/14}(\hat{\kappa}_{\text{PI}}/\kappa_* - 1) \xrightarrow{d} N(0, \sigma_{\text{PI}}^2)$$

が成り立つ。ただし分散 σ_{PI}^2 は $\sigma_{\text{PI}}^2 := 8W^{9/2}(S)G_{1,0}(S_4)\psi_0\psi_4^{-2}/25$ である。□

定理 4.4 の証明の詳細に関しては Appendix J を参照すること。

定理 4.2 は LSCV 推定量の収束レートが $O(n^{-1/10})$ であることを示している。定理 4.4 から DPI 推定量の収束レートはより速い $O(n^{-5/14})$ であることが分かる。したがって、DPI 推定量の理論的性能は LSCV 推定量よりも優れている。

4.4 シミュレーション

統計手法を有限標本のデータに適用する際には、分析結果が理論的性質とは異なる可能性に注意しなければならない。そのため、数値実験を通して、有限標本の下での LSCV 推定量と DPI 推定量の性能を調べる必要がある。DPI 推定量は 1 ステップ・プラグイン法の中で VM 分布を参照分布として採用している。そのため、真の分布として VM 分布を選ぶと DPI 推定量に有利である。したがって、VM 分布とは異なる分布の下で DPI 推定量の性能はどのように変化するかを明らかにしたい。また、LSCV 推定量は、その理論的性質から分布の選択の影響を受けないと予想できるが、有限標本下で LSCV 推定量がそのような特性を持つかどうか調査すべきである。

これらの疑問に答えるための方法として、VM 分布を含む柔軟な分布族を採用することが考えられる。ここでは、VM 分布や WC 分布を含み、さらに非対称な分布までも含む sine skewed Jones-Pewsey (SSJP) 分布 (Abe and Pewsey, 2011) を採用する。SSJP 分布 $\text{SSJP}(\mu, \tau, \psi, \lambda)$ の密度関数は次式で与えられる。

$$f(\theta; \mu, \tau, \psi, \lambda) := \frac{\cosh(\tau\psi)^{1/\psi}(1 + \tanh(\tau\psi)\cos(\theta - \mu))^{1/\psi}(1 + \lambda\sin(\theta - \mu))}{2\pi P_{1/\psi}(\cosh(\tau\psi))},$$

ただし、 $0 \leq \tau$ は、形状パラメータを、 $-1 \leq \lambda \leq 1$ は非対称パラメータを、 $P_{1/\psi}(\cosh(\tau\psi))$ は自由度 $(1/\psi)$ の 0 次第 1 種ルジェンドル陪関数をそれぞれ表す。ここで、 $\tau = 0$ のとき、明らかに SSJP 分布は円周上の一様分布であるので、以下の議論では、常に $0 < \tau$ とおくことにする。

SSJP 分布が含むいくつかの円周上の確率分布について述べる。 $\lambda = 0$ のとき、SSJP 分布は対称分布となることに注意されたい。SSJP 分布は、 $\lambda = 0$ の下で $\psi = 0, -1, 1$ のとき、それぞれ、VM 分布・WC 分布・

ハート形分布となる。ハート形分布の密度関数は単純で

$$f(\theta; \mu = 0, \tau, \psi = 1, \lambda = 0) = (1 + \tau \cos(\theta)) / (2\pi)$$

で与えられる。また、SSJP 分布は、 $\lambda \neq 0$ かつ $\psi = 0, -1$ のとき sine skewed VM 分布・sine skewed WC 分布と呼ばれ、 $0 < \lambda \leq 1$ かつ $\psi = 1$ のとき sine skewed ハート形分布と呼ばれる。

今挙げた SSJP 分布の 6 つのサブクラス (Model.1-6) を用いて数値実験を行うことにする。

Model.1 VM 分布: SSJP($\mu = 0, \tau = 1, \psi = 0, \lambda = 0$).

Model.2 ハート形分布: SSJP($\mu = 0, \tau = 1, \psi = 1, \lambda = 0$).

Model.3 WC 分布: SSJP($\mu = 0, \tau = 1, \psi = -1, \lambda = 0$).

Model.4 Sine-skewed VM 分布: SSJP($\mu = 0, \tau = 1, \psi = 0, \lambda = 0.6$).

Model.5 Sine-skewed ハート形分布 : SSJP($\mu = 0, \tau = 1, \psi = 1, \lambda = 0.6$).

Model.6 Sine-skewed : WC 分布 SSJP($\mu = 0, \tau = 1, \psi = -1, \lambda = 0.6$).

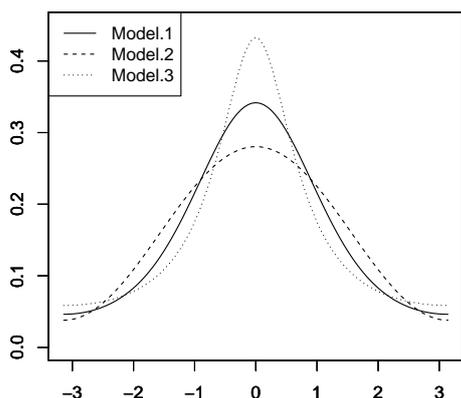


図 7: Model.1-3 の密度関数. Model.1-3 は VM・ハート形・WC 分布である.

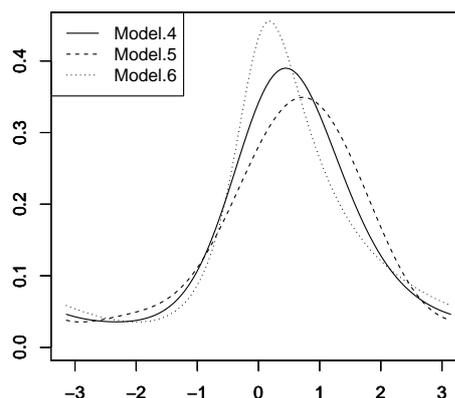


図 8: Model.4-6 の密度関数. Model.4-6 は sine skewed VM・sine skewed ハート形・sine skewed WC 分布の密度関数である.

実験 4.1. 数値実験の手順 :

1. Model.1 に関して次の手順に従って数値実験を行う.
 - (a) Model.1 の分布からサンプルサイズ n のランダム標本を発生させる.
 - (b) 真の分布が Model.1 の分布であるときの最適な集中度パラメータ κ_* を(3.4) から求める.
 - (c) 統計ソフト R の「circular」ライブラリーにある「bw.cv.mse.circular」関数を用いて LSCV 推定量 $\hat{\kappa}_{CV}$ を推定する.
 - (d) 1 ステップ・プラグイン法を用いて DPI 推定量 $\hat{\kappa}_{PI}$ を推定する.
 - (e) 次の 2 つの相対誤差 $Y_{CV} = \hat{\kappa}_{CV} / \kappa_* - 1$ と $Y_{PI} = \hat{\kappa}_{PI} / \kappa_* - 1$ を計算する.

- (f) (a)–(e) の手順を 1000 回繰り返し、 Y_{CV} と Y_{PI} の標本平均と標本標準偏差を与える。
 2. (a)–(f) の手順を Model.2–6 に対しても実行する。

表 7: 数値実験 4.1 の結果 (DPI). 各セルの値は, Model.1–6 からサンプルサイズ n の標本を 1000 回生成して求めた DPI 推定量の相対誤差 ($\hat{\kappa}_{PI}/\kappa_* - 1$) の標本平均 (mean) と標準偏差 (sd) である. ただし, $n = 50, 100, 200, 500, 1000$ である.

n	Model.1		Model.2		Model.3		Model.4		Model.5		Model.6	
	mean	sd										
50	0.11	0.31	0.38	0.41	-0.39	0.20	0.07	0.30	0.17	0.34	-0.35	0.19
100	0.07	0.25	0.29	0.3	-0.38	0.15	0.03	0.22	0.13	0.25	-0.35	0.14
200	0.05	0.19	0.24	0.22	-0.37	0.12	0.01	0.17	0.09	0.18	-0.33	0.11
500	0.03	0.13	0.19	0.14	-0.33	0.09	-0.01	0.12	0.06	0.13	-0.29	0.08
1000	0.02	0.10	0.16	0.11	-0.30	0.07	-0.01	0.09	0.05	0.10	-0.26	0.06

表 8: 数値実験 4.1 の結果 (LSCV). 各セルの値は, Model.1–6 からサンプルサイズ n の標本を 1000 回生成して求めた LSCV 推定量の相対誤差 ($\hat{\kappa}_{CV}/\kappa_* - 1$) の標本平均 (mean) と標準偏差 (sd) である. ただし, $n = 50, 100, 200, 500, 1000$ である.

n	Model.1		Model.2		Model.3		Model.4		Model.5		Model.6	
	mean	sd										
50	1.72	5.16	2.72	7.62	0.66	2.66	1.48	4.33	1.82	5.10	0.71	2.74
100	1.37	3.97	2.2	5.98	0.31	1.78	1.03	3.15	1.44	3.98	0.47	2.00
200	1.01	3.00	1.70	4.6	0.27	1.51	0.76	2.39	0.88	2.73	0.22	1.33
500	0.67	2.10	0.96	2.8	0.13	0.96	0.53	1.70	0.63	1.99	0.08	0.81
1000	0.39	1.40	0.66	1.95	0.06	0.72	0.40	1.30	0.43	1.38	0.06	0.64

表 7 と 8 から, Model.1–6 のいずれの場合でも, DPI 推定量は LSCV 推定量よりも標準偏差が小さいので, DPI 推定量は安定した推定量であることが分かる. Model.1 (VM 分布)・Model.2 (ハート形分布)・Model.4 (Sine skewed VM 分布)・Model.5 (Sine skewed ハート形分布) の下では, DPI 推定量は LSCV 推定量よりも平均値が 0 に近く, LSCV 推定量よりもその性能は優れている. しかし, Model.3 (WC 分布)・Model.6 (Sine skewed WC) 分布に関しては平均値を比較すると, サンプルサイズ n に応じて両者の性能の優位性は変化する. Model.3 の場合は $n = 50$ のとき, DPI 推定量はより良い性能は持つが, $n \geq 100$ のとき, LSCV 推定量の性能の方が優れている. また, Model.6 の場合は $n = 50, 100$ のとき, DPI 推定量の性能がより優れているが, $n \geq 200$ のときは LSCV 推定量の性能の方が高い.

図 7 を見ると, Model.1 と Model.2 は山の形が穏やかであり, 裾の減少が速い形状をしているが, Model.3 は山の頂上付近が急なカーブとなっていて, 裾が厚い形状をしている (図 8 は Model.4–6 についても同じ傾向があることを示している). DPI 推定量は VM 分布やハート形分布といった山の形が穏やかな分布では優れた性能を持つが, WC 分布のような裾が厚い分布に関してはバイアスを持っている. これは 1 ステップ・プラグイン法の参照分布として VM 分布を採用しているためと考えられる. Model.4–6 の結果から, LSCV

推定量と DPI 推定量は，分布の非対称性の影響をほとんど受けていないことが分かる．

LSCV 推定量の性能は不安定であるためにサンプルサイズ n が小さいときは DPI 推定量を用いるべきである．サンプルサイズ n が十分に大きいとき，LSCV 推定量と DPI 推定量の性能の優位性は分布の形状の影響を受けているので，どちらの推定量を採用するかは，分析者が 2 つの推定量を用いた KDE を構築したうえで両者の結果を比較して判断すべきである．

4.5 先行研究との比較

4 章では Hall 型 KDE に関して LSCV 推定量と DPI 推定量の漸近正規性とそれらの収束レートを導出し，Di Marzio 型 KDE の未解決問題であった集中度パラメータ推定量の理論的性質の解明に成功した．LSCV 推定量の収束レート $O(n^{-1/10})$ と DPI 推定量の収束レート $O(n^{-5/14})$ は，標準的な KDE に関する両者の収束レートに一致している．

5 Di Marzio 型局所多項式回帰 [Tsuruta and Sagae (2017d)]

5章では, Di Marzio et al. (2009) が与えた Di Marzio 型 LPR について議論する. また, Tsuruta and Sagae (2017d) が明らかにした Di Marzio 型 LPR の理論的性質や VM カーネルと WC カーネルを適用した場合の理論的性質について述べる.

5.1 Di Marzio 型局所多項式回帰の定義とその理論的性質

ランダムな 2 変数標本 $(Y_1, \Theta_1), \dots, (Y_n, \Theta_n)$ について目的変数 $Y_i \in \mathbb{R}$ と説明変数 $\Theta_i \in \mathbb{T} = [-\pi, \pi)$ との関係推定問題を考える. ただし, Y_i は次の仮定を満たすとする.

$$Y_i = m(\Theta_i) + v(\Theta_i)^{1/2} \varepsilon_i, \quad i = 1, \dots, n, \quad (5.1)$$

ここで, 回帰関数 m は周期性 $m(\theta) = m(\theta + 2\pi)$ を持つ. (5.1) についてその他の仮定に関しては (1.23) と同様であるとする. $m(\theta)$ の推定量は周期性を持つことが望ましいだろう. しかし, 1.3 節で定義した標準的な LPR は, (1.22) や (1.23) を見ても分かるように, 有限標本の下では周期性を持たない.

説明変数が角度変数 Θ_i であるときは, 周期性を持つ新しいノンパラメトリック回帰が必要となる. Di Marzio et al.(2009) は, 周期性を満たすノンパラメトリック回帰として sine 型局所多項式回帰 (S-LPR: sine local polynomial regression) を提案した. 彼らは, $\theta = \sin^{-1}(\sin(\theta))$ と考えれば $\theta \simeq \sin(\theta)$ で近似できることから, sine 級数型のテイラー展開:

$$m(\Theta_i) = m(\theta) + \sum_{j=1}^p m^{(j)}(\theta) \sin(\Theta_i - \theta)^j \quad (5.2)$$

が成り立つと主張した. 円周上の非負カーネルと sine 級数型の多項式 $\beta_0 + \sum_{j=1}^p \beta_j \sin(\cdot - \theta)^j$ を用いて, S-LPR $\hat{m}(\theta; p, \kappa)$ は, 次の重み付二乗誤差:

$$\sum_{i=1}^n \{Y_i - \beta_0 - \beta_1 \sin(\Theta_i - \theta) - \dots - \beta_p \sin(\Theta_i - \theta)^p\}^2 K_\kappa(\Theta_i - \theta) \quad (5.3)$$

を最小にする $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T$ の切片 $\hat{\beta}_0$ で与えられる. したがって, $\hat{m}(x; p, \kappa)$ は

$$\hat{m}(\theta; p, \kappa) := \mathbf{e}_1^T (\mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{Y} \quad (5.4)$$

となる. ただし, \mathbf{e}_1 は第 1 要素は 1 でそれ以外は 0 となるような $(p+1) \times 1$ ベクトル, $\mathbf{Y} = (Y_1, \dots, Y_n)$ は目的変数ベクトル:

$$\mathbf{S}_\theta := \begin{bmatrix} 1 & \sin(\Theta_1 - \theta) & \dots & \sin^p(\Theta_1 - \theta) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \sin(\Theta_n - \theta) & \dots & \sin^p(\Theta_n - \theta) \end{bmatrix}$$

は $n \times (p+1)$ デザイン行列, $\mathbf{W}_\theta := \text{diag}\{K_\kappa(\Theta_1 - \theta), \dots, K_\kappa(\Theta_n - \theta)\}$ は $n \times n$ 重み付き対角行列を表す.

S-LPR に定義 2.1 で与えた Di Marzio 型カーネルを適用したものを本稿では Di Marzio 型 LPR $\hat{m}(x; p, \kappa)$ と呼ぶことにする.

例えば, $p = 0$ のとき, (5.4) から $\hat{m}(\theta; 0, \kappa)$ は

$$\hat{m}(\theta; 0, \kappa) = \sum_i K_\kappa(\Theta_i - \theta) Y_i / \sum_i K_\kappa(\Theta_i - \theta) \quad (5.5)$$

となる. また, $p = 1$ のとき, (5.4) から $\hat{m}(\theta; 1, \kappa)$ は次式に等しい.

$$\hat{m}(\theta; 1, \kappa) := n^{-1} \sum_i \frac{\{\hat{s}_2(\theta; \kappa) - \hat{s}_1(\theta; \kappa) \sin(\Theta_i - \theta)\} K_\kappa(\Theta_i - \theta) Y_i}{\hat{s}_2(\theta; \kappa) \hat{s}_0(\theta; \kappa) - \hat{s}_1(\theta; \kappa)^2}, \quad (5.6)$$

ただし, $\hat{s}_l(\theta; \kappa) := n^{-1} \sum_i \sin(\Theta_i - \theta)^l K_\kappa \sin(\Theta_i - \theta)$ である. (5.5) と (5.6) から $\hat{m}(\theta; 0, \kappa)$ と $\hat{m}(\theta; 1, \kappa)$ は周期性を持つことが示される. 同様に考えれば, $p \geq 2$ のときも Di Marzio 型 LPR は周期性の性質を持つことは容易に分かる.

高次の Di Marzio 型 LPR の理論的性質を導出することは難しいので, 5 章では $p = 1$ のときの理論的性質のみを議論することにする.

$\Theta_n := \{\Theta_1, \dots, \Theta_n\}$ とおく. また, 条件付バイアスを $\text{Bias}_Y[\hat{m}(\theta; \kappa)|\Theta_n] := E_Y[\hat{m}(\theta; \kappa)|\Theta_n] - m(\theta)$ とし, 条件付分散を $\text{Var}_Y[\hat{m}(\theta; \kappa)|\Theta_n]$ とする.

Di Marzio et al. (2009) は次の定理として条件付バイアスと条件付分散を与えた.

定理 5.1. 次の 4 つの仮定:

- i) $\lim_{n \rightarrow \infty} n^{-1} R(K_\kappa) = 0$.
- ii) $\lim_{n \rightarrow \infty} \gamma_j(\kappa) = 1$.
- iii) 周辺密度 $f(\theta)$ は連続微分可能である. ただし, 任意の θ について $f(\theta) > 0$ である.
- iv) 導関数 $m''(\theta)$ と条件付分散 $v(\theta)$ はそれぞれ連続である.

を満たすならば, そのとき, 条件付バイアスは

$$\text{Bias}_Y[\hat{m}(\theta; 1, \kappa)|\Theta_n] \simeq \eta_2(K_\kappa) \frac{m''(\theta)}{2!} \quad (5.7)$$

となり, 条件付分散は

$$\text{Var}_Y[\hat{m}(\theta; 1, \kappa)|\Theta_n] \simeq R(K_\kappa) \frac{v(\theta)}{nf(\theta)} \quad (5.8)$$

となる. □

2 章で議論したのと同様な理由で $\eta_2(K_\kappa)$ と $R(K_\kappa)$ に関して κ とカーネル K を分離するのは難しいので, 一般的な条件付 MSE の収束レートを求めるのは困難である.

定理 5.1 と中心局限定理から $\hat{m}(\theta; 1, \kappa)$ の漸近正規性が成り立つことを示す.

定理 5.2. 定理 5.1 のすべての仮定が成り立つと仮定する. そのとき, $n \rightarrow \infty$ ならば,

$$\sqrt{n/R(K_\kappa)}[\hat{m}(\theta; 1, \kappa) - E_Y[\hat{m}(\theta; 1, \kappa)|\Theta_n]] \xrightarrow{d} N(0, v(\theta)/f(\theta)), \quad (5.9)$$

が成り立つ. □

定理 5.2 の証明の詳細に関しては Appendix K を参照すること.

5.2 フォン・ミーゼスカーネルと巻き込みコーシーカーネル

5.2.1 フォン・ミーゼスカーネル

フォン・ミーゼスカーネルを Di Marzio 型 LPR に適用したときの理論的性質を示す．重み付条件付 MISE を $\text{MISE}_Y[\hat{m}(\theta; 1, \kappa)|\Theta_n] := \mathbb{E}_Y[\int_{-\pi}^{\pi} \{\hat{m}(\theta; \kappa) - m(\theta)\}^2 f(\theta) d\theta | \Theta_n]$ として定義する．定理 5.1 と補題 2.1 から VM カーネルのバイアス・分散・MISE は次の定理として与えられる．

定理 5.3. 定理 5.1 の仮定に加えて $n \rightarrow \infty$, $\kappa \rightarrow \infty$ かつ $n^{-1}\kappa^{1/2} \rightarrow 0$ が成り立つならば, そのとき, バイアスは

$$\text{Bias}_Y[\hat{m}(\theta; 1, \kappa)|\Theta_n] \simeq \frac{1}{2\kappa} m''(\theta) \quad (5.10)$$

となり, 分散は

$$\text{Var}_Y[\hat{m}(\theta; 1, \kappa)|\Theta_n] \simeq \frac{\kappa^{1/2} v(\theta)}{2\pi^{1/2} n f(\theta)} \quad (5.11)$$

となる．(5.10) と (5.11) より, 漸近的な MISE は次式として与えられる．

$$\text{AMISE}_Y[\hat{m}(\theta; 1, \kappa)|\Theta_n] = \frac{1}{4\kappa^2} \int_{-\pi}^{\pi} m''(\theta)^2 f(\theta) d\theta + \frac{\kappa^{1/2} \int_{-\pi}^{\pi} v(\theta) d\theta}{2\pi^{1/2} n}. \quad (5.12)$$

そのとき, (5.12) を最小にする最適な集中度パラメータ κ_* は

$$\kappa_* = \left[\frac{2\pi^{1/2} \int_{-\pi}^{\pi} m''(\theta)^2 f(\theta) d\theta}{\int_{-\pi}^{\pi} v(\theta) d\theta} \right]^{2/5} n^{2/5} \quad (5.13)$$

である．(5.12) と (5.13) から $\text{AMISE}_Y[\hat{m}(\theta; 1, \kappa_*)|\Theta_n]$ は $O_p(n^{-4/5})$ となる． \square

定理 5.2 と定理 5.3 から VM カーネルを適用した場合の漸近正規分布は次の定理のようになる．

定理 5.4. 定理 5.3 の条件がすべて成り立つと仮定する． $\kappa = cn^\alpha$ とおく, ただし, c と α はそれぞれ定数である．もし, $\alpha > 2/5$ かつ $n \rightarrow \infty$ ならば, そのとき,

$$n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; 1, \kappa) - m(\theta)] \xrightarrow{d} N(0, v(\theta) / \{2\pi^{1/2} f(\theta)\})$$

が成り立つ． \square

定理 5.4 の証明の詳細に関しては Appendix L を参照すること．

5.3 巻き込みコーシーカーネル

$h = 1 - \rho^2$, $0 < h < 1$ に注意すると, そのとき, 定理 5.1 と補題 2.2 から WC カーネルを適用したときのバイアス・分散・MISE は次の定理として与えられる．

定理 5.5. 定理 5.1 のすべての条件に加えて, $n \rightarrow \infty$, $h \rightarrow 0$ かつ $nh \rightarrow \infty$ を仮定する．そのとき, バイアスは

$$\text{Bias}_Y[\hat{m}(\theta; 1, h)|\Theta_n] \simeq h \frac{m''(\theta)}{4} \quad (5.14)$$

となり、分散は

$$\text{Var}_Y[\hat{m}(\theta; 1, h)|\Theta_n] \simeq (nh)^{-1} \frac{v(\theta)}{\pi f(\theta)} \quad (5.15)$$

となる。(5.14) と (5.15) から漸近的な MISE は次式に等しいことが示される。

$$\text{AMISE}_Y[\hat{m}(\theta; 1, h)|\Theta_n] = \frac{h^2 \int_{-\pi}^{\pi} m''(\theta)^2 f(\theta) d\theta}{16} + \frac{\int_{-\pi}^{\pi} v(\theta) d\theta}{\pi nh}. \quad (5.16)$$

(5.16) を最小にする最適な h_* は

$$h_* = \left\{ \frac{8 \int_{-\pi}^{\pi} v(\theta) d\theta}{\pi \int_{-\pi}^{\pi} m''(\theta)^2 f(\theta) d\theta} \right\}^{1/3} n^{-1/3} \quad (5.17)$$

で与えられる。(5.16) と (5.17) から $\text{AMISE}_Y[\hat{m}(\theta; 1, h_*)|\Theta_n] = O_p(n^{-2/3})$ が導かれる。□

定理 5.3 と 5.5 を比較すると、この 2 つの結果から Di marzio 型 LPR は、採用したカーネルに応じて漸近的な MISE の収束レートが異なる。この性質は標準的な LPR は $p=1$ の場合、カーネルに依存することなく MISE の収束レートが $O_p(n^{-4/5})$ となるという性質とは異なっている。

定理 5.2 と定理 5.5 から WC カーネルを適用した場合の漸近正規分布は次の定理として与えられる。

定理 5.6. 任意の定数 c と α について $h = cn^\alpha$ とおく。定理 5.5 のすべての条件が成り立つと仮定する。もし、 $\alpha < -1/3$ かつ $n \rightarrow \infty$ ならば、そのとき、

$$(nh)^{1/2}[\hat{m}(\theta; 1, h) - m(\theta)] \xrightarrow{d} N(0, v(\theta)/\{\pi f(\theta)\})$$

が成り立つ。□

定理 5.6 の証明の詳細に関しては Appendix M を参照すること。

5.4 シミュレーション

前節では VM カーネルと WC カーネルの理論的性質について議論した。実データ分析の観点から、数値実験を通して有限標本の場合の両者の性能を調べることにする。また、一般に誤差項の標準偏差が大きくなるほど、回帰の性能は悪化するので、誤差項の標準偏差が変化するとき VM カーネルと WC カーネルの性能に変化があるか確かめる。

ここで次のようなモデルを考える。

$$Y_i = m(\Theta_i) + v^{1/2}(\Theta_i)\varepsilon_i, \quad \varepsilon \sim N(0, 1), \quad v(\theta) = t^2,$$

ただし、回帰関数は $m(\theta) = 2 + 3 \cos(\theta) + 2 \sin(3\theta)$ であり (図 9 を参照)、 Θ_i は円周上の一様分布: $f_{CU}(\theta) = 1/(2\pi)$ $\theta \in [-\pi, \pi)$ に従う。

上記の設定の下での最適な集中度パラメータを用いて VM カーネルと WC カーネルを適用した場合の比較を行う。最適な集中度パラメータを用いる理由は $m(\theta)$ の推定に関して集中度パラメータの推定誤差の影響を取り除くためである。VM カーネルの集中度パラメータは

$$\tilde{\kappa} = [333n/(2\pi^{1/2}t^2)]^{2/5}$$

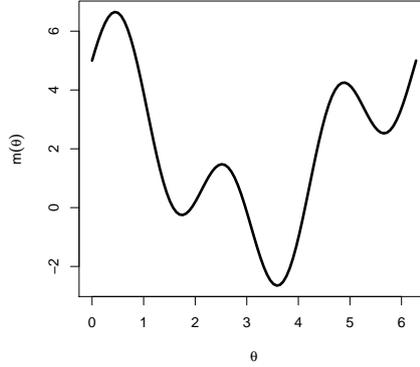


図 9: 回帰関数 : $m(\theta) = 2 + 3 \cos(\theta) + 2 \sin(3\theta)$.

であり, 巻き込みコーシーカーネルの集中度パラメータは

$$\tilde{h} = [32t^2/(333n)]^{2/5}$$

である. 小標本では $\text{MISE}_Y[\hat{m}(\theta; 1, \cdot) | \Theta_n]$ は説明変数 Θ_n のサンプリングに強く依存する. Θ_n のランダム性を弱めるために Θ_n について 100 回のサンプリングを行い, MISE の算術平均: $\text{Ave.MISE} = \sum_{j=1}^{100} \text{MISE}_{j,Y}[\hat{m}(\theta; 1, \cdot) | \Theta_n] / 100$ を計算する.

数値実験の手順は次の通りである.

実験 5.1. 数値実験の手順 :

1. 円周上の一様分布に従うサンプルサイズ n のランダム標本 $\{\Theta_1, \dots, \Theta_n\}$ を生成する.
2. 正規分布 $N(0, t^2)$ からサンプルサイズ n のランダム標本 $\{\varepsilon_1, \dots, \varepsilon_n\}$ を生成する.
3. 1-2 からランダム標本 $\{Y_1, \dots, Y_n\}$ を生成する.
4. VM カーネルを適用した $\hat{m}(\theta; 1, \tilde{\kappa})$ と WC カーネルを適用した $\hat{m}(\theta; 1, \tilde{h})$ をそれぞれ計算する.
5. 数値積分 $\text{ISE}_{\text{VM}} := \int_{-\pi}^{\pi} \{\hat{m}(\theta; 1, \tilde{\kappa}) - m(\theta)\}^2 / (2\pi) d\theta$ と $\text{ISE}_{\text{WC}} := \int_{-\pi}^{\pi} \{\hat{m}(\theta; 1, \tilde{h}) - m(\theta)\}^2 / (2\pi) d\theta$ をそれぞれ計算する.
6. 2 から 5 の手順を 1000 回繰り返して, $\text{MISE}_{\text{VM}} := \sum_j \text{ISE}_{\text{VM},j} / 1000$ と $\text{MISE}_{\text{WC}} := \sum_j \text{ISE}_{\text{WC},j} / 1000$ をそれぞれ計算する.
7. 1 から 6 の手順を 1000 回繰り返して, $\text{Ave.MISE}_{\text{VM}} := \sum_j \text{MISE}_{\text{VM},j} / 100$ と $\text{Ave.MISE}_{\text{WC}} := \sum_j \text{MISE}_{\text{WC},j} / 100$ をそれぞれ計算する.

表 9 と表 10 は, $n \leq 20$ のとき, WC カーネルの性能は VM カーネルのものよりも優れていることを示している. しかし, $n \geq 30$ のとき, VM カーネルは WC カーネルよりも優れた性能を持つことが分かる. 表 11 と表 12 は小標本の場合では WC カーネルの MISE の標準偏差は VM カーネルのものよりも小さいことを示している. つまり, WC カーネルは小標本の下では VM カーネルよりも安定している. これらの結果は WC カーネルが小標本の下では優れた性能を持つことを示唆している. $n \geq 30$ のとき, VM カーネルは WC カーネルよりも優れた性能を持つ.

誤差項の標準偏差 t が大きくなると VM・WC カーネルともに性能は悪化する. ただし, t が大きい場合で

も n が大きくなるにつれて両者の性能は改善されている。

数値実験 5.1 は 1 つの回帰関数の場合のみを対象にしており，VM カーネルと WC カーネルの実データ分析における性能の評価としては限定的な結果である．今後は，別の回帰関数を用いるなどした複数の設定の下での数値実験を追加すべきである。

表 9: 実験 5.1 の結果 (VM カーネル). 各セルの値は VM カーネルの重み付条件付 MISE: $\text{Ave.MISE}_{\text{VM}} := \sum_j \text{MISE}_{\text{VM},j}/100$ である. n はサンプルサイズ, t は誤差項の標準偏差を表す.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
$t = 0.5$	200.667	3.938	0.363	0.246	0.132	0.057
$t = 1$	11.314	1.441	0.678	0.401	0.315	0.158
$t = 1.5$	7.575	2.008	0.987	0.733	0.566	0.291
$t = 2$	12.122	2.805	1.470	1.076	0.840	0.447

表 10: 実験 5.1 の結果 (WC カーネル). 各セルの値は WC カーネルの重み付条件付 MISE: $\text{Ave.MISE}_{\text{WC}} := \sum_j \text{MISE}_{\text{WC},j}/100$. n はサンプルサイズ, t は誤差項の標準偏差を表す. 太字は, 同じ (n, t) のとき, VM カーネルよりも値が小さいことを表す.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
$t = 0.5$	2.090	1.075	0.594	0.421	0.319	0.148
$t = 1$	2.627	1.258	0.846	0.597	0.519	0.292
$t = 1.5$	2.713	1.650	1.165	0.938	0.815	0.476
$t = 2$	4.161	2.256	1.602	1.254	1.061	0.657

表 11: 実験 5.1 の結果 (VM カーネル). 各セルの値は VM カーネルの重み付条件付 MISE の標準偏差を表す. n はサンプルサイズ, t は誤差項の標準偏差を表す.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
$t = 0.5$	1784.428	28.477	0.390	0.377	0.040	0.009
$t = 1$	27.064	1.708	0.473	0.086	0.055	0.012
$t = 1.5$	18.149	2.043	0.407	0.303	0.086	0.021
$t = 2$	31.531	2.960	0.405	0.226	0.112	0.026

表 12: 実験 5.1 の結果 (WC カーネル). 各セルの値は WC カーネルの重み付条件付 MISE の標準偏差を表す. n はサンプルサイズ, t は誤差項の標準偏差を表す. 太字は, 同じ (n, t) のとき, VM カーネルよりも値が小さいことを表す.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
$t = 0.5$	1.197	0.814	0.395	0.306	0.137	0.051
$t = 1$	1.267	0.648	0.359	0.154	0.155	0.046
$t = 1.5$	1.080	0.593	0.325	0.228	0.228	0.063
$t = 2$	3.093	0.848	0.345	0.226	0.143	0.073

5.5 先行研究との比較

表 13 は標準的な LRR と Di Marzio 型 LPR の性質を比較したものである. 標準的な LRR は, 2 次の非負カーネルのみを用いることを前提とし多項式の次数 p を上げることで MISE の収束レートを改善する. また, カーネルの選択は MISE の定数項に影響を与えるが, MISE の収束レートには影響しない.

しかし, Di Marzio 型 LPR に関してカーネルの選択は, MISE の収束レートに大きな影響を与える. 5 章で示したように $p = 1$ のとき, VM カーネルを適用した場合の収束レートは $O_p(n^{-4/5})$ であるが, WC カーネルを適用した場合の収束レートは $O_p(n^{-2/3})$ に過ぎない. また, $p = 1$ のとき以外の Di Marzio 型 LPR は MISE の収束レートがどのくらいの速さとなるのかという疑問が未解決問題として残っている.

表 13: 標準的な LRR と Di Marzio 型 LPR の性質の比較. ○ : 当てはまる, × : 当てはまらない, △ : VM カーネルのみ当てはまる.

LPR	標準的な LRR	Di Marzio 型 LPR
標本空間	実数直線: \mathbb{R}	円周: \mathbb{T}
$p = 1 \Rightarrow \text{MISE} = O_p(n^{-4/5})$	○	△
高次の $p \Rightarrow \text{MISE}$ の収束レートの改善	○	×

6 Hall 型局所多項式回帰

6.1 はじめに

Hall 型 KDE に関して p 次オーダーカーネルを適用した場合の MISE の収束レートを導出可能なことを 3 章で示した. この結果から方向統計学における局所多項式回帰である sine 型局所多項式回帰 (S-LPR) にも Hall 型 LPR を適用することによって, 一般的な次数 p に関しての MISE の収束レートを導出できるのではないかという予想が成り立つ.

S-LPR に Hall 型カーネルを適用したものを Hall 型 LPR と呼ぶことにする. 6 章では Hall 型 LPR に関する理論的性質として MISE の収束レートと漸近正規性を導出する. また, 有限標本下の Hall 型 LPR の性質を調べるために数値実験を行う.

6.2 Sine 級数展開と Hall 型カーネル関数の理論的性質

記号の簡便化のために, ガウス記号を $[x]$ を導入する. $[x]$ は, 実数 x に対して, $n \leq x < n+1$ なる整数 n がただ一つ存在し, $n = [x]$ であることを意味する.

関数 $\sin^{-1}(u)$ のテイラー展開を与える.

$$\sin^{-1}(u) = \sum_{s=0}^{\infty} b_s u^{2s+1}, \quad |u| < 1, \quad (6.1)$$

ただし, $b_s := \{(2s-1)!!\} / \{(2s)!!(2s+1)\}$ である.

(6.1) を $m(\Theta_i)$ のテイラー展開:

$$m(\Theta_i) = \sum_{q=0}^{p+2} \frac{m^{(q)}(\theta)}{q!} (\Theta_i - \theta)^q + o_p((\Theta_i - \theta)^{p+2}) \quad (6.2)$$

に適用すると sine 級数 $\sin(\Theta_i - \theta)^q$ を用いた $m(\Theta_i)$ のテイラー展開を次の補題として得る.

補題 6.1.

$$M_q(\theta) := \sum_{t=1}^q m^{(t)}(\theta) B_q(p, t)$$

とする. ただし, $B_q(p, t)$ は

$$B_q(p, t) = \sum_B \frac{\prod_{i=0}^{[(p+1)/2]} b_i^{t_i}}{\prod_{m=0}^{[(p+1)/2]} t_m!},$$

ここで, \sum_B は, $\sum_{s=0}^{[(p+1)/2]} t_s = t$ かつ $\sum_{s=0}^{[(p+1)/2]} (2s+1)t_s = q$ を満たす $t_1, \dots, t_{[(p+1)/2]}$ についての和を意味する.

関数 $m(\theta)$ は次式となる.

$$m(\Theta_i) = m(\theta) + \sum_{q=1}^{p+2} M_q(\theta) \sin(\Theta_i - \theta)^q + o_p(\sin(\Theta_i - \theta)^{p+2}), \quad |\theta| < \pi/2. \quad (6.3)$$

□

補題 6.1 は, $m(\theta)$ を $m(\Theta_i) - \sum_{q=1}^p \beta_q \sin(\Theta_i - \theta)^q$ で近似できることを意味する. したがって, S-LPR の多項式として sine 級数型多項式 $\sum_{q=0}^p \beta_q \sin(\Theta_i - \theta)^q$ を採用することが正当化される.

3 章と 4 章で Hall 型カーネル族の平滑化パラメータは集中度パラメータ κ であったが 6 章では Hall 型カーネル族の平滑化パラメータとしてバンド幅 $h = \kappa^{-1/2}$ を採用する. $n \rightarrow \infty$ のとき, バンド幅 h は $h \rightarrow 0$ を満たすと仮定する. Hall 型カーネル族 $K_h(\theta) : [0, 2\pi) \rightarrow \mathbb{R}$ は次の定義に従う.

定義 6.1. Hall 型カーネル族 $K_h(\theta)$ を $K_h(\theta) := C_h^{-1}(L)L_h(\theta)$ とする. ただし, 関数 $L_h(\theta) := L(h^{-2}\{1 - \cos(\theta)\})$, 基準化定数 $C_h(L) := \int_{-\pi}^{\pi} L_h(\theta)d\theta$ である.

また, 関数 $L(r)$ は次の 2 つの条件:

(h.1) 関数 $L(r)$ は連続的微分可能である.

(h.2) r が十分大きいとき, 任意の正の偶数 $v \geq 0$ について $L(r) = o(r^{-(v+2)})$ が成り立つ.

□

(h.1') は定義 3.1 で与えた条件 (h.1) と同じであり, (h.2') は定義 3.1 の (h.2) の v について $v = p+1$ とおいたものである.

カーネル関数 $\bar{L}(z) := L(z^2/2)$ は条件 (h.1') と (h.2') を満たす実数直線上に台を持つ関数とする. このとき, 関数 $L_h(\theta)$ と $L(z)$ の関係について次のような性質が成り立つ.

補題 6.2. 関数 $L_h(\theta)$ は 次の 3 つの性質を満たす.

(性質 1) n が十分大きいとき, $\theta = hz$ とおくことで実数直線上のカーネル関数 \bar{L} で近似できる.

$$L_h(hz) = \bar{L}(z) + O(h^2), \quad z \in \mathbb{R}. \quad (6.4)$$

(性質 2) $n \rightarrow \infty$ のとき, $L_h(hz)$ は, 実数直線上の非負カーネル \bar{L} に分布収束する.

$$L_h(hz) \rightarrow \bar{L}(z), \quad z \in \mathbb{R}. \quad (6.5)$$

(性質 3) カーネル関数 \bar{L} のモーメントを $\alpha_t(\bar{L}) := \int_{\mathbb{R}} \bar{L}(z)z^t dz$ とする. ただし, $\alpha_0(\bar{L}) = 1$ とする. $t \leq 2p+2$ について, $\alpha_t(\bar{L})$ は有界であり, かつ, $\int_{-\pi/h}^{\pi/h} \bar{L}(z)z^t dz = \alpha_t(L) + o(h)$ である.

□

補題 6.2 の証明の詳細に関しては Appendix O を参照すること. 補題 6.2 より次の補題が導かれる.

補題 6.3. 関数 $K_h(\theta)$ は, n が十分大きいとき次の 3 つの性質を満たす.

(性質 1) 基準化定数 $C_h(L)$ は

$$C_h(L) := h + o(h^2) \quad (6.6)$$

となる.

(性質 2) $K_h(hz)$ は $L(\bar{z})$ で近似できる.

$$K_h(hz) = h^{-1}[\bar{L}(z) + O(h^2)], \quad z \in [-\pi/h, \pi/h]. \quad (6.7)$$

(性質 3) $K_h(\theta)$ の裾のスピードは

$$K_h(\theta) = o(h^{p+2}), \quad |\theta| \geq \pi/2 \quad (6.8)$$

で与えられる.

補題 6.3 の証明の詳細に関しては Appendix P を参照すること.

補題 6.2 と補題 6.3 から VM カーネル $K_h(hz)$ は正規カーネル $h^{-1}\bar{L}(z)$ で近似できることが分かる. 1.4 節で述べたように, VM カーネルが正規カーネルに分布収束することはよく知られている.

6.3 Hall 型局所多項式回帰の理論的性質

補題 6.1, 補題 6.2 と補題 6.3 によって Hall 型 LPR の条件付バイアスと条件付分散は次式で与えられる.

定理 6.1. 次の仮定:

1. $h = h(n)$ とする. $h(n) \rightarrow \infty$ かつ $nh(n) \rightarrow \infty$ である.
2. 導関数 $m^{(p+2)}(\theta)$ と条件付分散 $v(\theta)$ はそれぞれ連続である.
3. 周辺密度 $f(\theta) > 0$ は連続的微分可能である.

の下で条件付バイアスは, 次数 p が奇数のとき

$$\text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n] = h^{p+1}M_{p+1}(\theta)a_{p+1}(\bar{L}_{(p)}) + o_p(h^{p+1})$$

となり, 次数 p が偶数のとき

$$\text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n] = h^{p+2}\{M_{p+1}(\theta)f'(\theta)f(\theta)^{-1} + M_{p+2}(\theta)\}a_{p+2}(\bar{L}_{(p)}) + o_p(h^{p+2})$$

となる. また, 条件付分散はどちらの場合でも

$$\text{Var}_Y[\hat{m}(\theta; p, h)|\Theta_n] = (nh)^{-1}\frac{\alpha_0(\bar{L}_{(p)}^2)}{f(\theta)}v(\theta) + o_p((nh)^{-1})$$

となる. □

定理 6.1 の証明の詳細に関しては Appendix Q を参照すること.

定理 6.1 の条件付バイアスに関して, 隣り合った偶数と奇数の次数の場合バイアスの収束レートは同じである. 例えば, $p = 0$ と $p = 1$ のとき $\text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n] = O_p(h^2)$ となり同じバイアスになる. また, 偶数次の場合はバイアスが未知の周辺密度 $f(\theta)$ に依存しているが, 奇数次の場合はバイアスが $f(\theta)$ に依存していない. したがって, Hall 型 LPR に関しては優れた性質を持つ奇数次のものを用いるべきである.

p が奇数の場合, Hall 型 LPR の重み付条件付 MISE は次の系で与えられる.

系 6.1. 定理 6.1 のすべての条件が成り立つと仮定する. p が奇数の場合, MISE は

$$\text{MISE}_Y[\hat{m}(\theta; p, h)|\Theta_n] = \text{AMISE}_Y[\hat{m}(\theta; p, h)|\Theta_n] + o_p(h^{2(p+1)} + (nh)^{-1}), \quad (6.9)$$

ただし,

$$\text{AMISE}_Y[\hat{m}(\theta; p, h)|\Theta_n] = h^{2(p+1)}\alpha_{p+1}^2(\bar{L}_{(p)})\int_{-\pi}^{\pi}M_{p+1}(\theta)^2f(\theta)d\theta + (nh)^{-1}\alpha_0(L_{(p)}^2)\int_{-\pi}^{\pi}v(\theta)d\theta \quad (6.10)$$

である. (6.10) を最小にする最適な h_* は

$$h_* = \left[\frac{(p+1)(p!)^2\alpha_0(L_{(p)}^2)\int_{-\pi}^{\pi}v(\theta)d\theta}{2\alpha_{p+1}^2(L_{(p)})\int_{-\pi}^{\pi}M_{p+1}(\theta)^2f(\theta)d\theta} \right]^{1/(2p+3)} n^{-1/(2p+3)} \quad (6.11)$$

として与えられる．(6.9), (6.10) と (6.11) より, 最小化された MISE は $\text{MISE}[\hat{m}(\theta; p, h_*) | \Theta_n] = O_p(n^{-(2p+1)/(2p+3)})$ となる．同様に考えれば, p が偶数の場合の最小化された MISE は $\text{MISE}[\hat{m}(\theta; p, h_*) | \Theta_n] = O_p(n^{-(2p+4)/(2p+5)})$ となる． \square

定理 6.1 と標準的な中心極限定理より, Hall 型 LPR の漸近正規性が次の定理として与えられる．

定理 6.2. 任意の定数 c と α について $h = cn^\alpha$ とおく．定理 6.1 のすべての条件が成り立つと仮定する．このとき, p が奇数の場合, もし $\alpha < -1/(2p+3)$ かつ $n \rightarrow \infty$ ならば, そのとき,

$$n^{1/2}h^{-1/2}[\hat{m}(\theta; p, \kappa) - m(\theta)] \xrightarrow{d} N(0, v(\theta)\alpha_0(L_{(p)}^2)/f(\theta)), \quad (6.12)$$

が成り立つ．また, p が偶数の場合, もし $\alpha < -1/(2p+5)$ かつ $n \rightarrow \infty$ ならば, そのとき, (6.12) が成り立つ． \square

定理 6.2 の証明の詳細に関しては Appendix Q を参照すること．

6.4 シミュレーション

有限標本の下で 1 次の Hall 型 LPR $\hat{m}(\theta; 1, h)$ と 3 次の Hall 型 LPR $\hat{m}(\theta; 3, h)$ の性能を比較するための数値実験を行う．ここでは次のモデルを考える．

$$Y_i = m(\Theta_i) + \varepsilon_i, \quad (6.13)$$

ただし, 真の回帰関数は $m(\theta) = 2 + 3 \cos(\theta) + 2 \sin(3\theta)$ (図 9 を参照), 誤差項は $\varepsilon_i \sim N(0, 1)$ である．また, 説明変数 Θ_i は円周上の一様分布 $f_{\text{CU}}(\theta)$ に従うとする．

バンド幅は上記の設定の下での最適値 h_* と h_* の LSCV 推定量 \hat{h}_{CV} を用いることにする．LSCV 推定量を用いたのは実際のデータ解析では最適値 h_* を知ることは不可能であるからである． \hat{h}_{CV} の定義は次式で与えられる．

$$\arg \min_{h>0} \text{CV}_m(h), \quad (6.14)$$

ただし, CV 関数は

$$\text{CV}_m(h) = \frac{1}{n} \sum_i \{Y_i - \hat{m}_{-i}(\Theta_i; p, h)\}^2, \quad (6.15)$$

ここで, $\hat{m}_{-i}(\theta; p, h)$ は i 番目の (Y_i, Θ_i) を除いた Hall 型 LPR である．

ここで以下のような数値実験を行った．

実験 6.1. 数値実験は次の手順に従う．

1. 一様分布 $f_{\text{CU}}(\theta)$ からサンプルサイズ n の説明変数の標本 Θ_n を生成する．
2. 標準正規分布 $N(0, 1)$ からサンプルサイズ n の誤差項標本 ε_n を生成する．
3. 2つの標本 Θ_n と ε_n からサンプルサイズ n の目的変数標本 \mathbf{Y}_n を生成する．
4. 生成した標本から 4 つの推定量 $\hat{m}(\theta; 1, h_*)$, $\hat{m}(\theta; 1, \hat{h}_{\text{CV}})$, $\hat{m}(\theta; 3, h_*)$ と $\hat{m}(\theta; 3, \hat{h}_{\text{CV}})$ を与える．
5. 4 つの推定量について, それぞれ積分二乗誤差 $\text{ISE} := n^{-1} \sum_{i=1}^n \{\hat{m}(\theta; p, h) - m(\theta)\}^2$ を求める．
6. 1 から 5 の手順を 1000 回繰り返してそれぞれの ISE の算術平均 $\text{MISE} = \sum_{j=1}^{1000} \text{ISE}_j / 1000$ を求める．

表 14: 実験 6.1 の結果. 各セルの値は MISE である. p の値は Hall 型 LPR の次数を, Opt は最適なバンド幅を採用したことを, CV は LSCV 推定量をバンド幅として採用したことをそれぞれ示している.

p	バンド幅	$n = 50$	$n = 100$	$n = 200$	$n = 500$
1	Opt	0.332	0.157	0.084	0.039
1	CV	0.366	0.173	0.090	0.040
3	Opt	0.252	0.116	0.058	0.025
3	CV	0.280	0.128	0.062	0.026

表 14 から 3 次の Hall 型 LPR は, バンド幅が最適値・LSCV 推定量のどちらの場合でも 1 次のものよりもすべての n に関して良い性能を示している. また, 3 次の Hall 型 LPR に関して LSCV 推定量を採用したものはどの n の場合でも 2 番目に良い性能となっている. これは LSCV 推定量は適切なバンド幅推定量であることを示唆している. 3 次の Hall 型 LPR が 1 次のものよりも性能が良いという結果は理論的な結果と一致しており, これは有限標本下でも次数 p を上げることで Hall 型 LPR の性能を改良できることを示している.

数値実験 6.1 は 1 モデルの場合の数値実験の結果に過ぎないことに注意すべきである. 回帰関数や Θ の条件付分布, 誤差項などの設定を変えた複数のモデルの下で 1 次と 3 次の Hall 型 LPR の性能を比較することが, 今後の課題である.

6.5 先行研究との比較

表 15 は, Hall 型 LPR と Di Marzio 型 LPR の性質をまとめたものである. Di Marzio 型 LPR は, 1 次のときカーネルの選択が MISE の収束レートに大きな影響を与える点や次数 p を上げても MISE の収束レートが改善できるか不明である点が標準的な LRR と異なっていた. Hall 型 LPR はこれらの欠点を克服した推定量である. つまり, 次数 p が奇数であればカーネルに依存することなく $\text{MISE}[\hat{m}(\theta; p, h_*) | \Theta_n] = O_p(n^{-(2p+4)/(2p+5)})$ となり, 偶数のときは $\text{MISE}[\hat{m}(\theta; p, h_*) | \Theta_n] = O_p(n^{-(2p+4)/(2p+5)})$ となる. Hall 型 LPR は次数 p を上げることで推定量の性能を向上できる良い性質を持つ. ただし, 6 章ではバンド幅 h の推定量として LSCV 推定量を用いたが, Hall 型 LPR に関する LSCV 推定量の一致性などの理論的性質は明らかではなく, これは今後の研究課題となっている.

表 15: Hall 型 LPR と Di Marzio 型 LPR の性質の比較. ○: 当てはまる, ×: 当てはまらない, △: VM カーネルのみ当てはまる.

LPR	Hall 型 LPR	Di Marzio 型 LPR
標本空間	円周: \mathbb{T}	円周: \mathbb{T}
$p = 1 \Rightarrow \text{MISE} = O_p(n^{-4/5})$	○	△
高次の $p \Rightarrow \text{MISE}$ の収束レートの改善	○	×

7 方向統計学における多変量ノンパラメトリック回帰

7.1 先行研究

7.1.1 先行研究の分類

方向統計学における多変量ノンパラメトリック回帰の研究としては説明変数がトーラス上で与えられるトーラス上のノンパラメトリック回帰 (Di Marzio et al., 2009) や実数・円周上のノンパラメトリック回帰 (Qin et al., 2011) がある。これらの研究は、1.4 節で述べた実数空間上の多変量ノンパラメトリック回帰 (Ruppert and Wand, 1994) をヒントにして行われた。これらの先行研究を説明変数が定義される空間に応じて分類したのが図 10 である。

これらの多変量ノンパラメトリック回帰の説明をする前にトーラス上の空間と実数・円周上の空間について説明する。 q 次元トーラス空間 \mathbb{T}^q とは、円周 \mathbb{T} の q 個の直積で与えることができ、 $\mathbb{T}^q := \mathbb{T} \times \dots \times \mathbb{T}$ と定義される。 q 次元の角度変数ベクトル $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_d)^T$ は q 次元トーラス \mathbb{T}^q 上の空間で与えられる。例えば、(風向, 海鳥の飛び立つ方向) からなる 2 次元角度変数ベクトルは 2 次元トーラス \mathbb{T}^2 上で与えられ、2 次元トーラス \mathbb{T}^2 は図 10 のようなドーナツのような形となる。実数・円周上の空間 $\mathbb{R}^d \times \mathbb{T}$ は実数空間 \mathbb{R}^d と円周 \mathbb{T} の直積で与えられる。方向統計学でよく用いられるのはシリンダー $\mathbb{R} \times \mathbb{T}$ 上の空間である。シリンダー上で定義される変数ベクトルの例は (風速, 風向) である。また、(風速, 気圧, 風向) からなる変数ベクトルは $\mathbb{R}^2 \times \mathbb{T}$ 上で定義されることがすぐに分かる。

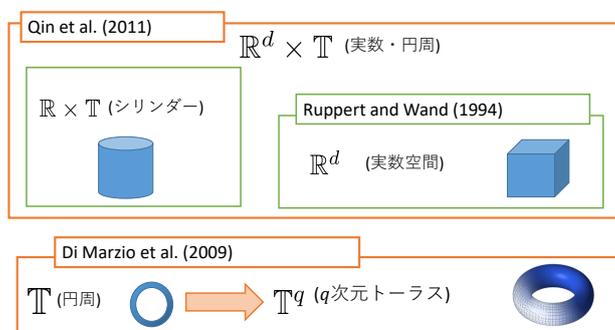


図 10: 方向統計学における多変量ノンパラメトリック回帰の分類.

7.1.2 トーラス上のノンパラメトリック回帰

Di Marzio et al. (2009) によるトーラス上のノンパラメトリック回帰の研究について述べる。独立同一分布からランダムに得られたデータ $(Y_1, \Theta_1), \dots, (Y_n, \Theta_n)$ を考える。 Θ_i は q 次元トーラス $\theta \in \mathbb{T}^q$ で値を持つ説明変数ベクトルであり、 Y_i は実数直線 \mathbb{R} 上で与えられたスカラーな目的変数とする。

我々は次のような回帰モデルを考える。

$$Y_i = m(\Theta_i) + v^{1/2}(\Theta_i)\varepsilon_i, \quad (7.1)$$

ただし, $m(\Theta_i) = E_Y[Y|\Theta = \theta]$ は真の回帰関数を, $\text{Var}_Y[Y|\Theta = \theta] = v(\theta)$ は条件付分散を表す. 集中度パラメータベクトルを $\kappa_q := \kappa I_q$ と定義する. このとき, トーラス上のカーネル $K_{\kappa_q}(\theta)$ は, 定義 2.1 を満たし, かつ, 非負の値を取る Di Marzio 型カーネルのプロダクトカーネルとして

$$K_{\kappa_q}(\theta) := \prod_{j=1}^q K_{\kappa}((\Theta_{ij} - \theta_j))$$

で定義される.

$\sin(\theta) := (\sin(\theta_1), \dots, \sin(\theta_n))^T$ とし, 回帰係数ベクトルを $\beta_q := (\beta_1, \dots, \beta_q)^T$ とし, 切片を α とおく. このとき, トーラス上の LLR(: local linear regression) $\hat{m}(\theta; \kappa_q)$ は, 最小二乗誤差:

$$\sum_{i=1}^n \{Y_i - \alpha - \beta_q^T \sin(\Theta_i - \theta)\}^2 K_{\kappa_q}(\Theta_i - \theta), \quad (7.2)$$

を最小にする $(\alpha, \beta_q^T)^T$ の切片 $\hat{\alpha}$ で与えられる. したがって, $\hat{m}(\theta; \mathbf{H})$ は

$$\hat{m}(\theta; p, h) := \mathbf{e}_1^T (\mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{Y} \quad (7.3)$$

となる. ただし, \mathbf{e}_1 は第 1 要素は 1 でそれ以外は 0 となるような $(q+1) \times 1$ ベクトル, $\mathbf{Y} = (Y_1, \dots, Y_n)$ は目的変数ベクトル,

$$\mathbf{S}_\theta := \begin{bmatrix} 1 & \sin(\Theta_1 - \theta)^T \\ \vdots & \vdots \\ 1 & \sin(\Theta_n - \theta)^T \end{bmatrix} = \begin{bmatrix} 1 & \sin(\Theta_{11} - \theta_1) & \cdots & \sin(\Theta_{1q} - \theta_q) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \sin(\Theta_{n1} - \theta_1) & \cdots & \sin(\Theta_{nq} - \theta_q) \end{bmatrix} \quad (7.4)$$

は $n \times (q+1)$ デザイン行列, かつ, $\mathbf{W}_\theta := \text{diag}\{K_{\kappa_q}(\Theta_i - \theta), \dots, K_{\kappa_q}(\Theta_i - \theta)\}$ は $n \times n$ 重み付き対角行列を表す.

トーラス上の重み付条件付 MISE を $\text{MISE}_Y[\hat{m}(\theta; \kappa_q)] := E[\int_{\mathbb{T}^q} \{\hat{m}(\theta; \kappa_q) - m(\theta)\}^2 f(\theta) d\theta]$ と定義する. トーラス上の LLR の重み付条件付 MISE を次の定理として与える.

定理 7.1. 次の仮定:

1. ヘッセ行列 $\mathcal{H}_m(\theta)$ は連続である.
2. 周辺密度 $f(\theta) > 0$ は連続微分可能であり, 条件付分散 $v(\theta)$ は連続である.
3. $n \rightarrow \infty$ のとき, $\eta_2(K_\kappa) \rightarrow 0$ かつ $R(K_\kappa) \rightarrow 0$ である.

を満たし, かつ, n が十分大きいならば, 漸近的な MISE は

$$\text{AMISE}_Y[\hat{m}(\theta; \kappa_q)] \simeq \frac{\eta_2(K_\kappa)^2 \int \{\nabla^2 m(\theta)\}^2 f(\theta) d\theta}{4} + \frac{R(K_\kappa)^q \int_{\mathbb{T}^d} v(\theta) d\theta}{n}$$

となる. □

定理 7.1 を見ると集中度パラメータ κ とカーネル K の分離が難しいので, Di Marzio 型 LPR と同様に一般的な MISE の収束レートを求めるのは難しい問題であることが分かる.

7.1.3 実数・円周上のノンパラメトリック回帰

Qin et al. (2011) が与えた実数・円周上のノンパラメトリック回帰の定義とその理論的性質について述べる．実数・円周 $\mathbb{R}^d \times \mathbb{T}$ 上の説明変数を $\mathbf{U}_i = (\mathbf{X}_i, \Theta_i)^T$ とする．ただし， \mathbf{X}_i は \mathbb{R}^d 上で定義される d 次元変数ベクトルを表す． (\mathbf{U}_i, Y_i) $i = 1, \dots, n$ はランダムな標本とする．このとき，我々は次のような回帰モデルを考える．

$$Y_i = m(\mathbf{U}_i) + v^{1/2}(\Theta_i)\varepsilon_i, \quad (7.5)$$

ただし， $m(\mathbf{u}) = E_Y[Y|\mathbf{U} = \mathbf{u}]$ は真の回帰関数を， $\text{Var}_Y[Y|\mathbf{U} = \mathbf{u}] = v(\mathbf{u})$ は条件付分散を表す．実数・円周上のカーネル $K_{\mathbf{H}\kappa}(\mathbf{U}_i - \mathbf{u})$ は以下のようなプロダクトカーネルとして定義される．

$$K_{\mathbf{H}\kappa}(\mathbf{U}_i - \mathbf{u}) = \frac{1}{h_1 \cdots h_n} \prod_{j=1}^d \bar{K}((X_{ij} - x_j)/h_j) \times K_\kappa(\theta_{ij} - \theta_j), \quad (7.6)$$

ただし， \bar{K} は通常非負カーネルであり， K_κ は定義 2.1 を満たしかつ非負の値を取る Di Marzio 型カーネルである．実数・円周上の LLR $\hat{m}(\mathbf{u}; \mathbf{H}, \kappa)$ は，最小二乗誤差：

$$\sum_{i=1}^n \{Y_i - \alpha - \beta_d^T(\mathbf{X}_i - \mathbf{x}) - \beta_{d+1} \sin(\Theta_i - \theta)\}^2 K_{\mathbf{H}\kappa}(\mathbf{U}_i - \mathbf{u}) \quad (7.7)$$

を最小にする $(\alpha, \beta_q^T, \beta_{d+1})^T$ の切片 $\hat{\alpha}$ で与えられる．ただし，

$$\mathbf{U}_u := \begin{bmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T & \sin(\Theta_1 - \theta) \\ \vdots & \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T & \sin(\Theta_n - \theta) \end{bmatrix} \quad (7.8)$$

は $n \times (q+1)$ デザイン行列を表す．したがって， $\hat{m}(\mathbf{u}; \mathbf{H}, \kappa)$ は

$$\hat{m}(\mathbf{u}; \mathbf{H}, \kappa) := \mathbf{e}_1^T (\mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{Y} \quad (7.9)$$

となる．ただし， \mathbf{e}_1 は第 1 要素は 1 でそれ以外は 0 となるような $(d+2) \times 1$ ベクトル， $\mathbf{Y} = (Y_1, \dots, Y_n)$ は目的変数ベクトル，かつ， $\mathbf{W}_u := \text{diag}\{K_{\mathbf{H}\kappa}(\mathbf{U}_1 - \mathbf{u}), \dots, K_{\mathbf{H}\kappa}(\mathbf{U}_n - \mathbf{u})\}$ は $n \times n$ 重み付き対角行列を表す．

実数・円周上の LLR の重み付条件付 MISE は次の定理として与えられる．

定理 7.2. 次の仮定：

1. $n \rightarrow \infty$ のとき， \mathbf{H} の各成分と $n^{-1}|\mathbf{H}|^{-1/2}$ は 0 に収束する．
2. ヘッセ行列 $\mathcal{H}_m(\mathbf{u})$ は，連続とである．
3. 周辺密度 $f(\mathbf{u}) > 0$ は連続微分可能であり，条件付分散 $v(\mathbf{u})$ は連続である．
4. $n \rightarrow \infty$ のとき， $\eta_2(K_\kappa) \rightarrow 0$ かつ $R(K_\kappa) \rightarrow 0$ である．

を満たし，かつ， n が十分大きいならば，漸近的な MISE は

$$\text{AMISE}_Y[\hat{m}(\mathbf{u}; \mathbf{H}, \kappa)] = \frac{\sum_{\lambda, \tau=1} C_{\lambda, \tau}(m) \tilde{h}_\lambda^2 \tilde{h}_\tau^2 \tilde{\alpha}_\lambda \tilde{\alpha}_\tau}{4} + \frac{R(\bar{K})^d R(K_\kappa) \int v(\mathbf{u}) d\mathbf{u}}{nh_1 \cdots h_d},$$

ただし,

$$C_{\lambda,\tau}(m) = \int (\partial/\partial u_\lambda^2)m(\mathbf{u})(\partial/\partial u_\tau^2)m(\mathbf{u})f(\mathbf{u})d\mathbf{u}, \quad \lambda, \tau = 1, \dots, d+1,$$

であり, ここで, \tilde{h}_λ と $\tilde{\alpha}_\lambda$ はそれぞれ

$$\tilde{h}_\lambda = \begin{cases} h_\lambda, & \lambda = 1, \dots, d, \\ 1, & \lambda = d+1, \end{cases}, \quad \tilde{\alpha}_\lambda = \begin{cases} \alpha_2(\bar{K}), & \lambda = 1, \dots, d, \\ \eta_2(K_\kappa), & \lambda = d+1 \end{cases}$$

を表す. □

定理 7.1 と同様な理由で定理 7.2 の漸近的な MISE の収束レートを求めることは難しい問題である. また, κ と h という異なる平滑化パラメータを一緒に使うことで MISE の形は非常に複雑な形となっている.

7.2 実数・トーラス上のノンパラメトリック回帰とその理論的性質

図 10 で分類した 3 つの多変量ノンパラメトリック回帰について, 説明変数が与えられる空間を $\mathbb{R}^d \times \mathbb{T}^q$ ま で拡張すれば図 11 を見ればわかるようにすべての多変量ノンパラメトリック回帰を含んだ一般的な定義を与えることができる. 本稿では, 実数・円周上とトーラス上の多変量ノンパラメトリック回帰に関する「MISE の収束レート」という未解決問題を実数・トーラス上のノンパラメトリック回帰の MISE の収束レートを導出する問題に帰着させることで解決することを目指す. MISE の収束レートを導出するために Hall 型カーネル族を実数・トーラス上のノンパラメトリック回帰に適用する.

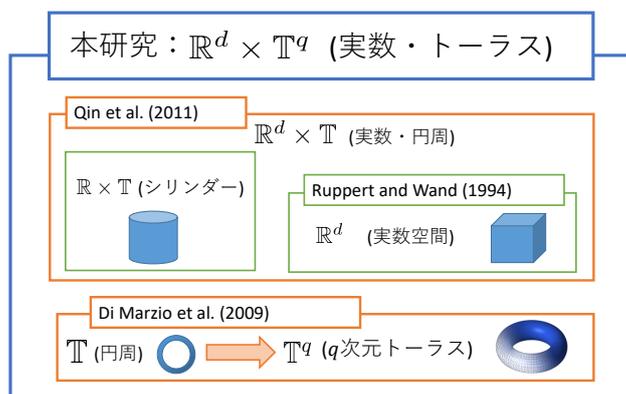


図 11: 実数・トーラス上の空間の概念図. 提案する実数・トーラス上のノンパラメトリック回帰が先行研究を含むことを表している.

説明変数を $\mathbf{U}_i = (\mathbf{X}_i^T, \boldsymbol{\Theta}_i^T)^T$ とする. ただし, \mathbf{X}_i は \mathbb{R}^d 上で定義される d 次元変数ベクトルとし, $\boldsymbol{\Theta}_i$ は \mathbb{T}^q 上で定義される q 次元角度変数ベクトルとする. (\mathbf{U}_i, Y_i) $i = 1, \dots, n$ はランダムな標本とする. このと

き、我々は次のような回帰モデルを考える。

$$Y_i = m(\mathbf{U}_i) + v^{1/2}(\mathbf{U}_i)\varepsilon_i, \quad (7.10)$$

ただし、 $m(\mathbf{u}) = \mathbb{E}_Y[Y|\mathbf{U} = \mathbf{u}]$ は真の回帰関数を、 $\text{Var}_Y[Y|\mathbf{U} = \mathbf{u}] = v(\mathbf{u})$ は条件付分散を表す。

$(d+q) \times (d+q)$ バンド幅行列 $\mathbf{H} := \text{diag}\{h_1^2, \dots, h_d^2, h_{d+1}^2, \dots, h_{d+q}^2\}$ は $\mathbf{H}_d := \text{diag}\{h_1^2, \dots, h_d^2\}$ と $\mathbf{H}_q := \text{diag}\{h_{d+1}^2, \dots, h_{d+q}^2\}$ に分割できるとする。

実数・トーラス上のカーネル $K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u})$ の定義を次式とする。

$$K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u}) = \bar{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \times \tilde{K}_{\mathbf{H}_q}(\Theta_i - \boldsymbol{\theta}),$$

ただし、 $\bar{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})$ は 1.4 節で与えた実数空間上のプロダクト・カーネルであり、 $\tilde{K}_{\mathbf{H}_q}(\Theta_i - \boldsymbol{\theta})$ は

$$\tilde{K}_{\mathbf{H}_q}(\Theta_i - \boldsymbol{\theta}) = |\mathbf{C}_q|^{-1/2} L_{\mathbf{H}_q}(\Theta_i - \boldsymbol{\theta}) = |\mathbf{C}_q|^{-1/2} \prod_{j=1}^q L_{h_j}(\Theta_{ij} - \theta_{ij}),$$

ただし、 $L_{h_j}(\Theta_{ij} - \theta_{ij})$ は、定義 6.1 で与えた Hall 型カーネルの関数 L であり、対角行列 $\mathbf{C}_q = \text{diag}\{C_{h_1}(L), \dots, C_{h_q}(L)\}$ の成分は基準化定数 $C_{h_j} = \int_{-\pi}^{\pi} L_{h_j}(\Theta_j - \theta_j) d\theta_j$ である。ここで、任意の整数 j, l について、 $\lim_{n \rightarrow \infty} h_j/h_l = O(1)$ とする。

実数・トーラス上の LLR $\hat{m}(\mathbf{u}; \mathbf{H})$ は、最小二乗誤差：

$$\sum_{i=1}^n \{Y_i - \alpha - \beta_d^T(\mathbf{X}_i - \mathbf{x}) - \beta_q^T \sin(\Theta_i - \boldsymbol{\theta})\}^2 K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u})$$

を最小にする切片 $(\alpha, \beta_d^T, \beta_q^T)^T$ の $\hat{\alpha}$ で与えられる。ただし、

$$\mathbf{U}_u := \begin{bmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T & \sin(\Theta_1 - \boldsymbol{\theta})^T \\ \vdots & \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T & \sin(\Theta_n - \boldsymbol{\theta})^T \end{bmatrix},$$

は $n \times (d+q+1)$ デザイン行列を表す。したがって、 $\hat{m}(\mathbf{u}; \mathbf{H})$ は

$$\hat{m}(\mathbf{u}; \mathbf{H}) := \mathbf{e}_1^T (\mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{Y} \quad (7.11)$$

となる。ただし、 \mathbf{e}_1 は第 1 要素は 1 でそれ以外は 0 となるような $(d+q+1) \times 1$ ベクトルを、 $\mathbf{Y} = (Y_1, \dots, Y_n)$ は目的変数ベクトルを、 $\mathbf{W}_u := \text{diag}\{K_{\mathbf{H}}(\mathbf{U}_1 - \mathbf{u}), \dots, K_{\mathbf{H}}(\mathbf{U}_n - \mathbf{u})\}$ は $n \times n$ 重み付き対角行列を表す。

補題 6.2 と補題 6.3 からトーラス上のカーネル $\tilde{K}_{\mathbf{H}_q}$ に関して次の性質が成り立つことを示せる。

補題 7.1. n が十分大きいとき、 $\tilde{K}_{\mathbf{H}_q}$ に関する次の 3 つの性質が成り立つ。

(性質 1) 行列 \mathbf{C}_q はバンド幅行列となる。

$$\mathbf{C}_q = \mathbf{H}_q^{1/2} \{1 + o(1)\}.$$

(性質 2) $\mathbf{z}_q = (z_1, \dots, z_q)^T$ とおく。カーネル $\tilde{K}_{\mathbf{H}_q}$ は実数空間上のプロダクトカーネル $\bar{L}(\mathbf{z}_q) := \prod_{j=1}^q \bar{L}(z_j)$ で近似できる。

$$\tilde{K}_{\mathbf{H}_q}(\mathbf{H}_q^{1/2} \mathbf{z}_q) \simeq |\mathbf{H}_q|^{-1/2} \bar{L}(\mathbf{z}_q) + O(\text{Tr}\{\mathbf{H}_q\}).$$

(性質 3) $\bar{L}(\mathbf{z}_q)$ は次式を満たす.

$$\int_{\mathbb{R}^q} \bar{L}(\mathbf{z}_q) d\mathbf{z}_q = 1, \quad \int_{\mathbb{R}^q} \bar{L}(\mathbf{z}_q) \mathbf{z}_q d\mathbf{z}_q = \mathbf{0}, \quad \text{かつ}, \quad \int_{\mathbb{R}^q} \bar{L}(\mathbf{z}_q) \mathbf{z}_q \mathbf{z}_q^T d\mathbf{z}_q = \alpha_2(\bar{L}) \mathbf{I}_q.$$

(性質 4) $\mathbb{J}^q = \{\mathbf{z}_q = (z_1, \dots, z_q)^T | z_j \in [-\pi/h_j, \pi/h_j]\}$ とする. このとき, $\bar{L}(\mathbf{z}_q)$ のモーメントについて,

$$\int_{\mathbb{J}^q} L(\mathbf{z}_q) d\mathbf{z}_q = 1 + (\text{Tr}\{\mathbf{H}_q^{1/2}\}), \quad \int_{\mathbb{J}^q} L(\mathbf{z}_q) \mathbf{z}_q d\mathbf{z}_q = \mathbf{0}, \quad \text{かつ},$$

$$\int_{\mathbb{J}^q} L(\mathbf{z}_q) \mathbf{z}_q \mathbf{z}_q^T d\mathbf{z}_q = \alpha_2(\bar{L}) \mathbf{I}_q + o(\mathbf{H}_q^{1/2})$$

が成り立つ.

(性質 5) すくなくとも 1 つの成分 θ_j が $\theta_j \geq \pi/2$ を満たせば,

$$\tilde{K}_{\mathbf{H}}(\mathbf{u}) = o_p(\text{Tr}\{H\})$$

が成り立つ.

□

補題 7.1 の証明は自明なので省略する. 補題 7.1 からトーラス上のカーネル $\tilde{K}_{\mathbf{H}_q}$ は漸近的には実数空間上のプロダクトカーネルで近似できることが分かる. つまり, 実数・トーラス上のカーネル $K_{\mathbf{H}}$ は $d+q$ 次元の実数空間上のプロダクト・カーネルと見なせる.

補題 7.1 から実数・トーラス上の LLR についての重み付条件付 MISE は, 次の定理として与えられる.

定理 7.3. 簡便化のために $\mathbf{H} = h\mathbf{I}_{d+q}$ とおく. 次の仮定:

1. $n \rightarrow \infty$ のとき, $h \rightarrow 0$ かつ $n^{-1}h^{d+q} \rightarrow 0$ が成り立つ.
2. ヘッセ行列 $\mathcal{H}_m(\mathbf{u})$ は, 連続である.
3. 周辺密度 $f(\mathbf{u}) > 0$ は連続微分可能であり, 条件付分散 $v(\mathbf{u})$ は連続である.

を満たし, かつ, n が十分大きいならば, 漸近的な MISE は次式となる.

$$\text{AMISE}_Y[\hat{m}(\mathbf{u}; \mathbf{H}) | \mathbf{U}] = \frac{h^4 \int_{\mathbb{R}^d \times \mathbb{T}^q} \text{Tr}\{\boldsymbol{\alpha} \mathcal{H}_m(\mathbf{u})\}^2 f(\mathbf{u}) d\mathbf{u}}{4} + \frac{R(\bar{K})^d R(\bar{L})^q \int_{\mathbb{R}^d \times \mathbb{T}^q} v(\mathbf{u}) d\mathbf{u}}{nh^{d+q}}, \quad (7.12)$$

ただし, $\boldsymbol{\alpha} := \text{diag}\{\alpha_2(\bar{K}), \dots, \alpha_2(\bar{K}), \alpha_2(\bar{L}), \dots, \alpha_2(\bar{L})\}$ である.

(7.12) を最小にする $h_* = O_p(n^{-1/(d+q+4)})$ を選ぶと $\text{MISE}_Y[\hat{m}(\mathbf{x}; \mathbf{H})] = O_p(n^{-4/(d+q+4)})$ となる. □

定理 7.3 の証明の詳細に関しては Appendix S を参照すること.

定理 7.3 は, 実数・トーラス上の LLR の MISE は次元の数 d が上がるにつれて収束レートが遅くなることを示している. つまり, 実数・トーラス上の LLR は次元の呪いの性質を持っている.

7.3 実データ分析への応用

前節で述べた実数・トーラス上の LLR を実データ分析に応用する. 実データ分析に適用したときの実数・トーラス上の LLR の性質を調べるために, 実数・トーラス上の LLR とパラメトリック回帰 (Johnson and Wehrly (1978) の回帰モデル) を比較する.

ここで用いる実データは、金沢大学の角間キャンパスで 2015 年 7 月と 2016 年の 7 月の 1 か月間 (休日を除く) に 1 時間毎に計測された電力需要データ ($n = 1008$) である。電力需要データは次のような 1 つの目的変数と 2 つの説明変数を含む。

1. 目的変数 電力需要量 (kWh) Y .
2. 説明変数:
 - (a) 気温 ($^{\circ}\text{C}$) X .
 - (b) 時間 (1~24 時) Z . ただし, Z の周期の長さを 2π に合わせるために, Z を $\Theta = \pi Z/12$ に変換する.

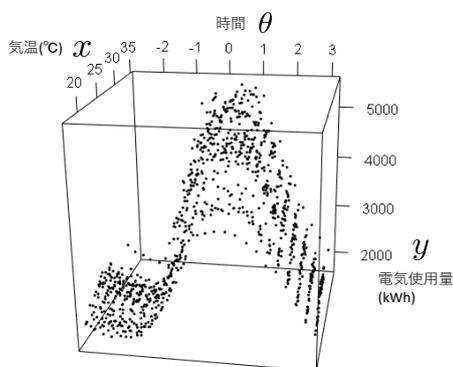


図 12: 電力需要データのプロット.

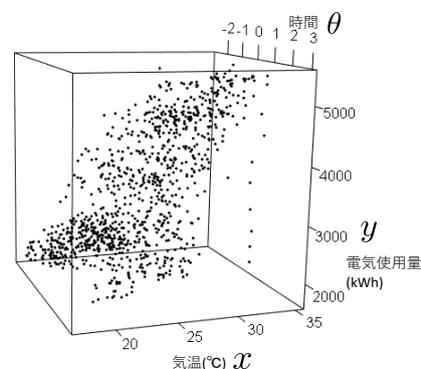


図 13: 電力需要データのプロット.

図 12 と図 13 は電力需要データをプロットした図である。図 12 は電力需要量 Y は時間 Θ に応じて周期的に変動していることを示している。また、図 13 から気温 X と電力需要量 Y は比例関係であることが予想される。

次に分析で使用するパラメトリック回帰は、Johnson and Wehrly (1978) の回帰モデル:(1.33) の回帰係数を最小二乗法で推定した次式で与えられる。

$$\hat{Y}_i = \hat{a} + \hat{\beta}_1 X_i + \hat{\beta}_2 \cos(\Theta_i) + \hat{\beta}_3 \sin(\Theta_i), \quad (7.13)$$

ただし, $\hat{a} = 1351.10$, $\hat{\beta}_1 = 76.94$, $\hat{\beta}_2 = 839.29$ と $\hat{\beta}_3 = 879.53$ である。

次に分析で使用する実数・トーラス上の LLR $\hat{m}(x, \theta; h\mathbf{I}_2)$ (これは (7.11) に関して $d = 1, q = 1$ としたと気に対応する) は以下の設定の下で与えられる。

1. カーネル関数は正規カーネルと VM カーネルから構成される 2 変量プロダクトカーネルを採用する。
2. バンド幅の推定量は LSCV 推定量 $\tilde{h}_{CV} = 0.236$ を用いる (\tilde{h}_{CV} は (6.15) の $\hat{m}_{-i}(\theta; p, h)$ を $\hat{m}_i(X_i, \Theta_i; h\mathbf{I}_2)$ に置き換えた上で (6.14) と同様な方法で求める)。
3. 求める推定量は $\hat{m}(x, \theta; \tilde{h}_{CV}\mathbf{I}_2)$ で与えられる。

図 14 と 15 を比較すると、パラメトリック回帰は時間による周期的変動と気温・電力需要量の比例関係を表現できている。これは電力需要データの全体的な傾向を捉えることに成功したと言える。実数・トーラス上の

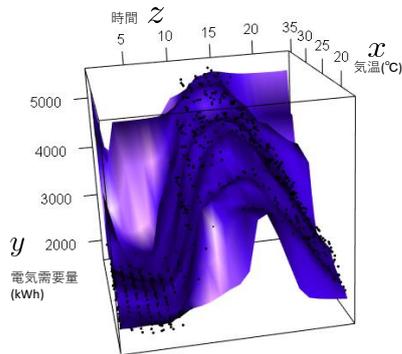


図 14: 電力需要データに実数・トーラス上の LLR $\hat{m}(x, \theta; \tilde{h}_{CV} \mathbf{I}_2)$ を適用した結果. 直感的に理解できるように時間変数 Θ を Z に変換した.

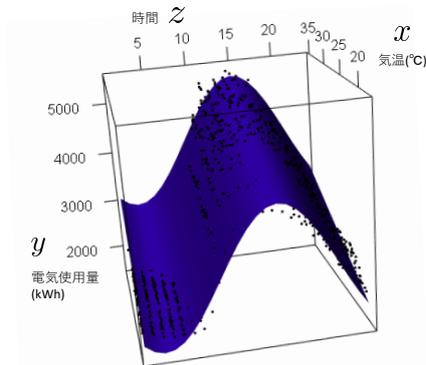


図 15: 電力需要データに (7.13) で与えたパラメトリック回帰を適用した結果. 直感的に理解できるように時間変数 Θ を Z に変換した.

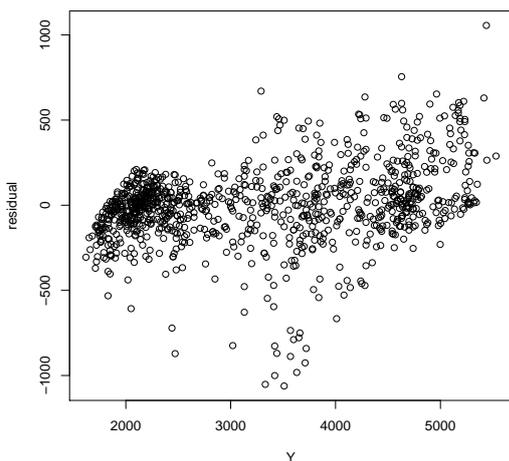


図 16: 実数・トーラス上の LLR $\hat{m}(x, \theta; \tilde{h}_{CV} \mathbf{I}_2)$ と目的変数 Y の残差. 決定係数: $R^2 = 0.957$.

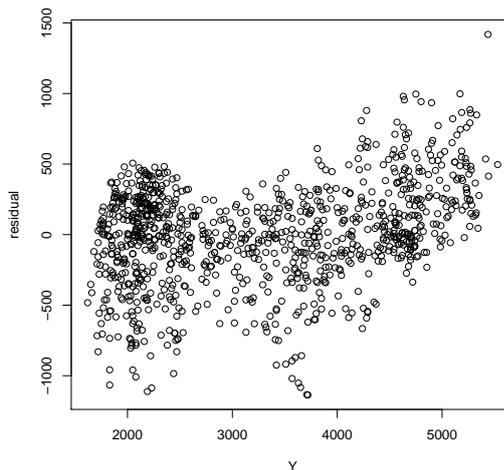


図 17: パラメトリック回帰 \hat{Y} と目的変数 Y の残差. 決定係数: $R^2 = 0.897$.

回帰はパラメトリック回帰のような全体的な傾向に加えて局所的な傾向も表現できていると考えられる. 例えば, 図 14 は朝方はほぼ一定であった電力需要量が大学の授業が始まる 9 時から 10 時頃にかけて急激に上昇していることが分かる. また, 図 14 には 12 時頃かつ気温が約 30 度のとき, 電力需要量の局所的な山がある. これは, 冷房は設定温度まで室温を下げようとして使用開始時に最も電気使用量が高くなるので, 暑くなって冷房を一斉に入れることで一時的に電力需要量が高くなったと解釈できる.

図 14 と 15 は実数・トーラス上の回帰またはパラメトリック回帰と電力需要量 Y の残差をプロットした図である. これらの図を比較すると, パラメトリック回帰と比べて, 実数・トーラス上の回帰は, 電力需要量が約 3000kWh 以下のときに残差が小さくなっていることが分かる.

以上の議論は直感的なものであるので, 回帰モデルの決定係数の値を比較することで両者の回帰モデルの説

明力を見ていくことにする.

$\bar{Y} = n^{-1} \sum_i Y_i$ とおくと, パラメトリック回帰 \hat{Y} の決定係数 R^2 は

$$R^2 := \frac{\left\{ \sum_{i=1}^n (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y}) \right\}^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2 \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \quad (7.14)$$

と定義される. 実数・トーラス上の LLR の決定係数 R^2 は (7.14) の \hat{Y}_i を $\hat{m}(X_i, \Theta_i; \tilde{h}_{CV} \mathbf{I}_2)$ で置き換えることで得られる. 実数・トーラス上の LLR の決定係数は $R^2 = 0.957$ であるが, パラメトリック回帰の決定係数は $R^2 = 0.897$ である. つまり, 実数・トーラス上の LLR はより説明力が高く, データの変動を説明できている. 実数・トーラス上の LLR の決定係数が高いのは, 上で述べたように実数・トーラス上の LLR はデータの複雑な構造を表現しているからである.

7.4 先行研究との比較

定理 7.3 は実数・トーラス $\mathbb{R}^d \times \mathbb{T}^q$ 上の LLR が $\text{MISE}_Y[\hat{m}(\mathbf{x}; \mathbf{H})] = O_p(n^{-4/(d+q+4)})$ を達成することを示している. この収束レートから説明変数ベクトルに角度変数を含む LLR も次元の呪いの特徴を持つことが分かる. そして, この結果から表 16 で示したように, トーラス上の LLR や実数・円周上の LLR に関する MISE の収束レートも導ける.

表 16: 多変量局所線形回帰 (LLR) の MISE の収束レート.

多変量 LLR の種類	標本空間	MISE の収束レート
実数・トーラス上の LLR	$\mathbb{R}^d \times \mathbb{T}^q$	$O_p(n^{-4/(d+q+4)})$
トーラス上の LLR	\mathbb{T}^q	$O_p(n^{-4/(q+4)})$
実数空間上の LLR	\mathbb{R}^d	$O_p(n^{-4/(d+4)})$
実数・円周上の LLR	$\mathbb{R}^d \times \mathbb{T}$	$O_p(n^{-4/(d+5)})$

8 結論

8.1 はじめに

本稿の研究目的は、方向統計学におけるノンパラメトリック統計解析の理論を整備することである。そのためにカーネル密度推定量 (KDE) とノンパラメトリック回帰の一種である局所多項式回帰 (LPR) の理論的性質に関する議論を行った。本稿の主要なアイデアは Hall et al.(1987) が与えた Hall 型カーネル族 (定義 3.1) をカーネル関数として採用したことである。本稿では、実数空間上の標準的なノンパラメトリック統計解析の理論と Hall 型カーネル族を組み合わせることで、この 2 つの手法に関する未解明だった多くの性質を導いた。2 節では方向統計学における KDE に関する主要な結果をまとめ、3 節では、方向統計学における LPR に関する主要な結果を整理する。最後に今後の研究の展望と課題について議論する。

8.2 方向統計学におけるカーネル密度推定量

2 章から 4 章では方向統計学における KDE に関する議論を行った。その結果を表 17 でまとめている。(1) から (3) が議論したい理論的性質である。1.2 節で説明したが、実数直線上で与えられた標準的な KDE に

表 17: 方向統計学における KDE の性質の比較。○：当てはまる，×：当てはまらない，△：VM カーネルのみ当てはまる。

KDE	標準	Di Marzio 型	Hall 型
標本空間	\mathbb{R} (実数)	\mathbb{T} (円周)	\mathbb{T}
(1) p 次オーダーカーネル \Rightarrow $\text{MISE} = O(n^{-2p/(2p+1)})$	○	×	○
(2) 2 次オーダーカーネルから高次オーダー KDE を構成	○	△	○
(3) 平滑化パラメータ推定量の収束レートを導出可能	○	×	○

関する理論的性質を 3 つ簡単に述べる。1 つ目はカーネル K のモーメント $\alpha_j(K)$ に関して $0 < j < p$ 次の $\alpha_j(K)$ が 0 となるような p 次オーダーカーネルを用いることで、KDE の平均積分二乗誤差 (MISE) は $\text{MISE} = O(n^{-2p/(2p+1)})$ を達成可能なことである。ちなみに、MISE とは KDE と真の密度の間の誤差を測る誤差基準の一つを指す。本稿では $\text{MISE} = O(n^{-2p/(2p+1)})$ を達成可能な KDE を p 次オーダー KDE と呼んでいる。2 つ目は Jones and Foster (1993) による加法型構成法や Terrell and Scott (1980) による乗法型構成法を用いれば、2 次オーダーカーネル (または KDE) から、高次オーダー KDE を構築できることである。3 つ目は KDE の平滑化パラメータに関する推定量の収束レートが導出でき、最適なパラメータへの一致性が保証されている点である。収束レートが導出されている代表的な推定量として最小二乗クロスバリデーション法 (LSCV) やダイレクト・プラグイン法 (DPI) が挙げられる。

2 章では Di Marzio et al.(2011) が提案した Di Marzio 型 KDE の理論的性質に関する議論を行った。Di Marzio 型 KDE とは、定義 2.1 で与えた Di Marzio 型カーネル族を採用した KDE のことを言う。その特徴はカーネルのモーメントを sine 型モーメント $\eta_j(K_\kappa)$ とし、低次の sine 型モーメントが 0 となる sine 型 p 次オーダーカーネルを用いた点にある。

2 章の中心的成果は、代表的な円周上の密度関数の一つである巻き込みコーシー (WC) カーネルを採用した KDE の漸近正規性とその MISE を導出したことである。Sine 2 次オーダーカーネルである WC カーネル

は、その MISE の収束レートが $O(n^{-2/3})$ となる。これは 0 次オーダー KDE の収束レートに過ぎない。同じ sine 2 次オーダーカーネルに含まれるフォン・ミーゼス (VM) カーネルの MISE の収束レートは $O(n^{-4/5})$ であり、2 次オーダー KDE の収束レートである。両者の収束レートの違いは、同じ sine 2 次オーダーカーネルであってもその MISE の収束レートは異なることを示している。Di Marzio et al.(2011) 自身も高次の sine 型 p 次オーダーカーネルの中にバイアスを改良できないものが含まれていることを指摘している。この 2 つのことから表 17 の (1) の性質が成り立たないことが分かる。また、Di Marzio et al.(2011) は VM カーネルに関しては 4 次以上の高次オーダー KDE を構築可能であることを示している。しかし、ほかのカーネルに関しても (2) の性質が成り立つかは分かっていない。方向統計学の KDE において平滑化パラメータに関する理論的性質を議論した研究は、筆者の知る限り存在しない。

3 章と 4 章の目的は、表 17 で示した 3 つの性質を満たす方向統計学における KDE を考案することである。筆者は Di Marzio 型 KDE の問題点は Di Marzio 型カーネルと sine p 次オーダーカーネルを採用したことにあると考えて、Hall 型カーネル族 (定義 3.1) を採用した Hall 型 KDE を提案した。

3 章は Hall 型 KDE の MISE の改良方法について議論した。Hall 型カーネル族の良い性質は、KDE のバイアスを高次の項まで展開したとき、その各項が平滑化パラメータ κ の級数 κ^l と Hall 型カーネル族のモーメント $\mu_l(L)$ の積から構成される点である。したがって、 $0 < l < p$ 次のモーメント $\mu_l(L)$ が 0 となるような新しい p 次オーダーカーネルを定義することで、 p 次オーダーカーネルに関するバイアスを改良できる。そのため、 p 次オーダーカーネルを採用したときの MISE は $\text{MISE} = O(n^{-2p/(2p+1)})$ となる。また、Hall 型 KDE は、加法型構成法と乗法型構成法を 2 次オーダーカーネル (KDE) に適用することで高次オーダー KDE を構築できる。以上の議論から Hall 型 KDE は表 17 の (1) と (2) の性質を満たすことが分かる。

4 章では Hall 型 KDE に関する平滑化パラメータ推定量の理論的性質を議論した。4 章で扱った推定量は LSCV 推定量と DPI 推定量である。LSCV 推定量の収束レートは $O(n^{-1/10})$ となり、DPI 推定量の収束レートは $O(n^{-4/15})$ となることを明らかにした。3 章と 4 章の議論から Hall 型 KDE は表 17 の 3 つの性質をすべて満たすことが示された。

8.3 方向統計学におけるノンパラメトリック回帰

5 章から 7 章までは方向統計学における LPR の議論を行った。その概要を表 18 に載せている。1.3 節で述べた実数空間上で与えられた標準的な LRR に関する理論的性質について簡単に述べたい。標準的な LRR は誤差基準として重み付条件付 MISE を用いているが、重み付条件付 MISE のことを単に MISE と表すことにする。まず、標準的な LRR に関して説明変数がスカラーであるときの理論的性質を述べる。標準的な LRR は多項式の次数 p を上げることで MISE の収束レートを改善できる。1 次のときの LPR を局所線形回帰 (LLR) と呼ぶ。LLR の MISE の収束レートは $O_p(n^{-4/5})$ に過ぎないが、例えば次数 p が奇数のときは $\text{MISE} = O_p(n^{-(2p+2)/(2p+3)})$ となり、次数 p を上げることで MISE を改良可能である。詳細は省略するが、次数 p が偶数のときも MISE を改良できる。標準的な LRR は説明変数が d 次元ベクトルであるとき、その MISE の収束レートは $O_p(n^{-4/(d+4)})$ となる。つまり、説明変数の次元 d が増えるごとに MISE の収束レートは遅くなる。この性質のことを一般的に次元の呪いと呼ぶ。

5 章では方向統計学における局所多項式回帰の一つである sine 型 LPR を扱った。主に、Di Marzio et al. (2009) が与えた Di Marzio 型 LPR の理論的性質について議論している。Di Marzio 型 LPR は、Di Marzio 型カーネルを採用した sine 型 LPR のことを指す。特に、5 章は説明変数がスカラーな角度変数であるときの Di Marzio 型 LLR ($p = 1$ の場合) に VM カーネルと WC カーネルを適用したときの理論的性質を与えた。

表 18: 標準的な LLR と Di Marzio 型 LPR の性質の比較. ○ : 当てはまる, × : 当てはまらない, △ : VM カーネルのみ当てはまる.

LPR	標準	Di Marzio 型	Hall 型
(4) $p = 1 \Rightarrow \text{MISE} = O_p(n^{-4/5})$	○	△	○
(5) 高次の $p \Rightarrow \text{MISE}$ の収束レートの改善	○	×	○
(6) 次元の呪い (MISE が次元数に依存)	○	×	○

Di Marzio 型 LPR は VM カーネルを適用した場合に $\text{MISE} = O_p(n^{-4/5})$ となり, このとき, 表 18 の (4) の性質を満たす. 一方, WC カーネルを適用したときの MISE の収束レートは $O_p(n^{-2/3})$ となり, これは (4) の性質を満たさない. また, 高次の場合に関する Di Marzio 型 LPR の MISE の収束レートは未解決問題であり, (5) の性質を満たさない. 7.1 節で議論したように Di Marzio 型 LLR は, 説明変数がトーラス \mathbb{T}^q 上で与えられるベクトルや実数・円周 $\mathbb{R}^d \times \mathbb{T}$ で与えられるとき, その MISE の収束レートは不明であり, (6) の性質を満たさない.

6 章では, sine 型 LPR に非負の値を取る Hall 型カーネル族を適用した Hall 型 LPR の理論的性質を議論した. 6 章では 5 章と同様に説明変数がスカラーな角度変数であるものを扱った. 本稿では, Hall 型カーネル族が実数直線上のカーネルに分布収束する性質を明らかにした. この性質を用いることで Hall 型 LPR の MISE を導出し, その収束レートを計算できる. 実際に次数 p が奇数のとき, MISE の収束レートは $O_p(n^{-(2p+2)/(2p+3)})$ となる (p が偶数の場合は省略). つまり, $p = 1$ のときは $\text{MISE} = O_p(n^{-4/5})$ であるが, 高次の p を選ぶことで MISE の収束レートを改善できる. したがって, Hall 型 LPR は表 18 の (4) と (5) の性質を満たす.

7 章では方向統計学における多変量 LLR の理論的性質の解明に取り組んだ. 対象としたのは, 説明変数が実数・トーラス $\mathbb{R}^d \times \mathbb{T}^q$ 上の $d + q$ 次元ベクトルある実数・トーラス上の LLR である. 実数・トーラス上の LLR は, Di Marzio et.al(2009) のトーラス \mathbb{T}^q 上の LLR や Qin et al. (2011) の実数・円周 $\mathbb{R}^d \times \mathbb{T}$ 上の LLR を含む一般的な定義であるという特徴がある. 実数・トーラス上の LLR 上に Hall 型カーネル族を適用した LLR の MISE の収束レートは $O_p(n^{-4/(d+q+4)})$ となる. 実数・トーラス上の LLR は一般的な定義であるので, この結果は, 表 16 でまとめたように, トーラス上・実数空間上・実数円周上の LLR の MISE の収束レートを含んだ一般的な形である. つまり, 方向統計学における多変量 LLR も次元の呪いの性質を持つ. したがって, Hall 型 LPR は (6) の性質も満たすことが分かる.

また, 7.3 節で実数・トーラス上の LLR を金沢大学角間キャンパスの電力需要データに応用する実データ分析を行った. 比較するために Johnson and Wehrly (1978) のパラメトリック回帰を採用した. 実数・トーラス上の LLR は, パラメトリック回帰と比べて, データが持つ局所的な傾向も表現できる柔軟なモデルを与えることができた. そのため, 実数・トーラス上の LLR は, パラメトリック回帰よりも決定係数 R^2 の値が高い. この結果から, 実数・トーラス上の LLR は, パラメトリック回帰よりも説明力が高い回帰モデルを構築可能なことが分かる.

8.4 今後の研究の展望と課題

本稿では, Hall 型カーネル族を用いることで方向統計学におけるノンパラメトリック統計解析手法の理論の整備を行った. しかし, Hall 型 LPR の平滑化パラメータ推定量の理論的な性質を与えていないので, この性

質を導出することが一番の課題である。また、KDE の課題として、本稿で挙げていない方法で MISE を改良することが考えられる。例えば、平滑化パラメータを観測値の集中具合に応じて可変的に動かすなどの方法があるだろう。

他の理論的な課題として KDE や LPR は各点の推定値を計算するために $O(n)$ の計算時間がかかることが挙げられる。データを $m(m < n)$ 個のビンに分けたヒストグラムに変換し、 m 個のビンに対してカーネル関数を適用することで両者の計算時間のオーダーを $O(m)$ に改善できると予想される。

本稿で議論した理論的特性から、方向統計学におけるノンパラメトリック統計解析手法は、周期的な変動を持つ経済データ分析全般に適用できることが期待される。例えば、24 時間稼働の物流拠点における注文数の周期的な変動を解明することで人員の最適な配置を与えたり、コンビニなどの小売りにおける来店者数や購買数の周期的な変動を予測することで廃棄ロスを減少させたりすることが挙げられる。本稿で提案した手法を周期的な変動を持つ経済データの解析に適用することで、方向統計学におけるノンパラメトリック統計解析手法がどのような経済現象の実証分析に貢献できるかを示すことが必要であろう。

本稿で議論した電力需要データや今述べた例は同じ周期の中で変動する時系列データである。周期的な変動を持つ経済データの多くはこのような周期性を持つ時系列データであろう。計量経済学の視点から見たときの今後の研究課題は、方向統計学におけるノンパラメトリック統計解析手法を周期性を持つ時系列データの分析に応用できる可能性を、数理統計学的なアプローチで探ることである。

参考文献

- [1] Abe, T., Pewsey, A. (2011). Sine-skewed circular distributions. *Statistical Papers*, **52**, 683–707.
- [2] Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353–360.
- [3] Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, **31**, 131–145.
- [4] Chen, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, **52**, 471–480.
- [5] Davis, K. B. (1975). Mean square error properties of density estimates. *The Annals of Statistics*, **3**, 1025–1030.
- [6] Di Marzio, M., Panzera, A. and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statistics & Probability Letters*, **79**, 2066–2075.
- [7] Di Marzio, M., Panzera, A. and Taylor, C. C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, **141**, 2156–2173.
- [8] Feller, M. (1966). *An Introduction to Probability Theory and Its Applications. II (Second ed.)*, John Wiley & Sons, Inc.
- [9] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis*, **14**, 1–16.
- [10] Hall, P., and Marron, J. S. (1987). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probability Theory and Related Fields*, **74**, 567–581.
- [11] Hall, P., Watson, G. S. and Cabrera, J. (1987). Kernel density estimation with spherical data.

- Biometrika*, **74**, 751–762.
- [12] Igarashi, G., and Kakizawa, Y. (2014). Re-formulation of inverse Gaussian, reciprocal inverse Gaussian, and Birnbaum-Saunders kernel estimators. *Statistics & Probability Letters*, **84**, 235–246.
- [13] 五十嵐岳, 柿沢佳秀. (2012). ベータカーネル推定量, ガンマカーネル推定量, ベルンシュタイン推定量の改良について. 「第 13 回ノンパラメトリック統計解析とベイズ統計」予稿集, 83–92.
- [14] Jammalamadaka, S. R. and senGupta, A. (2001). *Topics in Circular Statistics*. World Scientific.
- [15] Jones, M. C. and Foster, P.J. (1993). Generalized jackknifing and higher order kernels. *Journal of Nonparametric Statistics* **3**, 81–94.
- [16] Johnson, R. A., and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, **73**, 602–606.
- [17] Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computational Statistics & Data Analysis*, **14**, 457–471.
- [18] 加藤昇吾. (2017). 円周上のコーシー分布と関連した統計モデル. 日本統計学会誌, **46**, 85–111.
- [19] Mardia, K. V. (1972). *Probability and mathematical statistics: statistics of directional data*. London: Academic.
- [20] Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. John Wiley & Sons, London.
- [21] Qin, X., Zhang, J. S., and Yan, X. D. (2011). A nonparametric circular-linear multivariate regression model with a rule-of-thumb bandwidth selector. *Computers & Mathematics with Applications*, **62**, 3048–3055.
- [22] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65–78.
- [23] Ruppert, D., and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The annals of statistics*, **22**, 1346–1370.
- [24] 清水邦夫 (2006). 方向統計学の最近の発展. 計算機統計学, **19**, 127–150.
- [25] 清水邦夫, 王敏真. (2013). 環境科学における方向統計学の利用. 統計数理, **61**, 289–305.
- [26] Scott, D. W., and Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, **82**, 1131–1146.
- [27] Sheather, S. J., and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, **53**, 683–690.
- [28] Terrell, G. R. and Scott, D. W. (1980). On improving convergence rates for nonnegative kernel Density Estimators. *The Annals of Statistics*, **8**, 1160–1163.
- [29] Tsuruta, Y. and Sagae, M. (2017a) Asymptotic Property of Wrapped Cauchy Kernel Density Estimation on the Circle. *Bulletin of Informatics and Cybernetics* (掲載決定).
- [30] Tsuruta, Y. and Sagae, M. (2017b) Higher order kernel density estimation on the circle. *Statistics & Probability Letters*, **131**, 46–50.
- [31] Tsuruta, Y. and Sagae, M. (2017c) Theoretical Properties of Bandwidth Selectors for Kernel Density Estimation on the Circle. *Discussion Paper, No.34 in school of Economics, Kanazawa University* (投稿中).
- [32] Tsuruta, Y. and Sagae, M. (2017d) Asymptotic Properties of Circular Nonparametric Regression by

applying Von Mises and Wrapped Cauchy Kernels. *Discussion Paper, No.35 in school of Economics, Kanazawa University.*

[33] Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall, London.

A 付録

A1 Appendix A

証明. (定理 2.4) Put $h = (1 - \rho^2)$, Write (3.8) and (3.12) as

$$\mathbb{E}_f[K_h(\theta - \Theta_1)] \simeq f(\theta) + \frac{1}{4}hf''(\theta) + o(1), \quad (\text{A.1})$$

$$\text{Var}_f[K_h(\theta - \Theta_1)] = \frac{f(\theta)}{\pi h} + o(h^{-1}). \quad (\text{A.2})$$

$$\sqrt{nh}[\hat{f}_h(\theta) - f(\theta)] = \sqrt{nh}\{\hat{f}_h(\theta) - \mathbb{E}[\hat{f}_h(\theta)]\} + \sqrt{nh}\text{bias}[\hat{f}_h(\theta)]. \quad (\text{A.3})$$

The first term of (A.3) is equal to

$$\sqrt{nh}\{\hat{f}_h(\theta) - \mathbb{E}_f[\hat{f}_h(\theta)]\} = \sqrt{n} \left\{ n^{-1} \sum_{i=1}^n h^{1/2} K_h(\theta - \Theta_i) - \mathbb{E}_f[h^{1/2} K_h(\theta - \Theta_1)] \right\}. \quad (\text{A.4})$$

$$\begin{aligned} \mathbb{E}_f[h^{1/2} K_h(\theta - \Theta_1)] &= h^{1/2} \mathbb{E}_f[K_h(\theta - \Theta_1)] \\ &\simeq h^{1/2} \left[f(\theta) + \frac{f''(\theta)}{4} h \right]. \end{aligned} \quad (\text{A.5})$$

It is shown that $0 \leq |\mathbb{E}[h^{1/2} K_h(\theta - \Theta_1)]| < \infty$ from (A.5).

From (A.2), We obtain the following form:

$$\begin{aligned} \text{Var}_f[h^{1/2} K_h(\theta - \Theta_1)] &= h \text{Var}[K_h(\theta - \Theta_1)] \\ &= h \left[\frac{f(\theta)}{\pi h} + o(h^{-1}) \right] \\ &= \frac{f(\theta)}{\pi} + o(1) \\ &\simeq \frac{f(\theta)}{\pi}. \end{aligned} \quad (\text{A.6})$$

It is shown that $\text{Var}_f[h^{1/2} K_h(\theta - \Theta_1)] < \infty$ from (A.6).

Since (A.4) satisfies the condition of Lindeberg (Feller (1968, p.244)) from (A.5) and (A.6), it is given as the follows

$$\sqrt{nh}\{\hat{f}_h(\theta) - \mathbb{E}_f[\hat{f}_h(\theta)]\} \xrightarrow{d} N(0, f(\theta)/\pi), \quad n \rightarrow \infty. \quad (\text{A.7})$$

The order of the second term of (A.3) is equal to,

$$\begin{aligned} \sqrt{nh}\text{bias}_f[\hat{f}_h] &= \sqrt{nh}O(h) \\ &= O(\sqrt{nh^3}). \end{aligned} \quad (\text{A.8})$$

With $h = cn^{-\alpha}$, we obtain the follows,

$$\begin{aligned}\sqrt{nh^3} &\sim n^{1/2}n^{-3\alpha/2} \\ &= n^{(1-3\alpha)/2}.\end{aligned}$$

When $\alpha > 1/3$ is chosen, then (A.8) is given as the following form:

$$\begin{aligned}\sqrt{nh}\text{bias}_f[\hat{f}_h(\theta)] &= O(\sqrt{nh^3}) \\ &= o(1).\end{aligned}\tag{A.9}$$

For $\alpha > 1/3$ and as $n \rightarrow \infty$, Theorem 2.4 completes the proof from (A.7) and (A.9). \square

A2 Appendix B

補題 B.1. Let $g_j(r/\kappa) := \{2 - r/\kappa\}^{(j-1)/2}$ for $j \geq 0$. Then, $C_\kappa(L)$ is given as

$$C_\kappa(L) = 2 \sum_{m=0}^{v/2} \frac{g_0^{(m)}(0)}{m!} \kappa^{-(2m+1)/2} \mu_{2m}(L) + O(\kappa^{-(v+3)/2}).\tag{B.1}$$

If K_κ is a p th-order kernel, then the term $C_\kappa(L)$ is reduced to

$$C_\kappa(L) = \kappa^{-1/2} 2^{1/2} \mu_0(L) + O(\kappa^{-(p+1)/2}).\tag{B.2}$$

証明. The Taylor expansion of $g_j(r/\kappa)$ is given by

$$\begin{aligned}g_j(r/\kappa) &= g_j(0) + g_j^{(1)}(0)r/\kappa + \dots + \frac{g_j^{(v/2)}(0)}{(v/2)!} (r/\kappa)^{v/2} \\ &\quad + \frac{g_j^{(v/2+1)}(0)}{(v/2+1)!} (r/\kappa)^{v/2+1} + \dots \\ &= \sum_{m=0}^{v/2} (m!)^{-1} g_j^{(m)}(0) (r/\kappa)^m + O(\kappa^{-(v+2)/2}).\end{aligned}\tag{B.3}$$

By combing (ii), (B.3), and the expression $d\theta/dr = \{r\kappa(2 - r/\kappa)\}^{-1/2}$, it follows that

$$\begin{aligned}C_\kappa(L) &= 2 \int_0^{2\kappa} L(r) (r\kappa)^{-1/2} \{2 - r/\kappa\}^{-1/2} dr \\ &= 2 \sum_{m=0}^{v/2} \frac{g_0^{(m)}(0)}{m!} \kappa^{-(2m+1)/2} \int_0^{2\kappa} L(r) r^{(2m-1)/2} dr + O(\kappa^{-(v+3)/2}) \\ &= 2 \sum_{m=0}^{v/2} \frac{g_0^{(m)}(0)}{m!} \kappa^{-(2m+1)/2} \mu_{2\kappa, 2m}(L) + O(\kappa^{-(v+3)/2}) \\ &= 2 \sum_{m=0}^{v/2} \frac{g_0^{(m)}(0)}{m!} \kappa^{-(2m+1)/2} \mu_{2m}(L) + O(\kappa^{-(v+3)/2}).\end{aligned}\tag{B.4}$$

If K_κ is a p th-order kernel, then it follows from (B.4) that $C_\kappa(L)$ is equal to

$$C_\kappa(L) = 2^{1/2} \mu_0(L) \kappa^{-1/2} + O(\kappa^{-(p+1)/2}).$$

\square

補題 B.2. Set $\alpha_j(K_\kappa) := \int_{-\pi}^{\pi} K_\kappa(\theta)\theta^j d\theta$ and $a_s := (2s-2)!/\{(2s-1)!!s\}$. Furthermore, let $z \geq 0$ be even and $2t \leq z \leq v$. Then, the term $\alpha_{2t}(K_\kappa)$ is given as

$$\alpha_{2t}(K_\kappa) = 2C_\kappa^{-1}(L)\kappa^{-1/2} \sum_{q=t}^{z/2} \sum_{m=0}^{z/2-q} \kappa^{-(q+m)} A_q(z,t)(m!)^{-1} g_{2q}^{(m)}(0)\mu_{2(q+m)}(L) + O(\kappa^{-(z+2)/2}), \quad (\text{B.5})$$

where $A_q(z,t) := \sum_{\sum_{s=1}^{z/2} t_s=t, \sum_{s=1}^{z/2} s t_s=q} \frac{t!}{t_1!t_2!\cdots t_{z/2}!} \prod_{l=1}^{z/2} a_l^{t_l}$.

If K_κ is a p th-order kernel, then the equation (B.5) is reduced to

$$\alpha_{2t}(K_\kappa) = b_{p,2t}\mu_0^{-1}(L)\mu_p(L)\kappa^{-p/2} + O(\kappa^{-(p+2)/2}), \quad 0 < 2t \leq p,$$

where $b_{p,2t} = 2^{1/2} \sum_{q=t}^{p/2} A_q(p,t)(\{p/2-q\}!)^{-1} g_{2q}^{(p/2-q)}(0)$.

The term $\alpha_{p+2}(K_\kappa) = O(\kappa^{-(p+2)/2})$ follows from (B.5).

証明. Note that $\sin^2(\theta) = r/\kappa(2-r/\kappa)$. The Taylor expansion of $\theta^2 = \arcsin^2(\{r/\kappa(2-r/\kappa)\}^{1/2})$ for $0 \leq \theta < \pi/2$ is given by

$$\theta^2 = \sum_{s=1}^{z/2} a_s \{r/\kappa(2-r/\kappa)\}^s + O(\kappa^{-(z+2)/2}), \quad 0 \leq \theta < \pi/2, \quad (\text{B.6})$$

where $a_s := (2s-2)!/\{(2s-1)!!s\}$. Taking the t th power of (B.6) gives

$$\begin{aligned} \theta^{2t} &= \left[\sum_{s=1}^{z/2} a_s \{r/\kappa(2-r/\kappa)\}^s + O(\kappa^{-(z+2)/2}) \right]^t \\ &= \sum_{q=t}^{z/2} A_q(z,t) \{r/\kappa(2-r/\kappa)\}^q + O(\kappa^{-(z+2)/2}), \quad 0 \leq \theta < \pi/2. \end{aligned} \quad (\text{B.7})$$

We show that $\int_{\pi/2}^{\pi} K_\kappa(\theta)\theta^j d\theta$ can be ignored, because

$$\begin{aligned} \int_{\pi/2}^{\pi} K_\kappa(\theta)\theta^j d\theta &< \pi^j \int_{\pi/2}^{\pi} K_\kappa(\theta) d\theta \\ &= \pi^j C_\kappa^{-1}(L) \int_{\kappa}^{2\kappa} L(r)\kappa^{-1/2} r^{-1/2} dr \{2 + O(\kappa^{-1})\}^{-1/2}. \end{aligned} \quad (\text{B.8})$$

It follows from (ii) that $\int_{\kappa}^{\infty} L(r)r^{-1/2} dr = O(\kappa^{-(v+2)/2})$. This leads to

$$\int_{\kappa}^{2\kappa} L(r)\kappa^{-1/2} r^{-1/2} dr = O(\kappa^{-(v+2)/2}). \quad (\text{B.9})$$

It follows from (B.1), (B.8), and (B.9) that

$$\int_{\pi/2}^{\pi} K_\kappa(\theta)\theta^j d\theta = O(\kappa^{-(v+2)/2}). \quad (\text{B.10})$$

By considering (B.2), (B.7), and (B.10), the term $\alpha_{2t}(K_\kappa)$ is reduced to

$$\begin{aligned}\alpha_{2t}(K_\kappa) &= 2 \int_0^{\pi/2} K_\kappa(\theta) \theta^{2t} d\theta + O(\kappa^{-(v+2)/2}) \\ &= 2C_\kappa^{-1}(L) \int_0^\kappa L(r) \sum_{q=t}^{z/2} A_q(z, t) \{r/\kappa(2 - r/\kappa)\}^q [r\kappa\{2 - r/\kappa\}]^{-1/2} dr + O(\kappa^{-(v+2)/2}) \\ &= 2C_\kappa^{-1}(L) \kappa^{-1/2} \sum_{q=t}^{z/2} \sum_{m=0}^{z/2-q} \frac{A_q(z, t) g_{2q}^{(m)}(0)}{\kappa^{(q+m)} m!} \mu_{2(q+m)}(L) + O(\kappa^{-(z+2)/2}).\end{aligned}$$

If K_κ is a p th-order kernel, then the equation (B.5) for $z = p$ is equal to

$$\begin{aligned}\alpha_{2t}(K_\kappa) &= 2C_\kappa^{-1}(L) \kappa^{-(p+1)/2} \sum_{q=t}^{p/2} A_q(p, t) (\{p/2 - q\}!)^{-1} g_{2q}^{(p/2-q)}(0) \mu_p(L) + O(\kappa^{-(p+2)/2}) \\ &= b_{p,2t} \mu_0^{-1}(L) \mu_p(L) \kappa^{-p/2} + O(\kappa^{-(p+2)/2}).\end{aligned}$$

□

補題 B.3. The terms $R(K(u)u^t)$ for $t = 0, 1$ are reduced to

$$R(K(u)u^t) := \kappa^{-(2t-1)/2} [d_{2t}(L) + o(1)], \quad (\text{B.11})$$

where $d_{2t}(L) := 2^{-1} \mu_0^{-2}(L) \delta_{2t}(L)$ and $d(L) := d_0(L)$.

証明. Set $\delta_{\kappa^{1/2}, 2t}(L) := \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} L^2(z^2/2) z^{2t} dz$. If κ is large, then $\delta_{\kappa^{1/2}, 2t}(L) = \delta_{2t}(L) + o(1)$, because as $\kappa \rightarrow \infty$, $\int_{\kappa^{1/2}\pi}^\infty L(z^2/2) z^{2t} dz = 0$ and $\delta_{2t}(L)$ is bounded from (i). Computing Taylor expansion of $\cos(\kappa^{-1/2}z) = 1 - z^2/(2\kappa) + O(\kappa^{-2})$ implies that $L(\kappa\{1 - \cos(\kappa^{-1/2}z)\}) = L(z^2/2) + O(\kappa^{-1})$. It follows these properties, and (B.2) that

$$\begin{aligned}R(K_\kappa(u)u^t) &= C_\kappa^{-2}(L) \kappa^{-(2t+1)/2} \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} [L(z^2/2) + O(\kappa^{-1})] z^{2t} dz \\ &= C_\kappa^{-2}(L) \kappa^{-(2t+1)/2} [\delta_{2t}(L) + o(1)] \\ &= \kappa^{-(2t-1)/2} [d_{2t}(L) + o(1)].\end{aligned} \quad (\text{B.12})$$

□

A3 Appendix C

証明. (定理 3.2)

It follows from (4) that

$$\sqrt{n\kappa^{-1/2}} [\hat{f}_\kappa(\theta) - f(\theta)] = \sqrt{n\kappa^{-1/2}} \{\hat{f}_\kappa(\theta) - \mathbb{E}_f[\hat{f}_\kappa(\theta)]\} + \sqrt{n\kappa^{-1/2}} \text{bias}[\hat{f}_\kappa(\theta)], \quad (\text{C.1})$$

where $\text{bias}_f[\hat{f}_\kappa(\theta)] = O(\kappa^{-p/2})$.

Note that

$$\sqrt{n\kappa^{-1/2}} \{\hat{f}_\kappa(\theta) - \mathbb{E}_f[\hat{f}_\kappa(\theta)]\} = \sqrt{n} \left\{ n^{-1} \sum_{i=1}^n \kappa^{-1/4} K_\kappa(\theta - \Theta_i) - \mathbb{E}_f[\kappa^{-1/4} K_\kappa(\theta - \Theta_1)] \right\}.$$

We now derive the expectation and the variance of $\kappa^{-1/4}K_\kappa(\theta - \Theta_1)$. It follows from (4) that

$$\mathbb{E}_f[\kappa^{-1/4}K_\kappa(\theta - \Theta_1)] = \kappa^{-1/4} \left[f(\theta) + \mu_0^{-1}(L)\mu_p(L)\kappa^{-p/2} \sum_{t=1}^{p/2} \frac{b_{p,2t}f^{(2t)}(\theta)}{2t!} + O(\kappa^{-(p+2)/2}) \right]. \quad (\text{C.2})$$

Then, it can be derived from (C.2) that $|\mathbb{E}[\kappa^{-1/4}K_\kappa(\theta - \Theta_1)]| < \infty$. It follows from (8) that

$$\text{Var}_f[\kappa^{-1/4}K_\kappa(\theta - \Theta_1)] = f(\theta)d(L) + o(1). \quad (\text{C.3})$$

From (C.3), we obtain that $\text{Var}_f[\kappa^{-1/4}K_\kappa(\theta - \Theta_1)] < \infty$. From (C.2) and (C.3), the first term of (C.1) satisfies Lindeberg's condition (Feller (1968)):

$$\sqrt{n\kappa^{-1/2}}\{\hat{f}_\kappa(\theta) - \mathbb{E}_f[\hat{f}_\kappa(\theta)]\} \xrightarrow{d} \text{N}(0, f(\theta)d(L)). \quad (\text{C.4})$$

With $\kappa = cn^\alpha$, the rate of the second term of (C.1) is given as

$$\sqrt{n\kappa^{-1/2}}\text{bias}_f[\hat{f}_\kappa] = O(n^{(2-\alpha(2p+1))/4}). \quad (\text{C.5})$$

If we choose $\alpha > 2/(2p+1)$, then the convergence rate of (C.5) is equal to

$$\sqrt{n\kappa^{-1/2}}\text{bias}_f[\hat{f}_\kappa] = o(1). \quad (\text{C.6})$$

For $\alpha > 2/(2p+1)$ and taking $n \rightarrow \infty$, Theorem 2 completes the proof with (C.4), and (C.6). \square

A4 Appendix D

証明. (命題 3.1) It follows from (ii) and (iv) that

$$\begin{aligned} \mu_j(L_{[p+2]}) &= \frac{p+1}{p}\mu_j(L_{[p]}) + \frac{2}{p}[L_{[p]}(r)r^{(j+1)/2}]_0^\infty - \frac{2}{p}\frac{j+1}{2}\int_0^\infty L_{[p]}(r)r^{(j-1)/2}dr \\ &= \left(\frac{p+1}{p} - \frac{j+1}{p}\right)\mu_j(L_{[p]}). \end{aligned} \quad (\text{D.1})$$

Because $K_{[p]}$ is a p th-order kernel, the equation (D.1) gives that $\mu_0(L_{[p+2]}) = \mu_0(L_{[p]})$, $\mu_j(L_{[p+2]}) = 0$ for $j = 2, 4, \dots, p$, and $\mu_{p+2}(L_{[p+2]}) = -\frac{2}{p}\mu_{p+2}(L_{[p]})$. Therefore, $K_{[p+2]}$ is a $(p+2)$ th-order kernel. \square

A5 Appendix E

証明. (定理 3.3)

Set $W := \hat{f}_\kappa(\theta) - \mathbb{I}_\kappa(\theta)$ and $Z := \hat{f}_{\kappa/4}(\theta) - \mathbb{I}_{\kappa/4}(\theta)$. Then, Terrell and Scott (1980) showed the following:

$$\mathbb{E}_f[\hat{f}_\kappa^{\text{TS}}(\theta)] \simeq \mathbb{I}_\kappa(\theta)^{4/3}\mathbb{I}_{\kappa/4}(\theta)^{-1/3}, \quad (\text{E.1})$$

$$\text{Var}_f[\hat{f}_\kappa^{\text{TS}}(\theta)] \simeq \text{Var}_f\left[\frac{4}{3}W - \frac{1}{3}Z\right]. \quad (\text{E.2})$$

When K_κ is a 2nd-order kernel and $z = 4$, the equation (B.5) is equal to

$$\alpha_{2t}(K_\kappa) = 2^{1/2}\mu_0^{-1}(L) \sum_{q=t}^3 \sum_{m=0}^{3-q} \frac{A_q(4, t)g_{2q}^{(m)}(0)\mu_{2(q+m)}(L)}{\kappa^{q+m}m!} + O(\kappa^{-3}). \quad (\text{E.3})$$

Then, we can write (E.3) as $\alpha_2(K_\kappa) = \alpha_{2,1}\kappa^{-1} + \alpha_{2,2}\kappa^{-2} + O(\kappa^{-3})$, $\alpha_4(K_\kappa) = \alpha_{4,1}\kappa^{-2} + O(\kappa^{-3})$, and $\alpha_6(K_\kappa) = O(\kappa^{-3})$, where $\alpha_{2,1} := 2^{1/2}\mu_0^{-1}(L)\mu_2(L)A_1(4,1)g_2(0)$, $\alpha_{2,2} := 2^{1/2}\mu_0^{-1}(L)\mu_4(L)[A_1(4,1)g_2^{(1)}(0) + A_2(4,1)g_4(0)]$, and $\alpha_{4,1} := 2^{1/2}\mu_0^{-1}(L) \times \mu_4(L)A_2(4,2)g_4(0)$. We set $I_\kappa(\theta) := E_f[\hat{f}_\kappa(\theta)]$ and $c_j := \frac{f^{(j)}}{j!}$. By the similar procedure that employed by Terrell and Scott (1980), $\log I_\kappa(\theta)$ is given as

$$\log I_\kappa(\theta) = \log f(\theta) + \frac{c_2\alpha_{2,1}}{f(\theta)\kappa} + \frac{(c_2\alpha_{2,2} + c_4\alpha_{4,1})f(\theta) - c_2^2\alpha_{2,1}^2/2}{f^2(\theta)\kappa^2} + O(\kappa^{-3}).$$

Taking exponential of $\{4\log I_\kappa(\theta)/3 - \log I_{\kappa/4}(\theta)/3\}$ gives the following:

$$I_\kappa(\theta)^{4/3}I_{\kappa/4}(\theta)^{-1/3} = f(\theta) + 4\frac{c_2^2\alpha_{2,1}^2/2 - (c_2\alpha_{2,2} + c_4\alpha_{4,1})f(\theta)}{f^2(\theta)\kappa^2} + O(\kappa^{-3}). \quad (\text{E.4})$$

It follows from (E.1) and (E.4) that $\text{bias}_f[\hat{f}_\kappa^{\text{TS}}(\theta)] = G(\theta)\kappa^{-2} + O(\kappa^{-3})$, where $G(\theta) = 4\{c_2^2\alpha_{2,1}^2/2 - (c_2\alpha_{2,2} + c_4\alpha_{4,1})f(\theta)\}/\{f(\theta)\}$. We set $\delta_{\kappa^{1/2}, 2t, 4}(L) := \int_{-\kappa^{1/2}\pi}^{\kappa^{1/2}\pi} L(z^2/2)L(z^2/8)z^{2t}dz$ and $\delta_{2t, 4}(L) := \int_{-\infty}^{\infty} L(z^2/2)L(z^2/8)z^{2t}dz$. Then, it is shown that $\delta_{2t, 4}(L) < \infty$ and $\delta_{\kappa^{1/2}, 2t, 4}(L) = \delta_{2t, 4}(L) + o(1)$, because for all z , $L(z^2/2) > L(z^2/8)$ and $\delta_{2t}(L) < \infty$. In the same manner as in Lemma 3, this leads to

$$\int_{-\pi}^{\pi} K_\kappa(u)K_{\kappa/4}(u)u^{2t}du = \kappa^{-(2t-1)/2}[\delta_{2t, 4}(L) + o(1)], \quad (\text{E.5})$$

where $d_{2t, 4}(L) := 2^{-2}\mu_0^{-2}(L)\delta_{2t, 4}(L)$. Then, (E.5) implies the following:

$$\begin{aligned} E_f[K_\kappa(\theta - Y)K_{\kappa/4}(\theta - Y)] &= \int_{-\pi}^{\pi} K_\kappa(u)K_{\kappa/4}(u)du f(\theta) + O\left(\int_{-\pi}^{\pi} K_\kappa(u)K_{\kappa/4}(u)u^2du\right) \\ &= \kappa^{1/2}[f(\theta)d_{0, 4}(L) + o(1)]. \end{aligned} \quad (\text{E.6})$$

It then follows from (4) and (E.6) that

$$\text{Cov}_f[WZ] = n^{-1}\kappa^{1/2}f(\theta)d_{0, 4}(L) + o(n^{-1}\kappa^{1/2}). \quad (\text{E.7})$$

By (8) and (E.7), the equation (E.2) is reduced to

$$\begin{aligned} \text{Var}_f[\hat{f}_\kappa^{\text{TS}}(\theta)] &\simeq \frac{16}{9}\text{Var}_f[W] - \frac{8}{9}\text{Cov}_f[ZW] + \frac{1}{9}\text{Var}_f[Z^2] \\ &= n^{-1}\kappa^{1/2}f(\theta)D(L) + o(n^{-1}\kappa^{1/2}), \end{aligned} \quad (\text{E.8})$$

where $D(L) := (33d(L) - 16d_{0, 4}(L))/18$. □

A6 Appendix F

証明. (定理 4.1)

We set $\gamma(y_{ij}) = \gamma_{ij}$ to ease of notation. First, we calculate the expectation of $\overline{\text{CV}}(\kappa)$, which is given by

$$E_f[\overline{\text{CV}}(\kappa)] = \frac{R(K_\kappa)}{n} + \frac{2}{n^2} \sum_{i < j} E_f[\gamma_{ij}] + \frac{2}{n} \sum_i E_f[f(\Theta_i)] - R(f). \quad (\text{F.1})$$

We set $\gamma_i = E_f[\gamma_{ij}|\Theta_i]$. Then, the conditional expectation γ_i is given by

$$\gamma_i = -f(\Theta_i) + f^{(4)}(\Theta_i)\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}). \quad (\text{F.2})$$

The details are presented in Appendix B in Tsuruta and Sagae (2017c). It follows from (F.2) that

$$\begin{aligned} E_f[\gamma_{ij}] &= E_f[\gamma_i] \\ &= -R(f) + R(f^{(2)})\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + O(\kappa^{-3}). \end{aligned} \quad (\text{F.3})$$

By considering Lemma B.3, (F.3), and $E_f[f(\Theta_i)] = R(f)$, we obtain that $E_f[\overline{\text{CV}}(\kappa)]$ is equivalent to (4.4).

We calculate the variance of $\overline{\text{CV}}(\kappa)$. That is,

$$\text{Var}_f[\overline{\text{CV}}(\kappa)] \simeq \frac{2}{n^2}\text{Var}_f[\gamma_{ij}] + \frac{4}{n}\text{Var}_f[f(\Theta_i)] + \frac{4}{n}\text{Cov}_f[\gamma_{ij}, \gamma_{ik}] + \frac{8}{n}\text{Cov}_f[\gamma_{ij}, f(\Theta_i)], \quad (\text{F.4})$$

where $j \neq k$. Let $I_1 := R((f^{(4)})^{1/2}f)$, $I_2 := R(f^{(2)})R(f)$, and $I_3 := R(f^{3/2}) - R(f)^2$. Each term of the right side regarding (F.4) are given by

$$\text{Var}_f[\gamma_{ij}] = \kappa^{1/2}[Q(L)R(f) + o(1)], \quad (\text{F.5})$$

$$\text{Var}_f[f(\Theta_i)] = I_3, \quad (\text{F.6})$$

$$\text{Cov}_f[\gamma_{ij}, \gamma_{ik}] = I_3 - 2\{I_1 - I_2\}\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + o(\kappa^{-2}), \quad (\text{F.7})$$

and

$$\text{Cov}_f[\gamma_{ij}, f(\Theta_i)] = -I_3 + \{I_1 - I_2\}\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-2} + o(\kappa^{-2}). \quad (\text{F.8})$$

The details of (F.5) and (F.6) are presented in Appendix C in Tsuruta and Sagae (2017c). By considering (F.4)–(F.6), we obtain that $\text{Var}_f[\overline{\text{CV}}(\kappa)]$ is equivalent to (4.5). \square

A7 Appendix G

証明. (系 4.1)

We set $c := \hat{\kappa}_{\text{CV}}/\kappa_*$. Then, it is derived from combining Theorem 3.1 and Theorem 4.1 that

$$\text{AMISE}(c\kappa_*)/\text{MISE}(c\kappa_*) \xrightarrow{p} 1, \quad (\text{G.1})$$

$$\overline{\text{CV}}(c\kappa_*)/\text{MISE}(c\kappa_*) \xrightarrow{p} 1, \quad (\text{G.2})$$

and

$$\text{AMISE}(c\kappa_*)/\text{AMISE}(\kappa_*) = \frac{1}{5c^2} + \frac{4c^{1/2}}{5}. \quad (\text{G.3})$$

The equation (G.3) is the convex function such as the minimum at $c = 1$. Thus, if $c \neq 1$ and n is large, then it follows from combining (G.1) and (G.3) that

$$\text{MISE}(c\kappa_*) > \text{MISE}(\kappa_*). \quad (\text{G.4})$$

Suppose that c does not converge to 1. Recall that it is necessary that $\overline{\text{CV}}(c\kappa_*) \leq \overline{\text{CV}}(\kappa)$ for any κ , because $\hat{\kappa}_{\text{CV}}$ is the minimizer of $\overline{\text{CV}}(\kappa)$. Also, if n is large, then it is shown that $\overline{\text{CV}}(\kappa)$ is the convex

function such as the minimum at $\kappa = c\kappa_*$, because we obtain that $\overline{\text{CV}}(\kappa)$ approximates $\text{AMISE}(\kappa)$ from Theorem 4.1. Therefore, it follows that

$$P(\overline{\text{CV}}(c\kappa_*) < \overline{\text{CV}}(\kappa_*)) \rightarrow 1, \quad (\text{G.5})$$

as $n \rightarrow \infty$. From (G.2) and (G.5), then it holds that

$$\text{MISE}(c\kappa_*) < \text{MISE}(\kappa_*), \quad (\text{G.6})$$

as $n \rightarrow \infty$. By contradiction between (G.4) and (G.6), this completes the proof. \square

A8 Appendix H

証明. (定理 4.2) If n is large, it follows from Lemma B.3 that

$$\text{CV}(\kappa) \simeq \frac{d(L)\kappa^{1/2}}{n} + \frac{2}{n^2} \sum_{i < j} \gamma(y_{ij}). \quad (\text{H.1})$$

The derivative of (H.1) is given by

$$\frac{d\text{CV}(\kappa)}{d\kappa} \simeq \frac{d(L)}{2n\kappa^{1/2}} + \frac{2}{n^2\kappa^{1/2}} \sum_{i < j} V_{ij}, \quad (\text{H.2})$$

where

$$\begin{aligned} V_{ij} &:= \kappa^{-1/2}[\gamma(y_{ij}) + \rho(y_{ij}) + 3/4\mu_0^{-1}(L)\mu_2(L)\kappa^{-1}\tau(y_{ij})], \\ \phi_\kappa(y_{ij}) &:= \kappa C_\kappa^{-1}(L) \frac{d}{d\kappa} L_\kappa(y_{ij}), \\ \rho(y_{ij}) &:= K_\kappa(y_{ij}) + \int_{-\pi}^{\pi} \{\phi_\kappa(w)K_\kappa(w + y_{ij}) + K_\kappa(w)\phi_\kappa(w + y_{ij})\}dw - 2\phi_\kappa(y_{ij}), \end{aligned}$$

and

$$\tau(y_{ij}) := \int_{-\pi}^{\pi} K_\kappa(w)K_\kappa(w + y_{ij})dw - K_\kappa(y_{ij}).$$

The details are presented in Appendix E in Tsuruta and Sagae (2017c). The selector $\hat{\kappa}_{\text{CV}}$ satisfies that $d\text{CV}(\kappa)/d\kappa|_{\kappa=\hat{\kappa}_{\text{CV}}} = 0$. This is equivalent to

$$2n^{-2} \sum_{i < j} V_{ij} \Big|_{\kappa=\hat{\kappa}_{\text{CV}}} = -d(L)/(2n). \quad (\text{H.3})$$

Note that $V_i := \text{E}_f[V_{ij}|\Theta_i]$. Then, we set $H_{ij} := V_{ij} - V_i - V_j + \text{E}_f[V_i]$ and $X_i := V_i - \text{E}_f[V_i]$. Then, we rewrite $2n^{-2} \sum_{i < j} \{V_{ij} - \text{E}_f[V_{ij}]\}$ as

$$2n^{-2} \sum_{i < j} V_{ij} - 2n^{-2} \sum_{i < j} \text{E}_f[V_{ij}] \simeq 2n^{-1} \sum_i X_i + 2n^{-2} \sum_{i < j} H_{ij},$$

where $2n^{-2} \sum_{i < j} H_{ij}$ is the degenerate U-statistic. We obtain the asymptotic normality for $2n^{-1} \sum_i X_i$ from the standard central limit theorem (CLT). That is,

$$\frac{2}{n} \sum_i X_i \xrightarrow{d} N(0, Bn^{-1}\kappa^{-5}), \quad (\text{H.4})$$

where, $B := 16\mu_2^4(L)\{R(f^{(4)}f^{1/2}) - R(f''^2)\}/\{\mu_0^4(L)\}$. The details are presented in Appendix F in Tsuruta and Sagae (2017c). We obtain the asymptotic normality for $2n^{-2}\sum_{i<j}H_{ij}$ from Lemma 1. that is,

$$\frac{2}{n^2}\sum_{i<j}H_{ij}\xrightarrow{d}N(0,2n^{-2}\kappa^{-1/2}M_{1,0}(L)R(f)). \quad (\text{H.5})$$

See Appendix-G in Tsuruta and Sagae (2017c) for details. It is derived from combining (H.4) and (H.5) that the asymptotically normal for $2n^{-2}\sum_{i<j}V_{ij}$ is

$$\frac{2}{n^2}\sum_{i<j}V_{ij}\xrightarrow{d}N\left(-2R(f'')\mu_0^{-2}(L)\mu_2^2(L)\kappa^{-5/2}, Bn^{-1}\kappa^{-5} + 2n^{-2}\kappa^{-1/2}M_{1,0}(L)R(f)\right). \quad (\text{H.6})$$

We take $\kappa = \hat{\kappa}_{\text{CV}}$ in (H.6). Then, we replace $\hat{\kappa}_{\text{CV}}$ in the variance to κ_* by Corollary 4.1. Thus, it follows from combining (H.3) and (H.6) that

$$-2R(f'')\mu_0^{-2}(L)\mu_2^2(L)\hat{\kappa}_{\text{CV}}^{-5/2}\xrightarrow{d}N\left(-\frac{d(L)}{2n}, Bn^{-1}\kappa_*^{-5} + 2n^{-2}\kappa_*^{-1/2}M_{1,0}(L)R(f)\right). \quad (\text{H.7})$$

the first term for the variance of (H.7) is ignored, because the convergence rate of the first term is $O(n^{-3})$, and that of the second term is $O(n^{-11/5})$ by using $\kappa_* = O(n^{2/5})$. From (3.4), we obtain that $R(f'')\mu_2^2(L)n/(d(L)\mu_0(L)) = \kappa_*^{5/2}$. Thus, (H.7) is reduced to

$$(\hat{\kappa}_{\text{CV}}/\kappa_*)^{-5/2}\xrightarrow{d}N\left(1, 8d(L)^{-2}M_{1,0}(L)R(f)\kappa_*^{1/2}\right). \quad (\text{H.8})$$

Let $g(x) = x^{-5/2}$. Then, it follows that $g(1)=1$ and $\{g'(1)\}^2 = 25/4$. We obtain the asymptotic normality for $\hat{\kappa}_{\text{CV}}/\kappa_*$ by applying the delta method to (H.8). That is,

$$\hat{\kappa}_{\text{CV}}/\kappa_*\xrightarrow{d}N\left(1, 50d(L)^{-2}M_{1,0}(L)R(f)\beta(L)^{-1/2}R(f'')^{-1/5}n^{-1/5}\right). \quad (\text{H.9})$$

Theorem 4.2 completes the proof from (H.9). \square

A9 Appendix I

証明. (定理 4.3)

Let $U_{ij} = T_g^{(4)}(\Theta_i - \Theta_j)$, and $U_i = E_f[U_{ij}|\Theta_i]$. The expectation of $\hat{\psi}_4(g)$ is given by

$$E_f[\hat{\psi}_4(g)] = n^{-1}T_g^{(4)}(0) + 2n^{-2}\sum_{i<j}E_f[U_{ij}]. \quad (\text{I.1})$$

It follows from (4.8) that

$$S_g^{(4)}(0) = 3g^2[S_g^{(2)}(0) + O(g^{-1})]. \quad (\text{I.2})$$

By combining (I.2) and Lemma B.1, we obtain that the first term of the right side of (I.1) is equal to

$$n^{-1}T_g^{(4)}(0) = \frac{3g^{5/2}[S_g^{(2)}(0) + O(g^{-1})]}{2^{1/2}\mu_0(S)n}. \quad (\text{I.3})$$

It follows from Lemma B. 2 that

$$\begin{aligned}
U_i &= \int_{-\pi}^{\pi} T_g^{(4)}(\theta_j - \Theta_i) f(\theta_j) d\theta_j \\
&= \int_{-\pi}^{\pi} T_g(\theta_j - \Theta_i) f^{(4)}(\theta_j) d\theta_j \\
&= \sum_{t=0}^{p/2} \frac{f^{(4+2t)}(\Theta_i)}{(2t)!} \alpha_{2t}(T_g) + O(\alpha_{p+2}(T_g)) \\
&= f^{(4)}(\Theta_i) + \mu_0^{-1}(S) \mu_p(S) g^{-p/2} \sum_{t=1}^{p/2} \frac{b_{p,2t} f^{(4+2t)}(\Theta_i)}{(2t)!} + O(g^{-(p+2)/2}),
\end{aligned} \tag{I.4}$$

The expectation $\mathbb{E}_f[U_{ij}]$ of (I.1) is given by the expectation of (I.4) over Θ_i . That is,

$$\begin{aligned}
\mathbb{E}_f[U_{ij}] &= \mathbb{E}_f[U_i] \\
&= \psi_4 + \mu_0^{-1}(S) \mu_p(S) g^{-p/2} \sum_{t=1}^{p/2} \frac{b_{p,2t} \psi_{4+2t}}{(2t)!} + O(g^{-(p+2)/2}).
\end{aligned} \tag{I.5}$$

We obtain the bias from combining (I.1), (I.3) and (I.5).

We derive the variance of $\hat{\psi}_4(g)$. We set $W_{ij} := U_{ij} - U_i - U_j + \mathbb{E}_f[U_i]$ and $Z_i := U_i - \mathbb{E}_f[U_i]$. Then, we obtain that $\mathbb{E}_f[W_{ij}] = 0$, $\mathbb{E}_f[Z_i] = 0$ and $\text{Cov}_f[Z_i W_{ij}] = 0$. By using W_{ij} and Z_i . We present $\hat{\psi}_4(g) - \mathbb{E}_f[\hat{\psi}_4(g)]$ as

$$\hat{\psi}_4(g) - \mathbb{E}_f[\hat{\psi}_4(g)] = \frac{2(n-1)}{n^2} \sum_i Z_i + \frac{2}{n^2} \sum_{i < j} W_{ij}. \tag{I.6}$$

Thus, the variance of $\hat{\psi}_4$ is equal to

$$\begin{aligned}
\text{Var}_f[\hat{\psi}_4(g)] &= \mathbb{E}_f \left[\left\{ \frac{2(n-1)}{n^2} \sum_i Z_i + \frac{2}{n^2} \sum_{i < j} W_{ij} \right\}^2 \right] \\
&= \frac{4(n-1)^2}{n^4} \sum_i \text{Var}_f[Z_i] + \frac{4}{n^4} \sum_{i < j} \text{Var}_f[W_{ij}].
\end{aligned} \tag{I.7}$$

By combining (I.4) and (I.5), $\text{Var}_f[Z_i]$ is reduced to

$$\begin{aligned}
\text{Var}_f[Z_i] &= \mathbb{E}_f[U_i^2] - \mathbb{E}_f[U_i]^2 \\
&= \int_{-\pi}^{\pi} f^{(4)}(\theta_i)^2 f(\theta_i) d\theta_i - \left[\int_{-\pi}^{\pi} f^{(4)}(\theta_i) f(\theta_i) d\theta_i \right]^2 + o(1) \\
&= \text{Var}_f[f^{(4)}(\Theta_i)] + o(1).
\end{aligned} \tag{I.8}$$

By considering (I.5), $\mathbb{E}_f[U_{ij}^2] = g^{9/2}[G_{1,0}(S_4)\psi_0 + o(1)]$, and $\mathbb{E}[U_i^2] = \mathbb{E}[U_i]^2 = O(1)$ (The details of $\mathbb{E}_f[U_{ij}^2]$ and $\mathbb{E}[U_i^2]$ are presented in Appendix D in Tsuruta and Sagae (2017c).), we obtain $\text{Var}_f[W_{ij}]$. That is,

$$\begin{aligned}
\text{Var}_f[W_{ij}] &= \mathbb{E}_f[U_{ij}^2] - 2\mathbb{E}_f[U_i^2] + \mathbb{E}_f[U_i]^2 \\
&= g^{9/2}[G_{1,0}(S_4)\psi_0 + o(1)].
\end{aligned} \tag{I.9}$$

We obtain (4.11) from combining (I.7) (I.8), and (I.9). \square

A10 Appendix J

証明. (4.4) The Taylor expansion $\hat{\kappa}_{\text{PI}} = \hat{\kappa}_{\text{PI}}(\hat{\psi}_4(g_*))$ is given by

$$\begin{aligned}\hat{\kappa}_{\text{PI}}(\hat{\psi}_4(g_*)) &\simeq \beta(L)n^{2/5}\psi_4^{2/5} + \frac{2}{5}\beta(L)n^{2/5}\psi_4^{-3/5}(\hat{\psi}_4(g_*) - \psi_4) \\ &= \kappa_*[1 + 2(\hat{\psi}_4(g_*) - \psi_4)/(5\psi_4)].\end{aligned}\tag{J.1}$$

Equation (J.1) is reduced to

$$\hat{\kappa}_{\text{PI}}/\kappa_* - 1 = \frac{2}{5\psi_4}(\hat{\psi}_4(g_*) - \psi_4).\tag{J.2}$$

Noting $W_{ij} := U_{ij} - U_i - U_j + \text{E}_f[U_i]$, and $Z_i := U_i - \text{E}_f[U_i]$, it follows that (I.6) becomes

$$\hat{\psi}_4(g) - \text{E}_f[\hat{\psi}_4(g)] \simeq 2n^{-1} \sum_i Z_i + 2n^{-2} \sum_{i<j} W_{ij},\tag{J.3}$$

where $2n^{-2} \sum_{i<j} W_{ij}$ is the degenerate U-statistic. From (I.8), we obtain the asymptotic normality distribution from the standard CLT. That is,

$$n^{-1/2} \sum_i Z_i \xrightarrow{d} N(0, \text{Var}_f[f(\Theta_i)]).\tag{J.4}$$

If we choose $g_* = W(S)n^{2/7}$, then applying Lemma 1 to $2n^{-2} \sum_{i<j} W_{ij}$, it is given by

$$\frac{2}{n^2} \sum_{i<j} W_{ij} \xrightarrow{d} N(0, 2n^{-2}g_*^{9/2}G_{1,0}(S_4)\psi_0),\tag{J.5}$$

as $n \rightarrow \infty$. The details are presented in Appendix H in Tsuruta and Sagae (2017d). By combining (J.4) and (J.5), we obtain the asymptotic distribution of (J.3). That is,

$$\hat{\psi}_4(g_*) - \text{E}_f[\hat{\psi}_4(g_*)] \xrightarrow{d} N(0, 4n^{-1}\text{Var}_f[f(\Theta_i)] + 2n^{-2}g_*^{9/2}G_{1,0}(S_4)\psi_0).\tag{J.6}$$

Corollary 4.2 shows that the rate of $\text{Var}_f[\hat{\psi}_4(g_*)]$ is the order $n^{-5/7}$. Thus, the equation (J.6) is reduced to

$$n^{5/14}\{\hat{\psi}_4(g_*) - \text{E}_f[\hat{\psi}_4(g_*)]\} \xrightarrow{d} N(0, 2W^{9/2}(S)G_{1,0}(S_4)\psi_0).\tag{J.7}$$

The main term $\hat{\psi}_4(g_*) - \psi_4$ of the right side for (J.2) is equivalent to

$$n^{5/14}\{\hat{\psi}_4(g_*) - \psi_4\} = n^{5/14}\{\hat{\psi}_4(g_*) - \text{E}_f[\hat{\psi}_4(g_*)]\} - n^{5/14}\text{Bias}_f[\hat{\psi}_4(g_*)].\tag{J.8}$$

We show that $\text{Bias}_f[\hat{\psi}_4(g_*)] = O(n^{-4/7})$ from Corollary 4.2. Then, we obtain that $n^{5/14}\text{Bias}_f[\hat{\psi}_4(g_*)]$ is $O(n^{-3/14})$. Thus, if n is large, then this term is ignored. Therefore, the asymptotic normal distribution for $n^{5/14}\{\hat{\psi}_4(g_*) - \psi_4\}$ is given by

$$n^{5/14}\{\hat{\psi}_4(g) - \psi_4\} \xrightarrow{d} N(0, 2W^{9/2}(S)G_{1,0}(S_4)\psi_0).\tag{J.9}$$

Therefore, as $n \rightarrow \infty$, Theorem 4.4 completes the proof from (J.9) and (J.2). \square

A11 Appendix K

証明. (定理 5.2)

We use the Lindeberg's CLT; for example, see Feller (1966) for the details.

補題 K.4. Suppose $\{X_1, \dots, X_n\}$ is a sequence of independent random variables, each with the finite mean μ_i and the finite variance σ_i^2 . Put $S_n^2 = \sum_{i=1}^n \sigma_i^2$. Put $S_n^2 := \sum_{i=1}^n \sigma_i^2$, and let I_A denote indicator function. If, for any $\varepsilon > 0$, the Lindeberg's condition:

$$\lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{i=1}^n \mathbb{E}[(X_i - \mu_i)^2 \mathbf{I}_{\{|X_i - \mu_i| > \varepsilon S_n\}}] = 0, \quad (\text{K.1})$$

is satisfied, then it holds that

$$\frac{1}{S_n} \sum_i (X_i - \mu_i) \xrightarrow{d} N(0, 1),$$

as $n \rightarrow \infty$.

From (5.1), we rewrite S-LLR as

$$\hat{m}(\theta; \kappa) = n^{-1} \mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta^T)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{Y}. \quad (\text{K.2})$$

Put the vector $\mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta^T)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta = (c_1, \dots, c_n)$, where c_i are any constants. Then, from (K.2) S-LLR is given by the average of $c_i Y_i$. That is,

$$\hat{m}(\theta; \kappa) = n^{-1} \sum_{i=1}^n c_i Y_i. \quad (\text{K.3})$$

From combining (5.8) and (K.3), we obtain the sum of variances of $c_i Y_i / \sqrt{R(K_\kappa)}$ is approximately equal to

$$\begin{aligned} S_n^2 &= \sum_{i=1}^n \text{Var}_Y [c_i Y_i / \sqrt{R(K_\kappa)} | \Theta_n] \\ &= n^2 R(K_\kappa)^{-1} \text{Var}_Y [\hat{m}(\theta; \kappa) | \Theta_n] \\ &\simeq n^2 R(K_\kappa)^{-1} R(K_\kappa) \frac{v(\theta)}{nf(\theta)} \\ &= nv(\theta) / f(\theta). \end{aligned} \quad (\text{K.4})$$

It follows from (K.4) that as $n \rightarrow \infty$, $S_n^2 \rightarrow \infty$. If n is large enough, then $\mathbb{E}_Y [(Y_i - \mathbb{E}_Y [Y_i])^2 \mathbf{I}_{\{(Y_i - \mathbb{E}_Y [Y_i | \Theta_n]) > \varepsilon S_n\}} | \Theta_n]$ is equal to

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{E}_Y [(Y_i - \mathbb{E}_Y [Y_i | \Theta_n])^2 \mathbf{I}_{\{(Y_i - \mathbb{E}_Y [Y_i | \Theta_n]) > \varepsilon S_n\}} | \Theta_n] \\ &= \text{Var}_Y [Y_i | \Theta_n] \\ &\quad - \lim_{n \rightarrow \infty} \mathbb{E}_Y [(Y_i - \mathbb{E}_Y [Y_i | \Theta_n])^2 \mathbf{I}_{\{(Y_i - \mathbb{E}_Y [Y_i | \Theta_n]) \leq \varepsilon S_n\}} | \Theta_n] \\ &= \text{Var}_Y [Y_i | \Theta_n] - \text{Var}_Y [Y_i | \Theta_n] \\ &= 0. \end{aligned} \quad (\text{K.5})$$

From combining (K.4) and (K.5), it follows that

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{i=1}^n \mathbb{E}_Y [(c_i Y_i / \sqrt{R(K_\kappa)} - \mathbb{E}_Y [c_i Y_i / \sqrt{R(K_\kappa)} | \Theta_n])^2 \mathbb{I}_{\{Y_i - \mathbb{E}_Y [Y_i | \Theta_n] > \varepsilon S_n\}} | \Theta_n] \\
&= \lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{i=1}^n c_i^2 R(K_\kappa)^{-1} \mathbb{E}_Y [(Y_i - \mathbb{E}_Y [Y_i | \Theta_n])^2 \mathbb{I}_{\{Y_i - \mathbb{E}_Y [Y_i | \Theta_n] > \varepsilon S_n\}} | \Theta_n] \\
&= 0.
\end{aligned} \tag{K.6}$$

From (K.6), we show that $c_i Y_i / \sqrt{R(K_\kappa)}$ satisfies Linderberg condition for any $\varepsilon > 0$. Therefore, from considering Lemma K.4, (K.3), and (K.4), we obtain the following asymptotic distribution:

$$\begin{aligned}
& \frac{n}{\sqrt{n R(K_\kappa) v(\theta) / f(\theta)}} [\hat{m}(\theta; \kappa) - \mathbb{E}_Y [\hat{m}(\theta; \kappa) | \Theta_n]] \\
&= \frac{n}{\sqrt{n v(\theta) / f(\theta)}} [n^{-1} \sum_{i=1}^n \{c_i Y_i / \sqrt{R(K_\kappa)} - \mathbb{E}_Y [c_i Y_i / \sqrt{R(K_\kappa)} | \Theta_n]\}] \\
&= \frac{1}{S_n} \sum_{i=1}^n \{c_i Y_i / \sqrt{R(K_\kappa)} - \mathbb{E}_Y [c_i Y_i / \sqrt{R(K_\kappa)} | \Theta_n]\} \\
&\xrightarrow{d} \mathbb{N}(0, 1),
\end{aligned} \tag{K.7}$$

as $n \rightarrow \infty$. Theorem 5.2 completes the proof from (K.7). \square

A12 Appendix L

証明. (定理 5.4)

From Theorem 5.2 and Lemma 2.1, we obtain the following asymptotically normal distribution:

$$\sqrt{\frac{n}{\kappa^{1/2} / (2\pi^{1/2})}} [\hat{m}(\theta; \kappa) - \mathbb{E}_Y [\hat{m}(\theta; \kappa) | \Theta_n]] \xrightarrow{d} \mathbb{N}(0, v(\theta) / f(\theta)). \tag{L.1}$$

Equation (L.5) is reduced to

$$n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - \mathbb{E}_Y [\hat{m}(\theta; \kappa) | \Theta_n]] \xrightarrow{d} \mathbb{N}(0, v(\theta) / \{2\pi^{1/2} f(\theta)\}). \tag{L.2}$$

We obtain that $n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - m(\theta)]$ is equal to

$$\begin{aligned}
n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - m(\theta)] &= n^{1/2} \kappa^{-1/4} [\hat{m}(\theta; \kappa) - \mathbb{E}_Y [\hat{m}(\theta; \kappa) | \Theta_n]] \\
&\quad + n^{1/2} \kappa^{-1/4} \text{Bias}_Y [\hat{m}(\theta; \kappa) | \Theta_n].
\end{aligned} \tag{L.3}$$

We put $\kappa = cn^\alpha$. Then, recalling that the equation (5.10) gives that $\text{Bias}_Y [\hat{m}(\theta; \kappa) | \Theta_n] = O(\kappa^{-1})$, it follows that

$$\begin{aligned}
n^{1/2} \kappa^{-1/4} \text{Bias}_Y [\hat{m}(\theta; \kappa) | \Theta_n] &\propto n^{1/2} \kappa^{-5/4} \\
&= O_p(n^{(2-5\alpha)/4}).
\end{aligned} \tag{L.4}$$

From (L.4), we show that α such as $n^{(2-5\alpha)/4} = o_p(1)$ is $\alpha > 2/5$. Hence, if $\alpha > 2/5$ and $n \rightarrow \infty$, then the second term of the right side in (L.3) is vanished. Therefore, from combining (L.2) and (L.3), it

holds that

$$\begin{aligned} n^{1/2}\kappa^{-1/4}[\hat{m}(\theta; \kappa) - m(\theta)] &\simeq n^{1/2}\kappa^{-1/4}[\hat{m}(\theta; \kappa) - \mathbf{E}_Y[\hat{m}(\theta; \kappa)|\Theta_n]] \\ &\xrightarrow{d} \mathbf{N}(0, v(\theta)/\{2\pi^{1/2}f(\theta)\}), \quad n \rightarrow \infty. \end{aligned} \quad (\text{L.5})$$

Theorem 5.4 completes the proof from (L.5). \square

A13 Appendix M

証明. (定理 5.6)

From Theorem 5.2 and Lemma 2.2, we obtain the following asymptotically normal distribution:

$$(nh)^{1/2}[\hat{m}(\theta; h) - \mathbf{E}_Y[\hat{m}(\theta; h)|\Theta_n]] \xrightarrow{d} \mathbf{N}(0, v(\theta)/\{\pi f(\theta)\}). \quad (\text{M.1})$$

We show that $(nh)^{1/2}[\hat{m}(\theta; h) - m(\theta)]$ is equal to

$$\begin{aligned} (nh)^{1/2}[\hat{m}(\theta; h) - m(\theta)] &= (nh)^{1/2}[\hat{m}(\theta; h) - \mathbf{E}_Y[\hat{m}(\theta; h)|\Theta_n] + \text{Bias}_Y[\hat{m}(\theta; h)|\Theta_n]] \\ &= (nh)^{1/2}[\hat{m}(\theta; h) - \mathbf{E}_Y[\hat{m}(\theta; h)|\Theta_n] + (nh)^{1/2}\text{Bias}_Y[\hat{m}(\theta; h)|\Theta_n]]. \end{aligned} \quad (\text{M.2})$$

We put $h = cn^\alpha$. Then, recalling that equation (5.14) gives that $\text{Bias}_Y[\hat{m}(\theta; \kappa)|\Theta_n] = O(h)$, it follows that

$$\begin{aligned} (nh)^{1/2}\text{Bias}_Y[\hat{m}(\theta; h)|\Theta_n] &\propto n^{1/2}h^{3/2} \\ &= O_p(n^{(1+3\alpha)/2}). \end{aligned} \quad (\text{M.3})$$

From (M.3), we show that α such as $n^{(1+3\alpha)/2} = o_p(1)$ is $\alpha < -1/3$. Hence, if $\alpha < -1/3$ and $n \rightarrow \infty$, then the second term of the right side in (M.2) is vanished. Therefore, from combining (M.1) and (M.2), it holds that

$$\begin{aligned} (nh)^{1/2}[\hat{m}(\theta; h) - m(\theta)] &\simeq (nh)^{1/2}[\hat{m}(\theta; h) - \mathbf{E}_Y[\hat{m}(\theta; h)|\Theta_n]] \\ &\xrightarrow{d} \mathbf{N}(0, v(\theta)/\{\pi f(\theta)\}), \quad n \rightarrow \infty. \end{aligned} \quad (\text{M.4})$$

Theorem 5.6 completes the proof from (M.4). \square

A14 Appendix N

証明. (補題 6.1)

By combining (6.1) and (6.2) we obtain

$$m(\Theta_i) = m(\theta) + \sum_{t=1}^{p+2} \left(\sum_{s=0}^{\lfloor (p+1)/2 \rfloor} b_s \sin^{2s+1}(\Theta_i - \theta) \right)^t + o_p(\sin^{p+2}(\Theta_i - \theta)). \quad (\text{N.1})$$

From the polynomial theorem, the terms $\left(\sum_{s=0}^{\lfloor(p+1)/2\rfloor} b_s \sin^{2s+1}(\Theta_i - \theta)\right)^t$ of (N.1) is given by

$$\begin{aligned}
\left(\sum_{s=0}^{\lfloor(p+1)/2\rfloor} b_s \sin^{2s+1}(\Theta_i - \theta)\right)^t &= \sum_{\sum_{s=0}^{\lfloor(p+1)/2\rfloor} t_s} \frac{t!}{\prod_{m=0}^{\lfloor(p+1)/2\rfloor} t_m!} \prod_{l=0}^{\lfloor(p+1)/2\rfloor} b_l^{t_l} \sin(\Theta_i - \theta)^{\sum_{r=0}^{\lfloor(p+1)/2\rfloor} (2r+1)t_r} \\
&= t! \sum_{q=t}^{p+2} \sum_{\substack{\sum_{s=0}^{\lfloor(p+1)/2\rfloor} t_s = q \\ \sum_{s=0}^{\lfloor(p+1)/2\rfloor} (2s+1)t_s = q}} \frac{\prod_{l=0}^{\lfloor(p+1)/2\rfloor} b_l^{t_l}}{\prod_{m=0}^{\lfloor(p+1)/2\rfloor} t_m!} \sin^q(\Theta_i - \theta) \\
&= t! \sum_{q=t}^{p+2} B_q(p, t) \sin^q(\Theta_i - \theta). \tag{N.2}
\end{aligned}$$

It is shown by combining (N.1) and (N.2) that

$$\begin{aligned}
m(\Theta_i) &= m(\theta) + \sum_{t=1}^{p+2} \frac{m^{(t)}(\theta)}{t!} \sum_{q=t}^{p+2} t! B_q(p, t) \sin^q(\Theta_i - \theta) + o_p(\sin^{p+2}(\Theta_i - \theta)) \\
&= m(\Theta_i) = m(\theta) + \sum_{q=1}^{p+2} M_q(\theta) \sin^q(\Theta_i - \theta) + o_p(\sin^{p+2}(\Theta_i - \theta)), \quad |\theta| < \pi/2. \tag{N.3}
\end{aligned}$$

Lemma 5.2 completes the proof from (N.3). \square

A15 Appendix O

証明. (補題 6.2) We show the first property. The Taylor expansion of $\cos(hz)$ and the condition (h.1') of Definition 6.1 is given by

$$\begin{aligned}
L_h(hz) &= L(h^{-2}\{1 - \cos(\theta)\}) \\
&= L(h^{-2}\{1 - \cos(hz)\}) \\
&= L(h^{-2}[1 - \{1 - h^2 z^2/2 + O(h^4)\}]) \\
&= L(z^2/2) + O(h^2). \tag{O.1}
\end{aligned}$$

We obtain the first property from (O.1).

The condition (h.2') of Definition 6.1 gives

$$\begin{aligned}
\bar{L}(z) \times z^t &= o(z^{t-(2p+4)}) \\
&= o(z^{-1}), \tag{O.2}
\end{aligned}$$

for $t \leq 2p + 2$ and large z . the equation (O.2) gives the second property.

Let M and $\alpha > 0$ be any constants, respectively. If n is large enough, then the condition (h.3) show that the tails of $\alpha_t(\bar{L})$ can be ignored.

$$\begin{aligned}
\int_{\pi/h}^{\infty} \bar{L}(z) z^{t/2} dz &\leq \int_{\pi/h}^{\infty} \bar{L}(z) z^{2p+2} dz \\
&< M \int_{\pi/h}^{\infty} z^{-\alpha} dz \\
&= M[-(\alpha - 1)^{-1} z^{-\alpha+1}]_{\pi/h}^{\infty} \\
&= o(h). \tag{O.3}
\end{aligned}$$

The equation (O.3) provides the third property. \square

A16 Appendix P

証明. (補題 6.3) The three properties of Lemma 6.2 gives

$$\begin{aligned}
C_h(L) &= h \int_{-\pi/h}^{\pi/h} L_h(hz) dz \\
&= h \int_{-\pi/h}^{\pi/h} \{\bar{L}(z) + O_p(h^2)\} dz \\
&= h\{\mu_0(L) + o_p(h)\} \\
&= h + o_p(h^2).
\end{aligned} \tag{P.1}$$

The equation (P.1) provides the first property.

It is shown from the first and second properties of Lemma 6.2 that

$$\begin{aligned}
K_h(hz) &= C_h^{-1}(L)L_h(\theta) \\
&= C_h^{-1}(L)[\bar{L}(z) + O(h^2)] \\
&= (h + o(h^2))^{-1}[\bar{L}(z) + O(h^2)] \\
&= h^{-1}[\bar{L}(z) + O(h^2)].
\end{aligned} \tag{P.2}$$

We obtain the second property from (P.2).

Note that $1 - \cos(\theta) \geq 1$ for $|\theta| \geq \pi/2$. Let Q be any constant. Then, the (h.3) of definition 6.1 gives

$$\begin{aligned}
K_h(\theta) &= C_h(L)L(h^{-2}\{1 - \cos(\theta)\}) \\
&\leq C_h^{-1}(L)L(h^{-2}) \\
&< Q(h + o(h^2))^{-1}(h^{-2})^{-(p+\alpha/2)} \\
&= O(h^{2p+\alpha-1}) \\
&= o(h^{p+2}).
\end{aligned} \tag{P.3}$$

We obtain the third property from (P.3). \square

A17 Appendix Q

証明. (定理 6.1)

Put $\mathbf{M} := (m(\Theta_1), \dots, m(\Theta_n))^T$. Then, the bias is

$$\text{Bias}_Y[\hat{m}(\theta; p, h) | \Theta_n] = \mathbf{e}_1^T (\mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{M} - m(\theta). \tag{Q.1}$$

Applying \mathbf{M} to the Taylor series of Lemma 6.1 gives

$$\mathbf{M} = \mathbf{S}_\theta \begin{bmatrix} m(\theta) \\ M_1(\theta) \\ \vdots \\ M_p(\theta) \end{bmatrix} + \mathbf{T}_{m,\theta} + \mathbf{R}_{m,\theta}, \tag{Q.2}$$

where

$$\mathbf{T}_{m,\theta} = M_{p+1}(\theta) \begin{bmatrix} \sin(\Theta_1 - \theta)^{p+1} \\ \vdots \\ \sin(\Theta_n - \theta)^{p+1} \end{bmatrix} + M_{p+2}(\theta) \begin{bmatrix} \sin(\Theta_1 - \theta)^{p+2} \\ \vdots \\ \sin(\Theta_n - \theta)^{p+2} \end{bmatrix}, \quad (\text{Q.3})$$

and the remainder $\mathbf{R}_{m,\theta}$ is $\mathbf{R}_{m,\theta} = o_p(\mathbf{T}_{m,\theta})$.

From combining (Q.1) and (Q.2), the bias is given by

$$\begin{aligned} \text{Bias}_Y[\hat{m}(\theta; p, h) | \Theta_n] &= \mathbf{e}_1^T (\mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \{\mathbf{T}_{m,\theta} + \mathbf{R}_{m,\theta}\} \\ &= \mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \{\mathbf{T}_{m,\theta} + \mathbf{R}_{m,\theta}\}. \end{aligned} \quad (\text{Q.4})$$

Put $\alpha_j := \alpha(\bar{L})$, $\mathbf{A} := \text{diag}\{1, h, \dots, h^p\}$, and $\alpha_\kappa := (\alpha_\kappa, \alpha_{\kappa+1}, \dots, \alpha_{\kappa+p})^T$. Let \mathbf{Q}_p be the $(p+1) \times (p+1)$ matrix having the (i, j) entry equal to α_{i+j} . We show the asymptotic forms of $\mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1}$ and $n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{T}_{m,\theta}$ of (Q.4) as the two following lemmas, respectively.

補題 Q.5. The term $\mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1}$ is given by

$$\mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} = f^{-1}(\theta) [\mathbf{e}_1^T \mathbf{N}_p^{-1} - h f'(\theta) f^{-1}(\theta) \mathbf{e}_1^T \mathbf{N}_p^{-1} \mathbf{Q}_p \mathbf{N}_p^{-1} + o_p(h)] \mathbf{A}^{-1}.$$

補題 Q.6. The term $n^{-1} \mathbf{A}^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{T}_{m,\theta}$ is given by

$$\begin{aligned} n^{-1} \mathbf{A}^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{T}_{m,\theta} &= h^{p+1} \alpha_{p+1} f(\theta) M_{p+1}(\theta) + h^{p+2} \alpha_{p+2} \{f'(\theta) M_{p+1}(\theta) + f(\theta) M_{p+2}(\theta)\} + o_p(h^{p+2}). \end{aligned}$$

Proof of Lemma Q.5. $\hat{s}_l(\theta; h) = n^{-1} \sum_i K_h(\Theta_i - \theta) \sin(\Theta_i - \theta)^l$. $\sin(hz) = hz + O_p(h^3)$

Put $\hat{s}_l(\theta; h) = n^{-1} \sum_i K_h(\Theta_i - \theta) \sin(\Theta_i - \theta)^l$. If n is large enough, then $n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta$ is the $(p+1) \times (p+1)$ matrix having the (i, j) entry equal to $\hat{s}_{i+j}(\theta; h)$. Then, it is shown from combining Lemmas 6.2 and 6.3 that each of the entries is equal to

$$\begin{aligned} \hat{s}_l(\theta; h) &= \int_{-\pi}^{\pi} K_h(\theta_i - \theta) \sin(\Theta_i - \theta)^l f(\theta_i) d\theta_i + O_p(n^{-1}) \\ &= \int_{-\pi/h}^{\pi/h} K_h(hz) \sin(hz)^l f(\theta + hz) h dz + O_p(n^{-1}) \\ &= \int_{-\pi/h}^{\pi/h} h^{-1} \{\bar{L}(z) + O_p(h^2)\} \{hz + O_p(h^3)\}^l f(\theta + hz) h dz \\ &= \int_{-\pi/h}^{\pi/h} \{h^l \bar{L}(z) z^l + o_p(h^{l+1})\} [f(\theta) + h f'(\theta) z + o_p(h)] dz \\ &= h^l \left\{ f(\theta) \int_{-\pi/h}^{\pi/h} \bar{L}(z) z^l dz + h f'(\theta) \int_{-\pi/h}^{\pi/h} \bar{L}(z) z^{l+1} dz + o_p(h) \right\} \\ &= h^l \{f(\theta) \alpha_l(\bar{L}) + h f'(\theta) \alpha_{l+1}(\bar{L}) + o_p(h)\}. \end{aligned} \quad (\text{Q.5})$$

The equation (Q.5) provides

$$n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{T}_{m,\theta} = A [f(\theta) \mathbf{N}_p + h f'(\theta) \mathbf{Q}_p] A + o_p(h A I A). \quad (\text{Q.6})$$

Put $g(h f'(\theta) \mathbf{Q}_p) := [f(\theta) \mathbf{N}_p + h f'(\theta) \mathbf{Q}_p]^{-1}$. Then, the Taylor expansion of g is given by

$$g(h f'(\theta) \mathbf{Q}_p) = f^{-1}(\theta) \mathbf{N}_p^{-1} - h f'(\theta) f^{-2}(\theta) \mathbf{N}_p^{-1} \mathbf{Q}_p \mathbf{N}_p^{-1} + o_p(h). \quad (\text{Q.7})$$

From combining (Q.6), (Q.7), and $e_1^T \mathbf{A}^{-1} = e_1^T \mathbf{A} = e_1^T$, the matrix $e_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1}$ is equal to

$$\begin{aligned} e_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} &= e_1^T [\mathbf{A}^{-1} \{f^{-1}(\theta) \mathbf{N}_p^{-1} - h f'(\theta) f^{-2}(\theta) \mathbf{N}_p^{-1} \mathbf{Q}_p \mathbf{N}_p^{-1}\} \mathbf{A}^{-1}] + o_p(h \mathbf{A} \mathbf{I} \mathbf{A}) \\ &= e_1^T [f^{-1}(\theta) \mathbf{N}_p^{-1} - h f'(\theta) f^{-2}(\theta) \mathbf{N}_p^{-1} \mathbf{Q}_p \mathbf{N}_p^{-1}] \mathbf{A}^{-1} + o_p(h \mathbf{I} \mathbf{A}) \\ &= f^{-1}(\theta) [e_1^T \mathbf{N}_p^{-1} - h f'(\theta) f^{-1}(\theta) e_1^T \mathbf{N}_p^{-1} \mathbf{Q}_p \mathbf{N}_p^{-1} + o_p(h)] \mathbf{A}^{-1}. \end{aligned} \quad (\text{Q.8})$$

Lemma Q.5 completes the proof from (Q.8). \square

Proof of Lemma Q.6. The equation (Q.5) gives

$$\begin{aligned} n^{-1} \mathbf{A}^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta (\sin(\theta_1 - \theta)^k, \dots, \sin(\theta_n - \theta)^k)^T \\ &= \mathbf{A}^{-1} (\hat{s}_k(\theta; h), \dots, \hat{s}_{k+p}(\theta; h))^T \\ &= \mathbf{A}^{-1} \mathbf{A} [h^k f(\theta) \boldsymbol{\alpha}_k + h^{k+1} f'(\theta) \boldsymbol{\alpha}_{k+1} + o_p(h^{k+1})] \\ &= h^k f(\theta) \boldsymbol{\alpha}_k + h^{k+1} f'(\theta) \boldsymbol{\alpha}_{k+1} + o_p(h^{k+1}). \end{aligned} \quad (\text{Q.9})$$

Combining (Q.3) and (Q.9) gives

$$\begin{aligned} n^{-1} \mathbf{A}^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{T}_{m,\theta} \\ &= M_{p+1}(\theta) \{h^{p+1} f(\theta) \boldsymbol{\alpha}_{p+1} + h^{p+2} \boldsymbol{\alpha}_{p+2} f'(\theta) + o_p(h^{p+2})\} \\ &\quad + M_{p+2}(\theta) \{h^{p+2} f(\theta) \boldsymbol{\alpha}_{p+2} + o_p(h^{p+2})\} \\ &= h^{p+1} \boldsymbol{\alpha}_{p+1} f(\theta) M_{p+1}(\theta) + h^{p+2} \boldsymbol{\alpha}_{p+2} \{f'(\theta) M_{p+1}(\theta) + f(\theta) M_{p+2}(\theta)\} + o_p(h^{p+2}). \end{aligned} \quad (\text{Q.10})$$

The Lemma Q.6 completed the proof from (Q.10). \square

From combing (Q.4) and Lemmas Q.5 and Q.6, we obtain the bias that is

$$\begin{aligned} \text{Bias}_Y [\hat{m}(\theta; p, h) | \boldsymbol{\Theta}_n] \\ &= f^{-1}(\theta) [e_1^T \mathbf{N}_p^{-1} - h f'(\theta) f^{-1}(\theta) e_1^T \mathbf{N}_p^{-1} \mathbf{Q}_p \mathbf{N}_p^{-1} + o_p(h)] \mathbf{A}^{-1} \\ &\quad \times [h^{p+1} \boldsymbol{\alpha}_{p+1} f(\theta) M_{p+1}(\theta) + h^{p+2} \boldsymbol{\alpha}_{p+2} \{f'(\theta) M_{p+1}(\theta) + f(\theta) M_{p+2}(\theta)\} + o_p(h^{p+2})]. \\ &= h^{p+1} M_{p+1}(\theta) e_1^T \mathbf{N}_p^{-1} \boldsymbol{\alpha}_{p+1} + h^{p+2} \{M_{p+1}(\theta) f'(\theta) f^{-1}(\theta) + M_{p+2}(\theta)\} e_1^T \mathbf{N}_p^{-1} \boldsymbol{\alpha}_{p+2} \\ &\quad - h^{p+2} M_{p+1}(\theta) f'(\theta) f^{-1}(\theta) e_1^T \mathbf{N}_p^{-1} \mathbf{Q}_p \mathbf{N}_p^{-1} \boldsymbol{\alpha}_{p+1} + o_p(h^{p+2}). \\ &= h^{p+1} M_{p+1}(\theta) \sum_{j=1}^{p+1} (\mathbf{N}_p^{-1})_{1j} \boldsymbol{\alpha}_{p+j} \\ &\quad + h^{p+2} \{M_{p+1}(\theta) f'(\theta) f^{-1}(\theta) + M_{p+2}(\theta)\} \sum_{j=1}^{p+1} (\mathbf{N}_p^{-1})_{1j} \boldsymbol{\alpha}_{p+j+1} \\ &\quad - h^{p+2} M_{p+1}(\theta) f'(\theta) f^{-1}(\theta) e_1^T \mathbf{N}_p^{-1} \mathbf{Q}_p \mathbf{N}_p^{-1} \boldsymbol{\alpha}_{p+1} + o_p(h^{p+2}). \end{aligned} \quad (\text{Q.11})$$

For simplifying (Q.11), we employ the following Lemma given by Ruppert and Wand (1994).

補題 Q.7. It holds that

- (a) If p is odd, then $\alpha_j = 0$, otherwise $\alpha_j \neq 0$.
- (b) If $i + j$ is odd, then $(\mathbf{N}_p)_{ij} = (\mathbf{N}_p^{-1})_{ij} = 0$.
- (c) If $i + j$ is even, then $(\mathbf{Q}_p)_{ij} = 0$.

When p is odd, Lemma Q.7 shows that the first term in the right side of (Q.11) does not vanish. This leads to

$$\text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n] = h^{p+1}M_{p+1}(\theta) \sum_{j=1}^{p+1} (\mathbf{N}_p^{-1})_{1j} \alpha_{p+j} + o_p(h^{p+1}), \quad (\text{Q.12})$$

for p odd.

In the case where p is even, Combining (a) and (c) of Lemma Q.7 gives that the first term in that is zero. Also, it is shown that the last term in that vanishes with combining (b) and (c). Therefore, we obtain

$$\begin{aligned} & \text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n] \\ &= h^{p+2} \{M_{p+1}(\theta)f'(\theta)f^{-1}(\theta) + M_{p+2}(\theta)\} \sum_{j=1}^{p+1} (\mathbf{N}_p^{-1})_{1j} \alpha_{p+j+1} + o_p(h^{p+2}). \end{aligned} \quad (\text{Q.13})$$

Let an cofactor of the determinant $|\mathbf{N}_p|$ be c_{ij} . Then, noting $(\mathbf{N}_p^{-1})_{ij} = c_{ij}/|\mathbf{N}_p|$ and $|\mathbf{M}_p(z)| = \sum_j c_{1j} z^{j-1}/|\mathbf{N}_p|$, we obtain the following relation between the k -th moment of \bar{L}_p and $\sum_{j=1}^{p+1} (\mathbf{N}_p^{-1})_{1j} \alpha_{k-1+j}$ that is

$$\begin{aligned} \alpha_k(\bar{L}_p) &= \int_{\mathbb{R}} \bar{L}_p(z) z^k dz \\ &= \int_{\mathbb{R}} \sum_{j=1}^{p+1} \frac{c_{1j} z^{j-1}}{|\mathbf{N}_p|} \bar{L}(z) z^k dz \\ &= \sum_{j=1}^{p+1} (\mathbf{N}_p^{-1})_{1j} \alpha_{k-1+j}. \end{aligned} \quad (\text{Q.14})$$

Applying (Q.14) to (Q.12) and (Q.13) gives the two biases that is

$$\text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n] = h^{p+1}M_{p+1}(\theta)a_{p+1}(\bar{L}_{(p)}) + o_p(h^{p+1}), \quad (\text{Q.15})$$

for p odd and

$$\text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n] = h^{p+2} \{M_{p+1}(\theta)f'(\theta)f(\theta)^{-1} + M_{p+2}(\theta)\} a_{p+2}(\bar{L}_{(p)}) + o_p(h^{p+2}), \quad (\text{Q.16})$$

for p even.

we now consider the variance. Put $\mathbf{V} := \text{diag}\{v(\Theta_1), \dots, v(\Theta_n)\}$. Then the variance is given by

$$\text{Var}_Y[\hat{m}(\theta; p, h)|\Theta_n] = n^{-1} \mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{V} \mathbf{W}_\theta \mathbf{S}_\theta (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} \mathbf{e}_1. \quad (\text{Q.17})$$

Let \mathbf{T}_p be the $(p+1) \times (p+1)$ matrix having the (i, j) entry equal to $\alpha_{i+j}(\bar{L}^2)$. Then the matrix $n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{V} \mathbf{W}_\theta \mathbf{S}_\theta$ is given by the following Lemma.

補題 Q.8. The matrix $n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{V} \mathbf{W}_\theta \mathbf{S}_\theta$ is given by

$$n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{V} \mathbf{W}_\theta \mathbf{S}_\theta = \mathbf{A} \{h^{-1}v(\theta)f(\theta)\mathbf{T}_p + o_p(h^{-1})\} \mathbf{A}.$$

Proof of Lemma Q.8. Lemma 6.2 leads to

$$\int_{-\pi/h}^{\pi/h} \bar{L}(z)^2 z^l dz = \alpha_l(\bar{L}^2) + o_p(h). \quad (\text{Q.18})$$

Put $\hat{r}_l(\theta; h) := n^{-1} \sum_i K_h(\theta_i - \theta) \sin(\theta_i - \theta)^l v(\Theta_i)$ Then, the matrix $n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{V} \mathbf{W}_\theta \mathbf{S}_\theta$ is the matrix $(p+1) \times (p+1)$ matrix having the (i,j) entry to $\hat{r}_{i+j-2}(\theta; h)$. Combining Lemma 6.3 and (Q.18) shows that each of the entries $\hat{r}_l(\theta; h)$ is equal to

$$\begin{aligned} \hat{r}_l(\theta; h) &= \int_{-\pi}^{\pi} K_{(h)}^2(\theta_i - \theta) \sin(\theta_i - \theta)^l v(\theta_i) f(\theta_i) d\theta_i + O_p(n^{-1}) \\ &= \int_{-\pi/h}^{\pi/h} K_{(h)}^2(hz) \sin(hz)^l v(\theta + hz) f(\theta + hz) h dz + O_p(n^{-1}) \\ &= h^{-1} \int_{-\pi/h}^{\pi/h} \{\bar{L}(z) + O_p(h^2)\}^2 \{hz + O_p(h^3)\}^l v(\theta + hz) f(\theta + hz) h dz \\ &= h^{l-1} v(\theta) f(\theta) \int_{-\pi/h}^{\pi/h} [\bar{L}(z)^2 z^l dz + O_p(h)] \\ &= h^{l-1} v(\theta) f(\theta) \alpha_l(\bar{L}^2) + O_p(h^l). \end{aligned} \quad (\text{Q.19})$$

The equation (Q.19) derives that

$$n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{V} \mathbf{W}_\theta \mathbf{S}_\theta = \mathbf{A} \{h^{-1} v(\theta) f(\theta) \mathbf{T}_p + o_p(h^{-1})\} \mathbf{A}. \quad (\text{Q.20})$$

□

Lemma Q.8 completed from (Q.20). Combining Lemmas Q.5 and Q.8 gives

$$\begin{aligned} \text{Var}_Y[\hat{m}(\theta; p, h) | \Theta_n] &= f^{-1}(\theta) [e_1^T \mathbf{N}_p^{-1} + o_p(1)] \mathbf{A}^{-1} n^{-1} \mathbf{A} \{h^{-1} v(\theta) f(\theta) \mathbf{T}_p + o_p(h^{-1})\} \mathbf{A} \\ &\quad \times \mathbf{A}^{-1} [e_1 \mathbf{N}_p^{-1} + o_p(1)] f^{-1}(\theta) \\ &= n^{-1} h^{-1} v(\theta) f^{-1}(\theta) e_1^T \mathbf{N}_p^{-1} \mathbf{T}_p \mathbf{N}_p^{-1} e_1 + o_p(h^{-1}). \end{aligned} \quad (\text{Q.21})$$

The term $e_1^T \mathbf{N}_p^{-1} \mathbf{T}_p \mathbf{N}_p^{-1} e_1$ in the right side of (Q.21) is given by

$$e_1^T \mathbf{N}_p^{-1} \mathbf{T}_p \mathbf{N}_p^{-1} e_1 = \frac{1}{|\mathbf{N}_p|^2} \sum_{i=1}^{p+1} \sum_{j=1}^{p+1} c_{1i} c_{1j} \alpha_{i+j-2}(L^2). \quad (\text{Q.22})$$

$$\begin{aligned} \alpha_0(\bar{L}_{(p)}^2) &= \int_{\mathbb{R}} L_{(p)}^2(z^2/2) dz \\ &= \int_{\mathbb{R}} \left\{ \sum_{j=1}^{p+1} c_{1j} z^{j-1} |\mathbf{N}_p|^{-1} \bar{L}(z) \right\}^2 dz \\ &= \frac{1}{|\mathbf{N}_p|^2} \sum_{i=1}^{p+1} \sum_{j=1}^{p+1} c_{1i} c_{1j} \int_{\mathbb{R}} z^{i+j-2} \bar{L}(z) dz \\ &= \frac{1}{|\mathbf{N}_p|^2} \sum_{i=1}^{p+1} \sum_{j=1}^{p+1} c_{1i} c_{1j} \alpha_{i+j-2}(\bar{L}^2). \end{aligned} \quad (\text{Q.23})$$

Combining (Q.22) and (Q.23) gives that

$$\mathbf{e}_1^T \mathbf{N}_p^{-1} \mathbf{T}_p \mathbf{N}_p^{-1} \mathbf{e}_1 = \alpha_0(\bar{L}_{(p)}^2). \quad (\text{Q.24})$$

It is shown from combining (Q.21) and (Q.24) that

$$\text{Var}_Y[\hat{m}(\theta; p, h) | \Theta_n] = (nh)^{-1} \frac{\alpha_0(\bar{L}_{(p)}^2)}{f(\theta)} v(\theta) + o_p((nh)^{-1}). \quad (\text{Q.25})$$

Theorem 6.1 completes the proof from combining (Q.15), (Q.16), and (Q.25). \square

A18 Appendix R

証明. (証明 6.2). Put the vector $\mathbf{e}_1^T (n^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta \mathbf{S}_\theta)^{-1} \mathbf{S}_\theta^T \mathbf{W}_\theta = (c_1, \dots, c_n)$, where c_i are any constants. Then, $\hat{m}(\theta; p, h)$ is represented as the following average that is

$$\hat{m}(\theta; p, h) = n^{-1} \sum_{i=1}^n c_i Y_i. \quad (\text{R.1})$$

Let the sum of the conditional variance of $h^{1/2} c_i Y_i$ be $S_n^2 = \sum_{i=1}^n \text{Var}_Y[h^{1/2} c_i Y_i | \Theta_n]$. Then, it is shown from 6.1 shown that

$$\begin{aligned} S_n^2 &= \sum_{i=1}^n \text{Var}_Y[h^{1/2} c_i Y_i | \Theta_n] \\ &= n^2 h \text{Var}_Y \left[n^{-1} \sum_{i=1}^n \hat{m}(\theta; p, h) \middle| \Theta_n \right] \\ &\simeq n^2 h (nh)^{-1} v(\theta) \alpha_0(\bar{L}_{(p)}^2) f(\theta)^{-1} \{1 + o_p(1)\} \\ &= n v(\theta) \alpha_0(\bar{L}_{(p)}^2) f(\theta)^{-1} \{1 + o_p(1)\}. \end{aligned} \quad (\text{R.2})$$

The equation (K.4) gives $\lim_{n \rightarrow \infty} S_n = \infty$. For any ε , this leads to

$$\begin{aligned} &\lim_{n \rightarrow \infty} \text{E}_Y[(Y_i - \text{E}_Y[Y_i | \Theta_n])^2 \mathbf{I}_{\{Y_i - \text{E}_Y[Y_i | \Theta_n] > \varepsilon S_n\}} | \Theta_n] \\ &= \text{Var}_Y[Y_i | \Theta_n] \\ &\quad - \lim_{n \rightarrow \infty} \text{E}_Y[(Y_i - \text{E}_Y[Y_i | \Theta_n])^2 \mathbf{I}_{\{Y_i - \text{E}_Y[Y_i | \Theta_n] \leq \varepsilon S_n\}} | \Theta_n] \\ &= \text{Var}_Y[Y_i | \Theta_n] - \text{Var}_Y[Y_i | \Theta_n] \\ &= 0. \end{aligned} \quad (\text{R.3})$$

From combining (R.2) and (R.3) provides

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{i=1}^n \text{E}_Y[(h^{1/2} c_i Y_i - \text{E}_Y[h^{1/2} c_i Y_i | \Theta_n])^2 \mathbf{I}_{\{Y_i - \text{E}_Y[Y_i | \Theta_n] > \varepsilon S_n\}} | \Theta_n] \\ &= \lim_{n \rightarrow \infty} \frac{h}{S_n^2} \sum_{i=1}^n c_i^2 \text{E}_Y[(Y_i - \text{E}_Y[Y_i | \Theta_n])^2 \mathbf{I}_{\{Y_i - \text{E}_Y[Y_i | \Theta_n] > \varepsilon S_n\}} | \Theta_n] \\ &= 0. \end{aligned} \quad (\text{R.4})$$

The equation (R.4) shows that Linderberg condition for $h^{1/2}c_i Y_i$ holds for any $\varepsilon > 0$. Therefore, combining Lemma K.4, (R.1), and (R.2) gives

$$\begin{aligned}
& \frac{n^{1/2}h^{1/2}}{\sqrt{v(\theta)\alpha_p(\bar{L}_{(p)}^2)/f(\theta)}}[\hat{m}(\theta; p, h) - \mathbb{E}_Y[\hat{m}(\theta; p, h)|\Theta_n]] \\
&= \frac{h^{1/2}}{\sqrt{nv(\theta)\alpha_p(\bar{L}_{(p)}^2)/f(\theta)}} \left[\sum_{i=1}^n \{h^{1/2}c_i Y_i - \mathbb{E}_Y[h^{1/2}c_i Y_i|\Theta_n]\} \right] \\
&= \frac{1}{S_n} \sum_{i=1}^n \{h^{1/2}c_i Y_i - \mathbb{E}_Y[h^{1/2}c_i Y_i|\Theta_n]\} \\
&\xrightarrow{d} N(0, 1), \tag{R.5}
\end{aligned}$$

as $n \rightarrow \infty$.

It holds that

$$\begin{aligned}
n^{1/2}h^{1/2}[\hat{m}(\theta; p, h) - m(\theta)] &= n^{1/2}h^{1/2}[\hat{m}(\theta; p, h) - \mathbb{E}_Y[\hat{m}(\theta; p, h)|\Theta_n]] \\
&\quad + n^{1/2}h^{1/2} \text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n]. \tag{R.6}
\end{aligned}$$

If p is odd, then Theorem 6.1 gives that $n^{1/2}h^{1/2} \text{Bias}_Y[\hat{m}(\theta; p, h)|\Theta_n] = O_p(n^{\{1+\alpha(2p+3)\}/2})$. This means that the second term in the right side in (R.6) vanishes by choosing $\alpha < -1/(2p+3)$. Therefore, if $\alpha < -1/(2p+3)$ and $n \rightarrow \infty$, then it follows that

$$n^{1/2}h^{1/2}[\hat{m}(\theta; p, h) - m(\theta)] \xrightarrow{d} N(0, v(\theta)\alpha_0(\bar{L}_{(p)}^2)/f(\theta)). \tag{R.7}$$

We consider the case where p is even with the similar way as the case where p is odd. Then, if $\alpha < -1/(2p+5)$ and $n \rightarrow \infty$, then it follows that (R.7). \square

A19 Appendix S

証明. (定理 7.3)

Put $\mathbf{M} := (m(\mathbf{U}_1), \dots, m(\mathbf{U}_n))^T$. Then, the bias is given by

$$\begin{aligned}
\text{Bias}_Y[\hat{m}(\theta; p, h)|\mathbf{U}] &= \mathbf{e}_1^T (\mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{M} - m(\mathbf{u}) \\
&= \mathbf{e}_1^T (\mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{M} - m(\mathbf{u}). \tag{S.1}
\end{aligned}$$

Put $S(\mathbf{u}) := (x_1, \dots, x_d, \sin(\theta_1), \dots, \sin(\theta_q))^T$, and let the vector $\mathbf{Q}_m(\mathbf{u})$ be

$$\mathbf{Q}_m(\mathbf{u}) := [S(\mathbf{u}_1 - \mathbf{u})^T \mathcal{H}_m(\mathbf{u}) S(\mathbf{u}_1 - \mathbf{u}), \dots, S(\mathbf{u}_n - \mathbf{u})^T \mathcal{H}_m(\theta) S(\mathbf{u}_n - \mathbf{u})].$$

Then the Taylor's theorem and Lemma 6.2 gives

$$\mathbf{M} = \mathbf{S}_u \begin{bmatrix} m(\mathbf{u}) \\ \mathbf{D}_m(\mathbf{u}) \end{bmatrix} + \frac{1}{2} \mathbf{Q}_m(\mathbf{u}) + \mathbf{R}_m(\mathbf{u}), \tag{S.2}$$

where $\mathbf{R}_m(\mathbf{u}) := o_p(\mathbf{Q}_m(\mathbf{u}))$ is the vector of the remainder terms. Combining (S.2) and (S.1) provides

$$\begin{aligned} \text{Bias}_Y[\hat{m}(\theta; p, h)|\mathbf{U}] &= \mathbf{e}_1^T (\mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{U}_u^T \mathbf{W}_u \left[\mathbf{U}_u \begin{bmatrix} m(\mathbf{u}) \\ \mathbf{D}_m(\mathbf{u}) \end{bmatrix} + \frac{1}{2} \mathbf{Q}_m(\mathbf{u}) + \mathbf{R}_m(\mathbf{u}) \right] - m(\mathbf{u}) \\ &= \mathbf{e}_1^T (\mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{U}_u^T \mathbf{W}_u \left[\frac{1}{2} \mathbf{Q}_m(\mathbf{u}) + \mathbf{R}_m(\mathbf{u}) \right]. \end{aligned} \quad (\text{S.3})$$

The matrix $n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u$ is equal to

$$n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u = \begin{bmatrix} n^{-1} \sum_i K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u}) & n^{-1} \sum_i K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u})^T \\ n^{-1} \sum_i K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u}) & n^{-1} \sum_i K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u})^T \end{bmatrix}. \quad (\text{S.4})$$

Put $\sin(\mathbf{z}_q) := (\sin(z_1), \dots, \sin(z_q))^T$. The Taylor theorem implies that $\sin(\mathbf{H}_q^{1/2} \mathbf{z}_q) = \mathbf{H}_q^{1/2} \mathbf{z}_q + o_p(\mathbf{H}_q^{1/2})$. This provides

$$S(\mathbf{H}^{1/2} \mathbf{z}) = \begin{bmatrix} \mathbf{H}_d^{1/2} \mathbf{z}_d \\ \sin(\mathbf{H}_q^{1/2} \mathbf{z}_q) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_d^{1/2} \mathbf{z}_d \\ \mathbf{H}_q^{1/2} \mathbf{z}_q + o_p(\mathbf{H}_q^{1/2}) \end{bmatrix} = \mathbf{H}^{1/2} (\mathbf{z} + o_p(\mathbf{I})). \quad (\text{S.5})$$

We calculate the four entries of (S.4) with Lemma 7.1. Then, it is shown that

$$\begin{aligned} & n^{-1} \sum_i K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u}) \\ &= \int_{\mathbb{R}^{d_{\mathbb{T}^q}} K_{\mathbf{H}}(\mathbf{u}_i - \mathbf{u}) f(\mathbf{u}_i) d\mathbf{u}_i + O_p(n^{-1}) \\ &= \int_{\mathbb{R}^{d_{\mathbb{J}^q}} K_{\mathbf{H}}(\mathbf{H}^{1/2} \mathbf{z}) f(\mathbf{u} + \mathbf{H}^{1/2} \mathbf{z}) |\mathbf{H}|^{1/2} d\mathbf{z} + O_p(n^{-1}) \\ &= \int_{\mathbb{R}^{d_{\mathbb{J}^q}} |\mathbf{H}|^{-1/2} \{ \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) + o_p(1) \} [f(\mathbf{u}) + o_p(1)] |\mathbf{H}|^{1/2} d\mathbf{z} \\ &= f(\mathbf{u}) + o_p(1), \end{aligned} \quad (\text{S.6})$$

$$\begin{aligned} & n^{-1} \sum_i K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u}) \\ &= \int_{\mathbb{R}^{d_{\mathbb{T}^q}} K_{\mathbf{H}}(\mathbf{u}_i - \mathbf{u}) S(\mathbf{u}_i - \mathbf{u}) f(\mathbf{u}_i) d\mathbf{u}_i + O_p(n^{-1}) \\ &= \int_{\mathbb{R}^{d_{\mathbb{J}^q}} |\mathbf{H}|^{-1/2} [\bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) + O_p(\text{Tr}\{\mathbf{H}\})] \mathbf{H}^{1/2} (\mathbf{z} + o_p(\mathbf{I})) \\ &\quad \times [f(\mathbf{u}) + \mathbf{H}^{1/2} \mathbf{z}^T D_f(\theta) + o_p(\mathbf{H}^{1/2} \mathbf{I})] |\mathbf{H}|^{1/2} d\mathbf{z} \\ &= \left\{ \mathbf{H}^{1/2} f(\mathbf{u}) \int_{\mathbb{R}^{d_{\mathbb{J}^q}} \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) \mathbf{z} d\mathbf{z} \right. \\ &\quad \left. + \mathbf{H} \int_{\mathbb{R}^{d_{\mathbb{J}^q}} \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) \mathbf{z} \mathbf{z}^T d\mathbf{z} \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) \right\} \{1 + o_p(1)\} \\ &= \mathbf{H} \int_{\mathbb{R}^{d_{\mathbb{J}^q}} \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) \mathbf{z} \mathbf{z}^T d\mathbf{z} D_f(\theta) + o_p(\mathbf{H} \mathbf{I}) \\ &= \boldsymbol{\mu} \mathbf{H} D_f(\theta) + o_p(\mathbf{H} \mathbf{I}), \end{aligned} \quad (\text{S.7})$$

and

$$\begin{aligned}
& n^{-1} \sum_i K_{\mathbf{H}}(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u})^T \\
& \simeq \int_{\mathbb{R}^{d_{\mathbb{T}^q}}} K_{\mathbf{H}}(\mathbf{u}_i - \mathbf{u}) S(\mathbf{u}_i - \mathbf{u}) S(\mathbf{u}_i - \mathbf{u})^T f(\mathbf{u}_i) d\mathbf{u}_i \\
& = \int_{\mathbb{R}^{d_{\mathbb{J}^q}}} K_{\mathbf{H}}(\mathbf{H}^{1/2} \mathbf{z}) (\mathbf{H}^{1/2} \\
& \quad \times (\mathbf{z} + o_p(\mathbf{I})) (\mathbf{H}^{1/2} (\mathbf{z} + o_p(\mathbf{I})))^T [f(\mathbf{u}) + o_p(\mathbf{I})] |\mathbf{H}|^{1/2} d\mathbf{z} \\
& = f(\mathbf{u}) \mathbf{H} \int_{\mathbb{R}^{d_{\mathbb{J}^q}}} \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) \mathbf{z} \mathbf{z}^T d\mathbf{z} + o_p(\mathbf{H}) \\
& = f(\mathbf{u}) \boldsymbol{\mu} \mathbf{H} + o_p(\mathbf{H}). \tag{S.8}
\end{aligned}$$

From combining (S.4), (S.6), (S.7), and (S.8), we show that the inverse matrix of $n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u$ is equal to

$$(n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} = \begin{bmatrix} f^{-1}(\mathbf{u}) + o_p(1) & -f^{-2}(\mathbf{u}) D_f^T(\mathbf{u}) + o_p(1) \\ -f^{-2}(\mathbf{u}) D_f(\mathbf{u}) + o_p(1) & f^{-1}(\mathbf{u}) \boldsymbol{\mu}^{-1} \mathbf{H}^{-1} + o_p(\mathbf{H}^{-1}) \end{bmatrix}. \tag{S.9}$$

Let the first entry of $n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{Q}_m(\mathbf{u})$ be

$$B_1 = n^{-1} \sum_i K_{(\mathbf{H})}(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u})^T \mathcal{H}_m(\mathbf{u}) S(\mathbf{U}_i - \mathbf{u}),$$

and the second entry of that be

$$B_2 = n^{-1} \sum_i \{K_{(\mathbf{H})}(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u})^T \mathcal{H}_m(\mathbf{u}) S(\mathbf{U}_i - \mathbf{u})\} S(\mathbf{U}_i - \mathbf{u}).$$

The first entry B_1 is given by

$$\begin{aligned}
B_1 & = \int_{\mathbb{R}^{d_{\mathbb{T}^q}}} K_{(\mathbf{H})}(\mathbf{u}_i - \mathbf{u}) S(\mathbf{u}_i - \mathbf{u})^T \mathcal{H}_m(\mathbf{u}) S(\mathbf{u}_i - \mathbf{u}) f(\mathbf{u}_i) d\mathbf{u}_i + O_p(n^{-1}) \\
& = \int_{\mathbb{R}^{d_{\mathbb{J}^q}}} \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) \{\mathbf{H}^{1/2} (\mathbf{z} + o_p(\mathbf{I}))\}^T \mathcal{H}_m(\mathbf{u}) \{\mathbf{H}^{1/2} (\mathbf{z} + o_p(\mathbf{I}))\} d\mathbf{z} \\
& \quad \times [f(\mathbf{u}) + o_p(1)] \\
& = f(\mathbf{u}) \text{Tr} \left(\mathbf{H}^{1/2} \mathcal{H}_m(\mathbf{u}) \mathbf{H}^{1/2} \int_{\mathbb{R}^{d_{\mathbb{J}^q}}} \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) \mathbf{z} \mathbf{z}^T d\mathbf{z} \right) + o_p(\text{Tr}\{\mathbf{H}\}) \\
& = f(\mathbf{u}) \text{Tr}(\boldsymbol{\mu} \mathbf{H} \mathcal{H}_m(\mathbf{u})) + o_p(\text{Tr}\{\mathbf{H}\}). \tag{S.10}
\end{aligned}$$

The first entry B_2 is given by

$$\begin{aligned}
B_2 & = n^{-1} \sum_i \{K_{(\mathbf{H})}(\mathbf{U}_i - \mathbf{u}) S(\mathbf{u}_i - \mathbf{u})^T \mathcal{H}_m(\mathbf{u}) S(\mathbf{u}_i - \mathbf{u})\} S(\mathbf{U}_i - \mathbf{u}) \\
& = \int_{\mathbb{R}^{d_{\mathbb{J}^q}} |\mathbf{H}|^{-1/2} \{\bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) (\mathbf{H}^{1/2} \mathbf{z})^T \mathcal{H}_m(\mathbf{u}) \mathbf{H}^{1/2} \mathbf{z}\} \mathbf{H}^{1/2} \mathbf{z} f(\mathbf{u} + \mathbf{H}^{1/2} \mathbf{z}) |\mathbf{H}|^{1/2} d\mathbf{z} \\
& = O_p(\mathbf{H}^{3/2} \mathbf{I}). \tag{S.11}
\end{aligned}$$

From combining (S.10) and (S.11), We obtain the vector $n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{Q}_m(\mathbf{u}) = (B_1, B_2)^T$ that is

$$n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{Q}_m(\mathbf{u}) = \begin{bmatrix} f(\mathbf{u}) \text{Tr}(\boldsymbol{\mu} \mathbf{H} \mathcal{H}_m(\mathbf{u})) + o_p(\text{Tr}\{\mathbf{H}\}) \\ O_p(\mathbf{H}^{3/2} \mathbf{I}) \end{bmatrix}. \tag{S.12}$$

It follows from combining (S.3), (S.9) and (S.12) that

$$\begin{aligned}
\text{Bias}_Y[\hat{m}(\mathbf{u}; \mathbf{H})|U] &= \frac{1}{2} \mathbf{e}_1^T \begin{bmatrix} f^{-1}(\mathbf{u}) + o_p(1) & -f^{-2}(\mathbf{u})D_f^T(\mathbf{u}) + o_p(1) \\ -f^{-2}(\mathbf{u})D_f(\mathbf{u}) + o_p(1) & f^{-1}(\mathbf{u})\boldsymbol{\mu}^{-1}\mathbf{H}^{-1} + o_p(\mathbf{H}^{-1}) \end{bmatrix} \\
&\quad \times \begin{bmatrix} f(\mathbf{u})\boldsymbol{\mu} \text{Tr}(\mathbf{H}\mathcal{H}_m(\mathbf{u})) + o_p(\text{Tr}\{\mathbf{H}\}) \\ O_p(\mathbf{H}^{3/2}\mathbf{I}) \end{bmatrix} \\
&= \frac{1}{2} \text{Tr}(\boldsymbol{\mu}\mathbf{H}\mathcal{H}_m(\mathbf{u})) + o_p(\text{Tr}\{\mathbf{H}\}). \tag{S.13}
\end{aligned}$$

Put $\mathbf{V} := \text{diag}\{v(\Theta_1), \dots, v(\Theta_n)\}$. Then, the variance is given by

$$\text{Var}_Y[\hat{m}(\mathbf{u}; \mathbf{H})|U] = n^{-1} \mathbf{e}_1^T (n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{V} \mathbf{W}_u \mathbf{U}_u (n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{e}_1. \tag{S.14}$$

We now derive the four entries of that $n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{V} \mathbf{W}_u \mathbf{U}_u$. Then, the (1,1) entry is given by

$$\begin{aligned}
&n^{-1} \sum_i K_{\mathbf{H}}^2(\mathbf{U}_i - \mathbf{u})v(\mathbf{u}_i) \\
&= \int_{\mathbb{R}^d \mathbb{T}^q} K_{\mathbf{H}}^2(\mathbf{u}_i - \mathbf{u})v(\mathbf{u}_i)f(\mathbf{u}_i)d\mathbf{u}_i + O_p(n^{-1}) \\
&= \int_{\mathbb{R}^d \mathbb{J}^q} K_{\mathbf{H}}^2(\mathbf{H}^{1/2}\mathbf{z})v(\mathbf{u} + \mathbf{H}^{1/2}\mathbf{z})f(\mathbf{u} + \mathbf{H}^{1/2}\mathbf{z})|\mathbf{H}|^{1/2}d\mathbf{z} + O_p(n^{-1}) \\
&= \int_{\mathbb{R}^d \mathbb{J}^q} |\mathbf{H}|^{-1} \{\bar{K}(\mathbf{z}_d)\bar{L}(\mathbf{z}_q) + o_p(1)\}^2 [v(\mathbf{u}) + o_p(1)][f(\mathbf{u}) + o_p(1)]|\mathbf{H}|^{1/2}d\mathbf{z} \\
&= |\mathbf{H}|^{-1/2}v(\mathbf{u})f(\mathbf{u}) \int_{\mathbb{R}^d} \bar{K}^2(\mathbf{z}_d)d\mathbf{z}_d \int_{\mathbb{J}^q} \bar{L}^2(\mathbf{z}_q)d\mathbf{z}_q \{1 + o_p(1)\} \\
&= |\mathbf{H}|^{-1/2}v(\mathbf{u})f(\mathbf{u})\alpha_0^d(\bar{K}^2)[\alpha_0(\bar{L}^2) + o_p(1)]^q \{1 + o_p(1)\} \\
&= |\mathbf{H}|^{-1/2}\alpha_0(\bar{K}^2)^d\alpha_0(\bar{L}^2)^q f(\mathbf{u})v(\mathbf{u}) + o_p(|\mathbf{H}|^{-1/2}). \tag{S.15}
\end{aligned}$$

The (1,2) entry is

$$\begin{aligned}
&n^{-1} \sum_i K_{\mathbf{H}}^2(\mathbf{U}_i - \mathbf{u})S(\mathbf{U}_i - \mathbf{u})v(\mathbf{U}_i) \\
&= \int_{\mathbb{T}^q} K_{\mathbf{H}}^2(\mathbf{u}_i - \mathbf{u})S(\mathbf{u}_i - \mathbf{u})v(\mathbf{u}_i)f(\mathbf{u}_i)d\mathbf{u}_i + O_p(n^{-1}) \\
&= |\mathbf{H}|^{-1/2}\mathbf{H}^{1/2} \int_{\mathbb{R}^d \mathbb{J}^q} \{\bar{K}(\mathbf{z}_d)\bar{L}(\mathbf{z}_q) + o_p(1)\}^2 \mathbf{z}[v(\mathbf{u})f(\mathbf{u}) + o_p(1)]d\mathbf{z} \\
&= O_p(|\mathbf{H}|^{-1/2}\mathbf{H}^{1/2}) \\
&= o_p(|\mathbf{H}|^{-1/2}\mathbf{I}), \tag{S.16}
\end{aligned}$$

and the (1,3) entry is also equal to (S.16). The (2,2) entry is given by

$$\begin{aligned}
& n^{-1} \sum_i K_{\mathbf{H}}^2(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u}) S(\mathbf{U}_i - \mathbf{u})^T v(\mathbf{U}_i) \\
&= \int_{\mathbb{R}^q \times \mathbb{T}^q} K_{\mathbf{H}}^2(\mathbf{u}_i - \mathbf{u}) S(\mathbf{u}_i - \mathbf{u}) S(\mathbf{u}_i - \mathbf{u})^T v(\mathbf{u}_i) f(\mathbf{u}_i) d\mathbf{u}_i \\
&= \int_{\mathbb{R}^d \mathbb{J}^q} |\mathbf{H}|^{-1} \{ \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) + o_p(1) \}^2 \{ \mathbf{H}^{1/2}(\mathbf{z} + o_p(\mathbf{I})) \} \{ \mathbf{H}^{1/2}(\mathbf{z} + o_p(\mathbf{I})) \}^T \\
&\times v(\mathbf{u} + \mathbf{H}^{1/2} \mathbf{z}) f(\mathbf{u} + \mathbf{H}^{1/2} \mathbf{z}) |\mathbf{H}|^{1/2} d\mathbf{z} + O_p(n^{-1}) \\
&= |\mathbf{H}|^{-1/2} \mathbf{H} \int_{\mathbb{R}^d \mathbb{J}^q} \{ \bar{K}(\mathbf{z}_d) \bar{L}(\mathbf{z}_q) + o_p(1) \}^2 \mathbf{z} \mathbf{z}^T [v(\mathbf{u}) + o_p(1)] [f(\mathbf{u}) + o_p(1)] d\mathbf{z} \\
&= o_p(|\mathbf{H}|^{-1/2} \mathbf{I}). \tag{S.17}
\end{aligned}$$

Considering (S.9), (S.14), (S.15), (S.16), and (S.17), we obtain

$$\begin{aligned}
\text{Var}_Y[\hat{m}(\mathbf{u}; \mathbf{H}) | \mathbf{U}] &= \mathbf{e}_1^T (\mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{V} \mathbf{W}_u \mathbf{U}_u (\mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{e}_1 \\
&= n^{-1} \mathbf{e}_1^T (n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{V} \mathbf{W}_u \mathbf{U}_u (n^{-1} \mathbf{U}_u^T \mathbf{W}_u \mathbf{U}_u)^{-1} \mathbf{e}_1 \\
&= n^{-1} \mathbf{e}_1^T \begin{bmatrix} f^{-1}(\mathbf{u}) + o_p(1) & -f^{-2}(\mathbf{u}) D_f^T(\mathbf{u}) + o_p(1) \\ -f^{-2}(\mathbf{u}) D_f(\mathbf{u}) + o_p(1) & f^{-1}(\mathbf{u}) \boldsymbol{\mu} \mathbf{H}^{-1} + o_p(\mathbf{H}^{-1}) \end{bmatrix} \\
&\times \begin{bmatrix} |\mathbf{H}|^{-1/2} \alpha_0 (\bar{K}^2)^d \alpha_0 (\bar{L}^2)^q f(\mathbf{u}) v(\mathbf{u}) + o_p(|\mathbf{H}|^{-1/2}) & o_p(|\mathbf{H}|^{-1/2} \mathbf{I}) \\ o_p(|\mathbf{H}|^{-1/2} \mathbf{I}) & o_p(|\mathbf{H}|^{-1/2} \mathbf{I}) \end{bmatrix} \\
&\times \begin{bmatrix} f^{-1}(\mathbf{u}) + o_p(1) & -f^{-2}(\mathbf{u}) D_f^T(\mathbf{u}) + o_p(1) \\ -f^{-2}(\mathbf{u}) D_f(\mathbf{u}) + o_p(1) & f^{-1}(\mathbf{u}) \boldsymbol{\mu} \mathbf{H}^{-1} + o_p(\mathbf{H}^{-1}) \end{bmatrix} \mathbf{e}_1 \\
&= n^{-1} |\mathbf{H}|^{-1/2} \alpha_0 (\bar{K}^2)^d \mu_0^q (\bar{L}^2) f(\mathbf{u}) v(\mathbf{u}) + o_p(n^{-1} |\mathbf{H}|^{-1/2}). \tag{S.18}
\end{aligned}$$

Theorem 7.3 completes the proof from (S.13) and (S.18). \square