

ノンパラメトリック回帰における バイアス縮小推定量の実際的側面

島根大学大学院総合理工学研究科* 吉 崎 正 浩

島根大学総合理工学部 内 藤 貫 太

要 旨 散布図平滑化へのノンパラメトリックアプローチとして代表的な Kernel 法について新たな議論を加える。Kernel を用いた局所多項式推定量はよく知られているが、それを初期推定量として用い、調整項を加えることにより得られる新たな推定量を提案する。提案される推定量に見られる特徴は、縮小バイアスを持つことであり、特に回帰関数の曲率の大きい箇所で顕著にその効果を見ることができる。この論文においては、Bandwidth 選択法の提案とその評価、推定量の漸近正規性に基づく回帰関数の近似信頼区間の構成、そして既存のノンパラメトリック推定量との大規模なシミュレーションでの比較といった、実際的側面を強調した議論を展開する。

1. はじめに

ノンパラメトリック回帰の一手法として、Kernel を用いた局所多項式推定量はよく知られており、性質の研究も盛んに行われ、平滑化の文献においても紹介されている (Wand and Jones (1995), Simonoff (1996)). その漸近挙動は Ruppert and Wand (1994) によって詳しく調べられ、局所的にあてはめる多項式の次数 p を大きくすることで、漸近バイアスのオーダーが小さくなることがわかっている。一方、 $p = 1$ の場合に対応する Fan (1992) の局所線形推定量はよく用いられるものの、回帰関数の曲率の大きいところでのバイアスが目立つことが知られており、そのバイアスを縮小するアプローチとして多項式 ($p \geq 2$) を用いる局所多項式推定量の提案は自然な発展であったと言える。しかしまた、次数 p を大きくすると推定量そのものの計算が煩雑になり、それに伴い多くの計算時間が必要となることも知られている事実である。したがって、 p を大きくすることによってバイアス縮小を得るアプローチとは異なったアプローチでバイアス縮小ができるか、という問題が浮かび上がる。この問題への 1 つのアプローチとして、Linton and Nielsen (1994) がある。彼らは Nadaraya-Watson 推定量を初期推定量とし、それを乗法的に調整することでバイアス縮小を導いている。

上の問題へのアプローチとして、この論文の第 2 節において、従来の局所多項式推定量を初期推定量として用い、加法的調整を施して得られる新たな推定量 (Additive Bias Reduced Estimator:

* 現在の所属：小野薬品工業

ABRE)を提案する．そして，第3節においてはABREの漸近バイアス，漸近分散を一般的枠組みで評価し，特に $p = 1$ の場合での推定量の漸近正規性を導く．これに基づき，回帰関数の近似信頼区間の構成についても述べる．同時に，Linton and Nielsen (1994)の乗法的調整と，この論文で議論される加法的調整とどちらが良いのかについても考察を与える．第4節では，ABREのPlug-In法に基づくBandwidth選択法の提案を行う．データに基づく平滑化パラメータの選択は平滑化手法全てにおいて極めて重要であるが，ここではRuppert et al. (1995)で議論された局所線形推定量のBandwidth選択へのアプローチをABREへ拡張する．その選択アルゴリズムを簡潔にまとめ，それより得られるBandwidthの漸近挙動についても述べる．また，近似信頼区間の構成のためのアルゴリズムについても述べる．第5節では，大規模なシミュレーションの結果を報告する．ABREと局所線形推定量，局所3次多項式推定量，Data Sharpeningに基づく推定量(Choi et al, 2000)との有限標本での比較をシミュレーションを通して行う．幾つかの回帰関数，デザインのもとで各推定量の振る舞いについて考察を与える．シミュレーションでは同時に，第4節で提案されたBandwidth選択法の評価も行う．幾つかの実データへの適用例を第6節で報告する．最後の第7節でまとめを行う．付録に理論的結果の証明の概略を与える．

2. 推定量の構成

サンプル $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ は回帰モデル

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

にしたがって得られているものとする．ここで， $m(x) = E[Y|X = x]$ は回帰関数であり， $\delta(x) = \text{Var}[Y|X = x]$ ， $\varepsilon_i (i = 1, \dots, n)$ は互いに独立で平均0，分散1である．また， $\varepsilon_i (i = 1, \dots, n)$ と $X_i (i = 1, \dots, n)$ は独立である．回帰関数 m の推定がここで論じる問題である．まず，Ruppert and Wand (1994)で議論された局所 p 次多項式推定量は

$$L(a_0, \dots, a_p|x) = \sum_{i=1}^n \left\{ Y_i - a_0 - a_1(x - X_i) - \dots - a_p(x - X_i)^p \right\}^2 K_h(x - X_i)$$

を (a_0, \dots, a_p) で最小化したときの切片項 $\tilde{a}_0 = \tilde{a}_0(x)$ として得られ，それは

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

と表現される．ここで， $\mathbf{Y} = (Y_1, \dots, Y_n)^T : n$ 次元ベクトル， $K_h(u) = K(u/h)/h$ ，

$$\mathbf{X}_x = \begin{pmatrix} 1 & x - X_1 & \dots & (x - X_1)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x - X_n & \dots & (x - X_n)^p \end{pmatrix}, \quad \mathbf{W}_x = \text{diag} \{ K_h(x - X_1), \dots, K_h(x - X_n) \}$$

であり， \mathbf{e}_j は第 j 成分のみ1で，それ以外は全て0であるような $(p+1)$ 次元ベクトルである．また， K は原点対称密度関数であり， h は平滑化パラメータ，Bandwidthである． $\tilde{m}_p(x)$ は特別な場合としてよく知られたNadaraya-Watson推定量 [Nadaraya (1964), Watson (1964), $p = 0$] や，局所線形推定量 [Fan (1992), $p = 1$] を含んでいる．本論文では特に， $\tilde{m}_1(x)$ をLL， $\tilde{m}_3(x)$ をLCと表記する．この \tilde{m}_p を初期推定量として， $\widehat{m}_p = \tilde{m}_p + \xi$ と，加法的調整項を加えることによって

得られる新たな推定量を考える。調整項 ξ は、各 x について

$$L(\xi_0, \dots, \xi_p | x) = \sum_{i=1}^n \left\{ Y_i - \bar{m}_p(X_i) - \xi_0 - \xi_1(x - X_i) - \dots - \xi_p(x - X_i)^p \right\}^2 K_h(x - X_i)$$

を (ξ_0, \dots, ξ_p) について最小化したときの切片項 $\tilde{\xi}_0 = \tilde{\xi}_0(x)$ を用いるものとする。上記の局所 p 次多項式推定量の導出を参照すると、調整項は“局所 p 次多項式推定量の残差の局所 p 次多項式推定量”であることがわかる。 $\mathbf{M} = (Y_1 - \bar{m}_p(X_1), \dots, Y_n - \bar{m}_p(X_n))^T$ とすると調整項は

$$\tilde{\xi}_0(x) = \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{M}$$

と表現される。これより、新たに提案する ABRE

$$\widehat{m}_p(x) = \bar{m}_p(x) + \tilde{\xi}_0(x) = \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{R} \quad (1)$$

を得る。ただし、 $\mathbf{R} = (2Y_1 - \bar{m}_p(X_1), \dots, 2Y_n - \bar{m}_p(X_n))^T$ である。

ここで、関連する推定量について注意を与えておく。先に述べたように、Linton and Nielsen (1994) は $\bar{m}_0(x)$ を初期推定量とした乗法的な調整項に基づく推定量 (MBRE) を提案している。また、Choi et al. (2000) は共変量の次元が $d(\geq 1)$ という設定のもと、Data Sharpening (DS と略記する) と呼ばれる方法に基づく、縮小バイアスを持つ推定量を幾つか提案している。 $d = 1$ の場合、Choi et al. (2000) で提案・議論された幾つかの推定量のうちの 1 つは、本論文の \widehat{m}_1 に他ならない (Choi et al. (2000) における、 $\widehat{g}_{LL,2}$ を指している)。Choi et al. (2000) においては、 \widehat{m}_1 をはじめ、幾つかの推定量を議論しているが、そもそも推定量の導出方法に関しては直感的であって、推定量の導出・構成といった、推定量を得る方法については述べられていなかった。一方で、ABRE \widehat{m}_p の導出は上に述べたように極めて明快であり、Data sharpening に基づく推定量が何を捉えようとした推定量なのか、その狙いも同時に説明していると言えよう。さらに、Choi et al. (2000) では一切なされなかった Bandwidth 選択の議論についてもこの論文では正面から取り上げている (第 4 節)。

さて、Choi et al. (2000) では、 \widehat{m}_1 の“Leave-one-out”タイプの利用も可能であろうと述べている。つまりそれは、 $\bar{m}_1^{(-i)}(X_i)$ を (X_i, Y_i) 以外のデータから作られた $\bar{m}_1(x)$ の $x = X_i$ での値として、 $\check{Y} = (\check{Y}_1, \dots, \check{Y}_n)^T$, $\check{Y}_i = 2Y_i - \bar{m}_1^{(-i)}(X_i)$ を構成し、 $(X_i, \check{Y}_i)(i = 1, \dots, n)$ に基づいて得られる局所線形推定量である。本論文ではこのようにして得られる推定量を特に、“Leave-one-out Data sharpening 推定量”と呼ぶことにし、 \widehat{m}_{DS} で表すことにする。 \widehat{m}_1 と \widehat{m}_{DS} の違いは Leave-one-out かそうでないかだけでなく、もちろんこの 2 つの推定量の漸近挙動は同じであることは言うまでもない。しかしながら Leave-one-out かそうでないかは、有限標本での推定量の振る舞いに影響を及ぼすことが確認される (第 5 節)。ABRE, MBRE, そして DS で得られる推定量の性質については第 3 節で述べる。

3. 推定量の漸近挙動

$f_X(x)$ を X の密度とする。核関数 G について、 $\mu_\ell(G) = \int z^\ell G(z) dz$, $R(G) = \int G(z)^2 dz$ という表記を用いる。ABRE の条件付きバイアス、条件付き分散の漸近評価は次のようになる。

定理 1 p を奇数とする. $m(x)$ は $(2p+2)$ 回連続微分可能, $f_X(x) > 0$ とすると, $n \rightarrow \infty, h \rightarrow 0$ のとき,

$$\begin{aligned} \text{Bias}[\widehat{m}_p(x)|X_1, \dots, X_n] &= -\frac{1}{\{(p+1)!\}^2} h^{2p+2} \mu_{p+1}(K_{(p)})^2 m^{(2p+2)}(x) + o_p(h^{2p+2}), \\ \text{Var}[\widehat{m}_p(x)|X_1, \dots, X_n] &= \frac{\delta(x)}{nhf_X(x)} R(K_{(p)}^*) + o_p\left(\frac{1}{nh}\right). \end{aligned}$$

ただし, N_p は (i, j) 成分に $\mu_{i+j-2}(K)$ を持つ $(p+1) \times (p+1)$ 行列で, $M_p(u)$ は N_p の 1 列目を $(1, u, \dots, u^p)^T$ に入れ替えた行列であり,

$$K_{(p)}(u) = \frac{|M_p(u)|}{|N_p|} K(u), \quad K_{(p)}^*(u) = 2K_{(p)}(u) - \int K_{(p)}(u-z)K_{(p)}(z)dz$$

である.

p が偶数の場合の条件付きバイアス, 条件付き分散についても同様に求めることが可能であるが, 条件付きバイアスは極めて複雑な形をしておりここでは省略する. 条件付き分散は p が奇数の場合と同じである. 推定量の Asymptotic Mean Integrated Squared Error (AMISE) は

$$\text{AMISE}[\widehat{m}_p] = h^{4p+4} \frac{\mu_{p+1}(K_{(p)})^4}{\{(p+1)!\}^4} \int \{m^{(2p+2)}(u)\}^2 f_X(u) du + \frac{R(K_{(p)}^*)}{nh} \int \delta(u) du \quad (2)$$

となる. Linton and Nielsen (1994) が提案した MBRE の条件付き分散は ABRE の $p=1$ のものと等しく, 条件付きバイアスは

$$-\frac{1}{4} h^4 \mu_2(K)^2 m(x) \frac{d^2}{dx^2} \left\{ \frac{m''(x)}{m(x)} \right\} + o_p(h^4)$$

である. したがって MBRE と $p=1$ の ABRE を比べると, ABRE の方が条件付きバイアスの形が単純であることがわかる. このことは ABRE の方がその性質の理論的評価がしやすく, AMISE に基づいた Bandwidth 選択のアルゴリズムの構成も容易であると考えられる.

境界での条件付きバイアス, 条件付き分散に関しては次が成り立つ.

定理 2 p を奇数とする. f_X の定義域を $[0, 1]$, Kernel の定義域を $[-1, 1]$ とする. $m(x)$ は $(2p+2)$ 回微分可能で $f_X(x) > 0$ とすると, 0 の境界近くの点 ($x = ch, 0 < c < 1$) では, $h \rightarrow 0$ のとき,

$$\begin{aligned} \text{Bias}[\widehat{m}_p(x)|X_1, \dots, X_n] &= -\frac{1}{\{(p+1)!\}^2} h^{2p+2} \mu_{p+1}(K_{(p)}, c)^2 m^{(2p+2)}(x) + o_p(h^{2p+2}), \\ \text{Var}[\widehat{m}_p(x)|X_1, \dots, X_n] &= \frac{\delta(x)}{nhf_X(x)} \int_{-1}^c \widetilde{K}_{(p)}(u, c)^2 du + o_p\left(\frac{1}{nh}\right). \end{aligned}$$

ただし, $s_\ell(c) = \int_{-1}^c u^\ell K(u) du$, $N_p(c)$ は (i, j) 成分に $s_{i+j-2}(c)$ を持つ $(p+1) \times (p+1)$ 行列で, $M_p(u, c)$ は N_p の 1 列目を $(1, u, \dots, u^p)^T$ に入れ替えた行列であり,

$$\begin{aligned} K_{(p)}(u, c) &= \frac{|M_p(u, c)|}{|N_p(c)|} K_{(p)}(u), \\ \mu_\ell(K, c) &= \int_{-1}^c u^\ell K(u) du, \end{aligned}$$

$$\tilde{K}_{(p)}(u, c) = 2K_{(p)}(u, c) - \int_{-1}^c K_{(p)}(z, c)K_{(p)}(u - z, c)dz.$$

ここで MBRE と ABRE とを比較してみる. Linton and Nielsen (1994) は, MBRE の内点での条件付きバイアスのオーダーが $O(h^4)$ となることを示している. 一方, 境界での条件付きバイアスのオーダーは $O(h^2)$ となることが示される [具体的な表示は, 内藤・吉崎 (2002)]. つまり MBRE はいわゆる Boundary Effect を受ける. ABRE はと言うと, 定理 1 と定理 2 からわかるように Boundary Effect を受けない. ABRE の MBRE に対する優越性がわかるだろう.

定理 3 $p = 1$ の ABRE において, $m(x)$ は 4 回連続微分可能とし, $h = \gamma_n n^{-1/9}$ ($\gamma_n \rightarrow \gamma < \infty$) とすると, $x \in [h, 1 - h]$ において

$$n^{4/9} \{\widehat{m}_1(x) - m(x)\} \xrightarrow{D} N(b(x), v(x)).$$

ここで,

$$b(x) = -\frac{1}{4}\mu_2(K)^2\gamma^4 m^{(4)}(x), \quad v(x) = \frac{\delta(x)}{f_X(x)\gamma} R(K_{(1)}^*).$$

この漸近正規性に基づいて, 各点 x において $m(x)$ の近似信頼区間の構成が可能となる. すなわち,

$$\begin{aligned} 1 - \alpha &\doteq P\left(-z_{\alpha/2} \leq \frac{n^{4/9}(\widehat{m}_1(x) - m(x)) - b(x)}{\sqrt{v(x)}} \leq z_{\alpha/2}\right) \\ &= P\left(\widehat{m}_1(x) - n^{-4/9} \{b(x) + z_{\alpha/2} \sqrt{v(x)}\} \leq m(x) \leq \widehat{m}_1(x) - n^{-4/9} \{b(x) - z_{\alpha/2} \sqrt{v(x)}\}\right) \end{aligned}$$

より

$$I_{1-\alpha} = \left[\widehat{m}_1(x) - n^{-4/9} \{b(x) + z_{\alpha/2} \sqrt{v(x)}\}, \widehat{m}_1(x) - n^{-4/9} \{b(x) - z_{\alpha/2} \sqrt{v(x)}\} \right]$$

が $m(x)$ の信頼係数 $1 - \alpha$ の近似信頼区間となる. ここで, z_α は標準正規分布の上側 $100\alpha\%$ 点である. データに基づき実際に信頼区間をどのように構成するのかについては, Bandwidth 選択のアルゴリズムの議論と関連して, 次の第 4 節で詳しく述べる. 実データへの適用例は第 6 節で報告される.

4. Bandwidth の選択

従来の Kernel 推定量と同様に, ABRE を用いる際の避けて通れない問題が Bandwidth の選択の問題である. ここでは, Ruppert et al. (1995) が提案した Plug-In 法を拡張して ABRE に用いることとする. 彼らが提案した方法は局所線形推定量に対してであるが, 比較的簡単な拡張によって ABRE に適用可能になる. ここからは等分散性 $\delta(x) = \sigma^2$ を仮定し, f_X の定義域を $[a, b]$ とする.

まず, $p = 1$ の ABRE について最適な Bandwidth は, (2) より

$$h_{\text{AMISE}} = \left[\frac{2R(K_{(1)}^*)\sigma^2(b-a)}{n\mu_2(K_{(1)})^4\theta_{44}} \right]^{1/9}$$

である。ただし、

$$\theta_{rs} = \int m^{(r)}(x)m^{(s)}(x)f_X(x)dx \quad r, s > 0, r + s : \text{even}$$

である。そこで、 θ_{44} と σ^2 の推定を考える。

4.1. θ_{rs} の推定量

θ_{rs} の推定量として

$$\hat{\theta}_{rs}(g) = \frac{1}{n} \sum_{i=1}^n \widehat{m}^{(r)}(X_i; g) \widehat{m}^{(s)}(X_i; g)$$

を用いる。ここで

$$\widehat{m}^{(r)}(x; g) = r! \mathbf{e}_{r+1}^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{Y}$$

であり、これは Bandwidth g を持つ $m^{(r)}(x)$ の推定量で、適当な方法で g を決める必要がある。

定理 4 (Ruppert et al. (1995)) r, s は共に偶数とし、 p は $p > r, p > s$ である奇数とする。 $m(x)$ と $f_X(x)$ は共に $(p+1)$ 回微分可能とし、 $n \rightarrow \infty, g \rightarrow 0$ のとき、

$$\begin{aligned} \text{Bias}[\hat{\theta}_{rs}(g)|X_1, \dots, X_n] &\simeq \frac{1}{(p+1)!} \left\{ \mu_{p+1}(K_{s,p})g^{p-s+1}\theta_{r,p+1} + \mu_{p+1}(K_{r,p})g^{p-r+1}\theta_{s,p+1} \right\} \\ &\quad + n^{-1}g^{-r-s-1} \int K_{r,p}(u)K_{s,p}(u)du\sigma^2(b-a), \\ \text{Var}[\hat{\theta}_{rs}(g)|X_1, \dots, X_n] &\simeq 2n^{-2}g^{-2r-2s-1}R(K_{r,p} * K_{s,p})\sigma^2 \int \{\sigma^2 + 2m(x)^2\} dx \\ &\quad + 4\sigma^2 n^{-1} \int (mf_X)^{(r+s)}(x)^2 f_X(x)^{-1} dx. \end{aligned}$$

ただし、 $K_{r,p}(u) = r! \{ |M_{r,p}(u)| / |N_p| \} K(u)$ であり、 $M_{r,p}(u)$ は N_p の $(r+1)$ 列目を $(1, u, \dots, u^p)^T$ に入れ替えた行列である。

定理 4 から、 $\text{AMSE}[\hat{\theta}_{44}(g)|X_1, \dots, X_n]$ を最小にする g は

$$g_{\text{AMSE}} = C_1(K) \left[\frac{\sigma^2(b-a)}{n|\theta_{46}|} \right]^{\frac{1}{11}} \quad (3)$$

である。ここで、

$$\begin{aligned} C_1(K) &= \left[\frac{360R(K_{4,5})}{\mu_6(K_{4,5})} \right]^{\frac{1}{11}}, \quad \theta_{46} < 0, \\ &= \left[\frac{1620R(K_{4,5})}{\mu_6(K_{4,5})} \right]^{\frac{1}{11}}, \quad \theta_{46} > 0. \end{aligned}$$

4.2. σ^2 の推定量

以下においては、Bandwidth が λ である局所 p 次多項式推定量を $\widetilde{m}_p(x; \lambda)$ と記す。

$$\text{E} \left[\sum_{i=1}^n \{Y_i - \widetilde{m}_p(X_i; \lambda)\}^2 | X_1, \dots, X_n \right] = v\sigma^2$$

より, σ^2 の推定量として,

$$\hat{\sigma}_p^2(\lambda) = \nu^{-1} \sum_{i=1}^n \{Y_i - \bar{m}_p(X_i; \lambda)\}^2$$

を考える. ただし,

$$w_{ij} = \mathbf{e}_i^T (\mathbf{X}_{X_i}^T \mathbf{W}_{X_i} \mathbf{X}_{X_i})^{-1} \mathbf{X}_{X_i}^T \mathbf{W}_{X_i} \mathbf{e}_j, \quad \nu = n - 2 \sum_{i=1}^n w_{ii} + \sum_{i=1}^n \sum_{j=1}^n w_{ij}^2$$

である.

定理 5 (Ruppert et al. (1995)) p を奇数とする. $n \rightarrow \infty, \lambda \rightarrow 0$ のとき,

$$\begin{aligned} \text{Bias}[\hat{\sigma}_p^2(\lambda)|X_1, \dots, X_n] &\simeq \lambda^{2p+2} \left[\frac{\mu_{p+1}(K_{(p)})}{(p+1)!} \right]^2 \theta_{p+1,p+1}, \\ \text{Var}[\hat{\sigma}_p^2(\lambda)|X_1, \dots, X_n] &\simeq n^{-1} \sigma^4 \text{Var}[\varepsilon^2] + 2(n^2 \lambda)^{-1} \sigma^4 (b-a) R(K_{(p)}^*). \end{aligned}$$

これより, $\text{AMSE}[\hat{\sigma}_p^2(\lambda)|X_1, \dots, X_n]$ を最小にする λ は,

$$\lambda_{\text{AMSE}} = C_2(K) \left[\frac{\sigma^4 (b-a)}{n^2 \theta_{p+1,p+1}^2} \right]^{\frac{1}{4p+5}} \quad (4)$$

となる. ここで,

$$C_2(K) = \left[\frac{\{(p+1)!\}^4 R(K_{(p)}^*)}{2(p+1)\mu_{p+1}(K_{(p)})^4} \right]^{\frac{1}{4p+5}}$$

である. 本論文では, $p=1$ を採用する.

4.3. $m^{(4)}(x)$ の推定量

信頼区間を構成する際には, 第2節で示したように $m^{(4)}(x)$ の推定も必要となることがわかる. そこで, Ruppert and Wand (1994) の Theorem 4.2 において, $r=4, p=5$ の推定量を用いるとし, このための Bandwidth を β とすれば,

定理 6 (Ruppert and Wand(1994)) $m(x)$ は 6 回微分可能とし, $n \rightarrow \infty, \beta \rightarrow 0$ のとき,

$$\begin{aligned} \text{Bias}[\widehat{m}^{(4)}(x; \beta)|X_1, \dots, X_n] &\simeq \frac{1}{720} \beta^2 \mu_6(K_{4,5}) m^{(6)}(x), \\ \text{Var}[\widehat{m}^{(4)}(x; \beta)|X_1, \dots, X_n] &\simeq \frac{\sigma^2}{n \beta^9 f_X(x)} R(K_{4,5}). \end{aligned}$$

これより,

$$\text{AMISE}[\widehat{m}^{(4)}|X_1, \dots, X_n] = \frac{1}{720^2} \beta^4 \mu_6(K_{4,5})^2 \int m^{(6)}(x)^2 f_X(x) dx + \frac{1}{n \beta^9} R(K_{4,5}) \sigma^2 (b-a)$$

となり, これを最小にするという意味で最適な β は

$$\beta_{\text{AMISE}} = C_3(K) \left[\frac{\sigma^2 (b-a)}{n |\theta_{66}|} \right]^{\frac{1}{13}} \quad (5)$$

となる。ただし、

$$C_3(K) = \left[\frac{1166400R(K_{4,5})}{\mu_6(K_{4,5})^2} \right]^{\frac{1}{13}}$$

である。

4.4. DPI 法の構成

g_{AMSE} , λ_{AMSE} , β_{AMISE} の中には $\theta_{22}, \theta_{46}, \theta_{66}, \sigma$ があり、これらの値を推定しなければ実際に用いることができない。この問題へのアプローチとして、いわゆる Rule Of Thumb 推定量 (以下, ROT 推定量) を用いる。まず, n 個の標本を N 個のブロックに分けることを考える。 χ_j を j 番目のブロック内の X_i の集合とする。 n が N で割り切れる場合は $t = n/N$ とし, $\chi_j = \{X_{(j-1)t+1}, \dots, X_{jt}\}$ と分ける。割り切れない場合は上手く調整する。 $\widehat{m}_j^Q(x)$ を j 番目のブロックに k 次多項式をあてはめて得られた関数とする。ここで θ_{rs} の blocked estimator は、

$$\hat{\theta}_{rs}^Q(N, k) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N (\widehat{m}_j^Q)^{(r)}(X_i) (\widehat{m}_j^Q)^{(s)}(X_i) 1\{X_i \in \chi_j\}$$

であり ($1\{\cdot\}$ は定義関数), σ^2 の blocked estimator は、

$$\hat{\sigma}_Q^2(N, k) = (n - (k+1)N)^{-1} \sum_{i=1}^n \sum_{j=1}^N \{Y_i - \widehat{m}_j^Q(X_i)\}^2 1\{X_i \in \chi_j\}$$

となる。ここで N の選択が問題となるが, $RSS(N) = (n - (k+1)N)\hat{\sigma}_Q^2(N, k)$ とし、

$$C_p(N, k) = \frac{RSS(N)(n - (k+1)N_{max})}{RSS(N_{max})} - (n - 2(k+1)N)$$

が最小になるような \hat{N} を $N = 1, 2, \dots, N_{max}$ の中から選ぶ。また, N_{max} は、

$$N_{max} = \max\{\min([20/n], 5), 1\}$$

で決めることとする。この \hat{N} を用いて $\hat{\theta}_{46}^Q(\hat{N}, k)$, $\hat{\sigma}_Q^2(\hat{N}, k)$ を計算する。また、高次導関数の推定は境界付近でのバイアスが大きくなることが知られているので、 $100\alpha\%$ の境界のデータを切り捨てることを考え、

$$\hat{\theta}_{rs}^{[a]}(g) = \frac{1}{n} \sum_{i=1}^n \widehat{m}_r(X_i; g) \widehat{m}_s(X_i; g) 1\{(1-\alpha)a + \alpha b < X_i < \alpha a + (1-\alpha)b\}$$

とする。これらを用い、 h_{AMISE} の推定量 \hat{h}_{DPI} を以下の手順で導出する。

1. \hat{N} を求め、 $\hat{\theta}_{22}^Q(\hat{N}, k)$, $\hat{\theta}_{46}^Q(\hat{N}, k)$ および $\hat{\sigma}_Q^2(\hat{N}, k)$ を計算する。
2. (3), (4) の ROT 推定量、つまり次の2つを計算する。

$$\hat{g} = C_1(K) \left[\frac{\hat{\sigma}_Q^2(\hat{N}, k)(b-a)}{n|\hat{\theta}_{46}^Q(\hat{N}, k)|} \right]^{\frac{1}{11}}, \quad \hat{\lambda} = C_2(K) \left[\frac{\hat{\sigma}_Q^4(\hat{N}, k)(b-a)}{n^2\hat{\theta}_{22}^Q(\hat{N}, k)^2} \right]^{\frac{1}{9}}.$$

3. Bandwidth を

$$\hat{h}_{\text{DPI}} = \left[\frac{2\hat{\sigma}_1^2(\hat{\lambda})(b-a)R(K_{(1)}^*)}{n\mu_2(K)^4\hat{\theta}_{44}^{[0.05]}(\hat{g})} \right]^{\frac{1}{9}}$$

で求める.

また, f_X を $[a, b]$ 上の一様分布として, $m(x)$ の近似信頼区間は次のようにして構成できる.

1. \hat{N} を求め, $\hat{\theta}_{66}^Q(\hat{N}, k)$ を計算する.
2. (5) の ROT 推定量, すなわち

$$\hat{\beta} = C_3(K) \left[\frac{\hat{\sigma}_Q^2(\hat{N}, k)(b-a)}{n|\hat{\theta}_{66}^Q(\hat{N}, k)|} \right]^{\frac{1}{13}}$$

を計算する.

3. 信頼区間を $\hat{I}_{1-\alpha} = [\hat{I}_{1-\alpha}^L, \hat{I}_{1-\alpha}^U]$ とする. ここで,

$$\begin{aligned} \hat{I}_{1-\alpha}^L &= \hat{m}_1(x; \hat{h}_{\text{DPI}}) - n^{-4/9} \left\{ b(x; \hat{\beta}) + z_{\alpha/2} \sqrt{v(x; \hat{\lambda})} \right\}, \\ \hat{I}_{1-\alpha}^U &= \hat{m}_1(x; \hat{h}_{\text{DPI}}) - n^{-4/9} \left\{ b(x; \hat{\beta}) - z_{\alpha/2} \sqrt{v(x; \hat{\lambda})} \right\} \end{aligned}$$

であり,

$$b(x; \hat{\beta}) = -\frac{1}{4}\mu_2(K)^2\gamma^4\hat{m}^{(4)}(x; \hat{\beta}), \quad v(x; \hat{\lambda}) = \frac{\hat{\sigma}_1^2(\hat{\lambda})(b-a)}{\gamma}R(K_{(1)}^*)$$

である.

このようにして得られた \hat{h}_{DPI} の漸近挙動について, 次が成り立つ.

定理 7 (Naito and Yoshizaki (2003), 式 (4.10)) $n \rightarrow \infty$ のとき,

$$\hat{h}_{\text{DPI}} = h_{\text{AMISE}} \left\{ 1 + O_p(n^{-2/11}) \right\}.$$

また, Ruppert and Wand (1994) にある結果から AMISE を導出することで, LC の最適な Bandwidth は

$$h_{\text{AMISE}} = \left[\frac{72R(K_{(1)})\sigma^2(b-a)}{n\mu_4(K_{(3)})^2\theta_{44}} \right]^{\frac{1}{9}}$$

となることがわかり, 含まれる未知量にその推定値を Plug-In することで, 上記のアルゴリズムと同様な Bandwidth 選択アルゴリズムが構成可能である.

5. シミュレーション

この節では有限標本における ABRE の振る舞いを調べるため, そして前節で提案された Bandwidth 選択アルゴリズムの良し悪しを調べるために行った大規模なシミュレーションの結果を報

告する.

4つの回帰関数を

$$m_1(x) = 2 \exp\left[-\frac{(x-0.1)^2}{0.09}\right] + 3 \exp\left[-\frac{(x-0.9)^2}{0.64}\right],$$

$$m_2(x) = 2 + x - 2x^2 + 3x^5,$$

$$m_3(x) = 2 + \sin(2\pi x),$$

$$m_4(x) = 2 + x + \cos(5\pi x^2)$$

とし(図1), それぞれの $m_j(j = 1, 2, 3, 4)$ に対して第2節の回帰モデルにしたがってデータを作成した. ただし, 分散関数は定数とし($\delta(x) = \sigma^2$), 各 $m_j(j = 1, 2, 3, 4)$ に対してその定数を

$$\sigma_j = \frac{1}{8} \{\max_x m_j(x) - \min_x m_j(x)\}, \quad j = 1, 2, 3, 4$$

とした. また ε としては標準正規乱数を用いた.

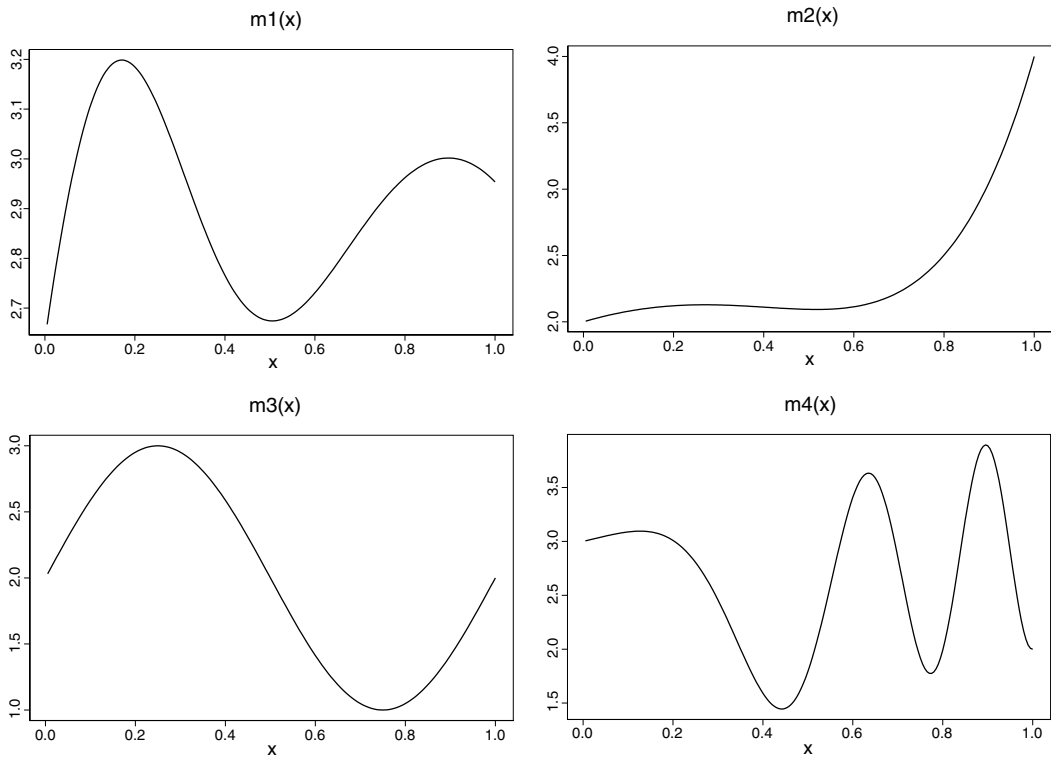


図1. 4つの回帰関数

シミュレーションで取り上げた推定量は, $p = 1$ の ABRE: \widehat{m}_1 , それと同じオーダーの漸近バイアス, 漸近分散を持つ LC: \widehat{m}_3 , そして DS: \widehat{m}_{DS} , および LL: \widehat{m}_1 である. また, Ruppert et al. (1995) による LL の Bandwidth 選択法で得られた Bandwidth と, 本論文の第4節で提案した \widehat{h}_{DPI}

の精度評価も行った。特にこのシミュレーションでは、ROT 推定量の構成で用いる多項式の次数 k についても考察を与えた。Ruppert et al. (1995) は ROT 推定量の構成において 4 次多項式 ($k = 4$) をあてはめたもののみを提案している。これは m の 4 回微分が必要であるため、最低 4 次多項式をあてはめる必要が生じるからである。このシミュレーションでは、 k の値による Bandwidth の値の変化と、それが推定量に与える影響を調べるために、 $k = 4, 6, 8$ を採用した。 $k = 4$ で得られた Bandwidth を用いて計算された LL を LL4, $k = 6$ とし得られた Bandwidth で計算された LL を LL6 とそれぞれ表記する。ABRE, LC, DS では最低 6 次多項式 ($k = 6$) が必要であるが、LL と同様に k の値に対応して ABRE6, ABRE8, LC6, LC8, DS6, DS8 という表記を用いる。なお、DS については、ABRE と同一の AMISE 表現を持つことが示されるので、ABRE と同じ Bandwidth 選択アルゴリズムが適用されることに注意しておく。したがって、以下の表中の Band-MSE の値は、DS6 は ABRE6 と、DS8 は ABRE8 と同じ値であるため明記していない。

全ての推定量でガウス型 Kernel を用い、デザインは $n = 25$ においては $[0, 1]$ の Equi-Spaced Fixed Design, $n = 200$ においては、 $[0, 1]$ の Equi-Spaced Fixed Design, 一様 Random Design, $B(4, 2)$ に従う Random Design である ($B(p, q)$ はパラメータ p と q のベータ分布)。計算の反復回数は 500 回とした。なお、表中の ISB, IV はそれぞれバイアス 2 乗の積分の推定値、分散の積分の推定値、MISE は ISE の平均値、Med-ISE は ISE の中央値、ISE-SE は ISE の標準誤差、Band-MSE は Bandwidth の MSE の推定値であり、全て実際の値を 10^6 倍した値で示している。

シミュレーション結果を見ていく際に、次のようなグルーピングをしておく。(LC6, ABRE6, DS6) の 3 つを RB6 グループと書くことにする。これは縮小バイアスが得られる推定量で、Bandwidth 選択の ROT 推定量の次数を理論的に考えられる最低次数 $k = 6$ としたものである。同様に、(LC8, ABRE8, DS8) の 3 つを RB8 グループと書く。Bandwidth 選択の ROT 推定量の次数だけに着目すれば LL4 は最低次数を用いているので RB6 グループと比較してよいだろう。また、LL6 は RB8 グループと比較される。全ての表において、このようなグルーピングでの比較が見易いように推定量が並べられている。

まず、Fixed design での結果が列挙された表 1 から表 4 までを見ていくことにしよう。ただちに明らかなのは、確かに縮小バイアスが生じていることである。表 1 から表 4 の全てにおいて、LL4 と RB6 グループの ISB, LL6 と RB8 グループの ISB を比べると著しいバイアス縮小が見取れる。しかしながら分散 IV は大きくなっている。ISB の縮小と IV の増大のトレードオフで決まる MISE で見ていくと、比較的なだらかな関数である m_1, m_2, m_3 に対しては LL4 の MISE が RB6 グループのそれより小さくなっており、LL6 と RB8 グループの MISE を比べてもそうである。ただし、 m_4 のような変動の激しい関数に対してはバイアス縮小の方が大きく、全く逆の結果となっている。なだらかな関数の推定に対してはバイアス縮小を持つ推定量を用いるまでもなく、従来の LL でも十分であるが、なだらかな関数に対してバイアス縮小が威力を発揮するわけである。

RB6 グループ、RB8 グループのグループ内での比較を試みよう。MISE で見ると $n = 25$ では ABRE が最も良い。これは DC や DS は小標本ではあまり威力を発揮しない可能性を示唆していると同時に、ABRE は標本数の少ない段階からでも利用可能であることを示していると言えよう。特に DS、すなわち Leave-one-out Data sharpening 推定量 \hat{m}_{DS} については、 n が小さければ、“Leave-one-out” することにより推定量の安定性に影響が現れるのは直感的には理解できるであろう。 $n = 200$ となるとグループ内で各推定量は競合し、優劣も一様ではない。

表 1. m_1 の場合の Fixed Design での結果

	LL4	LC6	ABRE6	DS6	LL6	LC8	ABRE8	DS8
$n = 25$								
ISB	3278	354	2231	3325	2820	122	360	4285
IV	16516	22134	14681	21868	17344	48101	23080	57861
MISE	19794	22488	16912	25193	20163	48223	23439	62145
MED-ISE	18553	18712	15404	20710	18598	30413	20811	38174
ISE-SE	432	687	407	796	440	2173	555	3052
Band-MSE	406	5621	3989	—	330	3532	2506	—
$n = 200$								
ISB	536	95	543	440	497	23	96	63
IV	2782	2300	1921	1946	2834	3026	2568	2625
MISE	3318	2394	2464	2386	3331	3049	2665	2687
MED-ISE	3169	2188	2301	2220	3179	2780	2588	2586
ISE-SE	54	48	48	47	53	60	48	50
Band-MSE	24	4546	3226	—	16	545	387	—

表 2. m_2 の場合の Fixed Design での結果

	LL4	LC6	ABRE6	DS6	LL6	LC8	ABRE8	DS8
$n = 25$								
ISB	17610	300	2371	3633	9301	484	649	1949
IV	140537	274605	179441	274630	178821	589187	297105	721033
MISE	158148	274904	181812	278264	188123	589671	297754	722982
MED-ISE	141730	207303	156319	201404	160435	396749	267538	469997
ISE-SE	4015	10927	5083	12667	5148	25601	7144	34125
Band-MSE	12829	18700	13268	—	4338	71220	50533	—
$n = 200$								
ISB	3347	197	1534	1386	2502	132	264	247
IV	23466	26171	21994	22257	25344	38737	32885	33604
MISE	26813	26368	23528	23643	27845	38868	33148	33851
MED-ISE	25067	23949	22102	22249	26207	35613	30886	31688
ISE-SE	527	589	496	500	549	777	627	646
Band-MSE	5272	5132	3642	—	3780	29149	20683	—

RB6 グループ, RB8 グループをグループ間で比べてみよう. 関数 m_1, m_2, m_3 に対しては標本数によらず RB6 の MISE の方が RB8 のそれよりも小さい. しかしながら, m_4 に対してはグループ間の優劣は逆になる. LL4 と LL6 の比較でも概ね同じ傾向が見て取れる. このことは, 複雑な関数の推定の際には ROT 推定量の次数 k を上げるべきであることを示唆していると言えよう.

次に Random design の場合, 表 5, 表 6 を見ていこう. この 2 つのケースにおいては, 境界領域での影響が出やすいため, MISE は X の密度の上側と下側それぞれ 5% を除いた区間で評価している. 一様分布では $[0.05, 0.95]$ であり, $B(4, 2)$ ではほぼ $[0.34, 0.92]$ という評価区間となる. このように評価区間が異なるため, 表 5 の結果と表 6 の結果の直接的な比較はフェアではなく, むしろ共通に見られる傾向について幾つか述べておく方が有益であろうと思われる. LL4 と RB6 グループ, LL6 と RB8 グループの ISB を比べると, Fixed design の場合よりも更に著しくバイアス縮小が達成されているのがわかる. 関数 m_1, m_2, m_3 では RB6 グループよりも RB8 グループ

表 3. m_3 の場合の Fixed Design での結果

	LL4	LC6	ABRE6	DS6	LL6	LC8	ABRE8	DS8
$n = 25$								
ISB	30701	577	1909	3200	14544	585	337	5477
IV	156596	292379	183547	294015	189655	655296	307722	815743
MISE	187297	292955	185455	297216	204200	655881	308060	821221
MED-ISE	169608	224318	163876	206031	184869	413531	281023	506126
ISE-SE	4301	10716	4633	12188	4859	29301	6989	38224
Band-MSE	1063	12183	8644	—	1194	52967	37582	—
$n = 200$								
ISB	7901	436	1126	817	4199	90	103	92
IV	26230	26808	22653	22894	27978	38652	32648	33373
MISE	34131	27244	23779	23711	32177	38741	32751	33464
MED-ISE	31544	25409	21782	21871	29899	35355	30005	30591
ISE-SE	783	610	543	543	623	802	660	677
Band-MSE	821	2429	1723	—	64	19270	13672	—

表 4. m_4 の場合の Fixed Design での結果

	LL4	LC6	ABRE6	DS6	LL6	LC8	ABRE8	DS8
$n = 25$								
ISB	1360928	307232	365667	238824	435037	36320	47738	49263
IV	270886	583134	362706	624398	343671	1077926	525368	1373512
MISE	1631815	890366	728373	863223	778708	1114246	573106	1422774
MED-ISE	1546409	802191	696188	754912	748783	815375	527048	984249
ISE-SE	21493	20095	12437	21986	11680	39309	11354	54494
Band-MSE	25547	11829	8392	—	4118	1084	769	—
$n = 200$								
ISB	148230	54096	75909	67987	43256	6004	9819	7207
IV	50180	58518	51203	52245	67459	84628	72983	75707
MISE	198409	112614	127112	120231	110715	90632	82802	82914
MED-ISE	195009	110099	125246	118804	110228	86322	79771	80308
ISE-SE	2423	1558	1599	1542	1463	1413	1157	1167
Band-MSE	2325	4137	2935	—	311	133	95	—

プが悪く、 m_4 に対してはそのグループ間優劣が逆転するといった Fixed design で見られた傾向は Random design でもその分布によらず見て取れる。なだらかな関数、特に m_2, m_3 では LL4, LL6 が良く、 m_4 では RB6 グループや RB8 グループの方が良いという傾向も同様である。RB6 グループと RB8 グループのグループ内での比較では、関数 m_1, m_2, m_3 では ABRE が概ね最良であるが、 m_4 では一転最悪となっており LC が最良である。そのような最悪の場合でも ABRE は LL よりはかなり良い。

表 1 から表 4 の $n = 200$ の場合と表 5 を比較してみよう。これは同じ一様デザインでありながら、それが Fixed か、Random かが推定量の振る舞いにどのような影響を与えたかを見ることになる。関数 m_1, m_2, m_3 に対しては概ね際立った違いは見られず、推定量の振る舞いにはデザインのランダムネスは影響を与えていない。関数 m_4 については違いが見られる。表 4 の MISE を見ると、RB6 グループ内では推定量は競合し、RB8 グループ内では ABRE が良いが、表 5 の

表 5. $n = 200$ の一様 Random Design での結果

	LL4	LC6	ABRE6	DS6	LL6	LC8	ABRE8	DS8
m_1								
ISB	827	62	367	300	525	9	37	22
IV	2103	1798	1753	1770	2332	2747	2620	2685
MISE	2930	1860	2120	2070	2857	2757	2657	2707
MED-ISE	2687	1703	1956	1909	2693	2557	2435	2503
ISE-SE	59	41	48	46	53	52	50	51
Band-MSE	318	3018	2140	—	41	2304	1634	—
m_2								
ISB	3169	13	258	220	1301	42	53	49
IV	18169	23122	20996	21238	23513	36892	34791	35645
MISE	21338	23135	21254	21458	24815	36934	34844	35694
MED-ISE	19590	20724	19342	19539	23047	33977	32748	33322
ISE-SE	468	521	476	481	512	701	656	674
Band-MSE	279	51720	36697	—	1010	104705	74292	—
m_3								
ISB	5187	151	374	216	3186	63	62	50
IV	21104	22680	20709	20932	24184	36036	34018	34857
MISE	26291	22831	21083	21149	27369	36100	34080	34907
MED-ISE	24514	20956	19355	19313	25837	32611	31141	31740
ISE-SE	543	513	468	472	549	707	676	698
Band-MSE	40	7953	5643	—	346	32568	23108	—
m_4								
ISB	223928	48947	71080	62437	49541	6723	9928	6740
IV	96740	53157	57435	58035	61412	65336	66131	67876
MISE	320668	102104	128515	120473	110952	72059	76059	74616
MED-ISE	213004	92830	111761	104687	104838	69463	72404	70738
ISE-SE	12747	2274	3114	2986	1924	1199	1302	1257
Band-MSE	5186	4191	2974	—	473	452	321	—

MISE では、RB6 グループ内、RB8 グループ内共に LC が ABRE を凌ぎ最良となっている。この原因は様々考えられるが、表 5 の IV を見ると ABRE の IV は LC の IV よりも大きくなってしまっている。このことは表 1 から表 4 では見られないことである。ランダムネスの影響が ABRE には大きく現れ、LC にはそれほど現れなかったケースであろう。結局、推定量の振る舞いに関するこのシミュレーションからの全体的な印象としては、『バイアス縮小では LC が ABRE や DS より勝るが、分散があまり増大しないという点では概ね ABRE が LC や DS よりも良く、したがって MISE でも ABRE が良いことが多い。ABRE が LC や DS より悪い場合であっても、LL よりが良い』と総括される。

Bandwidth 選択について見ていこう。 h_{AMISE} と \hat{h}_{DPI} との違いをグラフィカルに示したのが図 2 である。それぞれのパネルでは、各 m_j における LL4 の Bandwidth と ABRE6 の Bandwidth、LL6 の Bandwidth と ABRE8 の Bandwidth の比較が h_{AMISE} との比の Kernel density estimates で与えられている(この Kernel density estimates の計算のための Bandwidth は Normal reference rule, いわゆる正規分布による ROT 推定量で構成した)。比であるから、1.0 の近くに分布しているほど Plug-In 法によるアルゴリズムで h_{AMISE} に近い値の Bandwidth が計算されていることを意味する。

表 6. $n = 200$ の $B(4, 2)$ にしたがるデザインでの結果

	LL4	LC6	ABRE6	DS6	LL6	LC8	ABRE8	DS8
m_1								
ISB	556	11	22	17	191	6	7	6
IV	765	1038	985	1000	993	1651	1608	1660
MISE	1321	1050	1007	1017	1184	1657	1615	1666
MED-ISE	1146	904	857	866	1024	1468	1437	1484
ISE-SE	38	28	27	27	32	39	38	39
Band-MSE	3738	662	470	—	562	4572	3244	—
m_2								
ISB	1566	11	33	30	432	24	24	25
IV	9804	14150	13430	13671	14249	23182	22495	23400
MISE	11370	14161	13463	13701	14681	23206	22519	23424
MED-ISE	9856	12738	12232	12275	12973	21832	20931	21592
ISE-SE	298	354	340	348	354	517	492	521
Band-MSE	103	79278	56250	—	2601	125170	88812	—
m_3								
ISB	3387	28	51	35	1056	35	34	37
IV	9653	12964	12307	12506	13035	21104	20502	21167
MISE	13040	12992	12358	12542	14092	21139	20537	21204
MED-ISE	11398	11639	10839	11153	12613	18873	18556	19094
ISE-SE	348	348	335	341	359	501	477	493
Band-MSE	136	17554	12455	—	916	43444	30825	—
m_4								
ISB	86572	17955	28647	25725	34446	4768	7646	6081
IV	25475	25465	27207	27597	30790	31536	32405	33264
MISE	112047	43420	55853	53322	65236	36304	40051	39345
MED-ISE	107736	41231	51731	49181	61658	34563	37796	36870
ISE-SE	1525	791	1034	999	1054	736	795	787
Band-MSE	1452	2156	1530	—	267	203	144	—

Bandwidth 選択における ROT 推定量の構成での多項式の次数 k の影響に関しては、図 2 で見られるように、総じて LL では 4 次より 6 次、LC と ABRE に関しては 6 次より 8 次多項式をあてはめた方が小さい Bandwidth になり、その分布も概ね 1.0 の周りに近づいている。このことが m の推定量に及ぼす影響は様々であるが、表からもわかるとおり、多くの場合 m の推定量のバイアスは小さくなるものの分散は大きくなっている。関数 m_1, m_2, m_3 の場合は、どの推定量についてもバイアスの値が大きくなるに比べて、次数 k の影響はあるもののそれほど大きくはない。一方 m_4 においては、図 2 の (g) と (h) を見比べてわかるように、Bandwidth のバイアスが大きく改善され、これにより ABRE6 よりも ABRE8 の MISE の値が小さくなったと言えよう。

LL と ABRE の Band-MSE の比較では、多くの場合で LL の方が小さな値となっている。これは Ruppert et al. (1995, 式 (11)) と本論文の定理 7 にあるような収束の速さに関する結果を比較すると説明がつく。LL のための Plug-In 法で得られる Bandwidth の h_{AMISE} に対するいわゆる Relative rate of convergence は $O_p(n^{-2/7})$ であり、ABRE のための Plug-In 法でのそれは $O_p(n^{-2/11})$ である。しかし、このように最適 Bandwidth への収束が少々遅くても、LL より ABRE の方がバイアスを縮小し、MISE でも勝っている点は、ABRE の長所と言える。

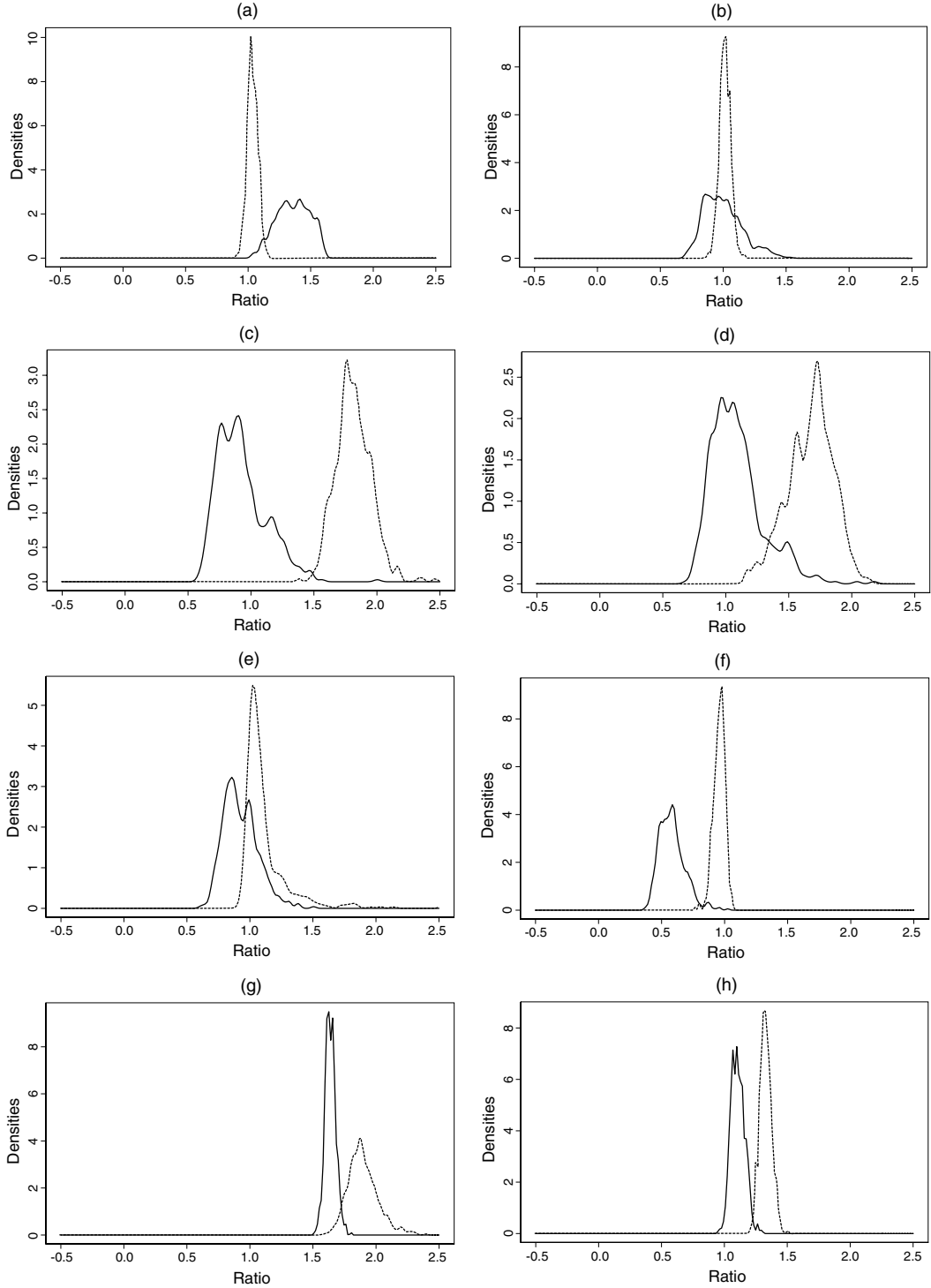


図 2. \hat{h}_{DPI}/h_{AMISE} の Kernel density estimates (Equi-spaced fixed design case). (a) m_1 の LL4 (点線) と ABRE6 (実線), (b) m_1 の LL6 (点線) と ABRE8 (実線). (c) m_2 の LL4 (点線) と ABRE6 (実線), (d) m_2 の LL6 (点線) と ABRE6 (実線). (e) m_3 の LL4 (点線) と ABRE6 (実線), (f) m_3 の LL6 (点線) と ABRE8 (実線). (g) m_4 の LL4 (点線) と ABRE6 (実線), (h) m_4 の LL6 (点線) と ABRE8 (実線).

図3は m_4 の場合で, LL4, ABRE6, ABRE8 のそれぞれで作成された 500 個の Bandwidth の分布において, それぞれ小さい方から 250 番目の Bandwidth を取り出し, その Bandwidth を出したデータで描いた回帰推定量である. 同じ 250 番目であっても LL4 は m_4 の曲率の大きい部分で過剰平滑化がされているのがわかるであろう. また, ABRE8 のための Bandwidth が ABRE6 のための Bandwidth より小さくなっており, その分だけ m_4 の曲率の大きい部分でもその変化を追うことができているのがわかる.

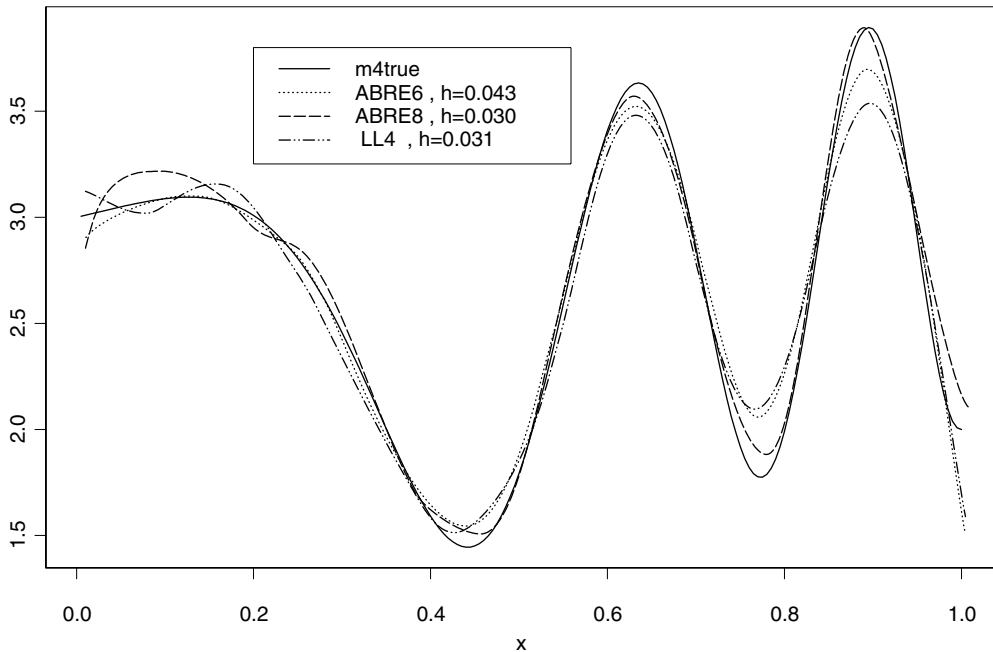


図3. m_4 (実線)と, ABRE6 の 250 番目の推定値(点線), ABRE8 の 250 番目の推定値(破線), LL4 の 250 番目の推定値(破点線) (Equi-spaced fixed design case)

m_4 の場合で ABRE8 に限定して, 図3と同様に小さい方から 50 番目, 250 番目, 450 番目を抽出し, 回帰推定量を描いたのが図4である(250 番目は図3と同じ). 50 番目に対応する回帰推定量はうまく m_4 を推定できているのがわかる. 450 番目に対応する回帰推定量についても, 図3における LL4 と同程度かそれよりわずかに良い精度で推定できていると言えるだろう.

6. 実データへの適用

この節では実データに LL6 と ABRE8 (共に $p = 1$) を適用した例を紹介する. Bandwidth は LL では Ruppert et al. (1995) が提案した方法, ABRE では本論文で提案した方法で求めている. 用いたデータの出典はいずれも Simonoff (1996) である. 図5はブドウ畑の 52 の畝における 3 年間の収量のデータ, 図6は 1940 年 1 月から 1947 年 12 月までの期間(96 ヶ月)のアメリカ合衆国における毎月の出生率のデータである. そして, 図7はある新生シロクジラの 6 時間毎の授乳時間のデータ(228 観測点)である.

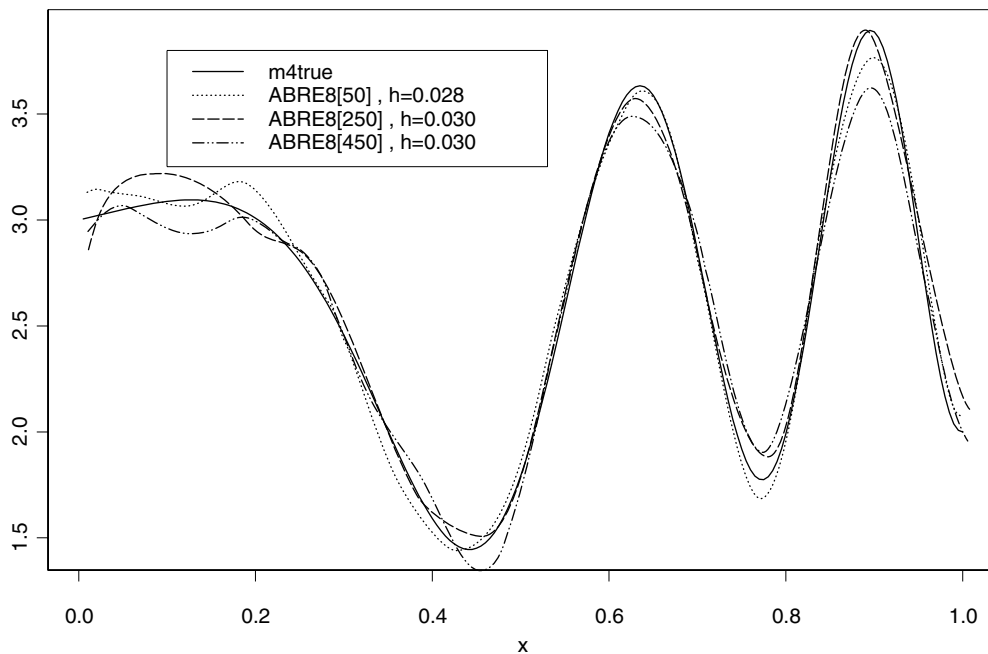


図 4. m_4 (実線)と、ABRE8 の 50 番目の推定値(点線), ABRE8 の 250 番目の推定値(破線), ABRE8 の 450 番目の推定値(破点線) (Equi-spaced fixed design case)

Vine Data(n=52)

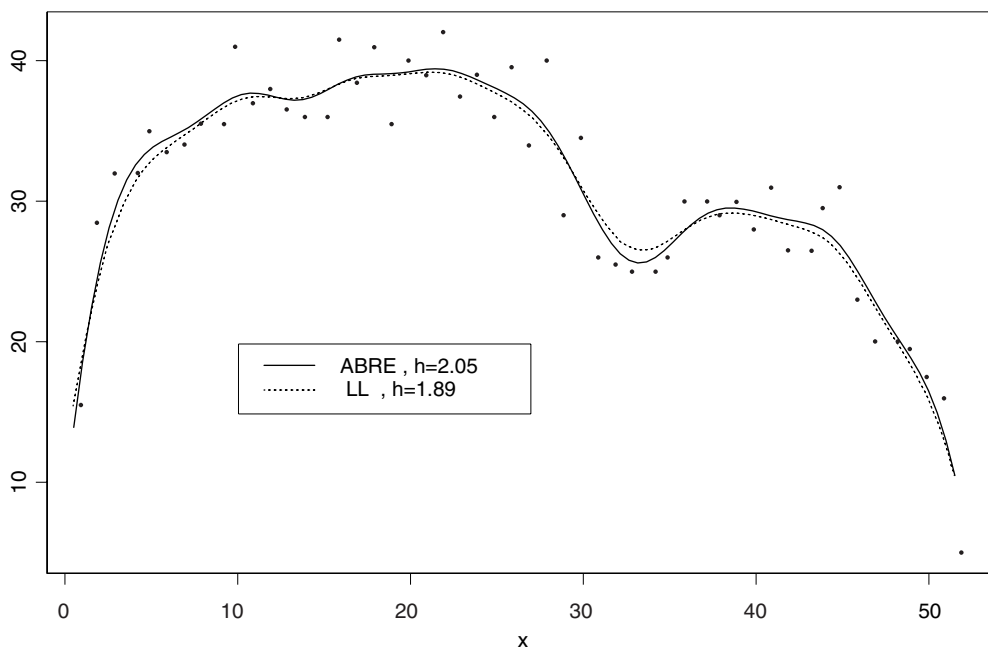


図 5. Vine データの ABRE8 と LL6 の推定値

Birth Rate Data(n=96)

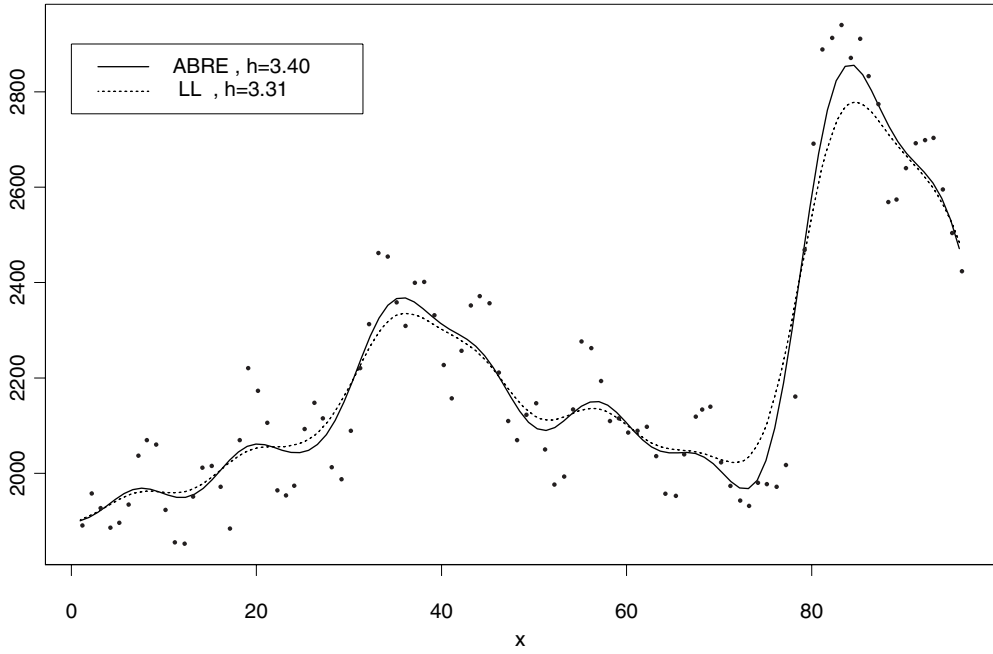


図 6. Birth Rate データの ABRE8 と LL6 の推定値

Whale Data(n=228)

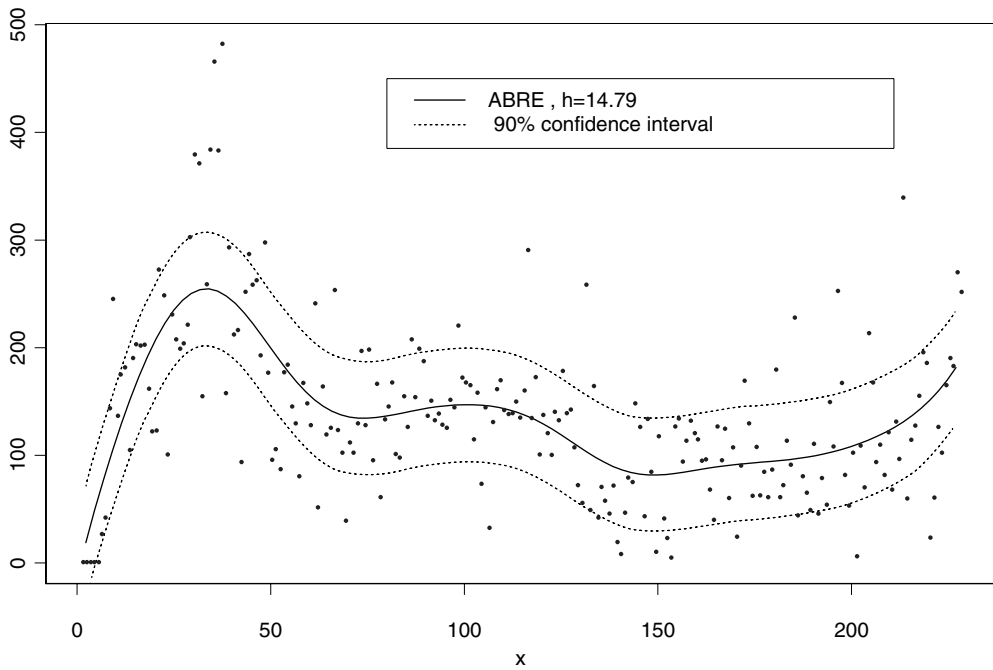


図 7. Whale データの ABRE8 の推定値と 90%信頼区間

個別に見ていくと、まず図 5 の場合は $30 \leq x \leq 35$ の辺りで見られる LL のバイアスを ABRE がきちんと縮小していることがわかるだろう。このようなバイアス縮小の効果は図 6 においても見ることができる。特に、 $80 \leq x \leq 85$ 付近での LL の過剰な平滑化が大きく改善されている。どちらにおいても、LL6 の Bandwidth よりも ABRE8 の Bandwidth の方が大きな値となっているにも関わらず ABRE8 の方がバイアスが縮小しているのは、ABRE の持つ長所の現れである。図 7 では ABRE8 の推定値と各点における 90%信頼区間を求めた。この際に $\gamma = 0.05$ を用いた。

7. おわりに

従来の局所多項式回帰推定量を初期推定量と見なし、その“残差”を局所多項式回帰推定量で平滑化し、それを初期推定量に加えて得られる推定量である ABRE について議論した。定理 1, 2, 3 にまとめられた理論的結果とあわせて、実際の側面を強調して議論を展開した。特に、ABRE の利用のための、Bandwidth の選択アルゴリズムを提案した。そのアルゴリズムで得られた Bandwidth での ABRE の実際の振る舞いを、他の関連する推定量との比較とあわせて大規模なシミュレーションで調べた。

ここで、定理 1 にある一般的結果の意味する事について注意しておこう。条件付きバイアスのオーダーは $O(h^{2p+2})$ であり、同じく p が奇数の場合の局所多項式回帰推定量の条件付きバイアスのオーダーは $O(h^{p+1})$ だから、例えば $p = 1$ の場合の ABRE と同じ条件付きバイアスのオーダーを達成するには、局所多項式回帰推定量単独では、 $p = 3$ を用いないといけないことになる。推定量に現れる行列 X_x のサイズを見れば p の増加による計算コストの増加がわかるであろう。第 5 節のシミュレーションにおいて、 $p = 1$ での ABRE の LC との競合性が確認されており、それは行列のサイズ変更を伴わない ABRE の効率の良さを示していると言える。また、 p が奇数である ABRE では条件付きバイアスに共変量密度は現れず、Fan (1992) の意味で Design-Adaptive となっている。このように、漸近理論、計算効率、そして有限標本でのパフォーマンスといった観点から見て、本論文で提案・議論された ABRE は有効な散布図平滑化手法であると言える。

謝 辞 2 人のレフェリーからの、Data sharpening 推定量の引用に関する指摘、シミュレーションから引き出される知見に関するコメント、推定量構成の記述に関する指摘、定理の証明に関する技術的なコメントはどれも大変有益なものであり、論文改訂に役立ちました。改めて謝意を表します。本研究は、文部科学省科学研究費補助金・若手研究(B)課題番号 15700234 の援助を受けている。

参 考 文 献

- Choi, E., Hall, P. and Rousson, V. (2000): Data sharpening methods for bias reduction in nonparametric regression. *Ann. Statist.* **28**, 1339–1355.
- Fan, J. (1992): Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87**, 998–1004.
- Linton, O. and Nielsen, J.P. (1994): A multiplicative bias reduction method for nonparametric regression. *Statist. Probab. Lett.* **19**, 181–187.
- Nadaraya, E.A. (1964): On estimating regression. *Theory of Probability and Its Applications* **9**, 141–142.
- Naito, K. and Yoshizaki, M. (2003): Nonparametric regression with additive adjustment: further discussion on data sharpening. *Technical Report* 03-11, *Statistical Research Group, Hiroshima University.*

Ruppert, D., Sheather, S.J. and Wand, M.P. (1995): An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90**, 1257–1270.

Ruppert, D. and Wand, M.P. (1994): Multivariate locally weighted least squares regression. *Ann. Statist.* **22**, 1346–1370.

Simonoff, J.S. (1996): *Smoothing Methods in Statistics*, Springer.

Wand, M.P. and Jones, M.C. (1995): *Kernel Smoothing*, Chapman & Hall.

Watson, G.S. (1964): Smooth regression analysis. *Sankhyā Ser. A* **26**, 359–372.

内藤貴太・吉崎正浩 (2002): Nonparametric regression with additive adjustment. 統計数理研究所共同研究リポート 155, 115–128.

付録：理論的結果の証明について

定理 1 の証明の概略：以下では条件付きの記号は省略する． $E[\widehat{m}_p(x)] = E[\widetilde{m}_p(x)] + E[\widetilde{\xi}_0(x)]$ であり，

$$E[\widetilde{m}_p(x)] = m(x) + Cm^{(p+1)}(x)h^{p+1} + \sum_{j=p+2}^{2p+2} R_j(x, h, K) + o_p(h^{2p+2})$$

である．ここで， $C = \mu_{p+1}(K_{(p)})/(p+1)!$ であり，

$$R_j(x, h, K) = \frac{(-1)^j}{j!} m^{(j)}(x) e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x \begin{pmatrix} (x - X_1)^j \\ \vdots \\ (x - X_n)^j \end{pmatrix} = O_p(h^j) = o_p(h^{p+1})$$

($j = p+2, \dots, 2p+2$)

である． $b_p(x) = E[\widetilde{m}_p(x)] - m(x)$ とすると，局所 p 次多項式推定量の残差 $M_i = Y_i - \widetilde{m}_p(X_i)$ について $E[M_i] = -b_p(X_i)$ である．ABRE の調整項 $\widetilde{\xi}_0(x)$ はデータ $(X_i, M_i) (i = 1, \dots, n)$ に対する局所 p 次多項式推定量であったから，

$$\begin{aligned} E[\widetilde{\xi}_0(x)] &= b_p(x) + Ch^{p+1}b_p^{(p+1)}(x) + \sum_{j=p+2}^{2p+2} R_j^*(x, h, K) + o_p(h^{2p+2}) \\ &= \left\{ -Cm^{(p+1)}(x)h^{p+1} - \sum_{j=p+2}^{2p+2} R_j(x, h, K) + o_p(h^{2p+2}) \right\} \\ &\quad + Ch^{p+1} \left\{ -Cm^{(2p+2)}(x)h^{p+1} - \sum_{j=p+2}^{2p+2} R_j^{(p+1)}(x, h, K) + o_p(h^{2p+2}) \right\} \\ &\quad + \sum_{j=p+2}^{2p+2} R_j^*(x, h, K) + o_p(h^{2p+2}) \\ &= -Ch^{p+1}m^{(p+1)}(x) - C^2h^{2p+2}m^{(2p+2)}(x) - \sum_{j=p+2}^{2p+2} R_j(x, h, K) + o_p(h^{2p+2}) \end{aligned}$$

を得る．ここで， $R_j^*(x, h, K)$ は $R_j(x, h, K)$ において $m^{(j)}$ を $b_p^{(j)}$ に置き換えたものであり， $o_p(h^{2p+2})$ である．これより，

$$\begin{aligned} E[\widehat{m}_p(x)] &= E[\widetilde{m}_p(x)] + E[\widetilde{\xi}_0(x)] \\ &= m(x) - \frac{1}{\{(p+1)!\}^2} \mu_{p+1}(K_{(p)})^2 h^{2p+2} m^{(2p+2)}(x) + o_p(h^{2p+2}) \end{aligned}$$

となり、条件付きバイアスが求まる。

次に条件付き分散を求める。まず、

$$\begin{aligned}
& \text{Var}[\widehat{m}_p(x)] \\
&= \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x M V M^T \mathbf{W}_x \mathbf{X}_x (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{e}_1 \\
&= 4\mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x V \mathbf{W}_x \mathbf{X}_x (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{e}_1 \\
&\quad - 2\mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x V L^T \mathbf{W}_x \mathbf{X}_x (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{e}_1 \\
&\quad - 2\mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x L V \mathbf{W}_x \mathbf{X}_x (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{e}_1 \\
&\quad + \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x L V L^T \mathbf{W}_x \mathbf{X}_x (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{e}_1 \\
&\equiv 4A_1 - 2A_2 - 2A_2 + A_3
\end{aligned}$$

となっていることに注意する。ただし、

$$M = 2I_n - L, \quad L = \begin{pmatrix} \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_n \end{pmatrix}, \quad \mathbf{L}_i = \mathbf{e}_1^T (\mathbf{X}_{X_i}^T \mathbf{W}_{X_i} \mathbf{X}_{X_i})^{-1} \mathbf{X}_{X_i}^T \mathbf{W}_{X_i}, \quad V = \text{diag} \{ \delta(X_1), \dots, \delta(X_n) \}$$

である。Ruppert and Wand (1994) にある計算と同様にして、

$$\begin{aligned}
A_3 &= \frac{\delta(x)}{nhf_X(x)} \int \{K_{(p)} * K_{(p)}(u)\}^2 du + o_p\left(\frac{1}{nh}\right), \\
A_2 &= \frac{\delta(x)}{nhf_X(x)} \int K_{(p)}(u) \{K_{(p)} * K_{(p)}(u)\} du + o_p\left(\frac{1}{nh}\right)
\end{aligned}$$

と求めることができ、そして、 A_1 は \widehat{m}_p の条件付き分散ゆえ

$$A_1 = \frac{\delta(x)}{nhf_X(x)} \int K_{(p)}(u)^2 du + o_p\left(\frac{1}{nh}\right).$$

まとめることで定理の条件付き分散の表現を得る。

定理 2 の証明の概略： Ruppert and Wand (1994) より、

$$\begin{aligned}
E[\widehat{m}_p(x)] &= m(x) + \frac{1}{(p+1)!} \mu_{p+1}(K_{(p)}, c) h^{p+1} m^{(p+1)}(x) + o_p(h^{p+1}) \\
&= m(x) + \frac{1}{(p+1)!} \mu_{p+1}(K_{(p)}, c) h^{p+1} m^{(p+1)}(x) + \sum_{j=p+2}^{2p+2} R_j(x, h, K) + o_p(h^{2p+2})
\end{aligned}$$

であり、内点での場合と同様にして

$$\begin{aligned}
E[\widetilde{\xi}_0(x)] &= -\frac{1}{(p+1)!} \mu_{p+1}(K_{(p)}, c) h^{p+1} \left\{ m^{(p+1)}(x) + \frac{1}{(p+1)!} \mu_{p+1}(K_{(p)}, c) h^{p+1} m^{(2p+2)}(x) \right\} \\
&\quad - \sum_{j=p+2}^{2p+2} R_j(x, h, K) + o_p(h^{2p+2}) \\
&= -\frac{1}{(p+1)!} \mu_{p+1}(K_{(p)}, c) h^{p+1} m^{(p+1)}(x) - \left[\frac{1}{(p+1)!} \mu_{p+1}(K_{(p)}, c) \right]^2 h^{2p+2} m^{(2p+2)}(x)
\end{aligned}$$

$$- \sum_{j=p+2}^{2p+2} R_j(x, h, K) + o_p(h^{2p+2}).$$

と求まるから,

$$\begin{aligned} E[\widehat{m}_p(x)] &= E[\widetilde{m}_p(x)] + E[\widetilde{\xi}_0(x)] \\ &= m(x) - \frac{1}{\{(p+1)!\}^2} \mu_{p+1}(K_{(p)}, c)^2 h^{2p+2} m^{(2p+2)}(x) + o_p(h^{2p+2}). \end{aligned}$$

次に条件付き分散を見ていく. 境界の場合も内点の場合と同様に

$$\text{Var}[\widehat{m}_p(x)] = 4A_1 - 4A_2 + A_3$$

という分解から計算を進めていく. 境界モーメントの計算に注意すれば, ほとんど内点の場合と同じ計算から定理の条件付き分散の表現を得る.

定理 3 の証明の概略: いま,

$$\begin{aligned} \widetilde{B}_x &= \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \begin{pmatrix} m(X_1) \\ \vdots \\ m(X_n) \end{pmatrix} - m(x), \\ \widetilde{V}_x &= \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \begin{pmatrix} \delta(X_1)^{1/2} \varepsilon_1 \\ \vdots \\ \delta(X_n)^{1/2} \varepsilon_n \end{pmatrix} \end{aligned}$$

とすると, $\widetilde{m}(x) - m(x) = \widetilde{B}_x + \widetilde{V}_x$ であり,

$$\widetilde{\xi}(x) = \widetilde{V}_x - \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \begin{pmatrix} \widetilde{B}_{X_1} + \widetilde{V}_{X_1} \\ \vdots \\ \widetilde{B}_{X_n} + \widetilde{V}_{X_n} \end{pmatrix}$$

であることがわかる. さらに,

$$\begin{aligned} A_n &= \left[2\widetilde{V}_x - \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \begin{pmatrix} \widetilde{V}_{X_1} \\ \vdots \\ \widetilde{V}_{X_n} \end{pmatrix} \right], \\ b_n &= \left[\widetilde{B}_x - \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \begin{pmatrix} \widetilde{B}_{X_1} \\ \vdots \\ \widetilde{B}_{X_n} \end{pmatrix} \right] \end{aligned}$$

とすると, $\widehat{m}_p(x) - m(x) = A_n(x) + b_n(x)$ であることがわかる.

これより, $p = 1$ とする. まず, $G(x) = \widehat{s}_2(x)\widehat{s}_0(x) - \widehat{s}_1(x)^2$, $H(x, y) = K_h(x - y) \{\widehat{s}_2(x) - \widehat{s}_1(x)(x - y)\}$ とする. ここで,

$$\widehat{s}_\ell(x) = \frac{1}{n} \sum_{j=1}^n (x - X_j)^\ell K_h(x - X_j)$$

である。これらを用いると、 $A_n(x) = \sum_{j=1}^n \varepsilon_j \delta(X_j)^{1/2} \zeta_j(x)$ となる。ただし、

$$\zeta_j(x) = \left[\frac{2H(x, X_j)}{nG(x)} - \frac{1}{n^2} \sum_{\ell=1}^n \frac{H(x, X_\ell)H(X_\ell, X_j)}{G(x)G(X_\ell)} \right].$$

Wand and Jones (1995, 5.3.2) の $\widehat{s}_\ell(x)$ の評価を用いることで、 ζ_j の漸近表現

$$\zeta_j(x) = \frac{f_X(x) - (x - X_j)f'_X(x)}{f_X(x)^2} \left[2K_h(x - X_j) - \frac{1}{h} \int K(z)K\left(\frac{x - X_j}{h} - z\right) dz \right]$$

を得る。これを用いて、あとは Lindeberg-Feller の中心極限定理を適用して

$$n^{4/9} A_n(x) \xrightarrow{D} N(0, v(x))$$

を得る。一方、簡単な計算から $n^{4/9} b_n(x) \xrightarrow{P} b(x)$ がわかるから、定理の漸近正規性が結論される。

(2003年6月4日受付 11月20日最終修正 2004年3月10日採択)

著者連絡先：〒 618-0015 大阪府三島郡島本町青葉 1-21-1-206
吉崎 正浩
E-mail: notfound@kiss.ac (Yoshizaki)

Practical Aspects of Bias Reducing Estimators in Nonparametric Regression

Masahiro Yoshizaki^{1,*} and Kanta Naito¹

¹ Department of Mathematics, Shimane University
1160 Nishi-Kawatsu, Matsue, 690–8504 Japan

Abstract

We discuss anew the kernel method, which is representative of approaches to nonparametric scatter plot smoothing. This paper proposes a new estimator obtained by adding an adjustment term to an initial estimator, where the initial estimator is the well known local polynomial estimator. An appealing feature of the proposed estimator is that it reduces bias; the effect can be observed especially when the true regression function has large curvature. In this paper, we emphasize practical aspects of the use of our proposal, such as introducing a reliable bandwidth selection method and its evaluation, constructing a pointwise approximate confidence interval for the true regression function based on asymptotic normality of the estimator, and comparing our proposal with existing estimators by conducting a large size simulation study.

Key words: bandwidth selection, data sharpening, kernel regression, local polynomial regression, reduced bias

* Corresponding author

E-mail address: notfound@kiss.ac (Yoshizaki)

Received June 4, 2003; Received in final form November 20, 2003; Accepted March 10, 2004.