Nonparametric Regression with Additive Adjustment

島根大学　総合理工学部　　内藤貫太
島根大学　総合理工学研究科　　吉崎正浩

1　はじめに

大容量データから瞬時に構造探索を行う際のツールとして，いまやノンパラメトリック回帰手法は欠かせないものとなっている．Kernel や Spline はすでに標準的な手法となっており [Simonoff(1996)，Wand and Jones(1995) などを参照]，最近では Wavelet の応用も盛んである [例えば Vidakovic(1999) を参照]．多変量回帰まで含めると実に様々なノンパラメトリック回帰手法が提案されていることがわかる．

このような探索的段階で用いられる様々な手法を，単なるデータ記述手法として理解するにとどまってしまうことは健全ではなく，それだけでは不十分である．そのデータ構造探索手法としての“ 振る舞い ”をきちんと理解し，精度評価を行うことで初めて意味ある探索ツールとなると思われる．そのような理論的評価が他の手法に比べて整備されているのが Kernel に基づく手法であろう．本稿の内容もいわゆる Kernel Regression に関するものである．Nadaraya-Watson 推定量 [Nadaraya(1964), Watson(1964)] や局所線形推定量 [Fan(1992)] などがよく知られているが，これらは推定すべき回帰関数の曲率が大きいところで比較的大きなバイアスを持つことが知られている．そこで，多少の分散の増大には目をつぶってでも，バイアスを縮小しようという議論が盛んで，局所多項式回帰推定量 [Ruppert and Wand(1994)] などもその成果の 1 つと見なすこともできる．本稿の目的は，推定量に用いる核関数のオーダーよりも高次のオーダーのバイアス（縮小バイアス）を持つような回帰推定量を提案し，その振る舞いを調べることである．従来の推定量を初期推定量と見なし，それを加法的に調整することで推定量を得るという構成法が特徴的である．以下において，推定量の構成を紹介し，推定量の漸近的性質，従来の推定量との比較，実データへの適用例などについて報告する．

2　推定量の構成：Fixed Design

サンプルを $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ とする．ここで，$X_i = i/n, i = 1, 2, \cdots, n$ の Equi-Spaced Fixed Design を考える．$m(X_i) = E[Y_i], \varepsilon_i = Y_i - m(X_i), i = 1, 2, \cdots, n$ は i.i.d. で平均 0，分散 σ^2 とする．問題は回帰関数 m の推定である．重み関数として

$$W_{xi} = \frac{1}{nh} K\left(\frac{x - X_i}{h}\right)$$

を定義する．ここで K は原点対称密度関数であり，h は Bandwidth である．

推定量は以下のように構成される．まず, Fan(1992) の局所線形推定量

$$\widetilde{m}(x) = \frac{\sum_{i=1}^{n} W_{xi} Y_i \left\{\widehat{s}_2(x) - \widehat{s}_1(x)(x - X_i)\right\}}{\widehat{s}_2(x)\widehat{s}_0(x) - \widehat{s}_1(x)^2}$$

を初期推定量として採用する．ただし，

$$\widehat{s}_\ell(x) = \sum_{i=1}^{n} W_{xi}(x - X_i)^\ell$$

である．次に $\widetilde{m}$ を $\widetilde{m}+\xi$ という形で調整することを試みる (Additive Adjustment)．調整項 ξ としては，各 x について，

$$L(\xi_0,\xi_1|x)=\sum_{i=1}^{n}W_{xi}\left\{Y_i-\widetilde{m}(X_i)-\xi_0-\xi_1(x-X_i)\right\}^2$$

を (ξ_0,ξ_1) について最小化したときの切片項 $\widehat{\xi}_0=\widehat{\xi}_0(x)$ を採用しよう．つまり調整項 $\xi=\xi(x)$ の実体は“ 局所線形推定量の残差の局所線形推定量 ”に他ならない．実際には

$$\xi(x)=\widehat{\xi}_0=\frac{\sum_{i=1}^{n}W_{xi}\left\{Y_i-\widetilde{m}(X_i)\right\}\left\{\widehat{s}_2(x)-\widehat{s}_1(x)(x-X_i)\right\}}{\widehat{s}_2(x)\widehat{s}_0(x)-\widehat{s}_1(x)^2}$$

と求まり，提案される推定量 (Additive Bias Reduction Estimator:ABRE)

$$\widehat{m}(x)=\widetilde{m}(x)+\xi(x)=\frac{\sum_{i=1}^{n}W_{xi}\left\{2Y_i-\widetilde{m}(X_i)\right\}\left\{\widehat{s}_2(x)-\widehat{s}_1(x)(x-X_i)\right\}}{\widehat{s}_2(x)\widehat{s}_0(x)-\widehat{s}_1(x)^2}$$

が導出される．

いわゆる回帰の分野で言われる“ 残差 ”とは加法的なものがほとんどであるという事を鑑みて，まず散布図を従来の平滑化推定量で平滑化し，その残差の平滑化推定量をもとの推定量に“ 加えれば ”よいだろうという，極めて単純な直感にもとづいて構成されたものがこの ABRE である．

3 漸近理論

漸近理論を考える上で，Bandwidth h については $h=h(n)=\gamma_n n^{-1/9}$ とする．ここで，$\{\gamma_n\}_{n=1}^{\infty}$ は $\gamma_n\to\gamma<\infty$ であるような positive sequence とする．

3.1 漸近正規性

Equi-Spaced Fixed Design の設定においては、次のように推定量の漸近正規性が得られる．

定理 1　$m(x)$ は 4 回連続微分可能とすると，

$$n^{4/9}\left\{\widehat{m}(x)-m(x)\right\}\stackrel{\mathcal{D}}{\Longrightarrow}N(b(x),v),$$

ここで，

$$\begin{aligned}
b(x) &= -\frac{1}{4}\mu_2(K)^2\gamma^4 m^{(4)}(x),\\
v &= \frac{\sigma^2}{\gamma}\int K^*(u)^2du,\\
\mu_\ell(K) &= \int u^\ell K(u)du,\\
K^*(u) &= 2K(u)-\int K(v)K(u-v)dv.
\end{aligned}$$

漸近分散に Twiced-Kernel K^* が現れるのが特徴的である．この分布論的結果は極めて有用である．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，

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

より

$$I_{1-\alpha} = [\widehat{m}(x) - n^{-4/9}\{b(x) + z_{\alpha/2}\sqrt{v}\}\ ,\ \widehat{m}(x) - n^{-4/9}\{b(x) - z_{\alpha/2}\sqrt{v}\}] \tag{3.1}$$

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．しかしながら実際に (3.1) を用いるためには m の 4 階微分などを推定する必要がある．これは平滑化パラメータのデータに基づく最適選択と同様な問題となっていて，別途考察が必要となる [Ruppert, Sheather and Wand(1996)].

3.2 理論的比較

Bandwidth の定義から，ABRE の漸近バイアスは

$$-\frac{1}{4}h^4\mu_2(K)^2 m^{(4)}(x)$$

であるから，局所線形推定量の漸近バイアス

$$\frac{1}{2}h^2\mu_2(K)m''(x)$$

と比較すると，同じ 2 階 kernel を用いているにも関わらず ABRE はバイアスの縮小が達成されていることがわかる．

一方，Linton and Nielsen(1994) では積的な調整項に基づく推定量 (Multiplicative Bias Reduction Estimator:MBRE) の漸近正規性を導出している．MBRE を $\widehat{m}_{LN}(x)$ と書くことにすると

$$n^{4/9}\left\{\widehat{m}_{LN}(x) - m(x)\right\} \stackrel{\mathcal{D}}{\Longrightarrow} N(b_{LN}(x), v) \tag{3.2}$$

が成立する．漸近分散は ABRE と同じであるが，漸近バイアスは，

$$b_{LN}(x) = -\frac{1}{4}\mu_2(K)^2\gamma^4 m(x)\left\{\frac{m''(x)}{m(x)}\right\}''$$

で与えられ，ABRE の漸近バイアス $b(x)$ に比べると複雑な構造をしているのがわかる．その意味で，ABRE の方が MBRE よりも安定していると言えるだろう．ABRE と MBRE の更なる比較は，Random Design における境界領域での挙動を調べる中でなされる．

4 シミュレーション

Fig.3 と Fig.4 は ABRE，MBRE，局所線形推定量の MSE の比較である．定義域 $[0,1]$ での Equi-Spaced Fixed Design を用いて Fig.3($n=200$) は $y_i = m_1(x_i) + 0.05\varepsilon_i$，Fig.4($n=100$) は $y_i = m_2(x_i) + \varepsilon_i$ から 1000 組の乱数を生成し，$h=0.2$ として得られた MSE の値を x に対してプロットしたものである．ただし，$\varepsilon_i \sim N(0,1)$ であり，

$$\begin{aligned} m_1(x) &= 2\exp\left[-\frac{(x-0.1)^2}{0.09}\right] + 3\exp\left[-\frac{(x-0.9)^2}{0.64}\right] \text{ (Fig.1)}, \\ m_2(x) &= \exp(4x) \text{ (Fig.2)} \end{aligned}$$

である．また，正規核関数を用いている．

Fig.3 では特に局所線形推定量と比較して x が 0.4 から 0.6 の辺りの $m_1(x)$ の曲率の大きな所での改善が見られる．Fig.4 で用いている $m_2(x)$ は比較的なだらかに変化する関数だが，x が大きくなるにつれて局所線形推定量との差が大きく見られる．

Fig.1 m1(x)

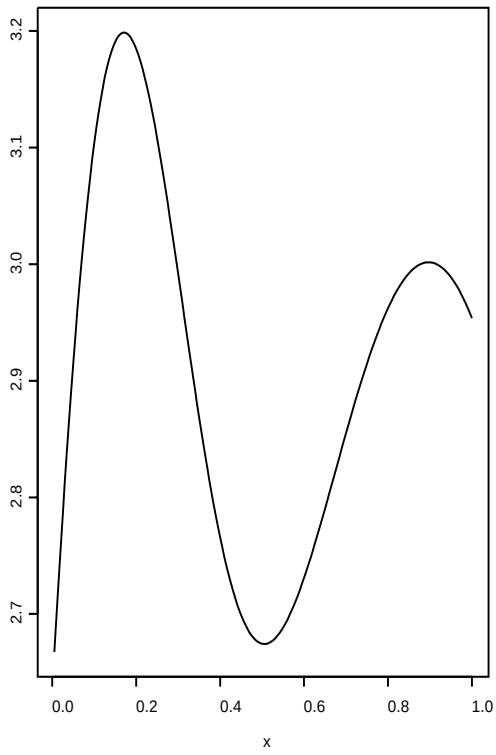


Fig.2 m2(x)

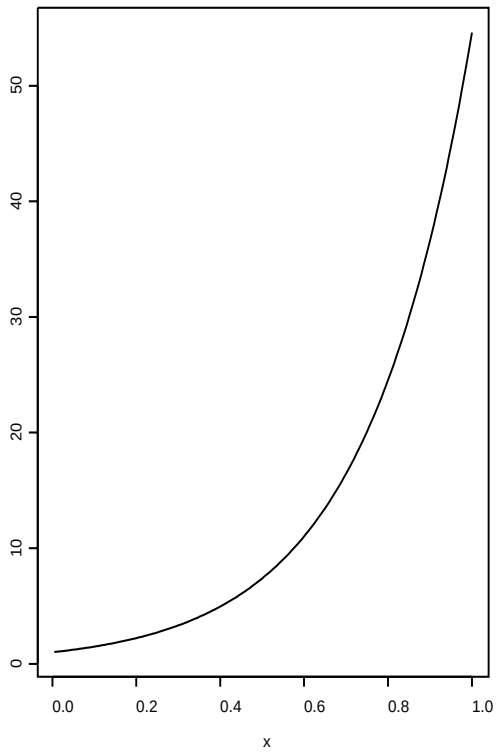


Fig.3 Simulation(n=200)

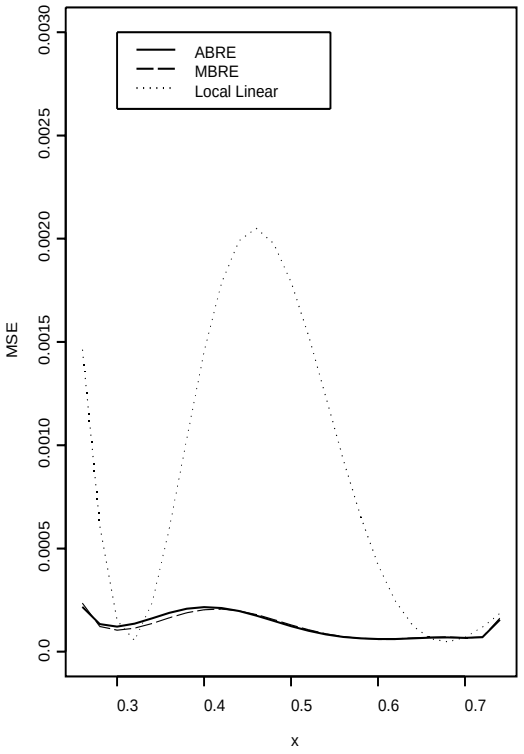
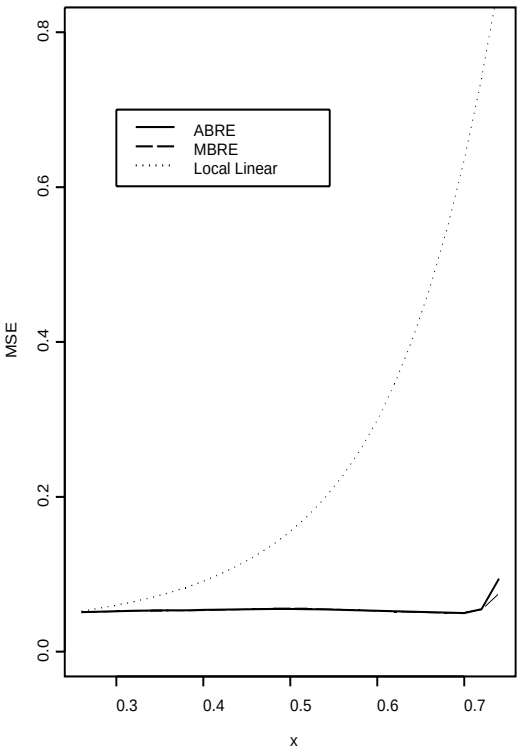


Fig.4 Simulation(n=100)



5 Random Design

次に $f(x,y)=f_{Y|X}(y|x)f_X(x)$ で，$(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)\sim f(x,y)$ である Random Design を考える．また，ε_i は i.i.d. で期待値は 0，分散は 1 であり，X とは独立であるとする．Random Design での ABRE の挙動は，条件付き期待値，条件付き分散で評価される．

5.1 漸近理論

ABRE の漸近的な条件付きバイアス，条件付き分散は次のようになる．

定理 2 $m(x)$ は 4 回連続微分可能，$f_X(x)>0$ とすると，$n\to\infty$，$h\to 0$ のとき，

$$\begin{aligned}\mathrm{Bias}[\widehat{m}(x)|X_1,\cdots,X_n] &= -\frac{1}{4}h^4\mu_2(K)^2m^{(4)}(x)+o_p(h^4),\\ \mathrm{Var}[\widehat{m}(x)|X_1,\cdots,X_n] &= \frac{v(x)}{nhf_X(x)}\int K^*(u)^2du+o_p\left(\frac{1}{nh}\right).\end{aligned}$$

Fixed Design での推定量の振る舞いとの違いは，分散に Design の密度が現れるのみである．

5.2 境界付近での挙動

Kernel を用いた推定量の特色として，境界バイアスが挙げられる．境界での漸近的な条件付きのバイアス，分散に関しては次が成り立つ．

定理 3 f_X の定義域を $[0,1]$，Kernel の定義域を $[-1,1]$ とし，$m(x)$ は 4 回連続微分可能，$f_X(x)>0$ とすると，0 の境界近くの点 $(x=ch,0<c<1)$ では，$h\to 0$ のとき，

$$\begin{aligned}\mathrm{Bias}[\widehat{m}(x)|X_1,\cdots,X_n] &= -\frac{1}{4}h^4\alpha(c)^2m^{(4)}(x)+o_p(h^4),\\ \mathrm{Var}[\widehat{m}(x)|X_1,\cdots,X_n] &= \frac{v(x)}{nhf_X(x)Q(c)}\int_{-1}^{c}K^+(u,c)^2du+o_p\left(\frac{1}{nh}\right).\end{aligned}$$

ただし，

$$\begin{aligned}s_\ell(c) &= \int_{-1}^{c}u^\ell K(u)du,\\ Q(c) &= s_2(c)s_0(c)-s_1(c)^2,\\ \alpha(c) &= \frac{s_2(c)^2-s_3(c)s_1(c)}{s_2(c)s_0(c)-s_1(c)^2},\\ \beta(u,c) &= \{s_2(c)-us_1(c)\}K(u),\\ K^+(u,c) &= 2\beta(u,c)-\frac{1}{Q(c)}\int_{-1}^{c}\beta(z,c)\beta(u-z,c)dz.\end{aligned}$$

ここで積的調整に基づく推定量 MBRE との比較を行おう．MBRE は Fixed Design でのみ提案された推定量であった．その Random Design への拡張は容易になされ，内点での条件付きバイアス，条件付き分散は (3.2) の漸近バイアス，漸近分散に対応するものとなっている．また，境界での条件付きバイアスは

$$\mathrm{Bias}[\widehat{m}_{LN}(x)|X_1,\cdots,X_n]=-h^2\frac{s_1(c)^2}{s_0(c)^2}m(x)\left[\frac{m'(x)}{m(x)}\right]'+o_p(h^2) \tag{5.1}$$

となり，境界での条件付き分散は

$$\begin{aligned}\mathrm{Var}[\widehat{m}_{LN}(x)|X_1,\cdots,X_n] &= \frac{v(x)}{nhf_X(x)}\int_{-1}^{c} K_{LN}^{+}(z,c)^2dz + o_p\left(\frac{1}{nh}\right),\\ K_{LN}^{+}(z,c) &= 2K(z) - \int_{-1}^{c} K(y)K(z-y)\left\{\int_{-1}^{c} K(u-y)du\right\}^{-1}dy\end{aligned}$$

と求まる．定理 3 のバイアスの式と (5.1) を比べることで，ABRE の MBRE に対する優越性がわかるだろう．

5.3 一般化

ここまでは初期推定量を局所線形推定量とし，調整項を“ 局所線形推定量の残差の局所線形推定量 ”として ABRE を構成した．今，p を奇数とし，初期推定量を局所 p 次多項式推定量，調整項を“ 局所 p 次多項式推定量の残差の局所 p 次多項式推定量 ”として構成すると，次のように一般化できる．

$\boldsymbol{e}_1 = (1,0,\cdots,0)^T$：$(p+1)\times 1$ ベクトル，$\boldsymbol{Y} = (Y_1,\cdots,Y_n)^T$：$n\times 1$ ベクトル，

$$X_x = \begin{pmatrix} 1 & x-X_1 & \cdots & (x-X_1)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x-X_n & \cdots & (x-X_n)^p \end{pmatrix},\ W_x = \frac{1}{h}diag\left\{K\left(\frac{x-X_1}{h}\right),\cdots,K\left(\frac{x-X_n}{h}\right)\right\}$$

とすると，局所 p 次多項式推定量は

$$\widetilde{m}(x) = \boldsymbol{e}_1^T(X_x^TW_xX_x)^{-1}X_x^TW_x\boldsymbol{Y} \tag{5.2}$$

と表せる．この $\widetilde{m}$ を初期推定量として得られる ABRE は，

$$\boldsymbol{R} = \begin{pmatrix} 2Y_1 - \widetilde{m}(X_1) \\ \vdots \\ 2Y_n - \widetilde{m}(X_n) \end{pmatrix}$$

として，

$$\widehat{m}(x) = \boldsymbol{e}_1^T(X_x^TW_xX_x)^{-1}X_x^TW_x\boldsymbol{R}$$

となる．この $\widehat{m}$ の漸近的な条件付きバイアス，条件付き分散に関して次が成り立つ．

定理 4　p を奇数とする．$f_X(x)>0$，$m(x)$ は $2p+2$ 回微分可能とすると，f_X の内点では，$n\to\infty$，$h\to 0$　のとき，

$$\begin{aligned}\mathrm{Bias}[\widehat{m}(x)|X_1,\cdots,X_n] &= -\frac{1}{\{(p+1)!\}^2}h^{2p+2}\mu_{p+1}(K_{(p)})^2m^{(2p+2)}(x) + o_p(h^{2p+2}),\\ \mathrm{Var}[\widehat{m}(x)|X_1,\cdots,X_n] &= \frac{v(x)}{nhf_X(x)}\int K_{(p)}^{*}(u)^2du + o_p\left(\frac{1}{nh}\right).\end{aligned}$$

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 $(1,u,\cdots,u^p)^T$ に入れ替えた行列であり，

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

である．

pが偶数の場合の条件付きバイアス，条件付き分散も同様に求めることが可能であるが，特に条件付きバイアスは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．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 5　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，$h \to 0$　のとき，

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

ただし，$N_p(c)$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$(1, u, \cdots, u^p)^T$に入れ替えた行列であり，

$$\begin{aligned}
K_{(p)}(u, c) &= \frac{|M_p(u, c)|}{|N_p(c)|} K_{(p)}(u), \\
\mu_\ell(K, c) &= \int_{-1}^{c} u^\ell K(u) du, \\
\widetilde{K}_{(p)}(u, c) &= 2K_{(p)}(u, c) - \int_{-1}^{c} K_{(p)}(z, c) K_{(p)}(u - z, c) dz.
\end{aligned}$$

6　実データへの適用

Fig.5とFig.6は実データに局所線形推定量とABRE($p = 1$)を適用した例である．Fig.5では，$x = 40$や$x = 80$辺りを中心に局所線形推定量に見られるような“ 平滑化しすぎ ”によるバイアスが，ABREでは縮小されていることがわかる．また，Fig.6でも$30 \leq x \leq 35$の辺りでバイアスの縮小が見られると同時に，$x < 3$辺りの境界付近での境界バイアスの縮小も見ることができる．

Bandwidthは，局所線形推定量ではRuppert, Sheather and Wand(1995)による方法で得られたものを用いている．Ruppertらの方法は$m(x)$の漸近的なMISEの式に基づいて，まず$m(x)$の2階微分を推定し，それをPlug-Inして最適なBandwidthの推定量を構成するものである．その他，誤差の分散の推定量も必要となり，実装の上では幾分面倒なステップを要するが，使い物になる方法である．一方，ABREのBandwidthについては，Ruppertらの方法をABREに対応するように拡張した方法を用いて選択している．その拡張した方法を用いると，(3.1)の近似信頼区間も構成可能となる．

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で平滑化し，それを初期推定量に加えて得られる推定量について議論した．そのような加法的調整がこれまで提案されている積的調整に基づく推定量よりも特に境界挙動において優れていることを見た．

最後に注意しておきたいのは定理4にある一般的結果の意味する事である．Ruppert and Wand(1996)の局所多項式推定量の結果で知られているように，pが偶数の局所多項式推定量の漸近バイアスには共変量密度が現れるが奇数の場合は現れない．Fan(1992)の局所線形推定量は$p = 1$の場合だから漸近バイアスに共変量密度は現れず，それをもってDesign-Adaptiveと言ったのである．定理4の結果でもpが奇数であれば，得られる推定量はDesign-Adaptiveとなる．しかも，漸近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推定量の漸近バイアスのオーダーは$O(h^{p+1})$だから，

Fig.5 Birth Rate Data(n=96)

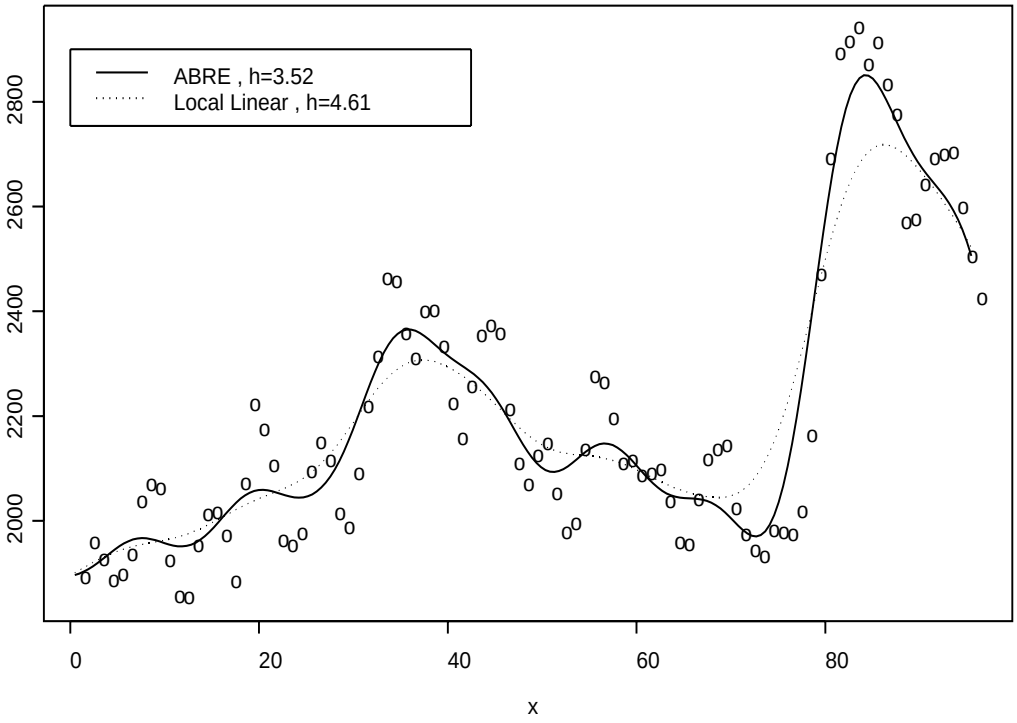
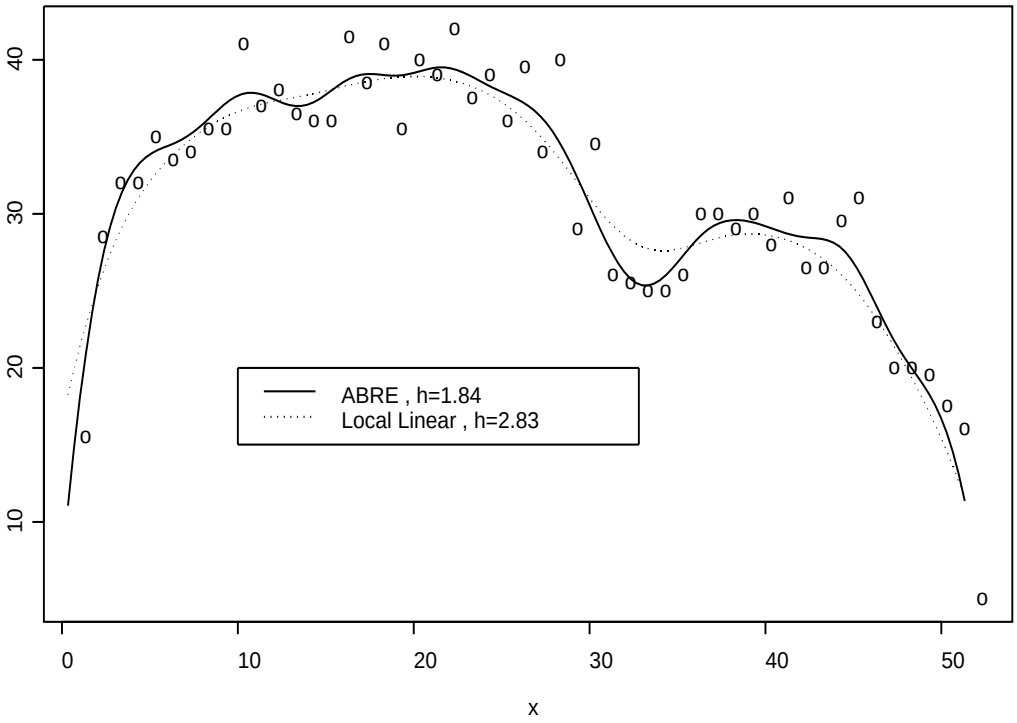


Fig.6 Vine Data(n=52)



例えば $p=1$ の場合，つまり局所線形推定量を局所線形で調整した推定量（定理 2）と同じ漸近バイアスのオーダーを達成するには，局所多項式推定量単独では，$p=3$ を用いないといけないことになる．(5.2) の局所多項式推定量の形からもわかるように，$p=1$ と $p=3$ の違いは計算コストの上では小さくはなく，その意味でも本稿の調整に基づく手法は有効であると言える．

参考文献

1. Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87**, 998-1004.
2. Linton, O. and Nielsen, J. P. (1994). A multiplicative bias reduction method for nonparametric regression. *Statist. Probab. Lett.* **19**, 181-187.
3. Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications* **9**, 141-142.
4. 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.
5. Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22**, 1346-1370.
6. Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Springer.
7. Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, Wiley.
8. Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman & Hall.
9. Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26**,359-372.

連絡先：〒 690-8504　松江市西川津町 1060　島根大学総合理工学部数理・情報システム学科
E-mail:naito@math.shimane-u.ac.jp（内藤貫太），s97499@math.shimane-u.ac.jp（吉崎正浩）