

A Kernel Smoothing Approach to Nonparametric Binomial Regression

中国短大 奥村英則
島根大 内藤貴太

1 はじめに

固体の反応の有, 無をそれぞれ 1, 0 で表す 2 値確率変数 Y は, 用量 (dose) x が与えられたときの条件付き確率

$$P(Y = 1|x) = p(x)$$

をもつと仮定する. 用量 x の関数 $p(x)$ は用量反応曲線 (dose response curve) 呼ばれる. 異なった分野, 例えば薬理学や毒物学で, 固体がある割合で反応する用量を調べることに関心がある. ここで関心ある用量とは, $0 < \alpha < 1$ とするとき, $100\alpha\%$ 有効量 ($ED_{100\alpha\%}$), $100\alpha\%$ 致死量 ($LD_{100\alpha\%}$) である. これらの用量を求めるために用量反応曲線の推定が行われる.

用量反応曲線の推定には 2 つの基本的アプローチがある. 1 つは, プロビット法やロジット法に代表されるパラメトリックアプローチである. この場合, $p(x)$ は有限個のパラメータで特徴付けられるパラメトリックモデルに属す, もしくは少なくともその近傍にあるという信念に基づいて推論される. これは $p(x)$ が仮定したモデルに含まれる場合には, 少ないデータで精度の良い推定ができるというメリットがある. もう 1 つのアプローチは, $p(x)$ に殆ど制約をしないで推論するノンパラメトリックアプローチである. 生体のある用量に対する反応のメカニズムが, 複雑で用量反応曲線の大部分わからないような場合, ノンパラメトリックアプローチはパラメトリックアプローチの代わりとなる. この場合, ノンパラメトリック回帰の種々の手法を使うことができる. ノンパラメトリック回帰の手法の概観は, Härdle(1990), Green and Silverman(1994) と Wand and Jones(1995) にみることができる. これらは, カーネル法, 平滑化スプラインと局所尤度法を含む. Fan, Heckman and Wand(1995) による局所ロジスティック回帰は 2 値反応変数を目的にされている. 2 値反応に対して Copas(1983) が最初に, カーネル推定量の 1 つである Nadaraya-Watson 推定量を使用した. 本稿では, 新たなカーネル推定量を提案し, その挙動について考察する. ここではデータは等間隔の各用量で固体の反応が繰り返し観測されるものとする. このとき, 各用量 x_i は一般性を欠くことなく,

$$x_i = (i - 1)h; i = 1, \dots, K$$

と書くことができる. 各用量 x_i における N_i 個の固体中反応した固体の和を Y_i とする. 各確率変数 Y_i は, パラメータ N_i と $p_i = p(x_i)$ をもつ 2 項分布 $B(N_i; p_i)$ に従うとし, $Y_i; i = 1, \dots, K$ は独立であるとする. このデータに対する Nadaraya-Watson 推定量は

$$\hat{p}_{NW}(x; h) = \frac{\sum_{i=1}^K Y_i w_i}{\sum_{i=1}^K N_i w_i} \quad (1)$$

で与えられる (Lloyd(1999)). ここで, $w_i = w(h^{-1}(x - x_i))$ であって, $w(x)$ はある滑らかなカーネル関数, h はバンド幅である. $w(x)$ については, 2 次カーネル, すなわち原点对称密度を用いるも

のとする。

各用量 x_i における反応確率 $p_i = p(x_i)$ の最尤推定量 (MLE) $Y_i = N_i$ を \hat{p}_i と書く。最も単純なノンパラメトリック推定量は、点 $(x_i; \hat{p}_i); i = 1; \dots; K$ を通る折れ線である。折れ線 $\hat{p}(x)$ は

$$\hat{p}(x) = (K - 1) \frac{f(x - \frac{i-2}{K-1})}{K-1} \hat{p}_i + \frac{i-1}{K-1} \hat{p}_{i+1}; x_{i-1} \leq x < x_i \quad (2)$$

で与えられる。ここで $x_0 = 0$ である。

次節では、データ $f(x_i; \hat{p}_i); i = 1; \dots; K$ をスムージングするという立場から構成した、あるカーネル推定量を提案する。その推定量は MLE \hat{p}_i の漸近分散推定量を含む形で構成される。第3節では、提案された推定量、Nadaraya-Watson 推定量と折れ線の漸近的性質が示される。そして、これらの推定量に対して、用量レベル K が小さいときの理論的比較が第4節で、数値的比較が第5節で行われる。

2 重み付きカーネル推定量

本節では、データ $f(x_i; \hat{p}_i); i = 1; \dots; K$ をスムージングする立場から、用量反応曲線 $p(x)$ のあるカーネル推定量を構成する。このデータに対して、曲線 $p(x)$ の局所的な推定量を考える。このとき、単純な1つの推定量として各用量 x で基準

$$\sum_{i=1}^K w_i (\hat{p}_i - \hat{a})^2$$

を \hat{a} について最小化する量 $\hat{a} = \hat{p}(x; h)$ が考えられるであろう。実際にはこれは

$$\hat{p}(x; h) = \frac{\sum_{i=1}^K \hat{p}_i w_i}{\sum_{i=1}^K w_i}$$

と表すことができる。各用量 x_i での観測個体数 N_i がすべて等しいとき、この推定量は $\hat{p}_{NW}(x; h)$ と等しくなるのがわかる。しかし \hat{p}_i は、分散 $v_i = N_i$ を持つ。ここで $v_i = p_i(1 - p_i)$ である。そこで各用量 x_i の分散 $v_i = N_i$ の大きさに応じて、重み w_i を変動させることを考える。これは局所的に“重み付き最小2乗法”を適応することによって解決される。その推定量 $\hat{p}_v(x; h)$ は基準

$$\sum_{i=1}^K \frac{N_i w_i}{v_i} (\hat{p}_i - \hat{a})^2$$

を最小化することによって得られ、それは

$$\hat{p}_v(x; h) = \frac{\sum_{i=1}^K Y_i w_i = v_i}{\sum_{i=1}^K N_i w_i = v_i}$$

と表せる。ここで v_i は未知であり、Berkson(1955) のルールを使用した p_i の推定量を

$$\hat{p}_i = \begin{cases} 1 = (2N_i); & Y_i = 0; \\ Y_i = N_i; & 0 < Y_i < N_i \\ (2N_i - 1) = (2N_i); & Y_i = N_i \end{cases}$$

で定義し、 v_i の推定量 \hat{v}_i として、 $\hat{p}_i(1 - \hat{p}_i)$ を採用する。最終的に、 $\hat{p}_v(x; h)$ の式に \hat{v}_i を代入して得られる

$$\hat{p}_v(x; h) = \frac{\sum_{i=1}^K Y_i w_i = \hat{v}_i}{\sum_{i=1}^K N_i w_i = \hat{v}_i} \quad (3)$$

が本稿で提案する推定量である。

用量 x_i での観測個体数 N_i 個中 j 番目の固体の反応を、 Y_{ij} で表す。式 (1) の $\hat{p}_{NW}(x; h)$ については、基準

$$\sum_{i=1}^K \sum_{j=1}^{X_j} w_i (Y_{ij} - \hat{a})^2$$

を最小化するものと解釈される。ここで、 $\hat{p}_{NW}(x; h)$ は誤差 $(Y_{ij} - \hat{a})^2$ に対して、重みの付け方が各用量 x_i で同じである。一方、 $\hat{p}_v(x; h)$ は誤差 $(\hat{p}_i - \hat{a})^2$ に対して、各用量 x_i で重みが変わっていることに注意する。

3 推定量の漸近的性質

用量レベル数 K を固定して、各用量での個体数を増したとき、3つのノンパラメトリック推定量 $\hat{p}_v(x; h)$, $\hat{p}_{NW}(x; h)$, $\hat{p}(x)$ の挙動について調べよう。用量レベル数 K とバンド幅 h を固定し、 $N = \sum_{i=1}^K N_i \rightarrow \infty$ としたとき $N_i \rightarrow \infty$ ($i=1, \dots, K$) である状況を考える。

このような状況での考察について少し述べておく。例えば用量反応曲線の推定量として折れ線を考えると、観測用量点以外では用量反応曲線の推定量としては不十分であることは明らかであろう。つまり、 $p(x)$ という関数の推定量としては、観測用量点以外の x では見当違いのものを推定している。とすると、ここで考えられる推定量も含めた、いわゆる用量反応曲線推定問題に用いられる推定量は、そもそも任意の x において何を推定しようとしているのかが疑問になる。それを考えるには各点 x において、推定量のできる限り漸近理論によらない厳密な期待値を見る必要があるだろう。そして、その期待値が $p(x)$ と近いかが問題となる。また、任意の x においての、推定量としての安定性を見るには、その分散が評価される必要がある。このように期待値、分散を評価する際に、 $\hat{p}_{NW}(x; h)$, $\hat{p}(x)$ については導出は容易だが、 $\hat{p}_v(x; h)$ の評価では $N \rightarrow \infty$ という漸近操作が止むを得ず必要となる。このように、できる限り厳密な期待値、分散を求めるという意味で、用量レベル数 K , バンド幅 h は固定、 $N \rightarrow \infty$ という求まる最低限の状況を考えている。1つの見方としては、ここでの用量レベル数 K , バンド幅 h を固定した上での評価は、推定量のノンパラメトリックな評価の一手手前の評価と言える。つまり、関数推定の問題ゆえ、 $h \rightarrow 0$, $K \rightarrow \infty$ といったノンパラメトリックな漸近理論の枠組みが常套である。そして、そのような状況でのみ一貫性などの推定量として持つべき漸近的性質が保証されるわけである。実際、 $h \rightarrow 0$, $K \rightarrow \infty$ の状況ではノンパラメトリック回帰の枠組みとなり、推定量の期待値の漸近展開の主項は $p(x)$ となる。しかし、以下でも述べるようにここでの $h \rightarrow 0$, $K \rightarrow \infty$ を考えない状況では一貫性などの性質は見えてこない。要するに、ここで考える用量レベル数 K , バンド幅 h は固定、 $N \rightarrow \infty$ の状況は、ノンパラメトリック推定量のパラメトリックな状況での評価とも言えるであろう。

このとき、提案した推定量 $\hat{p}_v(x; h)$ の平均、分散は適当な条件の下で以下ようになる。

$$E[\hat{p}_v(x; h)] = \frac{\sum_{i=1}^K N_i w_i (1 - \hat{p}_i)}{\sum_{i=1}^K N_i w_i} + \frac{1}{N} G + o\left(\frac{1}{N}\right); \quad (4)$$

$$V[\hat{p}_v(x; h)] = \frac{\sum_{i=1}^K N_i w_i^2 v_i}{\left(\sum_{i=1}^K N_i w_i\right)^2} + \frac{1}{N} H + o\left(\frac{1}{N}\right); \quad (5)$$

ここで

$$G = \frac{N}{D^2} \sum_{i \neq j} \frac{N_i w_i w_j}{v_i v_j^2} (p_j^3 - \hat{p}_i (1 - 3v_j)) - \frac{N}{D^3} \sum \frac{N_i N_j w_i w_j^2}{v_i v_j^3} (p_j - \hat{p}_i) (1 - 2p_j)^2;$$

$$H = \frac{2N}{D^3} \times \frac{N_i N_j w_i w_j^2}{v_i v_j^2} (p_j - p_i)(1 - 2p_j) \\ + \frac{N}{D^4} \times \frac{N_i N_j N_k w_i w_j^2 w_k}{v_i v_j^3 v_k} (1 - 2p_j)^2 (p_j - p_i)(p_j - p_k); \\ D = \sum_{i=1}^K \frac{N_i w_i}{v_i}$$

である．一方，簡単な計算から $\hat{p}_{NW}(x; h)$ と $\hat{p}(x)$ の平均，分散はそれぞれ

$$E[\hat{p}_{NW}(x; h)] = \frac{\sum_{i=1}^K N_i p_i w_i}{\sum_{i=1}^K N_i w_i}; \quad (6)$$

$$V[\hat{p}_{NW}(x; h)] = \frac{\sum_{i=1}^K N_i v_i w_i^2}{(\sum_{i=1}^K N_i w_i)^2}; \quad (7)$$

$$E[\hat{p}(x)] = (K - 1)f(x - \frac{i - 2}{K - 1})p_i + (\frac{i - 1}{K - 1} - x)p_{i-1}g; \quad x_{i-1} \hat{=} x \hat{=} x_i; \quad (8)$$

$$V[\hat{p}(x)] = (K - 1)^2 f(x - \frac{i - 2}{K - 1})^2 \frac{v_i}{N_i} + (\frac{i - 1}{K - 1} - x)^2 \frac{v_{i-1}}{N_{i-1}} g; \quad x_{i-1} \hat{=} x \hat{=} x_i \quad (9)$$

と求まる．

4 理論的比較

第3節で与えた3つの推定量 $\hat{p}_v(x; h)$, $\hat{p}_{NW}(x; h)$, $\hat{p}(x)$ の平均と分散の式(4)~(9)から，それらのMSEが導出できる：

$$MSE[\hat{p}_v(x; h)] = \frac{\sum_{i=1}^K N_i w_i (1 - p_i) - p(x)}{\sum_{i=1}^K N_i w_i = v_i} + \frac{2G}{N} \frac{\sum_{i=1}^K N_i w_i (1 - p_i) - p(x)}{\sum_{i=1}^K N_i w_i = v_i} \\ + \frac{\sum_{i=1}^K N_i w_i^2 = v_i}{(\sum_{i=1}^K N_i w_i = v_i)^2} + \frac{1}{N} H + o(\frac{1}{N}); \quad (10)$$

$$MSE[\hat{p}_{NW}(x; h)] = \frac{\sum_{i=1}^K N_i p_i w_i - p(x)}{\sum_{i=1}^K N_i w_i} + \frac{\sum_{i=1}^K N_i v_i w_i^2}{(\sum_{i=1}^K N_i w_i)^2}; \quad (11)$$

$$MSE[\hat{p}(x)] = (K - 1)f(x - \frac{i - 2}{K - 1})p_i + (\frac{i - 1}{K - 1} - x)p_{i-1}g - p(x) \\ + (K - 1)^2 f(x - \frac{i - 2}{K - 1})^2 \frac{v_i}{N_i} + (\frac{i - 1}{K - 1} - x)^2 \frac{v_{i-1}}{N_{i-1}} g; \quad (12) \\ x_{i-1} \hat{=} x \hat{=} x_i;$$

レベル数 $K = 4; 5$ と $N_i = 50; 70; 90$ に対して3つの推定量の $o(1/N)$ の項を無視したバイアスの2乗，分散，MSEの比較を行った．推定量のバンド幅の選択は，MPE(mean prediction error)の最小化で行った．ここで，推定量 $\hat{p}(x)$ のMPEは， $\sum_{i=1}^K MSE[\hat{p}(x_i)]$ で定義される．真の用量反応曲線として，線形変換したロジスティック分布

$$p(x) = \frac{1}{1 + \exp(4 - 8x)}$$

を使用した．カーネル w は標準正規分布の密度関数を用いた．レベル数 $K = 5$; $N_i = 50; 70; 90$ 推定量のバイアスの2乗，分散，MSEは図1, 2, 3で，推定量の最良MPEの値とそのときのバンド幅の値が表1で与えられる．

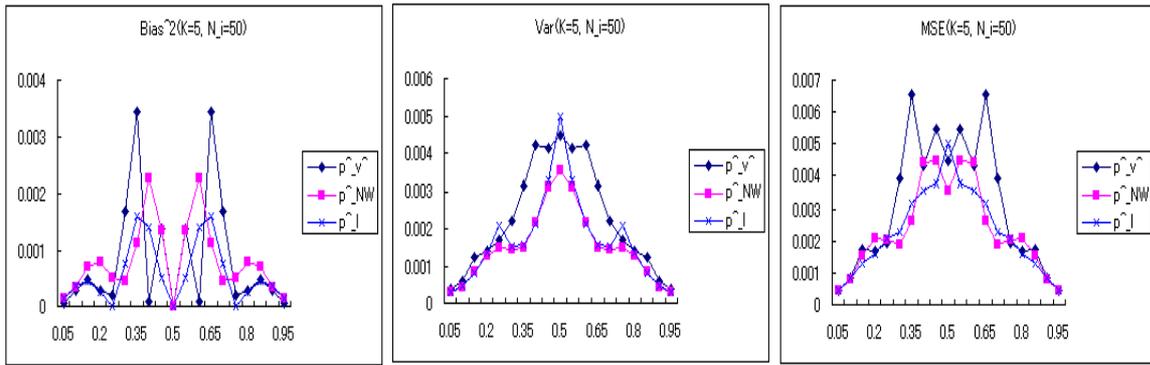


図 1 $N_i = 50$ のときの推定量のバイアスの 2 乗, 分散, MSE.

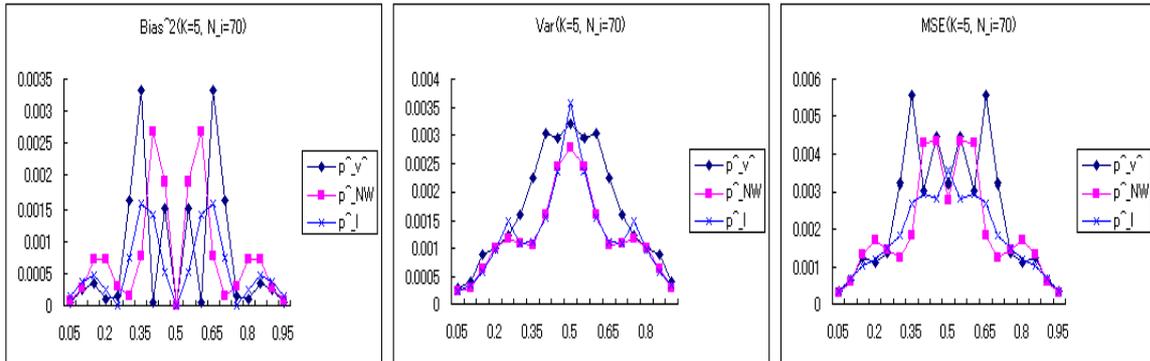


図 2 $N_i = 70$ のときの推定量のバイアスの 2 乗, 分散, MSE.

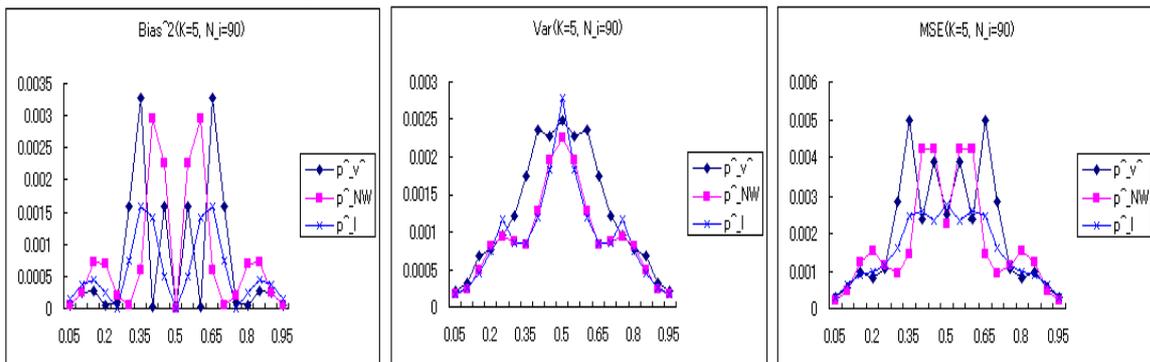


図 3 $N_i = 90$ のときの推定量のバイアスの 2 乗, 分散, MSE.

N_i	50	70	90
MPE [\hat{p}_v]	0.0091(0.085)	0.0065(0.085)	0.0050(0.085)
MPE [\hat{p}_{NW}]	0.0084(0.115)	0.0062(0.108)	0.0050(0.105)
MPE [\hat{p}_l]	0.0099	0.0071	0.0055

表 1 推定量の最良 MPE の値 (括弧内はバンド幅の値).

図 1, 2, 3 から, 2 つカーネル推定量は, バンド幅を各 N_i で選択しているのだから, それによる若干の値の変化がみられる. これらは用量の中央付近での挙動が激しい. しかし, 3 つの推定量の MSE の値は観測用量 $x = 0; 0.25; 0.5; 0.75; 1$ の付近では近くなる傾向がみてとれる. どの推定量も部分的に他の推定量より MSE が小さくなるが, 全体的に見たとき, 折れ線の挙動が安定している. また $K = 4$ の場合も 3 つの推定量は同様な挙動を示した.

5 数値的比較

第4節の比較と同じ実験をより小さい $N_i = 5; 10; 15$; $K = 4; 5$ に対して行った。ただし、 $\hat{\rho}_\psi(x; h)$ の MSE の計算は、

$$\sum_{y_1=0}^{K-1} \cdots \sum_{y_{K-1}=0}^{K-1} (\hat{\rho}_\psi(x; h) - \rho(x))^2 P(Y_j = y_j | x_j)$$

に基づき正確に算出した。 $N_i = 5; 10; 15$; $K = 5$ ときの推定量のバイアスの2乗, 分散, MSE は図4, 5, 6で与えられ, 推定量の最良 MPE の値とそのときのバンド幅の値が表2で与えられる。

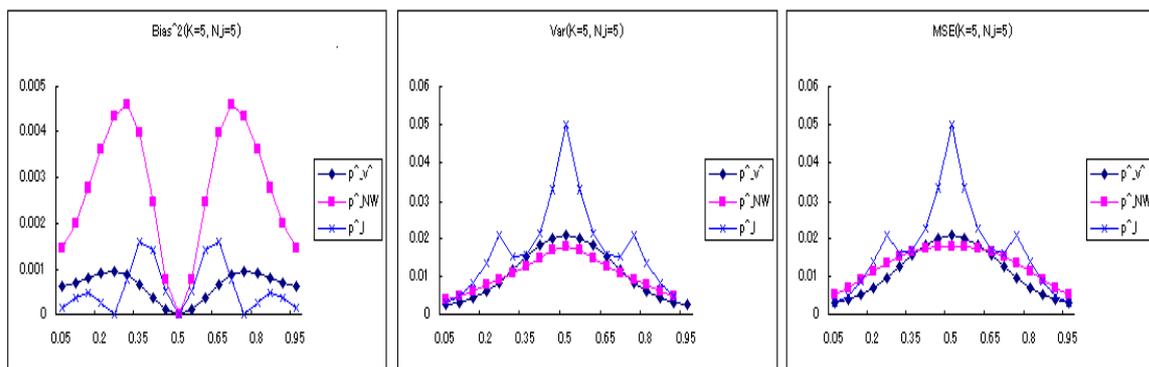


図4 $N_i = 5$ のときの推定量のバイアスの2乗, 分散, MSE.

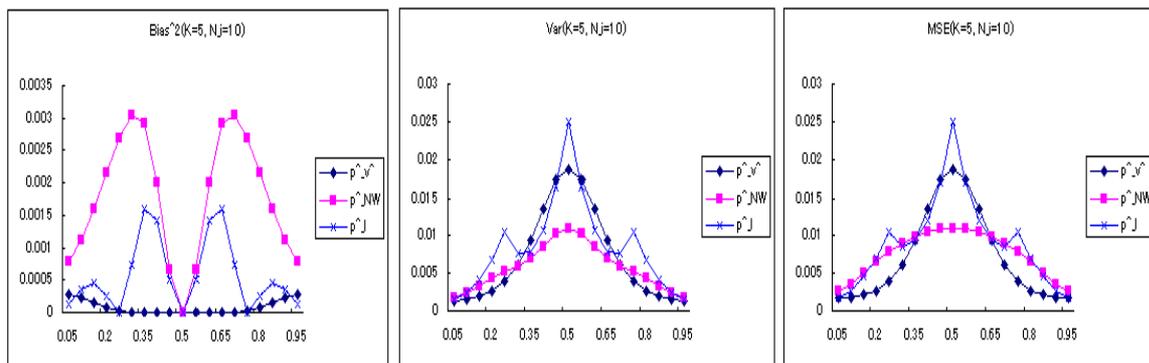


図5 $N_i = 10$ のときの推定量のバイアスの2乗, 分散, MSE.

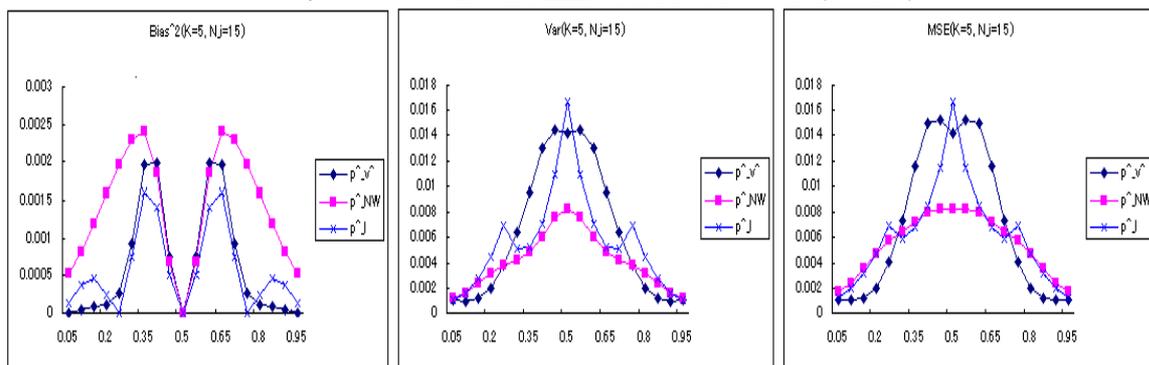


図6 $N_i = 15$ のときの推定量のバイアスの2乗, 分散, MSE.

N_i	5	10	15
MPE[$\hat{\rho}_\psi$]	0.0458(0.199)	0.0299(0.205)	0.0244(0.114)
MPE[$\hat{\rho}_{NW}$]	0.0538(0.176)	0.0310(0.155)	0.0225(0.144)
MPE[$\hat{\rho}_j$]	0.0991	0.0495	0.0330

表2 推定量の最良 MPE の値 (括弧内はバンド幅の値).

図 4 から, $N_i = 5$ のときカーネル推定量, 特に提案した推定量の MSE の値が全体的に小さいことがわかる. 図 5, 6 から $N_i = 10; 15$ とき, 用量の中央部では Nadaraya-Watson 推定量の MSE が小さく, 端の部分では, 提案した推定量の MSE が小さい. $K = 4$ の場合も同様な結果が得られた.

6 考察

第 4, 5 節では, 用量レベル数 K を固定したときの個体数 N_i の変化に対する 3 つ推定量の挙動を, 真の用量曲線を与えて, それと比較することで調べた. 第 4 節の比較的大きい $N_i = 50; 70; 90$ に対する比較では 2 つのカーネル推定量は $\hat{p}_v(x; h)$, $\hat{p}_{NW}(x; h)$ は観測用量 x_i 付近では, 折れ線に近い値をとるが, とくに中央部で激しい挙動を示めた. これは, バンド幅の選択を観測用量 x_i での誤差 (の 2 乗) を最小にする MPE で行ったためと考えられる. 各推定量は, 部分的に MSE が小さくなるが全体的に見ると折れ線が良い挙動を示した. 第 5 節の小さい大きい $N_i = 5; 10; 15$ に対する比較では, 提案した推定量が他の推定量と比べて, $N_i = 5$ では全体的に MSE が小さく, $N_i = 10; 15$ では特に用量の端の部分, すなわち反応確率の小さい所と大きい所で MSE が良い挙動を示した. Nadaraya-Watson 推定量と比べて, v_i で重み w_i を割ることによって, 用量の端の部分で, 推定の精度が上がるということがわかった.

7 おわりに

用量レベル数 K を固定し, 個体数 N_i を変化させたときの 3 つ推定量の挙動について調べた. 我々の提案した推定量は, 個体数が比較的小さく, 用量の端の部分, すなわち反応確率が小さいまたは大きい所で良い挙動を示した. 実際, \hat{a} が小さい ED100 \hat{a} %, LD100 \hat{a} % の計算が行われる. 我々の提案した推定量は, このような場面に適用できると思われる. 提案した推定量を実際に応用する場合, バンド幅 h の選択や推定量の推定区間の導出も重要である. これらに関しては本稿では, 扱わなかったが議論が必要である. また, v_i の推定量として何を選択するのが良いかという問題もある. K を固定し, 各用量 x_i で観測個体数 N_i を十分とっても, 3 つの推定量は真の曲線に, 式 (10),(11),(12) から収束しない. 従って, $N \rightarrow \infty; K \rightarrow \infty; NK \rightarrow \infty; h \rightarrow 0$ の状況での推定量の挙動についても, 調べておく必要がある. これについては当日報告する予定である.

参考文献

- Berkson, J.(1955). Maximum likelihood and minimum χ^2 estimates of the logistic function. J. Amer. Statist. Assoc. 50, 130-162.
- Copas, J. B.(1983). Plotting p against x . Applied Statistics 32, 25-31.
- Fan, J. Q., Heckman, N. E. and Wand, M. P.(1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. J. Amer. Statist. Assoc. 90, 141-150.
- Green, P. J. and Silverman, B. W.(1994). Nonparametric Regression and Generalized Linear Models. London: Chapman and Hall.
- Härdle, W.(1990). Applied Nonparametric Regression. Cambridge University Press.
- Lloyd, C. J.(1999). Statistical Analysis of Categorical Data. Wiley, New York.
- Wand, M. P. and Jones, M. C.(1995). Kernel Smoothing. London: Chapman and Hall.