

# ノンパラメトリック推定とその応用\*

---

講師: 長倉大輔

\* この資料は私のゼミにおいて使用するために作成した資料です。ゼミのWEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし間違いがあった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。間違いは発見次第、継続的に直していますが、まだ存在する可能性があります。間違いを見つけた場合 [nagakura@z7.keio.jp](mailto:nagakura@z7.keio.jp) までご連絡いただくと大変嬉しく思います。

# 今日の予定

---

1. ノンパラメトリック密度推定、カーネル法
2. ノンパラメトリック回帰分析、実証例
3. セミパラメトリック部分線形モデル、実証例
4. K nearest neighbor 法
5. まとめ

# ノンパラメトリック推定とその応用

---

## ■ パラメトリック、ノンパラメトリック推定

**パラメトリック推定**とは分布の仮定(最尤法)、関数形の仮定(最小二乗法)のようにモデルに強い仮定を置く必要がある。これらの仮定が実際のデータで満たされていない場合は、分析結果に重要な影響を及ぼす可能性がある。

**ノンパラメトリック推定**とは、モデルにこれらの仮定を置かないで推定する方法である。ただし、ノンパラメトリック推定はモデルの仮定が緩い分、後程見る「次元の呪い」と呼ばれる問題により、変数が増えるにつれて推定精度が低くなるという問題がある。

どちらも一長一短がある。

# ノンパラメトリック推定とその応用

---

## ■ ノンパラメトリック密度推定

$X$  は連続型確率変数とし、分布関数  $F(x)$ 、密度関数  $f(x)$  を持つとする。**ノンパラメトリック密度推定**とは、分布を仮定せずに、確率変数  $X$  の密度関数を推定することを目的とする。

この推定はこの後の**ノンパラメトリック回帰分析**の基礎となるものなので少し丁寧に説明する。

ここではノンパラメトリック密度推定として代表的な**カーネル推定法**について説明する。ヒストグラムは、最も代表的でよく知られているカーネル推定法である。

# ノンパラメトリック推定とその応用

## ■ 経験分布関数 (Empirical Distribution Function)

分布関数  $F(x)$  のノンパラメトリックな推定法として最もよく知られているのが**経験分布関数**である。

今、確率変数  $X$  のランダムサンプル  $X_1, \dots, X_n$  が観測されたとする。この時、経験分布関数は

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n 1(X_i \leq x)$$

と定義される。ここで  $1(A)$  は**指示関数**と呼ばれるもので、括弧の中の事象が起これば 1、それ以外の場合は 0 を取る関数である。 $1(X_i \leq x)$  の場合は  $X_i \leq x$  が成り立てば、 $1(X_i \leq x) = 1$  であり、成り立たない場合は 0 になる。

## ノンパラメトリック推定とその応用

---

ある数値  $x$  に対して  $\sum_{i=1}^n 1(X_i \leq x)$  は  $\{X_1, \dots, X_n\}$  のうち  $X_i \leq x$  となるデータの数を表しているので、経験分布関数は、分布関数  $F(x) = \Pr(X \leq x)$  を全てのデータ  $\{X_1, \dots, X_n\}$  のうち  $X_i \leq x$  となる割合で推定したもの、すなわち

$$\hat{F}(x) = \frac{\{X_1, \dots, X_n\} \text{のうち } X_i \leq x \text{ である } X_i \text{ の数}}{n}$$

で推定している。ある一点の値  $x$  における分布関数の値を推定していることに注意。

# ノンパラメトリック推定とその応用

---

例えば

$$\{X_1, \dots, X_{10}\} = \{3.5, -1.2, 5, 2, 6.2, 2.8, 5.8, 1, -2, -3.2\}$$

という 10 個の観測値の  $X$  の経験分布関数の値は

(1)  $x = 5$  という値に対しては

$$\hat{F}(5) = \frac{5 \text{ 以下の } X_i \text{ の数}}{10} = \frac{7}{10} = 0.7$$

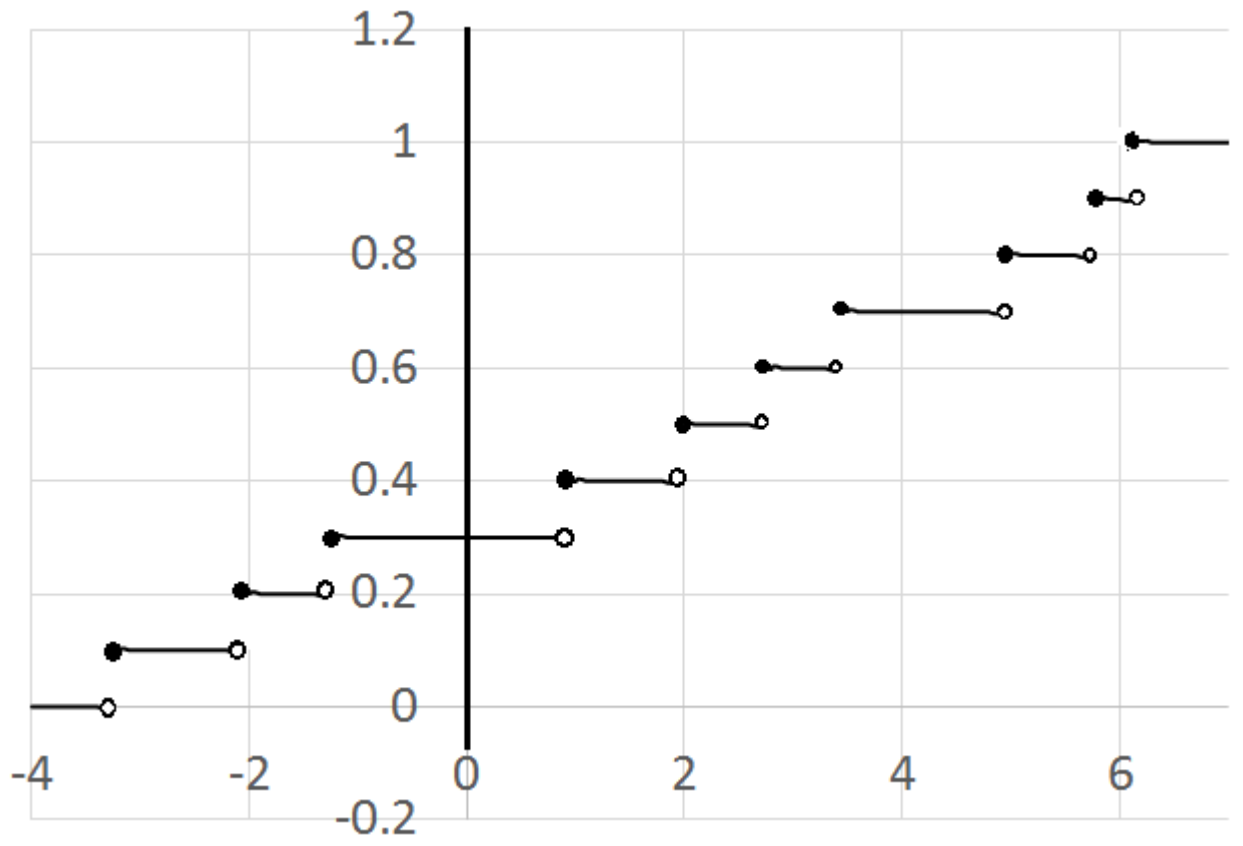
(2)  $x = 4$  という値に対しては

$$\hat{F}(4) = \frac{4 \text{ 以下の } X_i \text{ の数}}{10} = \frac{7}{10} = 0.7$$

となる。この10個の観測値に対する経験分布関数をグラフにすると以下のようなになる

# ノンパラメトリック推定とその応用

(経験分布関数のグラフ:  $Y$  軸が  $\hat{F}(x)$ 、 $X$  軸が  $x$ )





## ノンパラメトリック推定とその応用

---

この経験分布関数に対して、数値微分でよく用いられる**中心差分式**によって  $x$  における微分を推定することを考える。

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

ここで  $h > 0$  は定数(通常小さい値を取ると想定される)。

# ノンパラメトリック推定とその応用

---

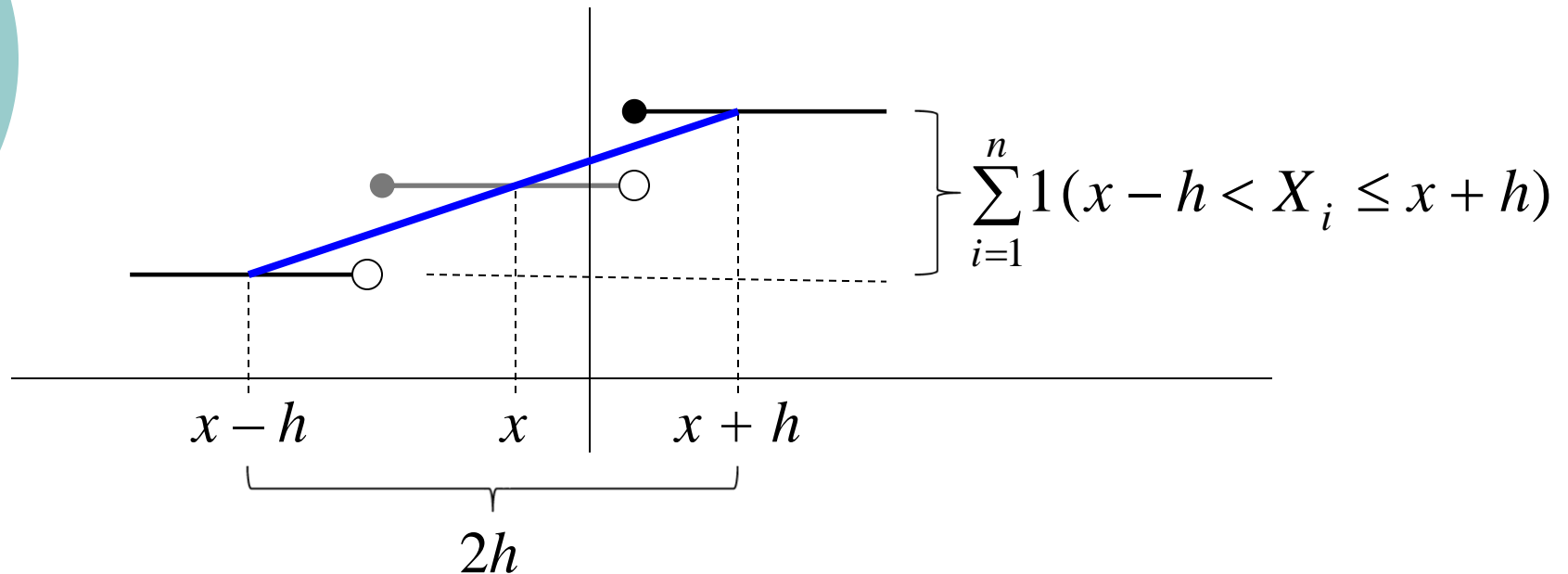
経験分布の定義式を代入すると

$$\begin{aligned}\hat{f}(x) &= \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h} \\ &= \frac{1}{2nh} \left( \sum_{i=1}^n 1(X_i \leq x+h) - \sum_{i=1}^n 1(X_i \leq x-h) \right) \\ &= \frac{1}{2nh} \sum_{i=1}^n 1(x-h < X_i \leq x+h) \\ &= \frac{1}{2nh} \sum_{i=1}^n 1\left(\frac{|X_i - x|}{h} \leq 1\right)\end{aligned}$$

となる。第3式よりこれは  $x$  を中心とした幅  $2h$  の区間に観測されたデータの割合を  $2h$  で割ったものである。

# ノンパラメトリック推定とその応用

(密度関数の推定のイメージ)



ここで  $\hat{f}(x)$  は上記の太い青線の傾きを  $n$  で割ったものである。上記の場合 2 つの観測値が区間  $(x-h, x+h]$  にあるので、 $\hat{f}(x) = (2/n) / 2h$  となる。  $x$  の位置、  $h$  の大きさに依存する。

# ノンパラメトリック推定とその応用

経験分布関数は観測数が増えると真の分布関数に収束する。よってその中心差分はその微係数、つまり密度関数の値を推定できそうだと直観的には思われる。

(バンド幅についてのバイアスと分散のトレードオフ)

$\hat{f}(x)$  は  $x$  を中心とした幅  $2h$  に含まれる観測値の割合を  $2h$  で割ったものである。幅を大きくとれば、より多くの観測値が含まれ(ただし分母の  $2h$  も大きくなる)、標本が変動した時の観測値が含まれる割合の変動が小さくなるため標本変動による  $\hat{f}(x)$  の変動は小さくなる。ただし幅を大きくすると数値微分としての精度は落ちるため、真の値とのかい離が大きくなる。**直観的には  $h$  が大きいと分散が小さくなるがバイアスが大きくなるといえよう**

# ノンパラメトリック推定とその応用

---

先ほどの式

$$\hat{f}(x) = \frac{1}{2nh} \sum_{i=1}^n 1 \left( \frac{|X_i - x|}{h} \leq 1 \right) = \frac{1}{nh} \sum_{i=1}^n \frac{1(|X_i - x|/h \leq 1)}{2}$$

は

$$k(u) = \begin{cases} \frac{1}{2}, & |u| \leq 1 \\ 0, & |u| > 1 \end{cases}$$

を用いると

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n k \left( \frac{X_i - x}{h} \right)$$

書き表すことができる。ここで  $k(u)$  は  $[-1, 1]$  上の一様分布の密度関数であることに注意。

# ノンパラメトリック推定とその応用

---

これは **カーネル推定量** の一つである。カーネル推定量は一般的に

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{X_i - x}{h}\right)$$

と書かれ、 $k(u)$  の形に応じて様々なカーネル推定量が考えられる。この時、関数  $k(u)$  は **カーネル関数** (kernel function) と呼ばれる。カーネル推定量において  $h$  は **バンド幅** (bandwidth) と呼ばれる。

先ほどの一様分布の密度関数は **一様カーネル** (uniform kernel) と呼ばれる。

# ノンパラメトリック推定とその応用

---

## ■ カーネル関数

カーネル関数は通常以下の性質を満たすと仮定される。

- (1)  $\int_{-\infty}^{\infty} k(u) du = 1$
- (2)  $k(u) \geq 0$  for all  $u$ ,

これらが満たされる場合は  $k(u)$  は**確率密度関数**である。  
さらに、ほとんどの場合は以下も仮定される。

- (3)  $k(u) = k(-u)$  (対称性)

以下ではこの3つの性質は常に満たされているとする。

# ノンパラメトリック推定とその応用

---

## ■ カーネル関数の例

上記の仮定を満たすカーネルでよくつかわれるものは

Uniform :  $k(u) = \frac{1}{2} 1(|u| \leq 1)$  ,

Epanechnikov :  $k(u) = \frac{3}{4} (1 - u^2) 1(|u| \leq 1)$  ,  $C(k) = 2.34$

Biweight :  $k(u) = \frac{15}{16} (1 - u^2)^2 1(|u| \leq 1)$  ,  $C(k) = 2.78$

Triweight :  $k(u) = \frac{35}{32} (1 - u^2)^3 1(|u| \leq 1)$  ,  $C(k) = 3.15$

Gaussian :  $k(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})$  ,  $C(k) = 1.06$

等があげられる。Epanechnikov と Gaussian カーネルが最もよく使われる。 $C(k)$ は後述する「経験則バンド幅」を計算するとき用いられる定数である。



# ノンパラメトリック推定とその応用

---

## ■ カーネル推定量の性質

条件 (1), (2), (3) を満たすカーネル推定量は以下の性質を持つ

$$(A): \hat{f}(x) \geq 0 \quad (B): \int_{-\infty}^{\infty} \hat{f}(x) dx = 1$$

$$(C): \hat{E}(x) = \int_{-\infty}^{\infty} x \hat{f}(x) dx = \bar{X} = n^{-1} \sum_{i=1}^n X_i$$

$$(D): \int_{-\infty}^{\infty} (x - \hat{E}(x))^2 \hat{f}(x) dx = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2 + h^2 v_2(k)$$

ここで  $v_2(k) = \int_{-\infty}^{\infty} x^2 k(x) dx$ 、つまりカーネルを密度関数とした**2次モーメント**である。

# ノンパラメトリック推定とその応用

---

性質 (A) と (B) は密度関数の推定量として望ましい性質である。

性質(C)はカーネル推定量によって推定した密度関数の期待値と標本平均が等しくなることを示しており、自然な性質である。

性質 (D)はカーネル推定量によって推定した密度関数の分散は標本分散よりも  $h^2v_2(k)$  だけ**大きくなる**ことを示している。

# ノンパラメトリック推定とその応用

## ■ カーネル推定量のバイアスと分散とMSE

性質 (1), (2), (3) を満たすカーネル関数による推定量は以下のバイアスと MSE をもつ。

$$Bias(\hat{f}(x)) = E[\hat{f}(x)] - f(x) \approx \frac{1}{2} f^{(2)}(x) h^2 v_2(k)$$

$$var(\hat{f}(x)) \approx \frac{f(x)R(k)}{nh}$$

$$\begin{aligned} MSE(\hat{f}(x)) &= E\{[\hat{f}(x) - f(x)]^2\} = [Bias(\hat{f}(x))]^2 + var(\hat{f}(x)) \\ &\approx \frac{v_2^2(k)}{4} [f^{(2)}(x)]^2 h^4 + \frac{f(x)R(k)}{nh} \end{aligned}$$

ここで  $f^{(2)}(x)$  は  $f(x)$  の  $x$  における2階微分、および  $R(k) = \int_{-\infty}^{\infty} k(u)^2 du$  である。

# ノンパラメトリック推定とその応用

---

$$AMSE(\hat{f}(x)) = \frac{v_2^2(k)}{4} [f^{(2)}(x)]^2 h^4 + \frac{f(x)R(k)}{nh}$$

を漸近MSE (Asymptotic MSE) という。これは各点  $x$  における  $n$  が十分に大きいときの  $\hat{f}(x)$  の MSE である。

これらの結果より、**バンド幅  $h$  が大きくなるとバイアスも大きくなり、 $nh$  が大きくなると分散は小さくなる**ことがわかる。MSEが  $n \rightarrow \infty$  につれて 0 に収束するには  $h \rightarrow 0$  および  $nh \rightarrow \infty$  となる必要がある。

カーネル関数の形状の影響は  $v_2(k)$  と  $R(k)$  を通してバイアス、分散、MSEに影響する。

# ノンパラメトリック推定とその応用

---

- 漸近的に最適なバンド幅

先ほどの AMSE を  $x$  について積分した

$$\begin{aligned} AMISE &= \int_{-\infty}^{\infty} AMSE(\hat{f}(x)) dx \\ &= \frac{v_2^2(k)}{4} h^4 \int_{-\infty}^{\infty} [f^{(2)}(x)] dx + \frac{R(k)}{nh} \end{aligned}$$

を Asymptotic Mean Integrated Squared Error という。これを最小にするバンド幅  $h$  は1階の条件を解くことによつて得られ、

$$h_0 = C(k, f) n^{-1/5}$$

となる。ここで  $C(k, f)$  はカーネル関数と  $f(x)$  に依存し決まる定数である。

# ノンパラメトリック推定とその応用

---

## ■ 漸近的に最適なカーネル

先ほどの  $h_0$  をAMISEに代入して漸近的に最小化されたAMISEを求めると

$$AMISE_0 = \frac{5}{4} [v_2^2(k)R(k)^4 \int_{-\infty}^{\infty} f^{(2)}(x)dx]^{1/5} n^{-4/5}$$

が得られる。これはカーネルに依存している。この  $AMISE_0$  を最小にするのは **Epanechnikov カーネル** であることを示すことができる。よって Epanechnikov カーネルがよく用いられる。

# ノンパラメトリック推定とその応用

---

## ■ 経験則バンド幅 (Rule-of-Thumb bandwidth)

漸近的に最適なバンド幅は

$$h_0 = C(k, f) n^{-1/5}$$

であった。しかしながら真の密度関数はわからないため ( $k$  は選択できるが)これは**実際には計算できない**。

よく用いられる方法は  $f(x)$  としてとりあえず**正規分布**を仮定する方法である。この時  $C(k, f)$  は

$$C(k, f) = \hat{\sigma} C(k) n^{-1/5}$$

とカーネルだけに依存した値になる。ここで  $C(k)$  はカーネル  $k$  だけに依存した定数、 $\hat{\sigma}$  は標本の標準偏差の推定値である(通常標本標準偏差を用いる)。

# ノンパラメトリック推定とその応用

---

このように計算したバンド幅を**経験則(親指ルール)バンド幅**という。

経験則バンド幅は真の密度関数が正規分布に近ければ最適なバンド幅のよい近似と考えられるが、もちろんそうかどうかはわからない。そういう意味で経験則バンド幅は「とりあえずの」バンド幅であるが、適当に決めるよりもまだましであろう。



# ノンパラメトリック推定とその応用

---

- クロスバリデーション (Cross-Validation)

経験則バンド幅よりもよりよいバンド幅の選択方法として**クロスバリデーション**という方法がある。

これは Mean Integrated Squared Error を推定し、その推定値はバンド幅に依存するので、その値を最小にするバンド幅を選択する方法である。

ここでは詳細の説明は省略する(詳細は Li and Racine (2006)を参照のこと)。

# ノンパラメトリック推定とその応用

---

## ■ カーネル密度推定量の漸近分布

カーネル密度推定量は漸近的に正規分布に従う、特にバンド幅を漸近的に最適なオーダーになるように選択した場合、つまり

$$h = Cn^{-1/5}$$

を満たすように  $h$  を選んだとき ( $C$  は定数)

$$\sqrt{nh}(\hat{f}(x) - f(x)) \xrightarrow{d} N\left(C^{5/2} \frac{v_2(k)}{2} f^{(2)}(x), f(x)R(k)\right)$$

となる。

# ノンパラメトリック推定とその応用

## ■ 多変量密度推定

$q$ 次元確率ベクトル  $\mathbf{X}_i = [X_{i1}, \dots, X_{iq}]'$  の密度関数  $f(\mathbf{x}) = f(x_1, \dots, x_q)$  を推定する問題を考える。

多変量カーネル推定量は次のように定義される。

$$\hat{f}(\mathbf{x}) = \frac{1}{n |\mathbf{H}|} \sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x}))$$

ここで  $\mathbf{H} = \text{diag}(h_1, \dots, h_q)$  (これは  $h_1, \dots, h_q$  を対角成分とする対角行列),  $h_1, \dots, h_q$  はバンド幅である。また関数  $K: R^q \rightarrow R$  は**多変量カーネル**と呼ばれる。 $|\mathbf{H}|$  は  $\mathbf{H}$  の行列式であり  $|\mathbf{H}| = h_1 h_2 \dots h_q$  である。

# ノンパラメトリック推定とその応用

---

$K(\cdot)$  は  $K(\mathbf{u}) \geq 1$ , および

$$\int_{R^q} K(\mathbf{u}) d\mathbf{u} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} K(u_1, \dots, u_q) du_1 \dots du_q = 1$$

を満たす。すなわちある多変量密度関数である。また $K(\cdot)$ として

$$K(\mathbf{u}) = k(u_1)k(u_2) \dots k(u_q)$$

がよく用いられる。ここで  $k(\cdot)$  は一変量カーネルである。このような多変量カーネルは**積型カーネル**と呼ばれる。

積型多変量カーネル推定量のバイアス、分散、MSEは

# ノンパラメトリック推定とその応用

$$\text{Bias}(\hat{f}(\mathbf{x})) \approx \frac{v_2(k)}{2} \sum_{j=1}^q \frac{\partial^2 f(\mathbf{x})}{(\partial x_j)^2} h_j^2, \quad \text{var}(\hat{f}(\mathbf{x})) \approx \frac{f(\mathbf{x})[R(k)]^q}{n |\mathbf{H}|}$$

$$\text{AMISE}(\hat{f}(\mathbf{x})) = \frac{v_2^2(k)}{2} \int_{R^q} \left[ \sum_{j=1}^q \frac{\partial^2 f(\mathbf{x})}{(\partial x_j)^2} h_j^2 \right]^2 d\mathbf{x} + \frac{[R(k)]^q}{n |\mathbf{H}|}$$

となる。  $h_1 = h_2 = \dots = h_q$  を仮定すると漸近的に最適なバンド幅  $h_0$  は

$$h_0 = \left[ \frac{4q[R(k)]^q}{4v_2^2(k) \int_{R^q} \left[ \sum_{j=1}^q \frac{\partial^2 f(x)}{(\partial x_j)^2} \right]^2 dx} \right]^{1/(4+q)} n^{-\frac{1}{2v+q}}$$

で与えられる。

## ノンパラメトリック推定とその応用

---

さらにこの最適バンド幅を  $AMISE$  に代入すると漸近的に最小な値、 $AMISE_0$  が求まる。

このようにして解かれた、 $AMISE_0$  には  $n^{-4/(4+q)}$  がかかっており、 $n$  が大きくなるにつれてこの項が 0 に近づいていく。 $q = 1$  の時には  $n^{-4/5}$  の速さで小さくなっていくのに比べて  $q$  が大きくなるにつれ、 $n^{-2/3}$  ( $q = 2$ の時)、 $n^{-4/7}$  ( $q = 3$ の時)と 0 に近づく速さが遅くなっていく。

このように  $q$  が大きくなると推定量の収束の速さが遅くなることは「**次元の呪い**」(**curse of dimensionality**) と呼ばれる。次元の呪いの影響を弱めるため 後述する Semiparametric Model などが用いられる。

# ノンパラメトリック推定とその応用

## ■ ノンパラメトリック回帰 – Nadaraya-Watson 回帰

$\{y_i, \mathbf{X}_i\}, i = 1, \dots, n$  を興味のある変数からの標本とする。  
ここで  $y_i$  はスカラー変数  $\mathbf{X}_i$  は  $q \times 1$  ベクトル変数とする。

ノンパラメトリック回帰とは

$$g(\mathbf{x}) = E(y_i | \mathbf{X}_i = \mathbf{x}) = \frac{\int_{-\infty}^{\infty} y f(y, \mathbf{x}) dy}{f(\mathbf{x})}$$

を推定することである。ここで  $f(y, \mathbf{x})$  は  $[y_i, \mathbf{X}_i']'$  の結合密度関数、 $f(\mathbf{x})$  は  $\mathbf{X}_i$  の周辺密度関数である。

# ノンパラメトリック推定とその応用

$g(x)$  の自然な推定量として、 $f(y, x)$  や  $f(x)$  をそのカーネル推定量で置き換えたものが考えられよう。

$f(y, x)$  と  $f(x)$  のカーネル推定量として

$$\hat{f}(y, \mathbf{x}) = \frac{1}{n |\mathbf{H}| h_y} \sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})) k\left(\frac{y_i - y}{h_y}\right),$$

および 
$$\hat{f}(\mathbf{x}) = \frac{1}{n |\mathbf{H}|} \sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})),$$

が考えられよう。またこれより

$$\int_{-\infty}^{\infty} y \hat{f}(y, \mathbf{x}) dy = \frac{1}{n |\mathbf{H}| h_y} \sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})) \int_{-\infty}^{\infty} y k\left(\frac{y_i - y}{h_y}\right) dy$$

$$= \frac{1}{n |\mathbf{H}|} \sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})) y_i$$

を得る。



# ノンパラメトリック推定とその応用

---

先ほどのページの結果は

$$y_i = \int_{-\infty}^{\infty} y \frac{1}{h_y} k\left(\frac{y_i - y}{h_y}\right) dy$$

を使っている(要確認)

よって、 $g(\mathbf{x})$  の推定量としてこの比をとった

$$\hat{g}(\mathbf{x}) = \frac{\int_{-\infty}^{\infty} y f(y, \mathbf{x}) dy}{f(\mathbf{x})} = \frac{\sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})) y_i}{\sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x}))}$$

を得る。これは **Nadaraya-Watson (NW) 推定量** と呼ばれる。

# ノンパラメトリック推定とその応用

## ■ 局所線形推定量

NW 推定量は分布の裾の方の推定値が**上方に大きなバイアスを持ってしまうという欠点**がある。また他にも、真の  $y_i$  と  $X_i$  の関係が確率的ではなく、正確に線形関係であったとしても、NW推定量の値はこの直線と一致しない。これらの欠点を補ったものとして**局所線形(Local Linear; LL)推定量**があり、以下のように定義される。

$$\begin{bmatrix} \hat{g}(\mathbf{x}) \\ \hat{g}^{(1)}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})) \mathbf{Z}_i \mathbf{Z}_i' \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})) \mathbf{Z}_i y_i \end{bmatrix}$$

ここで  $g^{(1)}(\mathbf{x})$  は  $\partial g(\mathbf{x}) / \partial \mathbf{x}$  を表し、 $\hat{g}^{(1)}(\mathbf{x})$  はその推定量、 $\mathbf{Z}_i$  は  $\mathbf{Z}_i = [1, (\mathbf{X}_i - \mathbf{x})']'$  である。**LL推定量は自動的に  $g^{(1)}(\mathbf{x})$  も推定する。**

# ノンパラメトリック推定とその応用

---

- LL 推定量は NW 推定量と比べて分布の端の周りでの性質がかなり改善されている。
- バイアスについてはどちらがよいかは状況によるが、おおむねLL推定量の方がよい性質を持っている。
- 両者の漸近分散は同じである。
- LL推定量は総合的にNW推定量より良い性質を持っているが、 $X_i$ についての局所的な傾きが小さい場合NW推定量の方が優れていることがあり、経済変数の場合はこれにあてはまることもよくあるので注意が必要である。

# ノンパラメトリック推定とその応用

---

- Bachmeier, Leelahanon, and Li (2007) の紹介

ここでは NW 推定量を用いたインフレ予測を扱った論文  
Bachmeier, L., Leelahann, S., and Li, Q. (2007)  
Macroeconomic Dynamic を紹介する。  
以下ではBLLと略す。

BLLは、短期のインフレ予測に線形VARモデルを使用した場合、money growth をモデルに加えても予測力があがらないという、従来の結果を nonparametric 回帰モデルを用いて検証した。

# ノンパラメトリック推定とその応用

---

BLL の主要な結論として、

- (i) 非線形モデルを用いると、線形モデルに比べて、予測力が大幅にアップする。
- (ii) 非線形モデルにおいてはMoney Growth をモデルに加えると予測力はわずかに向上する。Money Growth の代わりに Velocity を加えると予測力が大幅にアップする

が挙げられる。以下により詳しく説明する。

# ノンパラメトリック推定とその応用

---

BLL は St. Louis Federal Reserve bank の M1, M2, と M3, および対応する M1, M2, と M3 Divisia monetary service index および消費者物価指数の 1959.1 から 2002.5 までの月次データを用いた。

また Velocity として  $V = PQ/M$  も用いた。ここで PQ は industrial production、M は上の6つの monetary aggregates のうちの1つ(よってVは6つできることになる)。

以下では  $\pi_t$  :  $t$  時点のインフレ率、  
 $\Delta m_t$  :  $t$  時点の money growth

という表記を用いる。

## ノンパラメトリック推定とその応用

---

BLLはサンプル外予測によって予測力の比較を行った。具体的には1994年の1月から予測をはじめ2002年の4月までの100時点の予測を各モデルを用いて行い、それらの予測の mean squared prediction error (MSPE)

$$\frac{1}{100} \sum_{t=1994:1}^{2002:4} (\hat{\pi}_t - \pi_t)^2$$

を比較している。ここで  $\hat{\pi}_t$  は  $\pi_t$  の予測値である。

予測に用いるモデルは、予測を更新するたびごとに予測時点の1時点前までの全てのデータを用いて推定直し、予測を行っている。

# ノンパラメトリック推定とその応用

---

BLL はまず以下の線形のARおよびVARモデルを用いて  $s$  時点先予測 ( $s = 1, 6, 12, 24$ ) を行い、既存の結果を確認した。

線形VARモデル:

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + \alpha_3 \Delta m_{t-s} + \alpha_4 \Delta m_{t-s-1} + \varepsilon_t$$

線形ARモデル:

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + \varepsilon_t$$

VARの次数は BIC で 2 と選択し、それを AR にも適用  
次ページの表 (BLL の Table 1) はこれらのMSPE の比  
である。



# ノンパラメトリック推定とその応用

TABLE 1. Relative MSPE (Linear VAR/Linear AR)

Horizon	M1	M2	M3	M1D	M2D	M3D
1 Month	1.04 (0.44)	0.99 (0.04)	1.04 (0.02)	1.00 (0.91)	1.00 (0.25)	1.00 (0.05)
6 Months	1.09 (0.55)	1.05 (0.00)	1.13 (0.00)	0.98 (0.95)	1.01 (0.07)	1.05 (0.01)
12 Months	0.90 (0.54)	0.93 (0.00)	1.04 (0.00)	0.98 (0.92)	0.97 (0.08)	1.01 (0.02)
24 Months	1.05 (0.13)	0.76 (0.00)	1.07 (0.00)	0.99 (0.34)	0.87 (0.02)	0.94 (0.01)

表中の値は 線形VARモデルのMSPEを線形ARモデルのMSPE で割ったものである。数値が1より小さいと線形VARモデルのMSPEが小さいことになる。従来知られているように money growth をモデルに加えても予測力はほとんど改善していないことが確認できる

## ノンパラメトリック推定とその応用

次に線形性の仮定を緩め非線形回帰モデルを用いて予測を行うことを考える。ここでも  $\Delta m_t$  を含める場合と含めない場合を考える。

ノンパラメトリックVAR:  $\pi_t = f(\pi_{t-s}, \pi_{t-s-1}, \Delta m_{t-s}) + \varepsilon_t$

ノンパラメトリックAR:  $\pi_t = f(\pi_{t-s}, \pi_{t-s-1}) + \varepsilon_t$

これらのモデルに対して  $T + s$  時点の予測値を以下のように計算する(NW 推定量)

$$E(\pi_{T+s} | \mathbf{X}_T) = \frac{\sum_{i=1}^{T-s} K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{X}_T)) \pi_{t+s}}{\sum_{i=1}^{T-s} K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{X}_T))}$$

ここで  $K(\cdot)$  は積型ノーマルカーネル、 $\mathbf{H} = \text{diag}(h, h)$ 。  
 $\mathbf{X}_t$  は  $\mathbf{X}_t = [\pi_t, \pi_{t-1}, \Delta m_t]'$  もしくは  $\mathbf{X}_t = [\pi_t, \pi_{t-1}]'$  である。  
BLLは次ページの表の結果を得た(BLLのTable 2)。

# ノンパラメトリック推定とその応用

TABLE 2. Relative MSPE (Nonparametric VAR/Linear VAR)

Horizon	AR	M1	M2	M3	M1D	M2D	M3D
1 Month	0.88 (1.75)	0.94 (0.67)	0.84 (2.38)	0.84 (2.18)	0.88 (1.69)	0.87 (1.83)	0.86 (1.89)
6 Months	0.93 (2.12)	0.86 (1.15)	0.86 (2.43)	0.84 (2.22)	0.90 (2.08)	0.91 (1.92)	0.88 (2.22)
12 Months	0.93 (2.21)	0.99 (0.32)	0.91 (2.26)	0.88 (2.40)	0.94 (1.66)	0.95 (1.23)	0.91 (2.12)
24 Months	1.09 (-5.84)	1.00 (0.00)	0.94 (0.80)	0.91 (1.08)	1.07 (-3.91)	1.10 (-1.88)	1.05 (-0.83)

表中の値の定義は先ほどとほぼ同じである。これを見ると、非線形性を導入すると短期においても予測力が改善していることがわかる。次に money growth が予測力を改善したかどうかを見る(BLLのTable 3)

# ノンパラメトリック推定とその応用

TABLE 3. Relative MSPE (Nonparametric VAR/Nonparametric AR)

Horizon	M1	M2	M3	M1D	M2D	M3D
1 Month	1.12 (-0.97)	0.94 (1.78)	1.00 (0.06)	1.00 (0.01)	0.99 (0.46)	0.99 (0.78)
6 Months	1.00 (0.05)	0.98 (0.88)	1.02 (-0.66)	0.94 (1.80)	0.98 (0.79)	1.00 (0.24)
12 Months	0.95 (1.20)	0.90 (2.76)	0.98 (0.43)	0.99 (0.49)	0.99 (0.40)	0.98 (0.93)
24 Months	0.96 (0.69)	0.66 (6.43)	0.89 (1.10)	0.98 (1.46)	0.88 (4.33)	0.91 (1.86)

これを見ると、非線形性を導入した場合は、money growth による予測力の向上は、ややうかがえるもののそれほど顕著ではないことがわかる。

## ノンパラメトリック推定とその応用

---

またBLL はmoney growth の代わりに velocity を用いた場合に予測力が向上するかも調べている( P\*インフレーションモデルと呼ばれているそう。正式な理論的な基礎づけはないそう)

先ほどのモデルにおいて money growth の代わりに velocity を用いて同じことをしたものが以下の表である (BLL の Table 4 と Table 5)。

# ノンパラメトリック推定とその応用

TABLE 4. Relative MSPE (Nonparametric VAR/Parametric VAR)

Horizon	V1	V2	V3	V1D	V2D	V3D
1 Month	0.74 (2.24)	0.72 (3.04)	0.88 (1.62)	0.77 (2.05)	0.76 (2.42)	0.77 (2.10)
6 Months	0.84 (1.62)	0.90 (1.29)	0.93 (1.42)	0.87 (1.56)	0.85 (1.84)	0.84 (2.02)
12 Months	0.73 (3.10)	0.88 (1.77)	0.90 (2.07)	0.75 (2.87)	0.76 (3.56)	0.79 (3.14)
24 Months	0.83 (2.00)	0.82 (1.29)	1.20 (-3.14)	0.99 (0.09)	0.98 (0.54)	0.99 (0.16)

これを見ると、非線形性を導入するとこの場合もやはり予測力が改善することがわかる。さらにその改善度は money growth を用いた場合に比べてかなり大きい。

# ノンパラメトリック推定とその応用

TABLE 5. Relative MSPE (Nonparametric VAR/Nonparametric AR)

Horizon	V1	V2	V3	V1D	V2D	V3D
1 Month	0.94 (0.74)	0.90 (1.85)	0.99 (0.16)	0.97 (0.43)	0.87 (2.25)	0.87 (2.03)
6 Months	1.01 (-0.14)	0.93 (0.94)	0.93 (1.62)	1.03 (-0.43)	0.84 (2.01)	0.84 (2.29)
12 Months	0.93 (0.85)	0.83 (2.41)	0.83 (4.17)	0.93 (0.80)	0.71 (3.87)	0.71 (4.02)
24 Months	0.83 (2.57)	0.63 (4.12)	0.70 (3.31)	1.00 (-0.01)	0.60 (4.60)	0.59 (4.79)

これを見ると、非線形性を導入した場合にも、velocity を加えることによる予測力の向上は無視できないほど大きいことがわかる。

# ノンパラメトリック推定とその応用

---

これらの結果より、BLLは前述した主要結果を得た。

BLLではその他にもThreshold Model との比較など様々な統計的な検定を行っているが、ここではそれらについては省略する。



# ノンパラメトリック推定とその応用

---

## ■ セミパラメトリック推定と部分線形モデル

以下のようなモデルは**部分線形(Partially Linear)モデル**と呼ばれる。

$$y_i = \mathbf{X}_i' \boldsymbol{\beta} + g(\mathbf{Z}_i) + e_i$$

$$E(e_i | \mathbf{X}_i, \mathbf{Z}_i) = 0,$$

$$E(e_i^2 | \mathbf{X}_i = \mathbf{x}, \mathbf{Z}_i = \mathbf{z}) = \sigma^2(\mathbf{x}, \mathbf{z})$$

ここで、 $\mathbf{X}_i$  と  $\mathbf{Z}_i$  は説明変数ベクトル、 $g(\cdot)$  は  $\mathbf{Z}_i$  の未知の関数である。このような定式化においてしばしば、 $\mathbf{X}_i$  にはダミー変数を入れたり、 $\mathbf{Z}_i$  はスカラー変数としたりする。

ここでの目的は  $\boldsymbol{\beta}$  と  $g(\cdot)$  の推定である。

# ノンパラメトリック推定とその応用

---

線形の部分はパラメトリックであり、非線形の部分はノンパラメトリックなのでこのようなモデルは**セミパラメトリックモデル**と呼ばれる。

ここでは Robinson (1988) によって提案された方法の基本的なアイデアを簡単に紹介する。

回帰式の両辺を  $\mathbf{Z}_i$  の条件付期待値を取る(重複期待値の法則より  $E(e_i | \mathbf{Z}_i) = 0$ )。

$$\begin{aligned} E(y_i | \mathbf{Z}_i) &= E(\mathbf{X}_i' \boldsymbol{\beta} | \mathbf{Z}_i) + E(g(\mathbf{Z}_i) | \mathbf{Z}_i) + E(e_i | \mathbf{Z}_i) \\ &= E(\mathbf{X}_i | \mathbf{Z}_i)' \boldsymbol{\beta} + g(\mathbf{Z}_i) \end{aligned}$$

## ノンパラメトリック推定とその応用

---

この両辺を最初の回帰式から引くと  $g(\mathbf{Z}_i)$  がなくなる。

$$y_i - E(y_i | \mathbf{Z}_i) = [\mathbf{X}_i' - E(\mathbf{X}_i' | \mathbf{Z}_i)] \boldsymbol{\beta} + e_i.$$

ここで  $e_{yi} = y_i - E(y_i | \mathbf{Z}_i)$ ,  $\mathbf{e}_{xi} = \mathbf{X}_i' - E(\mathbf{X}_i' | \mathbf{Z}_i)$  とおくと  
上式は

$$e_{yi} = \mathbf{e}_{xi} \boldsymbol{\beta} + e_i$$

と書き直せる。この式は線形なので、もし  $e_{yi}$  と  $\mathbf{e}_{xi}$  がわか  
っていれば最小二乗法によって推定できる。

Robinson (1988) は  $E(y_i | \mathbf{Z}_i)$  と  $E(\mathbf{X}_i | \mathbf{Z}_i)$  を NW推定量でノ  
ンパラメトリック推定し、その残差  $\hat{e}_{yi}$ ,  $\hat{\mathbf{e}}_{xi}$  を使って  $\boldsymbol{\beta}$  を最  
小二乗法で推定することを提案している。

# ノンパラメトリック推定とその応用

---

## ■ Bachmeier and Li (2002) の紹介

ここでは部分線形モデルを用いてタームストラクチャーの推定をしている論文 Bachmeier, L., and Li, Q. (2007) *Applied Economics Letters* (以下BL)を紹介する。

BLは短期の利子率と長期の利子率との関係は誤差修正モデルに従っているとし、ベンチマークモデルとしてまず、以下の線形の誤差修正モデルを推定した。

$$\Delta R_t = \mu + \alpha w_{t-1} + \gamma_1 \Delta R_{t-1} + \gamma_2 \Delta r_{t-1} + u_t$$

ここで  $R_t$  は長期利子率、 $r_t$  は短期利子率、

$$w_{t-1} = R_{t-1} - \beta r_{t-1}$$

は誤差修正項、 $\beta$  は共和分係数である。

## ノンパラメトリック推定とその応用

また、このベンチマークモデルに対して、誤差修正項に非線形性を導入した部分線形モデル

$$\Delta R_t = \theta(w_{t-1}) + \gamma_1 \Delta R_{t-1} + \gamma_2 \Delta r_{t-1} + u_t$$

をセミパラメトリック推定し、推定したそれぞれのモデルの予測のパフォーマンスをサンプル外予測によって比較している。 $\gamma_1$  と  $\gamma_2$  の推定方法は Robinson (1988) および Li and Wooldridge (2002) による。(1時点先)予測を行うために  $E(\Delta R_{t+1} | w_t)$ ,  $E(\Delta R_t | w_t)$ ,  $E(\Delta r_t | w_t)$  をノンパラメトリック推定し、その推定値を用いて  $\theta(w_t)$  を

$$\hat{\theta}(w_t) = \hat{E}(\Delta R_{t+1} | w_t) - \hat{\gamma}_1 \hat{E}(\Delta R_t | w_t) - \hat{\gamma}_2 \hat{E}(\Delta r_t | w_t)$$

と推定し、これらを用いて  $\Delta R_{t+1}$  の予測値を

# ノンパラメトリック推定とその応用

---

$$\Delta \hat{R}_{t+1} = \hat{\theta}(w_t) + \hat{\gamma}_1 \Delta R_t + \hat{\gamma}_2 \Delta r_{t-1}$$

と計算した。予測のパフォーマンスはMSE

$$MSE = 80^{-1} \sum_{t=401}^{480} (\Delta \hat{R}_t - \Delta R_t)^2$$

を用いたとある。80個の予測値によって評価しているのがわかる。ただしBLにはパラメーターの推定をデータが増えるたびに更新しているのかそれとも一度推定したものを固定してやったのかが書いていないのでそのあたりはわからない。

長期の利子率として5年と10年、短期の利子率として1, 2, 3, 6, 12ヵ月を用いた場合の結果は以下のようなになった(BLのTable 1とTable 2)

# ノンパラメトリック推定とその応用

Table 1. *Average prediction MSE using 5-year interest rates*

	1 Month	2 Month	3 Month	6 Month	12 Month
Linear	0.316	0.160	0.140	0.152	0.160
Semiparametric	0.320	0.176	0.135	0.143	0.149

Table 2. *Average prediction MSE using 10-year interest rates*

	1 Month	2 Month	3 Month	6 Month	12 Month
Linear	0.323	0.157	0.136	0.149	0.159
Semiparametric	0.316	0.141	0.129	0.135	0.144

## ノンパラメトリック推定とその応用

---

これらより5年の長期利率に対しては(6か月と12か月を除いて)そこまでもないが、10年の長期利率に対しては非線形モデルを用いることによって予測力が増しているのがわかる。



# ノンパラメトリック推定とその応用

---

## ■ $K$ -Nearest Neighbor Method

カーネル推定量とは別のノンパラメトリック推定として  **$K$ -Nearest neighbor method** があげられる(以下では  $K$ -nn と略す)。近年マクロ経済学の実証の分野でもこの方法を用いる論文が出てきているため、ここで説明する。

この方法はカーネル推定量と比べてよい面もあれば悪い面もあり、どちらがいいとは一概には言えない。特に小標本での特性について今後の研究がまたれるところである。

## ノンパラメトリック推定とその応用

---

$K-nn$  法の基本的なアイデアは推定する点に応じて、バンド幅を自動的に変更していくというものである。これはカーネル法がバンド幅を推定する点に関係なく固定していたのと対照的である。

より具合的には 推定に用いられる観測数の数が常に “ $K$ ” 個 になるように **バンド幅を変更する**。これが  $K-nn$  法と呼ばれるゆえんである。

カーネル法ではバンド幅を観測数が増えるにつれて小さくしていったが、 $K-nn$  法では観測数が増えるにつれて  $K$  を大きくしていく。

# ノンパラメトリック推定とその応用

## ■ $K - nn$ 法による多変量密度関数の推定

ある点  $\mathbf{x} = [x_1, \dots, x_q]'$  と  $\mathbf{X}_i = [X_{1i}, \dots, X_{qi}]'$  の距離を

$$D_i = \|\mathbf{x} - \mathbf{X}_i\| = \sqrt{(x_1 - X_{1i})^2 + \dots + (x_q - X_{qi})^2}$$

と定義する。この距離を小さい順に並べる

$$0 \leq D_{(1)} \leq D_{(2)} \leq \dots \leq D_{(n)}$$

ここで  $D_{(k)}$  は  $D_i$  のうち  $k$  番目に小さいもの。

ここで“ $K$ ”番目に小さいものを  $R_x$  としよう。つまり

$$R_x = D_{(K)}$$

である。

## ノンパラメトリック推定とその応用

---

この時、 $\mathbf{X}_{(k)}$  を  $D_{(k)}$  に対応する  $\mathbf{X}_i$  とすると  $\mathbf{x}$  からの距離が  $R_x$  以下になるものは ( $\mathbf{X}_{(K)}$  も含めて)  $K$  個ある。

多変量 Uniform カーネルを

$$w(\|\mathbf{u}\|) = c_q^{-1} \mathbf{1}(\|\mathbf{u}\| \leq 1), \quad c_q = \frac{\pi^{q/2}}{\Gamma(q + 2/2)}$$

として、 $R_x$  をバンド幅としたカーネル推定量は

$$\begin{aligned} \tilde{f}(x) &= \frac{1}{nR_x^q} \sum_{i=1}^n c_q^{-1} \mathbf{1}(\|x - X_i\| \leq R_x) \\ &= \frac{1}{nR_x^q} \sum_{i=1}^n c_q^{-1} \mathbf{1}(D_i \leq R_x) \end{aligned}$$

である。

## ノンパラメトリック推定とその応用

---

ここで  $1 (D_i \leq R_x)$  となる観測数は必ず  $K$  個になることから、これはさらに

$$\tilde{f}(\mathbf{x}) = \frac{K}{nR_x^q c_q}$$

となる。これが  $K - nn$  法による密度関数の推定量である。これは  $R_x$  だけに依存している。上記の Uniform カーネルを使ったもの以外にも

$$\tilde{f}(\mathbf{x}) = \frac{1}{nR_x^q} \sum_{i=1}^n w\left(\frac{\|x - X_i\|}{R_x}\right)$$

という他のカーネルを用いたものも考えられる。ここで  $w$  は  $\int_{R^q} w(\|\mathbf{u}\|) d\mathbf{u} = 1$  を満たす関数である。

# ノンパラメトリック推定とその応用

---

$K-nn$  法による推定量についてもそのバイアス、分散、MSEを導くことができるがここでは省略する (Li and Racine (2006) を参照)

要点としては

- (1)  $K$  は  $n$  が増えるにつれて増やしていくが、漸近MSEが最小になる  $K$  の増加スピードは  $n^{4/(4+q)}$  である。
- (2) 最適な  $K$  の増加スピードの下では MSE の減少スピードは  $n^{-4/(4+q)}$  となる。これはカーネル推定の場合と同じである。

$K-nn$  を用いてノンパラメトリック回帰モデルを推定することもできる。これについても詳しい内容については (Li and Racine (2006)を参照のこと。

# ノンパラメトリック推定とその応用

---

## ■ Guégan and Rakotomarahy (2010) の紹介

ここでは  $K-nn$  法を 4半期GDPの Forecasting に利用した論文 Guégan and Rakotomarahy (2010) Ch. 8 in Nonlinear Modeling of Economic and Financial Time Series を紹介する。以下ではGRと略す。

GRは GDP 自体の予測はパラメトリックモデルを用いているが、GDPの予測式の中の変数(月次変数)について  $K-nn$  法を適用し予測値を計算し、その予測した変数を用いている。彼らの主張を要約すると、GDP速報値を出すときには予測モデルの中の月次変数で公表されていないものが多いので、それをうまく予測するとGDP予測が改善されるというものである。

## ノンパラメトリック推定とその応用

---

$Y_t$  を 4半期GDPの対数階差による変化率(対数階差  $\times 100$ )とする。GRでは  $Y_t$  の予測式として、まず以下の Bridge Equations という予測式を考える。

$$Y_t^{(i)} = g^{(i)}(\mathbf{X}_t^{(i)}) + \varepsilon_t, i=1, \dots, b$$

ここで  $Y_t^{(i)}$  は  $i$  番目の Bridge Equation からの予測値、 $g^{(i)}(\cdot)$  は  $i$  番目の Bridge Equation,  $\mathbf{X}_t^{(i)}$  をそこで使われた説明変数を表す。GRは 8つの方程式を考えている ( $b = 8$ )。GRでは  $g^{(i)}(\cdot)$  は全て線形の方程式である。実際の方程式の形および使われた変数のリストについては GR を参照のこと。最終的な予測値は 8つの予測値の平均をとっている(これはいわゆる予測平均)。



## ノンパラメトリック推定とその応用

GR は EABCN によって提供された real-time database を使用した。GR は  $X_t$  を  $nn$  法によって予測した場合と、VAR によって予測した場合とで予測の MSE を比べている。以下がその結果である。

Table 2: RMSE on GDP Growth from the Three Methods VAR,  $k$ -NN with  $d = 1$ , and  $k$ -NN with  $d > 1$

Horizon	VAR	$k$ -NN( $d = 1$ )	$k$ -NN( $d > 1$ )
6	0.225	<b>0.198</b>	0.214
5	0.224	0.203	<b>0.192</b>
4	0.214	0.202	<b>0.196</b>
3	0.192	0.186	<b>0.177</b>
2	0.181	<b>0.176</b>	0.177
1	0.173	0.174	<b>0.171</b>

$nn$  法によって予測力が改善されているのがわかる。

# まとめ

---

ノンパラメトリック推定について解説を行った。

ノンパラメトリック推定はメリットもあるがデメリットもある。特に次元が増えるにつれて推定精度が落ちる問題はかなり厳しい問題である。

ただしこの問題は部分的線形モデルやここでは紹介しなかった single index モデルなどのセミパラメトリックモデルを用いるとある程度緩和される。

いくつかの実証分析ではモデルに非線形を導入すると予測精度が上がることを示唆しており、非線形性がある場合はそれを考慮にいれることによってよりよい分析結果を得られる可能性がある。

## 追加参考文献

---

Robinson, P.M. (1988) Root-n consistent semiparametric regression, *Econometrica* 56, 931-954.