

9 ノンパラメトリック回帰

回帰モデル

$$y_i = X_i\beta + u_i$$

通常，線形を仮定 → この仮定を緩める

$$y_i = m(x_i) + u_i$$

$m(\cdot)$ が未知 → $m(\cdot)$ を推定 → ノンパラメトリック回帰

準備として，密度関数が未知で，密度関数を推定する

9.1 密度関数のノンパラメトリック推定

分布を仮定しない → 分布関数自体を推定 → ノンパラメトリック推定

ノンパラメトリック推定に関するテキスト：

- Pagan, A. and Ullah, A., (1999),

Nonparametric Econometrics, Cambridge University Press.

- Prakasa Rao, B.L.S., (1983),

Nonparametric Functional Estimation, Academic Press, Inc.

- Silverman, B.W., (1986),

Density Estimation for Statistics and Data Analysis (Monographs on Statistics and Applied Probability, No.26), Chapman & Hall.

同じ分布からの n 個の観測値 x_i , $i = 1, 2, \dots, n$, があるとする。

x_i の密度関数を $f(x)$ とする。

密度関数 $f(x)$ が未知であるとする。

密度関数 $f(x)$ の推定値を $\hat{f}(x)$ とする。

n 個のデータから密度関数を次のように推定することができる。

密度関数 $f(x)$ と分布関数 $F(x)$ の関係：

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x + \frac{h}{2}) - F(x - \frac{h}{2})}{h} \approx \frac{F(x + \frac{h}{2}) - F(x - \frac{h}{2})}{h}$$

$F(x) = \mathbf{P}(X \leq x) = x$ 以下の確率

$$\begin{aligned} \hat{f}(x) &= \frac{F(x + \frac{h}{2}) - F(x - \frac{h}{2})}{h} \\ &\approx \frac{1}{nh} [n \text{ 個の観測値 } x_1, \dots, x_n \text{ の中で } (x - \frac{h}{2}, x + \frac{h}{2}) \text{ の範囲に入っている個数}] \end{aligned}$$

$$= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (22)$$

ただし、 h はある小さな数で、バンド幅 (Band Width)、窓幅 (Windows Width)、平滑化定数 (Smoothing Parameter) などと様々な呼び名があるが、以降では、 h のことをバンド幅と呼ぶことにする。

また、(22)に含まれる $K(\cdot)$ は、

$$K(z) = \begin{cases} 1 & |z| < \frac{1}{2} \text{ のとき} \\ 0 & \text{その他} \end{cases}$$

と表される。→ この場合、 $K(\cdot)$ は一様分布 → Rectangle Kernel と呼ばれる。

$K(\cdot)$ はカーネル (Kernel) と呼ばれる。

しかし、より一般的には、 $\int K(t) dt = 1$ 、かつ、すべての t について $K(t) \geq 0$ となるような $K(\cdot)$ を選べばよい。

$K(\cdot)$ の選択として、代表的なものは標準正規分布である。

問題は h の選び方である。

h を大きくすると、分布関数は滑らかに近似され真の分布関数とは異なったものになってしまう。

逆に h を小さくすると、分布関数は必要以上に凸凹になり、真の分布関数とはかけ離れたものになる。

よって、適切な h が選ばれる必要がある。

密度関数 $f(x)$ の推定量は実現値 x_i をその確率変数 X_i で置き換えて、

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (23)$$

と書くことができる。

特に区別を必要としない限り、密度関数 $f(x)$ の推定値と推定量は同じ記号 $\hat{f}(x)$ を使うことにする。

さらに、 x が p 次元の場合は x の密度関数の推定値は次のように表される。

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

しかし、ここでは簡単化のため、 $p = 1$ の場合のみを扱うことにする。

9.2 バンド幅 h の選択

準備として、平均自乗誤差の積分 (Integrated Mean Square Error, IMSE) を考える。

それは次のように変形することができる。

$$\begin{aligned} \text{IMSE}(\hat{f}(x)) &\equiv \int \text{MSE}(\hat{f}(x)) \, dx \equiv \int \text{E}(\hat{f}(x) - f(x))^2 \, dx \\ &= \int (\text{E}(\hat{f}(x)) - f(x))^2 \, dx + \int \text{Var}(\hat{f}(x)) \, dx \end{aligned} \quad (24)$$

(24) に含まれる期待値，分散は，密度関数 $f(x)$ の推定量 $\hat{f}(x)$ に含まれる確率変数 X_i , $i = 1, 2, \dots, n$ に関する期待値，分散であることを注意せよ。

また， X_i は密度関数 $f(x)$ を持つ確率変数である。

よって，(23) を用いて， $E(\hat{f}(x))$, $\text{Var}(\hat{f}(x))$ は次のように書き換えられる。

$$\begin{aligned} E(\hat{f}(x)) &= E\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\right) = \frac{1}{n} \sum_{i=1}^n E\left(\frac{1}{h} K\left(\frac{x - X_i}{h}\right)\right) = E\left(\frac{1}{h} K\left(\frac{x - X}{h}\right)\right) \\ &= \int \frac{1}{h} K\left(\frac{x - y}{h}\right) f(y) dy \end{aligned} \quad (25)$$

$$\begin{aligned} \text{Var}(\hat{f}(x)) &= \text{Var}\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\right) = \sum_{i=1}^n \text{Var}\left(\frac{1}{nh} K\left(\frac{x - X_i}{h}\right)\right) \\ &= \frac{1}{n} \text{Var}\left(\frac{1}{h} K\left(\frac{x - X}{h}\right)\right) = \frac{1}{n} E\left(\frac{1}{h} K\left(\frac{x - X}{h}\right)\right)^2 - \frac{1}{n} \left(E\left(\frac{1}{h} K\left(\frac{x - X}{h}\right)\right)\right)^2 \\ &= \frac{1}{n} \int \frac{1}{h^2} K\left(\frac{x - y}{h}\right)^2 f(y) dy - \frac{1}{n} \left(\int \frac{1}{h} K\left(\frac{x - y}{h}\right) f(y) dy\right)^2 \end{aligned} \quad (26)$$

簡単化のために，カーネル $K(\cdot)$ は，

$$\int K(t) dt = 1 \quad \int tK(t) dt = 0 \quad \int t^2 K(t) dt = k_2 \neq 0 \quad (27)$$

を満たす左右対称な関数とする。

また，未知の密度関数 $f(x)$ はすべての次数で微分可能な連続関数とする。

$k_2 = 1$ と置くこともできる。

$y = x - ht$ と変数変換を行い，(25) を次のように書き直す。

$$E(\hat{f}(x)) = \int \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y) dy = \int K(t) f(x - ht) dt \quad (28)$$

$ht = 0$ の近傍で $f(x - ht)$ をテーラー展開すると，次のようになる。

$$f(x - ht) = f(x) - ht f'(x) + \frac{1}{2} h^2 t^2 f''(x) + \frac{1}{6} h^3 t^3 f'''(x) + O(h^4)$$

(28) に代入すると,

$$\begin{aligned} E(\hat{f}(x)) &= \int K(t) \left(f(x) - ht f'(x) + \frac{1}{2} h^2 t^2 f''(x) + \frac{1}{6} h^3 t^3 f'''(x) + O(h^4) \right) dt \\ &= f(x) + \frac{1}{2} h^2 f''(x) k_2 + O(h^4) \end{aligned} \quad (29)$$

が得られる (カーネル $K(\cdot)$ は左右対称と仮定しているので, 1 行目の式の右辺の第 1 項と第 3 項はゼロになることに注意)。

同様に, $y = x - ht$ と変数変換を行うと, (26) の第 1 項, 第 2 項は, それぞれ,

$$\begin{aligned} \frac{1}{n} \int \frac{1}{h^2} K\left(\frac{x-y}{h}\right)^2 f(y) dy &= \frac{1}{n} \int \frac{1}{h} K(t)^2 f(x - ht) dt \\ &= \frac{1}{n} \int \frac{1}{h} K(t)^2 \left(f(x) - ht f'(x) + O(h^2) \right) dt \\ &= \frac{1}{nh} f(x) \int K(t)^2 dt + O(n^{-1}) \end{aligned} \quad (30)$$

$$\begin{aligned} \frac{1}{n} \left(\int \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y) dy \right)^2 &= \frac{1}{n} (f(x) + \frac{1}{2} h^2 f''(x) k_2 + O(h^4))^2 = \frac{1}{n} (f(x)^2 + O(h^2)) \\ &= O(n^{-1}) \end{aligned} \quad (31)$$

となる (n が大きくなるにつれて, h は小さくなるを考える)。

よって, $\text{Var}(\hat{f}(x))$ は, (26) に (30) と (31) を代入して,

$$\text{Var}(\hat{f}(x)) = \frac{1}{nh} f(x) \int K(t)^2 dt + O(n^{-1}) = \frac{1}{nh} (f(x) \int K(t)^2 dt + O(h)) \quad (32)$$

と表される。(29) と (32) を用いて, $\text{MSE}(\hat{f}(x))$ は次のように近似される。

$$\begin{aligned} \text{MSE}(\hat{f}(x)) &= \text{Var}(\hat{f}(x)) + (\text{E}(\hat{f}(x)) - f(x))^2 \\ &\approx \frac{1}{nh} f(x) \int K(t)^2 dt + \frac{1}{4} h^4 k_2^2 f''(x)^2 \end{aligned} \quad (33)$$

また, (24) の $\text{IMSE}(\hat{f}(x))$ は, (33) の $\text{MSE}(\hat{f}(x))$ の近似を x について積分する

ことにより,

$$\begin{aligned}\text{IMSE}(\hat{f}(x)) &= \int \text{MSE}(\hat{f}(x)) \, dx \\ &\approx \frac{1}{nh} \int K(t)^2 \, dt + \frac{1}{4} h^4 k_2^2 \int f''(x)^2 \, dx\end{aligned}\quad (34)$$

が得られる。

(34) が最小になる h を \hat{h} とすると,

$$\hat{h} = k_2^{-2/5} \left(\int K(t)^2 \, dt \right)^{1/5} \left(\int f''(x)^2 \, dx \right)^{-1/5} n^{-1/5}\quad (35)$$

となる。

また, (27) の条件を満たし, (34) の $\text{IMSE}(\hat{f}(x))$ を最小にするカーネル $K(\cdot)$ は,

$$K(t) = \begin{cases} \frac{3}{4}(1-t^2) & -1 \leq t \leq 1 \\ 0 & \text{その他} \end{cases}\quad (36)$$

となり、Epanechnikov カーネルと呼ばれる (Epanechnikov (1969) を参照せよ)。ただし、(36) は $\text{IMSE}(\hat{f}(x))$ の近似 (34) を最小にしたものであり、必ずしも最適なカーネルとは限らないことに注意せよ。

以上をもとにして、バンド幅 h の選択として考えられるものとしては、大別して、(24) や尤度関数に基づくものと (35) に基づくものの 2 つがあげられる。前者はクロス・バリデーション、後者はプラグ・イン法として知られている。

ここでは、クロス・バリデーションに基づく方法を 2 つ、プラグ・イン法に基づく方法を 2 つの合計 4 つのバンド幅 h の推定を紹介する。

9.2.1 クロス・バリデーション (Cross-Validation)

クロス・バリデーション (Cross-Validation) には 2 つの種類がある。一つは (24) を最小にする h を求めるという最小自乗クロス・バリデーション (Least-Squares

Cross-Validation) と呼ばれるものであり、もう一つは尤度関数を最大にする h を求めるという尤度クロス・バリデーション (Likelihood Cross-Validation) である。2 つを以下に簡単に説明しておく。

最小自乗クロス・バリデーション (Least-Squares Cross-Validation): (24) の $\text{IMSE}(\hat{f}(x))$ は,

$$\text{IMSE}(\hat{f}(x)) = E \int \hat{f}(x)^2 dx - 2E \int \hat{f}(x)f(x) dx + \int f(x)^2 dx \quad (37)$$

と分解される。(37) の第 3 項は h に依存しないため、第 2 項までを最小にする h を求めればよい。期待値 $E(\cdot)$ を無視すると、第 1 項、第 2 項は次のようになる。

$$\text{CV}(h) \equiv \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{j=1}^n \int K\left(\frac{x_i - x_j}{h} - t\right) K(t) dt - 2 \frac{1}{n} \sum_{i=1}^n \hat{f}_{-i}(x_i) \quad (38)$$

ただし、 $\hat{f}_{-i}(x_i)$ は、

$$\hat{f}_{-i}(x_i) \equiv \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{x_i - x_j}{h}\right)$$

である。

(38) の第 1 項の評価は、モンテ・カルロ積分や数値積分が考えられる。モンテ・カルロ積分を利用すると、 $K(\cdot)$ から m 個の乱数 t_1, t_2, \dots, t_m を発生させて、次のように $CV(h)$ が計算される。

$$CV(h) = \frac{1}{n^2 h^2} \frac{1}{m} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m K\left(\frac{x_i - x_j}{h} - t_k\right) - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{x_i - x_j}{h}\right) \quad (39)$$

$K(\cdot)$ は、分析者によって特定される密度関数であるので、乱数生成の簡単な関数を選べばよい。

尤度クロス・バリデーション (**Likelihood Cross-Validation**): 対数尤度関数 $\log L$ は次のように近似される。

$$\log L = \sum_{i=1}^n \log f(x_i) \approx \sum_{i=1}^n \log \hat{f}_{-i}(x_i) = \sum_{i=1}^n \log \left(\frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{x_i - x_j}{h}\right) \right) \quad (40)$$

(40) が最大になるような h を選べばよい。(39) や (40) に基づいて h の推定値 \hat{h} を求める方法は、単純探査法 (Simple Grid Search) 等を用いる必要がある。

9.2.2 プラグ・イン法 (Plug-In Method)

(35) は、近似された $\text{IMSE}(\hat{f}(x))$ を最小にするような h である。問題は、(35) において未知の部分は $f''(x)$ であるということである。最も簡単な方法は、未知の密度関数 $f(\cdot)$ を正規分布 (分散を σ^2 とする) で近似してしまうことである。すなわち、

$$\hat{h} = k_2^{-2/5} \left(\int K(t)^2 dt \right)^{1/5} \left(\frac{3}{8\sqrt{\pi}} \sigma^{-5} \right)^{-1/5} n^{-1/5} \quad (41)$$

となる。ただし、 $k_2 \equiv \int t^2 K(t) dt = 1$ とする。 σ は x_1, x_2, \dots, x_n から得られる標本不偏分散の平方根 (すなわち、標本標準偏差) に置き換えられる。 $f(x)$ が分散 σ^2 の正規分布のとき、 $\int f''(x)^2 dx = \frac{3}{8\sqrt{\pi}}\sigma^{-5}$ と計算されることに注意せよ。

また、 $f(\cdot)$ を $\hat{f}(\cdot)$ で置き換えると、次のようになる。

$$\hat{h} = k_2^{-2/5} \left(\int K(t)^2 dt \right)^{1/5} \left(\int \hat{f}''(x)^2 dx \right)^{-1/5} n^{-1/5} \quad (42)$$

(42) の $\int \hat{f}''(x)^2 dx$ の評価について、カーネル $K(\cdot)$ を標準正規分布としたとき、

$$\begin{aligned} \int \hat{f}''(x)^2 dx &= \frac{1}{n^2 h^5} \sum_{i=1}^n \sum_{j=1}^n \int \left(\left(\frac{x_i - x_j}{h} - t \right)^2 - 1 \right) (t^2 - 1) K\left(\frac{x_i - x_j}{h} - t \right) K(t) dt \\ &\approx \frac{1}{n^2 h^5 m} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \left(\left(\frac{x_i - x_j}{h} - t_k \right)^2 - 1 \right) (t_k^2 - 1) K\left(\frac{x_i - x_j}{h} - t_k \right) \end{aligned} \quad (43)$$

となる。ただし、 t_1, t_2, \dots, t_m は $K(t)$ から生成された乱数とする。問題は、(43) が h に依存することである。次節では、 $K(\cdot)$ を標準正規分布とにおいて、(41) を用いて、 h を求めると、 $1.06\hat{\sigma}n^{-1/5}$ となるので、これを (43) の中の h とした。

(39) や (43) と同様に, $\int K(t)^2 dt$ は, モンテ・カルロ積分や数値積分で評価される。 $K(t)$ から生成された t_1, t_2, \dots, t_m をもとにして,

$$\int K(t)^2 dt = E(K(X)) \approx \frac{1}{m} \sum_{i=1}^m K(t_k) \quad (44)$$

と計算される。ただし, X は密度関数 $K(\cdot)$ を持つ確率変数とする。次節では。カーネルには (36) と標準正規分布を用いるので, $\int K(t)^2 dt$ を明示的に求めることができる。(44) のようなモンテ・カルロ積分による積分の近似は必要としない。すなわち, $K(\cdot)$ を (36) とすると, $\int K(t)^2 dt = 0.6$ となり, $K(\cdot)$ を標準正規分布とすると, $\int K(t)^2 dt = (2\sqrt{\pi})^{-1}$ となる。

(39), (43), (44) の積分値を求める場合に, 誤差を減らすためのテクニックを以下に記しておく。 $K(t)$ から生成された乱数 t_1, t_2, \dots, t_m は, 既に, 小さい順に並べ替えられているものとする。 t_1, t_2, \dots, t_m は乱数なので, 等確率で現れるもの

と離散近似することが出来る。すなわち、初期値 t_1 は

$$\int_{-\infty}^{t_1} K(t) dt = \frac{1}{2m}$$

を満たす値として求め、 t_2, t_3, \dots, t_m は

$$\int_{t_{k-1}}^{t_k} K(t) dt = \frac{1}{m} \quad k = 2, 3, \dots, m$$

を満たす値として順次求められる。言い換えると、 t_k は、密度関数 $K(\cdot)$ の $100 \times (k - 0.5)/m$ % 点に相当する。このようにして t_1, t_2, \dots, t_m の値を定めることによって、(39), (43), (44) の積分値を求めるためのシミュレーション誤差を減らすことができる。

以上のように、 h の推定としては、(39) を最小にする h , (40) を最大にする h , (41), (43) を用いた (42) の 4 通りを紹介した。(39) や (40) に基づいて h の推定値 \hat{h} を求める方法は、単純探索法 (Simple Grid Search) を用いる必要があるため、

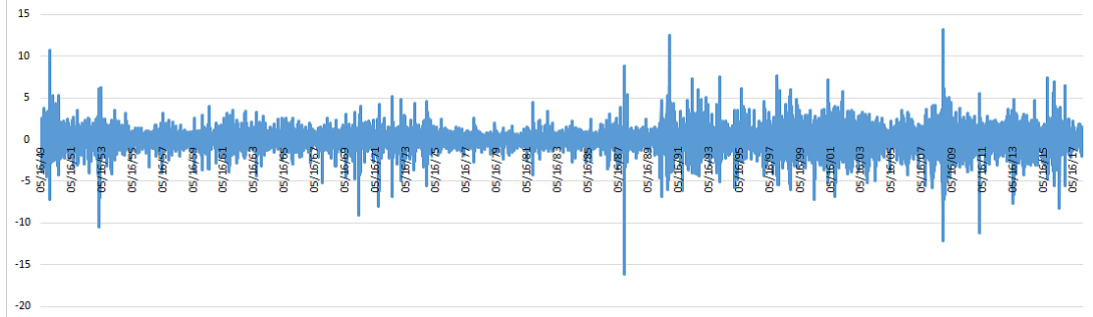
計算時間が非常にかかる。特に, (39) には, Σ が 3 つも含まれているため, 単純探査法による h の最適点を求めることはほとんど不可能に近い。



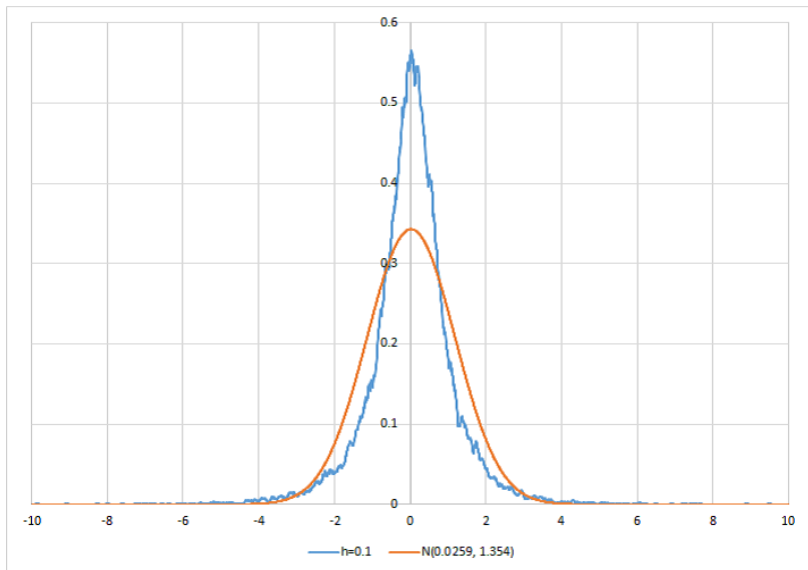
グラフの例：

日経平均株価指数 1949年5月16日～2017年12月29日の日時データ

变化率

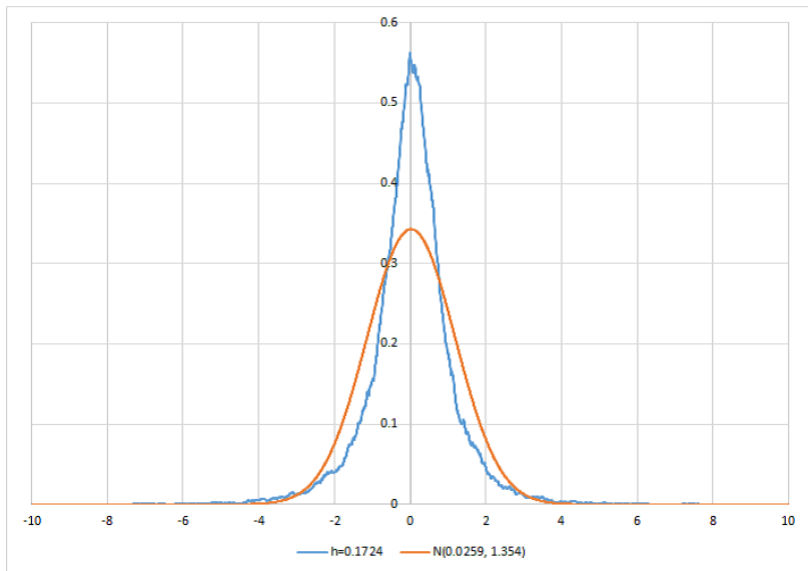


上昇率 $100 \ln(x_i/x_{i-1}), i = 1, 2, \dots, n, n = 18767$



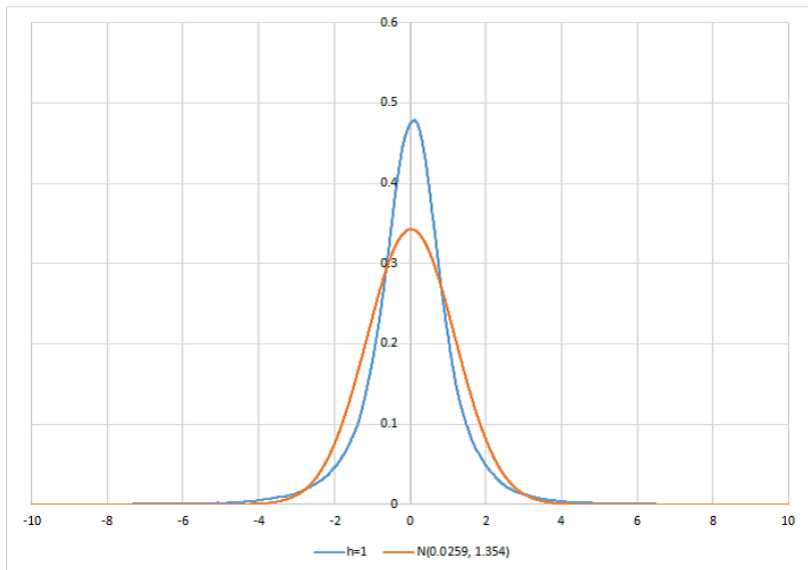
青線： $h = 0.1$ で密度関数を推定

赤線： $N(0.0259, 1.354)$ ← データからの平均と分散による正規分布



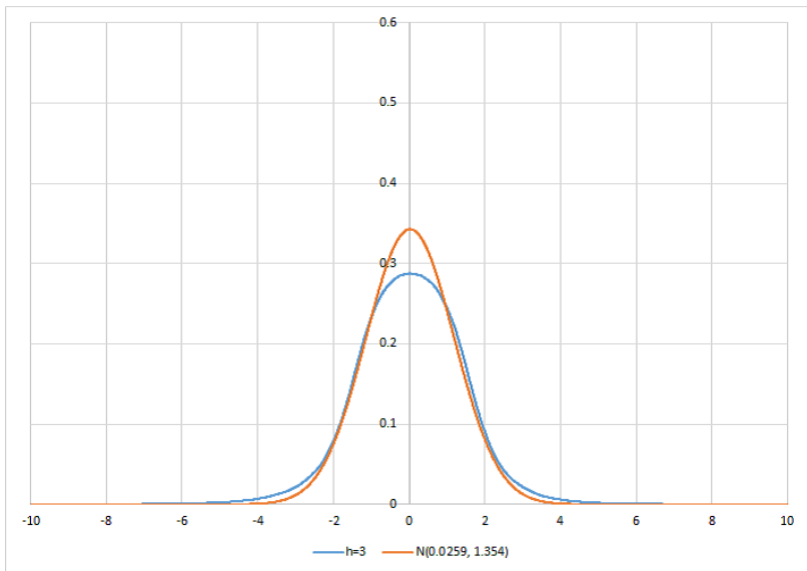
青線： $h = 0.17$ で密度関数を推定 ← $h = 1.06\hat{\sigma}n^{-1/5} \approx 0.17$, $\hat{\sigma}^2 = 1.354$

赤線： $N(0.0259, 1.354)$ ← データからの平均と分散による正規分布



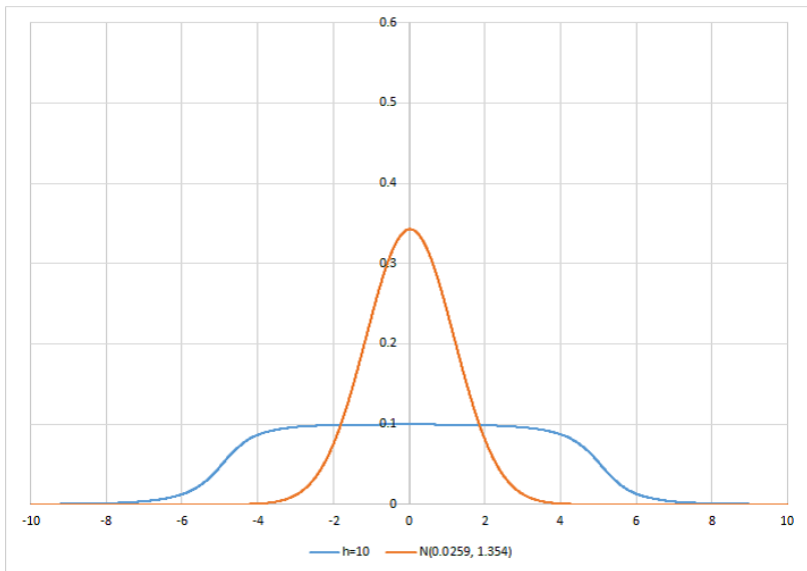
青線： $h = 1$ で密度関数を推定

赤線： $N(0.0259, 1.354)$ ← データからの平均と分散による正規分布



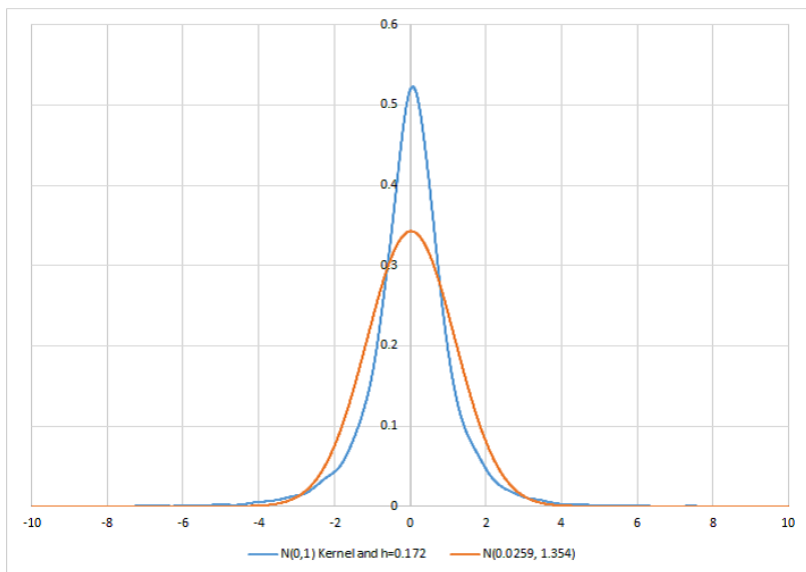
青線： $h = 3$ で密度関数を推定

赤線： $N(0.0259, 1.354)$ ← データからの平均と分散による正規分布



青線： $h = 10$ で密度関数を推定

赤線： $N(0.0259, 1.354)$ ← データからの平均と分散による正規分布



青線： $N(0, 1)$ Kernel, $h = 0.17$ で密度関数を推定 ← $h = 1.06\hat{\sigma}n^{-1/5} \approx 0.17$,
 $\hat{\sigma}^2 = 1.354$

赤線： $N(0.0259, 1.354)$ ← データからの平均と分散による正規分布

赤線, 青線のプログラム (Fortran 77)

```
implicit real*8 (a-h,o-z)
dimension y(100000)
open(1,file='r.txt')
read(1,*)
  do 1 i=1,100000
1 read(1,*,end=2) y(i)
2 close(1)
  n=i-1

  a=0.0
  v=0.0
  do 3 i=1,n
  a=a+y(i)/float(n)
3 v=v+y(i)*y(i)/float(n)
  se=sqrt(v-a*a)

  h=1.06*se*( float(n)**(-0.2) )
c h=1
```

```

write(1,4) h,a,se*se
4 format(3f15.10)

do 5 i=-1000,1000
x=float(i)/100.
sum=0.0
do 6 j=1,n
z=(x-y(j))/h
if( abs(z).le.0.5 ) sum=sum+1.
6 continue
f=sum/(float(n)*h)
fn=exp(-0.5*(x-a)*(x-a)/(se*se))/sqrt(2.*3.141592*se*se)
5 write(1,7) x,f,fn
7 format(f7.2,2f15.10)

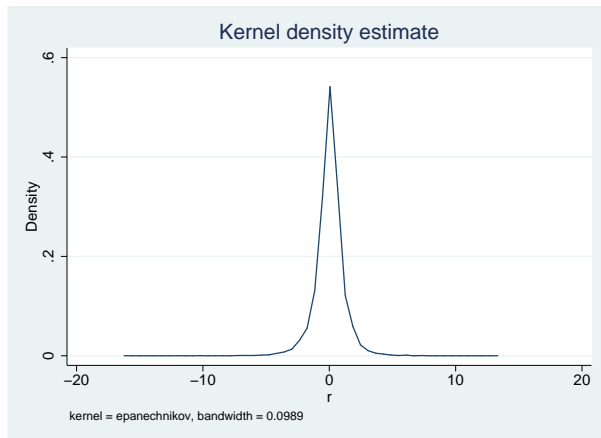
end

```

do 6 j=1,n から 6 continue の部分が、Rectangle Kernel に対応する。

Stata による結果 :

```
. kdensity r
```



デフォルトで (36) の Epanechnikov カーネルが選択され, $h = 0.0989$