ESTIMASI FUNGSI REGRESI MENGGUNAKAN METODE DERET FOURIER

Suparti[1] dan Sudargo[2]

[1)]Jurusan Matematika , FMIPA, Undip

[2)] Jurusan Pend. Matematika, FPMIPA, IKIP PGRI, Semarang

Abstrak

Ada beberapa metode dalam mengestimasi regresi non parametric. Diantaranya metode deret Fourier dan metode kernel.. Salah satu ukuran kebaikan suatu estimator dapat dilihat dari penurunan IMSE-nya (Integrated Mean Square Error). Estimator kernel mempunyai laju penurunan IMSE yang lebih cepat menuju nol dari IMSE deret Fourier. Oleh karena itu metode kernel lebih efektif dari metode deret Fourier.

Kata-kata kunci: regresi nonparametrik, estimator kernel.

PENDAHULUAN

Misalkan ada sebanyak n data pengamatan independen $\left\{(X_i, Y_i)\right\}_{i=1}^{n}$ dan akan dicari hubungan antara X dan Y yang memenuhi model

$$Y_i = f(X_i) + \varepsilon_i,\ i=1,2,...,n, \qquad (1)$$

dengan f fungsi regresi yang tidak diketahui dan ε_i variabel galat random independen dengan mean 0 dan varian σ^2 . Permasalahannya adalah bagaimana mengestimasi fungsi regresi f.

Dilihat dari variabel prediktornya, Hardle (1990) dan Antoniadis *et al* (1994) membedakan dalam dua kasus, yaitu kasus dimana variabel prediktornya sebagai variabel random dan variabel tetap. Jika variabel prediktornya merupakan variabel random maka model rancangannya dinamakan model rancangan stokastik (model random) tetapi jika variabel prediktornya merupakan variabel tetap maka model rancangannya dinamakan model rancangan non stokastik (model tetap). Sedangkan jika dilihat dari asumsi bentuk fungsinya maka ada dua pendekatan yang dapat dilakukan dalam mengestimasi fungsi regresi f, yaitu pendekatan parametrik dan non parametrik. Pendekatan parametrik dilakukan jika asumsi bentuk f diketahui tergantung dari suatu parameter sedangkan pendekatan non parametrik dilakukan jika asumsi bentuk f tak diketahui. Dalam hal ini, diasumsikan bahwa fungsi f termuat dalam kelas fungsi mulus artinya mempunyai turunan yang kontinyu atau dapat diintegralkan secara kuadrat.

Untuk pendekatan non parametrik, beberapa teknik yang dapat digunakan diantaranya adalah teknik pemulus deret orthogonal seperti deret Fourier, dan teknik pemulus kernel . Suatu ukuran kebaikan estimator dari f dapat dilihat dari tingkat kesalahannya. Semakin kecil tingkat kesalahannya semakin baik estimasinya. Salah satu ukuran kesalahan dapat dilihat dari IMSE (integral rata-rata kesalahan kuadrat) atau MSE (Rata-rata kesalahan kuadrat). Pada estimator fungsi regresi dengan deret Fourier diperoleh laju konvergensi IMSE optimal $\propto n^{-1/2}$ (Suparti dan Sudargo, 2005). Dalam makalah ini dibahas tentang penentuan estimator regresi non parametrik dengan metode kernel pada model rancangan tetap, berikut sifat-sifat asimtotis dan contoh simulasinya dengan menggunakan program S-Plus for Windows. Pada model rancangan tetap , variabel prediktor $\{(X_i)\}_{i=1}^{n}$ sering diambil dari distribusi yang sama pada interval [a,b] dan tanpa mengurangi keumuman dapat diambil pada interval [0,1] dan $X_i = i/n$.

METODE PENELITIAN

Penelitian ini merupakan kajian literatur yang kemudian dikembangkan dengan simulasi menggunakan software S-Plus For Windows. Dalam tulisan ini dilakukan pembahasan tentang pemulus kernel dan aplikasinya dalam mengestimasi fungsi regresi non parametrik dan sifat-sifat asimtotisnya serta simulasi pada data yang dibangkitkan dari suatu fungsi yang diberi error (noise). Dari data yang telah diberi error ini kemudian diestimasi fungsinya menggunakan teknik pemulus kernel. Hasil estimasi fungsi selanjutnya dibandingkan dengan fungsi aslinya. Semakin cepat mendekati fungsi aslinya semakin baik estimasinya.

HASIL DAN PEMBAHASAN

Diberikan data pengamatan independen $\{(X_i, Y_i)\}_{i=1}^{n}$ mempunyai model (1). Jika $\{W_{ni}(x)\}$ barisan bobot-bobot positip sehingga $n^{-1}\sum_{i=1}^{n} W_{ni}(x) = 1$, maka estimator kuadrat terkecil dari g adalah

$$\hat{g}(x) = n^{-1}\sum_{i=1}^{n} W_{ni}(x)Y_i . \tag{2}$$

Estimator kernel

Suatu fungsi K(.) disebut fungsi kernel jika K fungsi kontinu, berharga riil, simetris, erbatas dan $\int_{-\infty}^{\infty} K(y)dy = 1$. Jika K suatu kernel dengan sifat

1. $\int_{-\infty}^{\infty} x^j K(x)\,dx = 0$, untuk $j = 1,2,\ldots, r-1$.
2. $\int_{-\infty}^{\infty} x^r K(x)\,dx \neq 0$ atau ∞, maka K disebut kernel order r .

Secara umum estimator regresi kernel dari g adalah estimator kuadrat terkecil $\hat{g}(x) = n^{-1}\sum_{i=1}^{n} W_{ni}(x)Y_i$, dengan fungsi bobot $W_{ni}(x)$ tergantung pada kernel K. Jika densitas X tak diketahui, Hardle (1990) memberikan bobot $W_{ni}(x) = \dfrac{K_h(x - X_i)}{\hat{f}_h(x)}$ dengan $\hat{f}_h(x) = n^{-1}\sum K_h(x - X_i)$ dan $K_h(u) = h^{-1}K(\frac{u}{h})$,

sehingga estimator kernel dari regresi g adalah $\hat{g}_h(x) = \dfrac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i}{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)}$.

Selanjutnya, jika densitas variabel X diketahui, Greblicki (1974) *cit.* Hardle (1990) memberikan bobot $W_{ni}(x) = K_h(x-X_i)/f(x)$, sehingga estimator kernel dari regresi g adalah $\hat{g}_h(x) = \dfrac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i}{f(x)}$. Kemudian dalam model rancangan tetap dari ruang yang sama dengan $\{X_i\}_{i=1,\ldots,n}$ tetap pada [0,1],Priestley dan Chao (1972) cit. Hardle(1990) memberikan bobot $W_{ni}(x) = n(X_i - X_{i-1})K_h(x-X_i)$, $X_0 = 0$

dan $\hat{f}(x) = (n(X_i - X_{i-1}))^{-1}$ untuk $x \in (X_{i-1}, X_i)$, sehingga estimator kernel dari regresi g adalah $\hat{g}_h(x) = (nh)^{-1}\sum_{i=1}^{n} K_h(x - X_i)$.

Lema

Diberikan model rancangan tetap dengan variabel prediktor X satu dimensi dan didefinisikan $c_K = \int_0^1 K^2(u)du$ dan $d_K = \int_0^1 u^2 K(u)du$. Diambil $W_{ni}(x) = n(X_i - X_{i-1})K_h(x - X_i)$ dengan asumsi

1. K mempunyai support [-1,1] dengan K(-1) = K(1) = 0
2. $g \in C^2$
3. $X_i = i/n$, i = 1,2,...,n
4. $var(\varepsilon_i) = \sigma^2$, i = 1,2,...,n
5. $n \to \infty$, $h \to 0$ dan $nh \to \infty$,

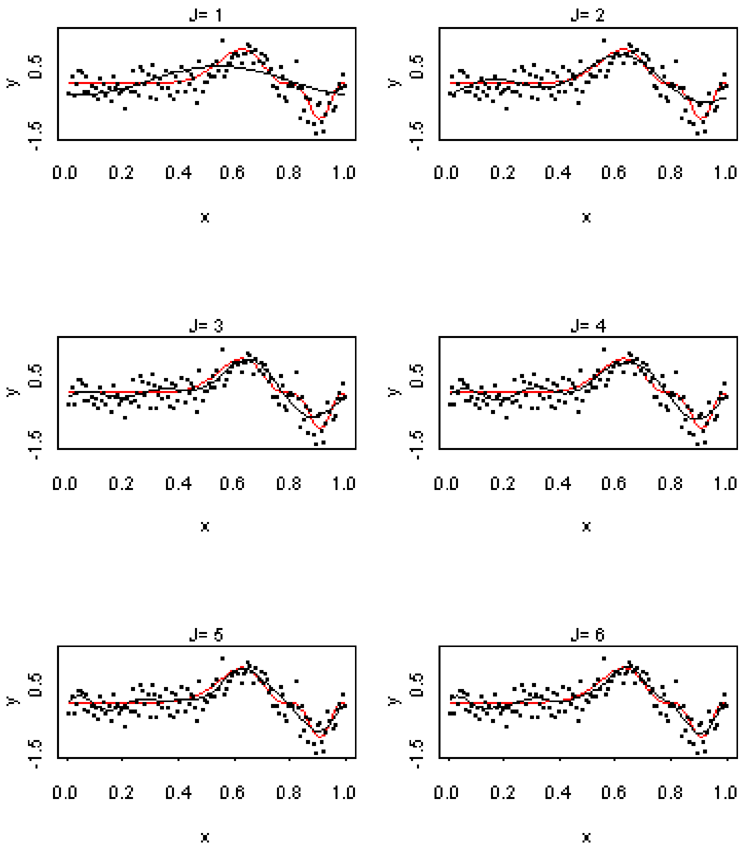
maka MSE($\hat{g}_h(x)$) = $(nh)^{-1}\ \sigma^2 c_K + h^4 d_K^2 [g''(x)]^2/4$ dan

IMSE($\hat{g}_h(x)$) = $\{(nh)^{-1}\ \sigma^2 c_K + 1/4\ [h^4 d_K^2 \int_0^1\ [g''(x)]^2]\ dx$

Dalam estimator kernel, tingkat kemulusan $\hat{g}_h$ ditentukan oleh fungsi kernel K dan lebar jendela h yang disebut parameter pemulus, tetapi pengaruh kernel K tidak sedominan parameter pemulus h. Nilai h kecil memberikan grafik yang kurang mulus sedangkan nilai h besar memberikan grafik yang sangat mulus. Oleh karena itu, perlu dipilih nilai h optimal untuk mendapatkan grafik optimal. Menurut Hardle (1990) parameter pemulus h optimal dapat diperoleh dengan meminimalkan IMSE dari $\hat{g}_h(x)$. Dengan cara ini $h_{opt} \propto n^{-1/5}$ dan IMSE opt $\propto n^{-4/5}$. Jika syarat (2) pada lema di atas diganti dengan $g \in C^r$, maka h optimal $\propto n^{-1/(2r+1)}$ dan IMSE opt $\propto n^{-2r/(2r+1)}$.

Contoh simulasi .

Diberikan suatu data yang dibangkitkan dari sebuah fungsi $g_i = \sin^3\left(2\pi X_i^{\,3}\right) + \varepsilon_i$, $(i = 1,2,...,128)$ dengan X_i berdistribusi uniform pada [0,1] dan ε_i variabel random independen berdistribusi N(0;0,3). Akan dibandingkan fungsi asli dengan fungsi hasil estimasi menggunakan metode Fourier. Data diolah dengan program S-Plus For Windows. Hasilnya sebagai berikut :



Gb.2. Estimasi regresi dengan deret Fourier

Keterangan

………… : diagram pencar (data)

________ : fungsi asli

________ : fungsi estimasi

Dari hasil simulasi dapat dilihat bahwa pada estimasi fungsi regresi dengan metode Fourier semakin kecil J semakin mulus estimasi fungsinya tetapi semakin jauh dari fungsi aslinya. Fungsi estimasi pada level resolusi J = 5 dan J = 6 hampir sama dan pada level J= 6 fungsi estimasi sudah mendekati fungsi aslinya.

KESIMPULAN

Untuk mengestimasi fungsi regresi yang sulit diprediksi bentuknya dapat dilakukan dengan pendekatan non parametrik. Salah satu pendekatan non parametrik adalah dengan metode deret Fourier . Tingkat kemulusan fungsi estimasi dengan deret Fourier ditentukan oleh level resolusi. Semakin kecil level resolusinya semakin mulus fungsi estimasinya dan sebaliknya. Oleh karena itu perlu dipilih level resolusi yang optimal. Level resolusi optimal sebanding dengan $n^{1/2}$.

DAFTAR PUSTAKA

Antoniadis, A., Gregore, G., and Mckeague, W., *Wavelet Methods for Curve Estimation*, JASA, Vol.89, No.428, 1994, 1340-1353.

Eubank, R.L., *Spline Smoothing and Nonparametric Regression,* Marcel Dekker.Inc, New York,1988.

Hardle, W., *Applied Nonparametric Regression,* Cambridge University Press., New York, 1990.

Suparti, *Estimasi Fungsi Mulus dengan Metode Wavelet,* Tesis S-2, UGM,Yogyakarta,1999.

Vetterli,M. and Kovacevic,J.,*Wavelets and Subband Coding*, Prentice Hall PTR, New Jersey,1995.