KOREKSI BIAS ESTIMATOR KERNEL DENGAN BOOTSTRAP

Oleh : Zulfikar
Alumni Magister Statistika ITS Surabaya
Dosen STMIK Bahrul Ulum Jombang

Abstrak

Algoritma resampling merupakan metode praktis dan simpel untuk mengatasi bias pada regresi kernel seperti pada kernel *Nadaraya-Watson* dan *Locally Linear* order dua. Penelitian ini berfokus untuk mendapatkan estimator kernel dengan menetapkan polinomial lokal dan estimasi β digunakan *least square* terbobot. Pada metode yang sama juga akan didapatkan persamaan bias, variansi dan *Mean Square Error* (*MSE)*.

Aplikasi kernel pada data penelitian Canadian Males oleh Murphy dan Welch (1990) menunjukkan bahwa dengan estimasi bootstrap akan menurunkan nilai bias, variansi dan *MSE* serta dengan *improved bootstrap* akan lebih memperkecil nilai–nilai tersebut. Kurva regresi yang dibentuk dari estimasi bootstrap akan membentuk permukaan yang smooth.

Kata Kunci dan Phrasa: Estimasi Nonparametrik, Improved Bootstrap dan Polinomial Lokal.

Pendahuluan

Dalam proses estimasi nonparametrik, estimator kernel menghasilkan bias yang sangat menggganggu inferensi. Bias tersebut memiliki dua tipe, yaitu tipe bias yang ditunjukkan sebagai kurva bias dan tipe kedua yang ditunjukkan sebagai *boundary bias* (Racine, 1990). Kurva bias bisa terlihat pada kernel order ke-*p* setelah diestimasi biasnya meningkat saling berkaitan dengan order berikutnya. Sedangkan *boundary bias* dimana kecepatan konvergensi menjadi lambat dan memiliki bias yang tinggi dekat *boundary* pada interval.

Sejumlah pendekatan reduksi bias yang dikenal yaitu: reduksi bias yang bertumpu pada ekspansi asimtotik, penggunaan kernel order tinggi, metode resampling dan reduksi bias *boundary*. Pendekatan reduksi bias tersebut memiliki kelemahan dimana hanya mereduksi bias tetapi tidak merubah bias. Hardle dan Bowman (1988) menunjukkan bahwa ekspansi asimtotik untuk reduksi bias hanya menggunakan ekspansi dan tidak menggunakan order lebih kecil. Pendekatan reduksi bias terkoreksi dengan bootstrap untuk estimator kernel yang dilakukan Hardle dan Marron (1991) hanya sebatas mereduksi bias dan tidak merubah bias. Bartlett (1963) memperkenalkan reduksi bias kernel order tinggi bermasalah bila syarat densitas (integral bernilai satu, non negatif) tidak terpenuhi pada pemilihan bandwidth dan diskontinuitas pada estimasi. Gasser dan Muller (1979) dan Rice (1984) menunjukkan bahwa estimator kernel Nadaraya-Watson memiliki kecepatan konvergensi yang lambat dan bias yang tinggi pada batas interval. Gasser dan Muller (1979) menunjukkan bahwa pada batas interval mendominasi perilaku asimtotik secara global dan pengaruhnya menjadi buruk untuk kernel order tinggi.

Dalam penelitian ini digunakan algoritma resampling yang diperbaiki (*improved bootstrap*) untuk uji bias regersi kernel menggunakan kernel order dua dan dapat diaplikasikan pada estimator kernel Nadaraya-Watson dan *Locally Linear*. Metode ini memberikan alternatif praktis dan simpel untuk pendekatan mereduksi bias dan dapat merubah bias untuk memperoleh estimasi bias terkoreksi pada kurva regresi yang *smooth*.

Estimator Kernel pada Fungsi Regresi

Asumsi dasar regresi nonparametrik adalah keberadaan fungsi penghalus *m*(.) dari hubungan respon *y* dan prediktor *x*, yaitu :

$$Y_i = m(x_i) + \varepsilon_i, i = 1,2,...,n \quad (1)$$

dimana $\varepsilon_i \sim N(\mu, \sigma^2{}_\varepsilon(x))$ dan $\sigma^2{}_\varepsilon(x) = \sigma^2{}_\varepsilon$ sebuah konstanta.

Fungsi regresi $m(x_i)$ pada model regresi nonparametrik dapat diestimasi dengan pendekatan kernel yang didasarkan pada fungsi densitas kernel, serta penghalus dengan pendekatan kernel ini selanjutnya dikenal sebagai penghalus kernel (*smoothing kernel*)(Hardle (1991)).

Untuk menduga kurva regresi $m(x_i)$ dapat digunakan pendekatan kernel. Dua estimator kernel terkenal adalah estimator Nadaraya-Watson dan *Locally Linear*. Metode yang digunakan adalah dengan penetapan polinomial lokal. Metode ini menetapkan estimasi *least square* terbobot :

$$\beta = (\beta_0, \beta_1, ..., \beta_p)^T \quad (2)$$

dengan koefisien polinomial *p* + 1 pada dejarat *p* dengan meminimumkan :

$$\sum_{i=1}^{n} \left[Y_i - \beta_0 - \beta_1(x_i - x) - ... - \beta_p(x_i - x)^p\right]^2 K_h(x_i - x) \quad (3)$$

dimana $K_h(x_i - x)$ sebagai pembobot kernel. Solusi standar adalah estimator (*p* + 1) x 1:

$$\hat{\beta} = (\mathbf{X}^{\mathbf{T}}{}_{\mathbf{x}}\mathbf{W}_{\mathbf{x}}\mathbf{X}_{\mathbf{x}})^{-1}\mathbf{X}^{\mathbf{T}}{}_{\mathbf{x}}\mathbf{W}_{\mathbf{x}}\mathbf{Y} \quad (4)$$

dalam bentuk matrik singular, dimana **Y** = $(Y_1,...,Y_n)^T$ adalah vektor respon,

$$\mathbf{X_x} = \begin{bmatrix} 1 & x_1 - x & . & . & . & (x_1 - x)^P \\ . & . & . & & & . \\ . & . & & . & & . & . \\ . & . & & & . & . \\ 1 & x_n - x & . & . & . & (x_n - x)^P \end{bmatrix}$$

adalah sebuah bentuk matrik *n x (p + 1)* dan $\mathbf{W_x} = diag\{K_h(x_1 - x),...,K_h(x_n - x)\}$ merupakan matrik diagonal *n x n* pada pembobot. Ketika estimator pada *m*(*x*) adalah koefisien intercept diperoleh :

$$\hat{m}(x; p, h) = \mathbf{e^T_1} \quad (\mathbf{X^T_x W_x X_x})^{-1} \mathbf{X^T_x W_x Y}$$

(5)

dimana $\mathbf{e_1}$ adalah vektor (*p* + *1*) *x 1* bernilai 1 dalam masukan pertama dan nol untuk yang lainnya. Zulfikar (2005) menyatakan bahwa ketika *p* = 0 secara eksplisit akan membentuk formula estimator Nadaraya-Watson dan *p* = 1 akan membentuk estimator *Locally Linear*. Untuk *p* = 0 diperoleh estimasi :

$$\hat{m}_{NW}(x;0,h) = n^{-1}\sum_{i=1}^{n} K_h(x_i - x)Y_i \Big/ \sum_{i=1}^{n} K_h(x_i - x)$$

(6)

sebagai bentuk estimator kernel Nadaraya-Watson. Sedangkan untuk *p* = 1 diperoleh estimasi $\hat{m}_{LL}(x;1,h) =$

$$= n^{-1}\sum Y_i \frac{\left[K_h(x_i - x)\sum(x_i - x)^2 K_h(x_i - x) - (x_i - x)K_h(x_i - x)\sum(x_i - x)K_h(x_i - x)\right]}{\sum\left[K_h(x_i - x)\sum(x_i - x)^2 K_h(x_i - x) - (x_i - x)K_h(x_i - x)\sum(x_i - x)K_h(x_i - x)\right]}$$

(7)

Jika $\hat{s}_j = \sum(x_i - x)^j K_h(x_i - x),\ j = 1,2$ maka diperoleh estimator kernel *Locally Linear*

$$\hat{m}_{LL}(x,1,h) = n^{-1}\frac{\sum K_h(x_i - x)\left[\hat{s}_2 - (x_i - x)\hat{s}_1\right]Y_i}{\sum K_h(x_i - x)\left[\hat{s}_2 - (x_i - x)\hat{s}_1\right]}$$

(8)

Untuk mendapatkan persamaan bias, dan variansi maka dibuat asumsi sebagai berikut :

(i) Fungsi m, m' dan m'' kontinu dalam [0,1].

(ii) Kernel K simetris sekitar nol.

Sehingga diperoleh bias dan variansi sebagai berikut :

Tabel 1. Bias dan Varisnsi Estimator Kernel Order Dua.

Kernel	**Bias**	**Variansi**
Nadaraya -Watson	$Bias\{\hat{m}(x)\} = h_n^2\left(\frac{1}{2}m''(x) + \frac{m'(x)f'(x)}{f(x)}\right)\int z^2K(z)dz$	$Var\{\hat{m}(x)\} = \frac{\sigma^2(x)}{nh_n}\int K^2(z)dz$
Locally Linear	$Bias\{\hat{m}(x)\} = \{\int z^2K(z)dz\}\frac{1}{(2)!}m''h^2$	$Var\{\hat{m}(x)\} = \int K^2(z)dz\frac{\sigma^2(x)}{nh}$

Koresksi Bias dengan Bootstrap

Mendapatkan bias terkoreksi dengan algoritma bootstrap sebagai berikut:

a. Dapatkan $\hat{m}(x_i)$.

b. Dapatkan persamaan residual $\hat{\varepsilon} = y_i - \hat{m}(x_i), i = 1,2,...,n.$

c. Dapatkan rasio $\eta = \frac{\hat{m}(x_i)}{y_i}, i = 1,2,...,n$,.

d. Resampling dengan pengembalian, kemudian transformasi $y_i^* = \eta.y_i$ dan $x_i^* = \eta.x_i$, untuk i = 1,2,…,n.

e. Definisikan sampel bootstrap $\{y_i^*, x_i^*\}, i = 1,2,...,n$ dengan $y_i^* = \hat{m}(x_i) + \varepsilon_i^*, i = 1,2,...,n.$

f. Lakukan aplikasi pada estimator kernel untuk mendapatkan $\hat{m}^*(x_i)$, dan proses ini diulang sebanyak B kali dan menghasilkan $\hat{m}_1^*(x_i),...,\hat{m}_B^*(x_i), i = 1,2,...,n.$

g. Lakukan estimasi bootstrap pada bias yang diberikan oleh $\hat{bias}[\hat{m}(x_i)] = \left[B^{-1}\sum_{j=1}^{B}\hat{m}_j^*(x_i)\right] - \hat{m}(x_i), i = 1,2,...,n.$

h. Lakukan estimasi bias terkoreksi yang diberikan oleh $\hat{m}^C(x_i) = \hat{m}(x_i) - \hat{bias}[\hat{m}(x_i)], i = 1,2,...,n.$

i. Untuk improved bootstrap

1. Lakukan resampling mean dalam model probablititas *p*[i] pada titik *x*[i] untuk setiap x dan probablitias bernilai nol untuk yang lainnya.
2. Dapatkan resampling rasio untuk menghitung estimator rasio data mean (*x*) dan mean (*y*) selanjutnya kembali ke langkah e.

Aplikasi Kernel dengan Koreksi Bias

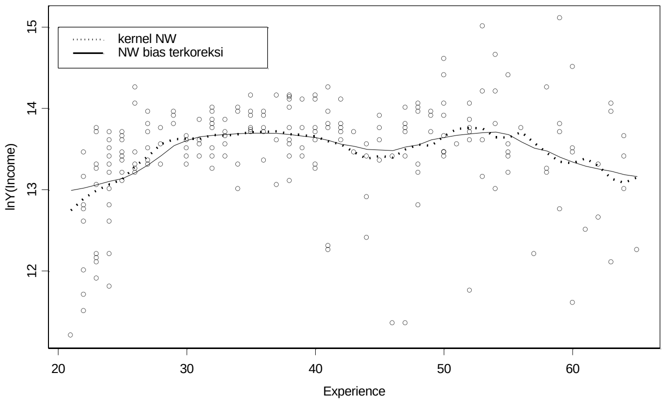
Untuk mengaplikasikan estimator kernel *Nadaraya-Watson* dan *Locally Linear* dengan koreksi bias digunakan data penelitian Canadian Males oleh Murphy dan Weclh (1990) dengan melihat hubungan antara pengalaman kerja (*experience*)(*x*) dan penghasilan (*Income*)(*y*). Bandwidth optimum diperoleh sebesar 7,2 dengan menggunakan *Generalized Cross Validation* (GCV).

Tabel 2. Perbandingan Nilai, Bias, Variance, MSE dan Koreksi Bias antara Dua Estimator

Kernel	**Bias**	**Variance**	**MSE**	**Koreksi Bias**
NW				
Ord 2	$6{,}7.10^{-3}$	$6{,}75.10^{-2}$	$6{,}75.10^{-2}$	
Boot	$8{,}9.10^{-5}$	$9{,}3.10^{-6}$	$9{,}3.10^{-6}$	$9{,}3.10^{-6}$
I.Boot	$9{,}1.10^{-6}$	$8{,}8.10^{-6}$	$8{,}8.10^{-6}$	$8{,}8.10^{-6}$
LL				
Ord 2	$8{,}2.10^{-2}$	$8{,}0.10^{-2}$	$8{,}7.10^{-2}$	
Boot	$7{,}7.10^{-5}$	$9{,}2.10^{-6}$	$9{,}2.10^{-6}$	$9{,}2.10^{-6}$
IBoot	$8{,}5.10^{-6}$	$8{,}5.10^{-6}$	$8{,}5.10^{-6}$	$8{,}5.10^{-6}$

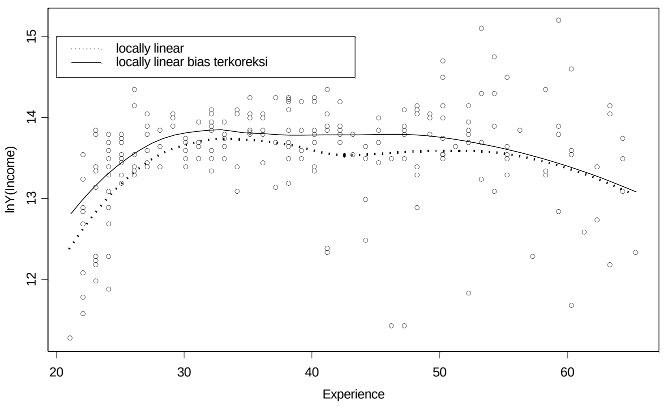
Dari tabel di atas terlihat bahwa nilai bias, variansi dan *MSE* kernel order dua lebih besar dibandingkan dengan bootstrap, dan dengan estimasi improved bootstrap terlihat nilai tersebut semakin kecil. Ini berarti metode estimasi bootstrap sangat efektif mengatasi bias pada estimator kernel.

Untuk melihat sejauh mana metode bootstrap dapat menjadikan kurva regresi lebih smooth ditunjukkan pada gambar berikut:



Gambar 1. Kernel Nadaraya-Watson dengan Bias Terkoreksi Bootstrap

Pada Gambar 1 terlihat bahwa kurva regresi kernel Nadaraya-Watson dengan bias terkoreksi bootstrap lebih *smooth* dibandingkan dengan estimasi kernel order dua. Hal yang sama juga terlihat pada estimator kernel *Locally Linear* seperti terlihat pada gambar berikut:



Gambar 2. Kurva Kernel *Locally Linear* dengan Bias Terkoreksi

Kesimpulan

1. Estimator kernel Nadaraya-Watson dan *Locally Linear* bisa diperoleh dengan menggunakan polinomial lokal, dengan metode yang sama juga didapatkan persamaan bias, variansi dan *MSE* dari kedua estimator.
2. Estimator kernel dengan bias terkoreksi bootstrap akan memperkecil nilai bias, variansi dan *MSE* baik pada kernel Nadaraya-Watson maupun kernel *Locally linear*.

3. Kurva regresi yang dibentuk oleh estimator kernel bias terkoreksi bootstrap memiliki permukaan kurva yang *smooth* dibandingkan kurva regresi yang dibentuk dari estimator kernel *Nadaraya-Watson* maupun *Locally Linear* order dua.

Daftar Pustaka

Bartlett, M. S. (1963), Statistical Estimation of Density Functions, *Sankhya* **25**, 245-254.

Gasser, T and Muller, H. G(1979), *Kernel estimation of Regression Functions, in Smoothing Techniques for Curve estimation*, Springer-Verlag, Berlin, Heildleberg, New York, pp. 23-68.

Hardle, W. (1991), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

Hardle, W. and Bowman, S. (1988), Bootstrapping in Nonparametric Regression Local Adaptive Smoothing and Confidence Bounds, *Journal of the American Statistics Assosiation* **83**, 102-110.

Hardle, W. and Maron, S. (1991), Bootstrapping Simultaneous ErrorBar for Nonparametric regression. *Annals of Statistics* **19**, 778-796.

Murphy, K. M. dan Welch, F. (1990), Empirical Age- Earning Profiles,*Journal of Labour Economics* **8**(2), 202-229.

Racine, J. (1998), *Bias-Corrected Kernel Regression*, Department of Economics, University of South Florida, Tampa, FL., USA 33620.

Rice, J. A. (1984), Boundary Modification for Kernel Regression, *Communication in Statistics* **13**, 893-900.

Zulfikar (2005), *Bias, Variance and MSE Component of Kernel Estimation on Nonparametric Regression*, (Thesis), Department of Statistica Science, ITS, Surabaya.