

ESTIMASI DENSITY KERNEL DENGAN BANDWITH YANG BERBEDA

THE ESTIMATED DENSITY KERNEL WITH A DIFFERENT BANDWITH

Harizahayu

PoliteknikNegeri Medan, Medan [Email: harizahayu@polmed.ac.id]

[§]Corresponding Author

Received Maret 2021; Accepted Juni 2021; Published Juni 2021;

Abstrak

Dalam estimator kernel, parameter penghalus h merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan kurva. Bentuk kurva kernel bergantung pada kurva penduga fungsi kepekatan peluang akan bergantung pada lebar jendela (h) yang merupakan parameter pemulus. Hasil simulasi menunjukkan *bandwith* (lebar jendela) lebih besar atau sama dengan satu telah memperlihatkan bentuk kurva normal untuk kernel quartic, gaussian, epanechnikov, dan kosinus.

Kata Kunci: Estimator Kernel, Bandwith, fungsi kepekatan peluang.

Abstract

In the kernel estimator, the smoothing parameter h is controlling the balance between the fit of the curve to the data and the smoothness of the curve. The shape of the kernel curve depends on the estimator curve. The probability density function will depend on the window width (h) which is the smoothing parameter. The simulation results show that the bandwidth (window width) greater than or equal to one has shown the normal curve shape for quartic, gaussian, epanechnikov, and cosine kernels.

Keywords: Kernel Estimator, Bandwidth, probability density function.

1. Pendahuluan

Pendekatan regresi dibedakan menjadi dua yaitu pendekatan parametrik dan nonparametrik. Pendekatan parametrik merupakan pemodelan regresi yang terikat dengan asumsi-asumsi dalam regresi seperti multikolinieritas, residual normalitas, homoskedasitas residual, dan nonautokorelasi. Sedangkan regresi nonparametrik dilakukan jika bentuk kurva regresinya tidak diketahui dan diasumsikan

smooth yang termuat dalam suatu ruang fungsi tertentu [1]. Salah satu metode pendekatan dalam regresi nonparametrik yang digunakan adalah regresi nonparametrik kernel. Dalam analisis regresi, tidak semua variabel penjelas dapat didekatin dengan pendekatan parametrik, karena tidak adanya informasi yang jelas bagaimana bentuk hubungan variabel independen dengan variabel dependen sehingga harus digunakan

pendekatan pendekatan nonparametrik. Tujuan analisis regresi adalah menentukan hampiran untuk kurva regresi m [3].

Jika diberikan data pengamatan independen $\{X_i, i = 1, 2, \dots, n\}$ untuk menentukan distribusi dari X ekuivalen dengan menentukan densitasnya. Untuk mengestimasi fungsi densitas f dapat dilakukan dengan dua pendekatan yaitu pendekatan parametrik dan nonparametrik. Pendekatan nonparametrik dilakukan jika asumsi bentuk f dapat dilakukan dengan dua pendekatan parametrik dan nonparametrik. Pendekatan nonparametrik dilakukan jika asumsi bentuk f dapat dilakukan jika asumsi bentuk f tidak diketahui. Dalam hal ini diasumsikan bahwa fungsi f termuat dalam kelas fungsi mulus yaitu mempunyai turunan kontinu atau terintegralkan secara kuadrat [1].

Salah satu teknik untuk mengestimasi fungsi mulus adalah teknik pemulus kernel [Hardle]. Metode yang paling sederhana adalah histogram. Teknik pemulus kernel pada estimator densitas merupakan pengembangan dari estimator histogram.

Penggunaan fungsi kernel yang berbeda yaitu fungsi kernel Triangle dan kernel Gaussian dengan bandwidth optimal menghasilkan estimasi kurva regresi yang hampir sama, sehingga dapat diperlihatkan bahwa pemilihan bandwidth lebih penting dibanding dengan pemilihan fungsi kernel [2].

Pemilihan bandwidth optimum dengan menggunakan data training didapatkan nilai 2 yang artinya untuk mengestimasi sebuah nilai maka diperlukan 2 data sebelum dan sesudah

titik estimasi tersebut, metode smoothing kernel dengan bandwidth optimum sebesar 2 dapat menggambarkan pola nilai ekspor kelapa sawit dengan baik, smoothing kernel statis lebih akurat dalam meramalkan data [6].

Berdasarkan Hurdle, dengan menggunakan pendekatan regresi nonparametrik untuk fungsi regresi m diperoleh estimator yang konsisten dengan menggunakan kernel gaussian, bandwidth 0,2; 1; 1,7; dan 2,5.

Estimator kernel lebih baik dibandingkan estimator spline dalam mengestimasi kurva regresi nonparametrik. Dalam estimator kernel, estimasi dilakukan pada setiap titik data sehingga menghasilkan kurva yang lebih mulus, sedangkan spline merupakan potongan-potongan polinomial yang digabungkan bersama pada setiap titik knot [5].

Berdasarkan latar belakang yang diraikan di atas, tujuan penelitian ini model regresi nonparametrik menggunakan estimator kernel, dengan fungsi kernel uniform, quartik, gaussian, Epanechnikov, Cosinus dengan program R 3.5.2. Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

1. Mengestimasi kurva regresi nonparametrik dengan estimator kernel uniform, quartic, gaussian, epanechnikov, cosinus.
2. Membandingkan hasil estimasi antar estimator kernel Quartic, Gaussian, dan Cosinus dengan ukuran bandwidth yang berbeda (0,2;0,8;1,4;4).

2. Landasan Teori

2.1 Estimator Histogram

Metode estimasi densitas secara nonparametrik yang paling populer adalah histogram. Namun sebenarnya, histogram ini bukanlah merupakan alat estimasi densitas yang baik, karena bentuknya yang sangat mudah dipengaruhi oleh jumlah kelas dan lokasi nilai tengahnya, dan juga estimasi densitas yang dihasilkan tidak kontinu pada batas kelas. Diketahui sampel random $\{X_i: i = 1, 2, \dots, n\}$ dari suatu populasi dengan fungsi densitas yang tidak diketahui dengan variabel f . Berdasarkan sampel random ini akan diestimasi fungsi densitasnya. Misalkan daerah nilai x dibagi menjadi disjoint interval dengan panjang $2h$. Peluang observasi yang masuk ke dalam $(X_0 - h, X_0 + h)$ adalah:

$$P\{X \in (X_0 - h, X_0 + h)\} = \int_{X_0 - h}^{X_0 + h} f(x) dx \quad (1)$$

Sehingga diperoleh estimator histogram untuk $f(x)$, yaitu (Hardle):

$$\hat{f}_h(x) = \frac{1}{2nh} \#\{X_i \in (X_0 - h, X_0 + h)\} \quad (2)$$

Untuk semua $X \in (X_0 - h, X_0 + h)$, hal ini berarti bahwa observasi yang masuk ke dalam interval yang tergantung h , yakni $(X_0 - h, X_0 + h)$ memberikan sokongan yang sama terhadap $f_h(x)$. Pemilihan lebar kelas h kecil, histogram memuat banyak batang kecil-kecil sedangkan untuk h besar histogram memuat sedikit batang besar-besar.

2.2 Estimator Densitas Kernel

Estimator density kernel merupakan pengembangan dari estimator histogram. Estimator kernel diperkenalkan oleh Rosenblatt (1956) dan Parzen (1962) sehingga disebut estimator densitas kernel Rosenblatt-Parzen [3].

Secara umum kernel K dengan bandwith h

didefinisikan sebagai berikut [3]:

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right), \text{ untuk } -\infty < x < \infty, h > 0 \quad (3)$$

Yang memenuhi syarat sebagai berikut:

- i. $K(x) \geq 0$, untuk semua x
- ii. $\int_{-\infty}^{\infty} K(x) dx = 1$
- iii. $\int_{-\infty}^{\infty} x^2 K(x) dx = \sigma^2 > 0$
- iv. $\int_{-\infty}^{\infty} x K(x) dx = 0$

maka estimator densitas kernel untuk fungsi densitas $f(x)$ adalah:

$$\hat{f}_h = \frac{1}{nh} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (4)$$

Berdasarkan persamaan (4) dapat dilihat bahwa $\hat{f}_h(x)$ tergantung pada fungsi kernel K dan parameter h . Bentuk bobot kernel ditentukan oleh fungsi kernel K , sedangkan ukuran pada bobotnya ditentukan oleh parameter pemulus h yang disebut bandwith. Peran bandwith seperti lebar interval pada histogram.

Beberapa jenis fungsi kernel antara lain:

1. Kernel Uniform :

$$K(x) = \frac{1}{2} I(|x| \leq 1)$$

2. Kernel Triangle:

$$K(x) = (1 - |x|) I(|x| \leq 1)$$

3. Kernel Evanechnikov: $K(x) =$

$$\frac{3}{4} (1 - x^2) I(|x| \leq 1)$$

4. Kernel Quartik:

$$K(x) = \frac{15}{16} (1 - u^2)^2 I(|x| \leq 1)$$

5. Kernel Triweight:

$$K(x) = \frac{35}{32} (1 - u^2)^3 I(|x| \leq 1)$$

6. Kernel Gaussian:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2\right)$$

7. Kernel Cosinus:

$$K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) I(|x| \leq 1)$$

dengan I adalah indikator. Pada tulisan ini penulis akan memberikan bentuk kernel dengan menampilkannya dalam bentuk gambar. Komputasi dan plot dari kernel ini didasarkan pada algoritma S seperti berikut ini yang mengarah ke grafik kernel.

3. Hasil Dan Pembahasan

3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data primer yang diambil dari laporan harian di Kampung ternak”. Adapun data yang digunakan untuk penelitian ini terdiri dari 8 (delapan) variabel, yaitu kandang satu sampai dengan kandang delapan yang terdiri dari 1116 data.

3.2 Metode Analisa Data

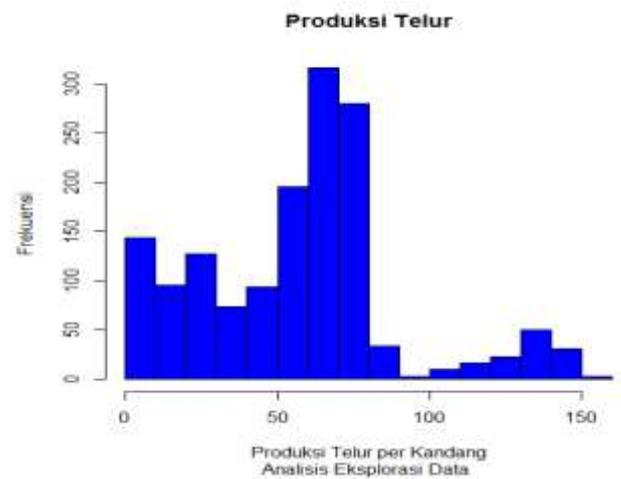
Gambaran umum data yang diolah dengan menggunakan software R 3.5.2 secara rinci dapat dilihat pada Tabel 1.

Tabel 1.Deskripsi Data

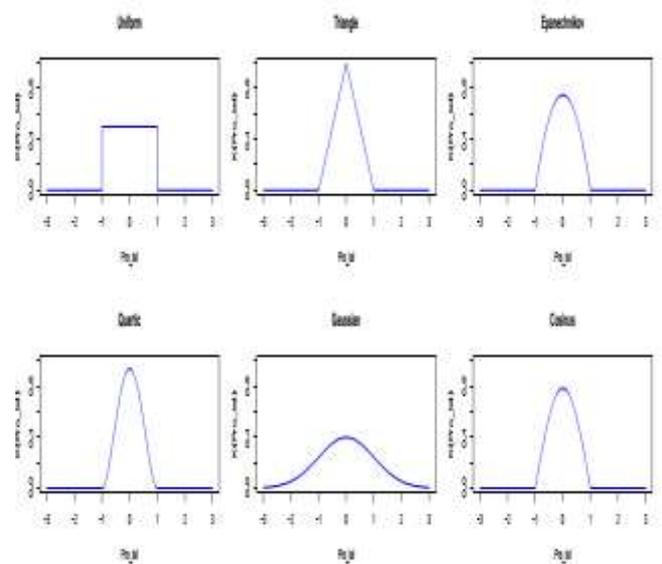
NV	N	Mi	Ma	Mean	(Q ₂)	Std	Var
.	n	x					
K1	186	37	85	66,59	68	11,88	141,22
K2	186	19	151	105,10	126	43,52	1893,90
K3	186	0	82	64,69	65	7,79	60,62
K4	186	25	73	14,42	0	23,56	555,34
K5	186	33	84	60,89	63	13,34	177,84
K6	186	33	86	63,32	68,5	14,03	196,83
K7	186	6	28	19,16	20	4,43	19,57
K8	186	22	81	59,47	64	15,21	231,28

Berdasarkan Tabel 1. Dari setiap kandang yaitu kandang 1 sampai dengan kandang 8 dengan

jumlah data 186 data, sehingga total data dari seluruh kandang berjumlah 1488. Adapun nilai data minimum dari kedelapan kandang tersebut adalah 33 butir yang terdapat pada kandang 5 dan 6, sedangkan data maksimum adalah 115 butir terdapat pada kandang 2. Selanjutnya nilai rata-rata (*mean*) dan kuartil 2 (Q_2) tertinggi terdapat pada kandang 2 yaitu 105,1 dan 126, sedangkan nilai rata-rata dan kuartil 2 (Q_2) terendah terdapat pada kandang 4 yaitu 14,42 dan 0. Tabel 1 juga menampilkan kuartil standar deviasi dan varians yang tertinggi terdapat pada kandang 2 dan yang terendah pada kandang 7.



Gambar 1. Histogram Hasil Produksi Telur

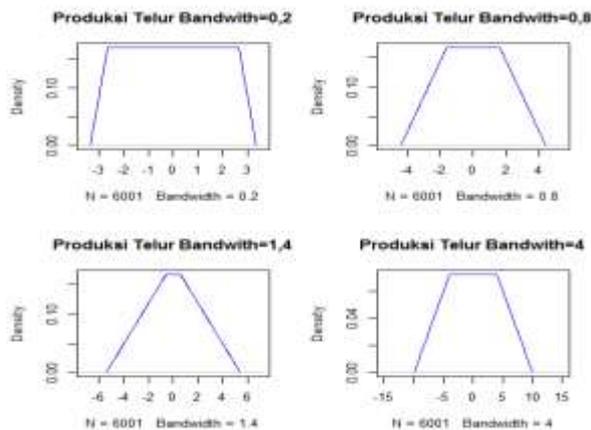


Gambar 2. Kurva Beberapa Fungsi Kernel

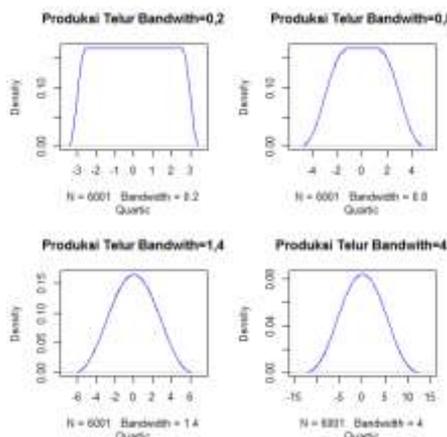
3.3 Pemilihan Bandwith

Bentuk kurva kernel bergantung pada kurva penduga fungsi kepekatan peluang akan bergantung pada lebar jendela (h) yang merupakan parameter pemulus [8]. Estimasi densitas kernel berdasarkan dua parameter, yaitu [2] : Bandwith h dan Fungsi densitas kernel k

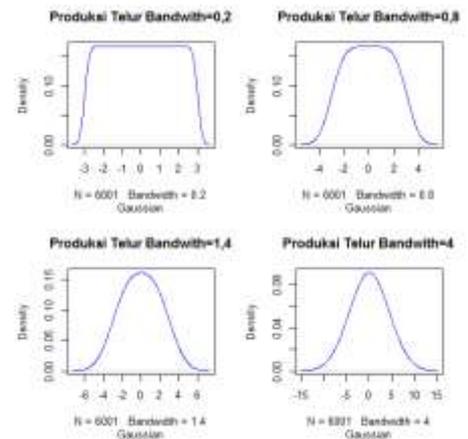
Dalam estimator kernel, parameter penghalus h merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan kurva. Berikut ini ilustrasi fungsi kepekatan peluang dengan fungsi kernel uniform, quartik, gaussian, Epanechnikov, Cosinus menggunakan jendela yang berbeda 0,2; 0,8; 1,4; dan 4:



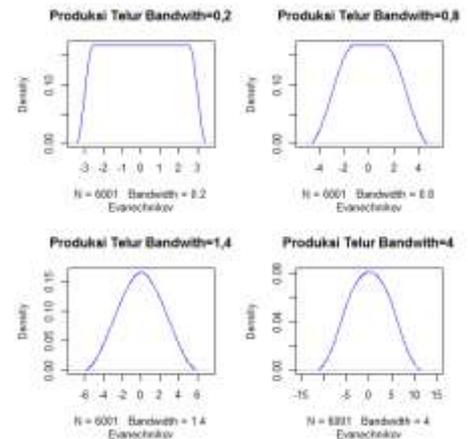
Gambar 3. Kernel Uniform



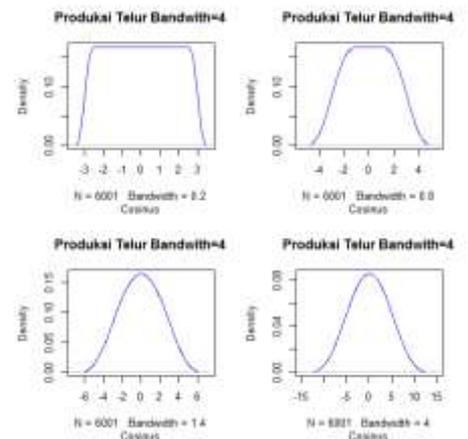
Gambar 3. Kernel Quartic



Gambar 4. Kernel Gaussian



Gambar 5. Kernel Epanechnikov



Gambar 6. Kernel Cosinus

Berdasarkan keenam gambar di atas apabila lebar jendela ($bandwith$) semakin kecil maka kurva fungsi kernel hampir sama untuk kernel quartic, gaussian, epanechnikov, dan cosinus yaitu berbentuk landai di bagian puncaknya. Sedangkan jika bandwith nya semakin besar yaitu lebih besar dari 1 (satu) maka kurva fungsi kernel

cenderung lebih mulus dan berbentuk kurva normal.

4. Kesimpulan Dan Saran

Berdasarkan uraian di atas dapat diambil kesimpulan pemilihan bandwith yang optimal sangat dalam regresi non paramaterik. Semakin besar nilai bandwith maka kurva fungsi kernel cenderung lebih mulus dan mendekati bentuk kurva normal. Berdasarkan simulasi yang telah dicoba oleh penulis, *bandwith* (lebar jendela) lebih besar atau sama dengan satu telah memperlihatkan bentuk kurva normal untuk kernel quartic, gaussian, epanechnikov, dan kosinus. Adapun saran dari penulis adalah perlu dilanjutkan pembahasan fungsi kernel dengan menggunakan kernel Triweight.

5. UcapanTerimaKasih

Pada penelitian ini penulis mengucapkan terimakasih banyak ke pada rekan-rekan yang telah membantu penulis dalam perbaikan tulisan dan saran revisi *source code* pemograman R sehingga penelitian ini dapat selesai tepat waktu.

Daftar Pustaka

- [1] Eubank, R. 1998. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker New York.
- [2] Hayati, Laila. 2010. *Pendekatan Model Kernel Densitas Mulus*. J. Pijar MIPA, Vol. V No.2, September : 81 - 85ISSN 1907-1744
- [3] Hardle, W. (1990). *Smoothing Techniquewith Implementation in S*.

- Springer Verlag,New York.
- [4] Hastie, T.J. and R.J. Tibshirani. 1990 *Generalized Additive Models*. Chapmanand Hall. New York. London.
- [5] I. K. G. Sukarsa and I. G. A. M. Srinadi, *Estimator Kernel dalam Model Regresi Nonparametrik*. *J. Mat.*, vol. 2, no. 1, pp. 19–30, 2012.
- [6] A. F. Arifin and E. T. Astuti. *Pemodelan Nilai Ekspor Kelapa Sawit Di Indonesia Menggunakan Smoothing Kernel (Modeling the Value of Palm Oil Exports in Indonesia Using Smoothing Kernel)*. *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 1094–1104, 2021, doi: 10.34123/semnasoffstat.v2020i1.597.
- [7] I. K. G. Sukarsa and I. G. A. M. Srinadi *et al.*, *Perbandingan Estimator Kernel Dan Estimator Spline Dalam*. KNM XVI, December, 2015.
- [8] Sartono, bagus dan Kholifatunnisa, amalia. <https://rpubs.com/bagusco/kerneldensity>. Diakses tanggal 28 Maret 2020.