

ESTIMASI DENSITAS KERNEL *ADJUSTED*: STUDI SIMULASI**Novita Eka Chandra**Universitas Islam Darul ‘Ulum Lamongan
novitaekachandra@gmail.com**Masriani Mahyuddin**

Universitas Islam Darul ‘Ulum Lamongan

Siti Alfiatur Rohmaniah

Universitas Islam Darul ‘Ulum Lamongan

ABSTRACT. *Kernel adjusted density estimation is a modification of the regular kernel density estimation. The modification is applied to a kernel function. This kernel function is derived from the location-scale transformation. Simulation study shows that this estimation have better results than the regular estimation because it has smaller MSE value. In addition, if normal kernel is used as a kernel function then the curve estimation will be smoother than other kernel function such as uniform kernel and Epachenikov kernel.*

Keywords: *estimation, density, kernel, location-scale transformation.*

ABSTRAK. Estimasi densitas kernel *adjusted* merupakan modifikasi dari estimasi densitas kernel biasa. Modifikasi dari estimasi ini dilakukan pada fungsi kernelnya. Fungsi kernel yang digunakan berasal dari transformasi skala-lokasi. Berdasarkan simulasi, estimasi ini memberikan hasil yang lebih baik dari estimasi densitas kernel yang biasa, karena mempunyai nilai MSE yang lebih kecil. Selain itu, jika kernel yang digunakan kernel normal maka hasil kurva estimasinya lebih halus dibandingkan kernel lainnya seperti kernel uniform dan kernel Epachenikov.

Kata Kunci: estimasi, densitas, kernel, transformasi skala-lokasi.

1. PENDAHULUAN

Analisis data bertujuan untuk memperoleh informasi dari suatu data, seperti pola sebaran data, maupun penyajian data supaya mudah dipahami. Pola sebaran data dapat diperiksa melalui bentuk fungsi densitasnya. Dalam prakteknya, bentuk suatu fungsi densitas biasanya belum diketahui. Untuk mengestimasi fungsi densitas tersebut, digunakan pendekatan nonparametrik, yaitu dengan fungsi kernel (Wand dan Jones, 1995).

Menurut Hardle (1991), pemilihan *bandwidth* (h) dan kernel (K) sangat penting dalam estimasi densitas kernel. Akan tetapi, pemilihan kernel tidak begitu berpengaruh dalam estimasi. Oleh sebab itu, Srihera dan Stute (2011) memodifikasi estimasi tersebut pada bentuk kernelnya dengan menggunakan transformasi skala-lokasi, dan selanjutnya disebut estimasi densitas kernel *adjusted*. Suatu estimasi yang baik dapat dilihat berdasarkan nilai *Mean Square Error* (MSE) yang minimum. Pada tulisan ini akan dibandingkan estimasi densitas kernel dengan estimasi densitas kernel *adjusted* berdasarkan nilai MSE.

2. ESTIMASI DENSITAS KERNEL

Diberikan data pengamatan dari variabel random $X_i, i=1,2,\dots,n$ berdistribusi independen identik (*i.i.d*) dengan densitas f . Estimasi densitas kernel tergantung pada parameter *bandwidth* (h) dan kernel (K). Semakin besar nilai *bandwidth*, maka semakin halus kurva estimasi yang dihasilkan. Secara umum, definisi kernel K adalah

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right).$$

Definisi 2.1 (Hardle, 1991) *Estimator densitas kernel untuk fungsi densitas f didefinisikan*

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i). \quad (1)$$

Berikut sifat tak bias asimtotik dari $\hat{f}_h(x)$. Karena $X_i, i=1,2,\dots,n$ *i.i.d*, maka

$$\begin{aligned} E[\hat{f}_h(x)] &= E\left[\frac{1}{n} \sum_{i=1}^n K_h(x - X_i)\right] \\ &= E[K_h(x - u)] \\ &= \int K(s)f(x + sh)ds. \end{aligned} \quad (2)$$

Dengan menggunakan perluasan deret Taylor untuk $f(x + sh)$, persamaan (2) menjadi

$$E[\hat{f}_h(x)] = f(x) + \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2) \quad (3)$$

dengan $\mu_2(K) = \int s^2 K(s) ds$. Untuk h konvergen ke nol, maka $E[\hat{f}_h(x)]$ konvergen ke $f(x)$. Artinya, estimator densitas kernel tak bias asimtotik (Bain dan Engelhardt, 1992).

Selanjutnya, dari persamaan (3), diperoleh bias dari $\hat{f}_h(x)$ sebagai berikut

$$\begin{aligned} \text{Bias}[\hat{f}_h(x)] &= E[\hat{f}_h(x)] - f(x) \\ &= \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2), h \rightarrow 0. \end{aligned}$$

Dari hasil tersebut terlihat bahwa bias merupakan fungsi kuadrat dalam h . Dengan demikian, untuk mengurangi bias dipilih nilai h yang kecil. Karena $X_i, i=1, 2, \dots, n$ *i.i.d*, maka

$$\begin{aligned} \text{Var}[\hat{f}_h(x)] &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[K_h(x - X_i)] \\ &= \frac{1}{n} \text{Var}[K_h(x - u)] \\ &= \frac{1}{n} \left[E(K_h^2(x - u)) - (E(K_h(x - u)))^2 \right] \\ &= \frac{1}{nh} \|K\|_2^2 f(x) + o\left(\frac{1}{nh}\right), h \rightarrow 0, nh \rightarrow \infty \end{aligned}$$

dengan $\|K\|_2^2 = \int K^2(s) ds$. Terlihat bahwa bila nh naik, maka variansi berkurang. Hal ini kontradiksi dengan bias. Untuk mengatasi hal tersebut digunakan MSE yang merupakan kombinasi dari variansi dan bias kuadrat dari $\hat{f}_h(x)$,

$$\text{MSE}[\hat{f}_h(x)] = \frac{1}{nh} f(x) \|K\|_2^2 + \frac{h^4}{4} (f''(x) \mu_2(K))^2 + o\left(\frac{1}{nh}\right) + o(h^4), h \rightarrow 0, nh \rightarrow \infty.$$

Nilai MSE tersebut akan konvergen ke nol jika $h \rightarrow 0, nh \rightarrow \infty$. Hal ini berarti estimator densitas kernel konsisten, yaitu $\hat{f}_h(x) \xrightarrow{p} f(x)$ (Bain dan Engelhardt, 1992).

Selanjutnya, nilai *bandwidth* optimal, h_{op} , diperoleh dari turunan pertama $MSE[\hat{f}_h(x)]$ terhadap h sama dengan nol, yaitu

$$h_{op} = \left(\frac{f(x) \|K\|_2^2}{n(f''(x)\mu_2(K))^2} \right)^{1/5}.$$

3. ESTIMASI DENSITAS KERNEL *ADJUSTED*

Pada tahun 2011, Srihera dan Stute memodifikasi bentuk kernel menggunakan transformasi skala-lokasi yang berkaitan dengan estimator densitas kernel (Chandra, dkk, 2015). Bentuk kernelnya menjadi

$$K^*(x) = \sigma \hat{f}_h(\sigma x + \theta) \quad (4)$$

dengan $\theta \in \mathbb{R}$ parameter lokasi dan $\sigma > 0$ parameter skala. Dengan menggunakan bentuk kernel baru Persamaan 4, maka diperoleh estimator densitas kernel *adjusted* berikut

$$\hat{f}_{ha}(x) = \frac{\sigma}{n^2 h^2} \sum_{j=1}^n \sum_{i=1}^n K \left(\frac{\sigma x - \sigma X_i - \theta h - h X_j}{h^2} \right). \quad (5)$$

Teorema 3. 1 (Srihera dan Stute, 2011) Diberikan bahwa K merupakan fungsi yang simetris dan f terdiferensial dua kali secara kontinu pada x , serta $E(X^2) < \infty$. Selanjutnya, jika $n \rightarrow \infty$ dan $h \rightarrow 0$, sedemikian sehingga $nh \rightarrow \infty$, maka untuk $\theta \in \mathbb{R}$ dan $\sigma > 0$,

$$\text{Bias}[\hat{f}_{ha}(x)] = \frac{f'(x)h}{\sigma^2} \int f(y)(\theta - y)dy + \frac{f''(x)h^2}{2\sigma^2} \int f(y)(\theta - y)^2 dy + o(h^2)$$

dan

$$nh\text{Var}[\hat{f}_{ha}(x)] = \sigma f(x) \int f^2(y)dy + o(1).$$

Bukti: Dengan menggunakan persamaan (5), diperoleh

$$E[\hat{f}_{ha}(x)] = \frac{\sigma}{h^2} E \left[K \left(\frac{\sigma x - \sigma z + h\theta - hy}{h^2} \right) \right]$$

$$= \iint K(-u) f(y) f\left(x + \frac{h\theta - hy + uh^2}{\sigma}\right) dy du.$$

Karena K simetris, maka dengan menggunakan perluasan deret Taylor, diperoleh

$$E\left[\hat{f}_{ha}(x)\right] = f(x) + \frac{hf'(x)}{\sigma} \int f(y)(\theta - y) dy + \frac{h^2 f''(x)}{2\sigma^2} \int f(y)(\theta - y)^2 dy + o(h^2).$$

Estimator densitas kernel *adjusted* tak bias asimtotik apabila h konvergen ke nol. Untuk menentukan bias dari estimator densitas kernel *adjusted*, diperoleh dengan selisih antara harga harapannya dengan estimator tersebut. Dengan demikian,

$$\text{Bias}\left[\hat{f}_{ha}(x)\right] = \frac{hf'(x)}{\sigma} \int f(y)(\theta - y) dy + \frac{h^2 f''(x)}{2\sigma^2} \int f(y)(\theta - y)^2 dy + o(h^2).$$

Selanjutnya, nilai variansinya diperoleh dengan cara berikut.

$$\text{Var}\left[\hat{f}_{ha}(x)\right] = \frac{\sigma^2}{n^3 h^4} \text{Var}\left[\sum_{i=1}^n \left(\frac{\sigma x - \sigma X_i + \theta h - hy}{h^2}\right)\right].$$

Dengan menggunakan definisi variansi, $\text{Var}(X) = E(X^2) - (E(X))^2$ (Bain dan Engelhardt, 1991), diperoleh

$$\begin{aligned} \text{Var}\left[\hat{f}_{ha}(x)\right] &= \frac{\sigma}{nh} \iiint K(u)K(v) f(y_1) f(y_1 + uh - hv) f\left(x + \frac{\theta h - hy_1 - h^2 u}{\sigma}\right) dy_1 dv du \\ &\quad - \frac{1}{n} (f(x) + o(h))^2 \\ &= \frac{\sigma f(x)}{nh} \int f^2(y) dy + o\left(\frac{1}{nh}\right), h \rightarrow 0. \end{aligned} \quad (q.e.d)$$

Selanjutnya, berdasarkan Teorema 3.1, nilai MSE dari estimator densitas kernel *adjusted* diperoleh sebagai berikut.

$$\begin{aligned} \text{MSE}\left[\hat{f}_{ha}(x)\right] &= \frac{\sigma f(x)}{nh} \int f^2(y) dy + \left[\frac{hf'(x)}{\sigma} \int f(y)(\theta - y) dy + \frac{h^2 f''(x)}{2\sigma^2} \int f(y)(\theta - y)^2 dy \right]^2 \\ &\quad + o\left(\frac{1}{nh}\right) + o(h^4), h \rightarrow 0, nh \rightarrow \infty. \end{aligned}$$

Estimator densitas kernel *adjusted* konsisten untuk $h \rightarrow 0, nh \rightarrow \infty$, karena MSEnya konvergen ke nol. Dengan demikian, dapat ditulis $\hat{f}_{ha}(x) \xrightarrow{p} f(x)$

(Bain dan Engelhardt, 1991). Dengan mengambil nilai $\theta = E(X)$, nilai MSEnya menjadi

$$MSE[\hat{f}_{ha}(x)] \approx \frac{\sigma f(x)}{nh} \int f^2(y)dy + \left(\frac{h^2 f''(x) \text{Var}(X)}{2\sigma^2} \right)^2. \quad (6)$$

Selanjutnya, diambil $a = \frac{h}{\sigma}$, akibatnya persamaan (6) menjadi

$$MSE[\hat{f}_{ha}(x)] \approx \frac{f(x)}{na} \int f^2(y)dy + \left(\frac{a^2 f''(x) \text{Var}(X)}{2} \right)^2. \quad (7)$$

Dengan melakukan turunan pertama pada persamaan (7) terhadap a sama dengan nol, diperoleh nilai *bandwidth* optimal dan σ optimal berikut

$$h_{opt} = n^{-1/5}$$

$$\sigma_{opt} = \left(\frac{(f''(x) \text{Var}(X))^2}{f(x) \int f^2(y)dy} \right)^{-1/5}.$$

4. SIMULASI

Simulasi dilakukan untuk melihat kinerja suatu estimator, yaitu dengan membandingkan nilai MSE antara estimator densitas kernel dengan estimator densitas kernel *adjusted*. Estimator yang memiliki nilai MSE lebih minimum merupakan estimator terbaik. Dalam simulasi ini, penulis menggunakan data acak yang berdistribusi Normal, dan tiga jenis kernel yaitu kernel Uniform, Epachnenikov, serta Normal (Gaussian). Selain itu, digunakan pula nilai $h_{optimal} = n^{-1/5}$, mean dari data adalah θ , dan standar deviasi dari data adalah σ . Hasil simulasi ditampilkan pada Tabel 1.

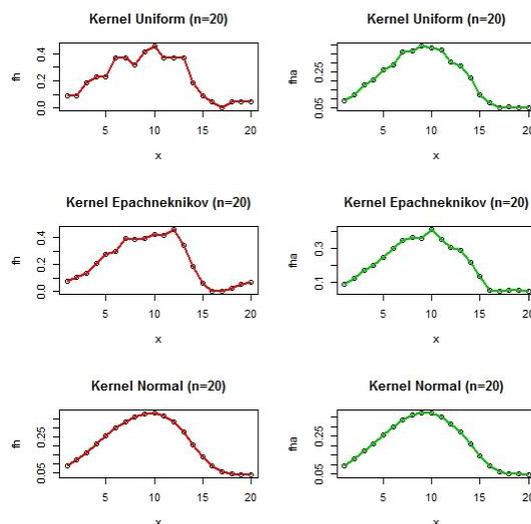
Tabel 1. Perbandingan Nilai MSE dari $\hat{f}_h(x)$ dan $\hat{f}_{ha}(x)$ dengan Kernel Berbeda

Jumlah Data	Kernel	$MSE[\hat{f}_h(x)]$	$MSE[\hat{f}_{ha}(x)]$
$n = 20$	Uniform	0,9526722	0,9379001
	Epachnenikov	0,9569618	0,9429582
	Normal	0,9421523	0,9365913
$n = 50$	Uniform	1,192391	1,177564
	Epachnenikov	1,192018	1,176661

	Normal	1,179788	1,175789
$n = 100$	Uniform	0,7387496	0,7373932
	Epachnenikov	0,7409351	0,7371
	Normal	0,7372438	0,7367734

Dari Tabel 1 terlihat bahwa nilai MSE untuk estimator densitas kernel *adjusted* lebih kecil dibandingkan nilai estimator densitas kernel. Lebih lanjut, kernel normal menghasilkan nilai MSE lebih kecil dibandingkan kernel uniform dan Epachnenikov.

Selain itu, berikut ini ditampilkan kurva estimasi densitas dari masing-masing kernel untuk $n = 20$ (kiri: estimasi densitas kernel dan kanan: estimasi densitas kernel *adjusted*).



Gambar 1. Kurva Estimasi Densitas untuk $n = 20$

Dari Gambar 1 terlihat bahwa estimator densitas kernel *adjusted* memberikan kurva estimasi yang lebih halus dibandingkan kurva estimasi densitas kernel. Selain itu, dengan kernel Normal, kurva estimasi yang dihasilkan lebih halus dibandingkan dengan menggunakan kernel lainnya.

5. KESIMPULAN

Estimator densitas kernel *adjusted* merupakan estimator yang diperoleh dengan memodifikasi estimator densitas kernel biasa. Estimator ini dapat direpresentasikan pada persamaan (5). Dari hasil simulasi untuk sampel sebanyak 20, 50, dan 100, diperoleh

bahwa estimator densitas kernel *adjusted* lebih baik dari estimator densitas kernel biasa karena mempunyai nilai MSE yang lebih kecil. Hasil simulasi juga menunjukkan bahwa penggunaan kernel normal dapat memberikan hasil yang lebih baik dibanding kernel lainnya seperti kernel uniform dan kernel Epachenikov.

DAFTAR PUSTAKA

- Bain, L. J. dan Englehardt, M., *Introduction to Probability and Mathematical Statistics*, Duxbury Press, 1992.
- Chandra, N. E., Haryatmi, S. dan Zulaela, *Regresi Nonparametrik Kernel Adjusted*, Jurnal Ilmiah Matematika dan Pendidikan Matematika, **7**(1) (2015), 1 – 10.
- Hardle, W., *Smoothing Techniques with Implementation in S*, Springer-Verlage, 1991.
- Srihera, R. dan Stute, W., *Kernel Adjusted Density Estimation*, Statistics and Probability Letters, **81** (2011), 571 – 579.
- Wand, M. P. dan Jones, M. C., *Kernel Smoothing*, Chapman and Hall, 1995.