Fraktal: Jurnal Matematika dan Pendidikan Matematika
Volume 2, No. 1, Mei 2021, Hal. 43-53
Available online at https://ejurnal.undana.ac.id/fraktal
Doi: https://doi.org/10.35508/fractal.v2i1.4036

PERBANDINGAN ESTIMATOR HISTOGRAM DAN ESTIMATOR KERNEL

Ch. Krisnandari Ekowati
Pendidikan Matematika, FKIP Universitas Nusa Cendana, Kupang
Email : ekowatichristine@gmail.com

Diterima (01April 2021); Revisi (15 April 2021); Diterbitkan (21 Mei 2021)

Abstrak

Salah satu hal penting dalam analisis statistik adalah prosedur estimasi suatu fungsi padat peluang yang biasa disebut fungsi densitas. Ada dua metode pendekatan yang biasanya digunakan, yaitu pendekatan parameter yang terkait dengan asumsi distribusi tertentu dan metode estimasi densitas secara non parametrik. Metode non parametrik yang sering kita jumpai adalah metode histogram. Beberapa kelemahan metode histogram menjadi acuan untuk dikembangkannya metode yang lain yaitu metode kernel, dimana estimator densitas kernel ini mempunyai parameter yang perlu diestimasi yaitu bandwidth h. Yang menjadi rumusan masalah dalam kajian pustaka ini adalah bagaimana memilih bandwidth dari estimator densitas kernel pada suatu fungsi densitas f di R dan membandingkannya dengan estimator histogram. Kesimpulan yang dapat diambil antara lain: (1) Estimasi densitas histogram adalah $f_h(x) = (nh)\sum_{i=1}^{n}\sum I(x_i \in B_j)$ dengan $B_j = \left[xo - \frac{h}{2}.xo + \frac{h}{2}\right]$, (2) Estimator densitas kernel adalah $f_h(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - x_1) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-x_1}{h}\right)$, (3) Estimator kernel mengatasi kelemahan- kelemahan dari estimator histogram, (4) Tingkat konvergensi dari estimator kernel lebih baik dari pada histogram, (5) Pemilihan bandwidth dengan metode unbiased cross validation (least cross validation) secara asymtotik menghasilkan bandwidth yang optimum

Kata Kunci : Estimator, Histogram, Kernel

Abstract

One of the important things in statistical analysis is the quarterly report of the probability density function which is called the density function. There are two approach methods that are usually used, namely the parameter approach associated with certain distribution assumptions and the non-parametric density calculation method. The non-parametric method that we often encounter is the histogram method. Some of the weaknesses of the histogram method become a reference for developing other methods, namely the kernel method, where the kernel density estimator has a parameter that needs to be estimated, namely the bandwidth h. The problem formulation in this literature review is how to choose the bandwidth from the kernel density estimator in a density function f in R and compare it with the histogram estimator. The conclusions that can be drawn include: (1) The estimated density of the histogram is $f_h(x) = (nh)\sum_{i=1}^{n}\sum I(x_i \in B_j)$ dengan $B_j = \left[xo - \frac{h}{2}.xo + \frac{h}{2}\right]$, (2) The kernel density estimator is $f_h(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - x_1) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-x_1}{h}\right)$, (3) The kernel estimator overcomes the weaknesses of the histogram estimator, (4) The convergence rate of the kernel estimator is better than the histogram, (5) The selection of bandwidth with

Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

the asymptotically unbiased cross validation (least cross validation) method produces the optimum bandwidth.
Keywords : Estimator, Histogram, Kernel

PENDAHULUAN

Estimasi fungsi densitas memegang peranan yang cukup penting dalam analisa statistik. Umumnya estimasi densitas dilakukan dengan menggunakan pendekatan parameter. Estimasi dapat pula dilakukan berdasarkan pendekatan yang tidak terikat dengan asumsi berdistribusi tertentu. Metode dengan pendekatan seperti ini dinamakan metode estimasi densitas secara non parametric (Ekowati,1998). Metode estimasi densitas secara non parametrik yang paling populer adalah histogram. Saat ini telah berkembang suatu metode yang memperbaiki kelemahan histogram yaitu metode kernel. Estimator densitas kernel ini mempunyai parameter yang perlu diestimasi yaitu bandwidth h (Bartlett,M.S, 1963).

Fungsi kernel (k) menentukan bentuk sumbangan dari setiap pengamatan. Bandwidth h merupakan parameter pemulus (smoothing). Semakin besar h akan makin mulus estimasi densitas yang diperoleh. Karena bandwidth sangat berpengaruh terhadap pola estimator densitas, maka besarnya harus ditentukan dengan cermat. Bandwidth yang diinginkan adalah Bandwidth optimum yang akan menghasilkan estimator densitas yang mewakili densitas sebenarnya. Suatu ukuran yang menunjukan seberapa baik estimator densitas yang didapatkan adalah Mean Integrated Square Error (MISE) (Bartlett, 1963).

Berdasarkan latar belakang di atas, maka masalah yang akan dirumuskan adalah bagaimana memilih bandwidth dari estimator densitas kernel pada suatu fungsi densitas f di R dan membandingkannya dengan estimator histogram. Tujuannya adalah untuk membahas metode-metode estimasi fungsi densitas secara non parametrik, yaitu:

1. Metode Histogram
2. Metode Kernel

dan selanjutnya membandingkan kedua metode di atas. Manfaat dari penulisan karya ilmiah ini adalah agar dapat melakukan estimasi fungsi densitas yang tidak mempunyai distribusi tertentu sebagai asumsi, agar dapat memilih estimator yang terbaik dari metode estimasi yang ada, sebagai tambahan pengetahuan bagi pembaca khususnya para penggemar statistika.

Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

METODE

Metode penulisan artikel ini adalah studi pustaka dimana pada pembahasan akan diuraikan satu per satu setiap metode estimasi dan kemudian membandingkannya.

HASIL DAN PEMBAHASAN

Metode Histogram

Diketahui sampel random $X_1, \dots, X_n$ dari suatu populasi dengan fungsi densitas tak diketahui f. Berdasarkan sampel random ini akan diestimasi fungsi densitasnya. Metode yang paling sederhana adalah metode histogram. Adapun estimator histogram untuk $f(x)$, yaitu (Hall,1982) :

$$f_h(x) = \frac{1}{nh} \# \left\{ x_1 \in \left[xo - \frac{h}{2}, xo \frac{h}{2} \right\} \right.$$

Untuk semua $x \in \left[xo - \frac{h}{2}, xo + \frac{h}{2} \right]$ dengan xo pusat bin dan h lebar bin.

Metode Kernel

Ide estimator densitas kernel dari fungsi f yang tidak diketahui adalah (Hall P,1982) :

$$f_h(x) = \frac{1}{n} \sum_{i=1}^{n} k_h\,(x - x_i) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h} \right) \dots\dots (2.1)$$

Dengan merupakan fungsi kernel.

Fungsi kernel K yang umum dipakai adalah fungsi densitas dan biasanya dilengkapi dengan asumsi-asumsi tertentu. Di sini diasumskan bahwa fungsi kernel K fungsi densitas kontinu, terbatas, simetris terhadap pusat interval dan mempunyai support yang finit (Berger, 1980).

Berikut ini analisis statistik yang akan memberikan pedoman bagaimana memilih h dan k untuk memperoleh estimator yang baik sesuai kriteria kebaikan yang tepat (Berliner,L.M, 1993).

Teorema 1

Jika $\hat{f}_h(x)$ diberikan oleh (2.1), maka untuk $n \ \to \infty$,

$E\left[\hat{f}_h(x)\right] \to f(k), h \to 0$

Bukti :

$$E\left[\hat{f}_h(x)\right] = \frac{1}{2} \sum_{i=1}^{n} E[K_h(x - x_i)]$$

$$= \ E[K_h(x - x_i)]$$

Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

$$= \int_{-10}^{10} K_h(x-u)f(u)du$$

$\int_{-\alpha}^{10} K(s)f(x+sh)ds$ (dengan substitusi $u = x + sh$)

Untuk $h \to 0, n \to \alpha$ diperoleh :

$$E[\hat{f}_h(x)] \to \int_{-10}^{10} k(s)f(x)ds = f(x)$$

Teorema 2

Jika $\hat{f}_h(x)$ diberikan oleh (2.1), maka

Bias $[\hat{f}_h(x)] \to \int_{-10}^{10} k(s)f(x)ds = f(x)$

$$\alpha(k) = \int_{-10}^{10} S^2 k(s)ds$$

Bukti :

Dari teorema 2 diperoleh :

Bias $[\hat{f}_h(x)] = \int_{-10}^{10} K(s)f(x+sh)ds - f(x)$

$$= \int_{-10}^{10} K(s)\left(f(x) + \frac{hs}{1!}f'(x) + \frac{h^2s^2}{2!}f''(x) + 0(h^2)\right)ds - f(x)$$

$$= f(x)\int_{-10}^{10} k(s)ds + hf'(x)\int_{-10}^{10} SK(s)ds + \frac{h^2}{2}f''(x)\int_{-10}^{10} S^2K(s)ds + o(h^2) - f(x)$$

$$= f(x) + \frac{h^2}{2}f''(x)\alpha(k) + o(h^2) - f(x)$$

$$= \frac{h^2}{2}f''(x)\alpha(k) + o(h^2)$$

$$= \frac{h^2}{2}f''(x)\alpha(k) + o(h^2)$$

Teorema 3

Jika $\hat{f}_h(x)$ diberikan oleh (2.1), maka

var $[\hat{f}_h(x)] = (nh)^{-1}f(x)\|k\|_2^2 + o(nh)^{-1}), nh \to \infty$

dengan $\|k\|_2^2 = \int_{-10}^{10} K^2(S)ds$

Fraktal: Jurnal Matematika dan Pendidikan Matematika *Ekowati, Perbandingan estimator……*
Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

Bukti :

$$Var\left[\hat{f}_h(x)\right] = \frac{1}{2^n}\sum_{i=1}^{n} Var[k_h(x-x_i)]$$

$$= \frac{1}{n} Var[k_h(x-x_i)]$$

$$= \frac{1}{n}\left\{E\left[k_h^2(x-x_i)\right] - (E[k_h(x-x_i)])^2\right\}$$

$$= \frac{1}{n}\left(h^{-2}\int_{-10}^{10} k^2\left(\frac{x-u}{h}\right) f(u)du - \left(h^{-1}\int_{-10}^{10} k\left(\frac{x-4}{h}\right) f(u)du\right)^2\right)$$

$$= \frac{1}{n}\left(h^{-1}\int_{-10}^{10} k^2(s) f(x+sh)ds - [f(x)+o(h)]^2\right)$$

$$= \frac{1}{nh} f(x)\int_{-10}^{10} k^2(s)ds + o((nh)^{-1})$$

$$= \frac{1}{nh} f(x)\|k\|_2^2 + o((nh)^{-1}), nh \to 10$$

Dari teorema 3 di atas, terlihat bahwa varians proporsional dengan $(nh)^{-1}$. Akibatnya untuk menurunkan variansi diperlukan h yang besar. Hal ini bertentangan dengan bias kuadrat.

Untuk mengatasi hal ini, diperhatikan harga Mean Square Error (MSE) yang memberikan kontrol antara bias kuadrat dengan variansi.

Teorema 4

Jika $\grave{f}_h(x)$ diberikan oleh (2.1), maka

$$\text{MSE}\left[\grave{f}_h(x)\right] = \frac{1}{nh} f(x)\|k\|_2^2 + \frac{h^4}{4}[f''(x)\,\alpha\,(k)]^2 + 0(h^4) + 0((nh)^{-1})$$

Bukti:

$$\text{MSE}\left[\grave{f}_h(x)\right] = \text{E}\left[\grave{f}_h(x) - f(x)^2\right]$$

$$= \text{E}\left[\grave{f}_h(x) - E\left[\grave{f}_h(x)\right] + E\left[\grave{f}_h(x)\right] - f(x)^2\right]$$

Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

$$= E\left[\left(\hat{f}_h(x) - E\left[\hat{f}_h(x)\right]\right)^2\right] + \left(E\left[\hat{f}_h(x)\right] - f(x)\right)^2$$

$$= var\left[\hat{f}_h(x)\right] + \left(bias\left[\hat{f}_h(x)\right]\right)^2$$

Dengan menggabungkan variansi bias kuadrat diperoleh:

$$= \frac{1}{nh} f(x) \|k\|_2^2 + \frac{h^4}{4} [f''(x)\ \alpha\ (k)]^2 + o((nh)^{-1})$$

Selanjutnya didefenisikan Badwidth Optimal h_{opt} sebagai berikut:

$$h_{opt} = arg_h^{min} MSE\left[\hat{f}_h(x)\right]$$

Teorema 5

Jika h → 0 dan nh → 10 maka:

i. $h_{opt} 0(n^{-\frac{1}{5}})$ dan

ii. MSE $[\grave{fh}_{opt}] = 0(n^{-4/5})$

Bukti :

Dari teorema 4 diperoleh :

$$MSE\left[\hat{f}_h(x)\right] \approx (nh)^{-1} A + \frac{h4}{4} B.$$

Dengan $A = f(x)\|k\|_2^2$ dan $B = f''(x)\alpha(k)^2$

$$\frac{JMSE}{Jh} \approx -n^{-1}h^{-2}A + h^2 B$$

Apabila diambil $\frac{JMSE}{Jh} = 0$, diperoleh :

$$h_{opt} = \left(\frac{A}{nB}\right) = 0(n^{-1})$$

Apabila nilai h_{opt} disubstitusikan pada $MSE\left[\hat{f}_h(x)\right]$

Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

Diperoleh :

$$MSE\left[t\grave{h}_{opt}\right] = \left(\frac{A}{n}\right)B + \left(\frac{A}{n}\right)\frac{B}{4}$$

$$= \left(\frac{A}{n}\right)B + \left(1 + \frac{1}{4}\right)$$

$$= 0(n)^{-4/3}$$

Selanjutnya akan digunakan kriteria MISE yang secara rinci disajikan dalam teorema berikut.

Teorema 6

Jika $\left[\grave{f}_{\mathrm{h}}(\mathrm{x})\right]$ diberikan oleh (2.2.1) maka

$$\mathrm{MISE}\left[\grave{f}_{\mathrm{h}}(\mathrm{x})\right] = (\mathrm{nh})^{-1}\|\mathrm{k}\|_2^2 + \frac{\mathrm{h4}}{4}\|\mathrm{f}''\|_2^2\, \alpha^2(\mathrm{k}) + 0((\mathrm{nh})^{-1}) + 0(\mathrm{h}^!)$$

Bukti :

$$\mathrm{MISE}\left[\grave{f}_{\mathrm{h}}(\mathrm{x})\right] = \int_{-10}^{10} \mathrm{MSE}\left[\grave{f}_{\mathrm{h}}(\mathrm{x})\right]\mathrm{dx}$$

$$= (nh)^{-1}\|k\|_2^2 \int_{-10}^{10} f(x)dx + \frac{h4}{4}\, a^2(k) \int_{-10}^{10} \left(f''(x)\right)^2 dx$$
$$+ 0(((nh)^{-1} + 0(h^4)$$

$$= (nh)^{-1}\|k\|_2^2 + \frac{h4}{4}\|f''\|_2^2\, a^2(k) + \ 0\ ((nh)^{-1}) + 0(h^4)$$

Apabila bagian yang berode tinggi dari sisi kanan teorema 2.6 diabaikan didefinisikan sebagai A-MISE, yaitu:

$$A - MISE\left(\grave{f}_h(x)\right) \approx (nh)^{-1}\|k\|_2^2 + \frac{h4}{4}\|f''\|_2^2\left(\int_{-10}^{10} S^2 k(s)ds\right)^2$$

Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

Kemudian kita tentukan bandwidth optimum h, dengan terlebih dehulu mendeferensialkan A-MISE terhadap parameter h, sehingga diperoleh:

$$\frac{JA-MISE}{Jh} \approx -\frac{1}{(nh)^2}\|k\|_2^2 + h^3\|f''\|_2^2\left(\int_{-10}^{10} S^2k(s)ds\right)^2$$

Diambil $\frac{JA-MISE}{Jh} = 0$, diperoleh:

$$h_{opt} = \left(\frac{\|k\|_2^2}{\|f''\|_2^2\left(\int_{-10}^{10} S^2k(s)ds\right)^2 .n}\right)^{1/5}$$

Berikut ini disajikan daftar kecepatan konvergensi untuk estimasi dengan histogram dan densitas karnel.

Estimator	Bias	Variansi	ho [MISE]	MISE$[f\grave{h}o]$
Histogram	$\sim h$	$\sim(nh)^{-1}$	$\sim n^{1/3}$	$\sim n^{-2/3}$
Karnel	$\sim h^2$	$\sim(nh)^{-1}$	$\sim n^{1/5}$	$\sim n^{-4/5}$

Least-Squares Cross- Validation

Pandang suatu ukuran jarak alternative antara $\hat{f}$ dan f. yaitu Ïntegrated Squared Error”(ISE) yang didefenisikan (Bowman, A, 1984):

$$ISE\ (h) = \int\left(\grave{f}_h - f\right)^2(x)\alpha x \qquad (3.1)$$

$$= \int\left[\grave{f}_h(x)\right]^2 dx - 2\int\left(\grave{f}_h f\right)(x)dx + \int f^2(x)dx \qquad (3.2)$$

Dari persamaan (3.2) diperoleh :

$\int \hat{f}_h\ (x)dx$ dapat dihitung dari data

$\int f^2\ (x)dx$ tidak bergantung pada h.

$\int\left(\hat{f}_h f\right)(x)dx$ satu-satunya bagian yang harus diestimasi dari data.

Apabila persamaan (3.2) dikurangi dengan bagian konstan, mengakibatkan peminimalan ISE dalam hubungannya dengan h yang ekuivalen dengan meminimalkan

Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

$$ISE(h) - \int f^2\,(x)dx = \int \hat{f}_h^2\,(x)dx - 2\int \left(\hat{f}_h f\right)(x)dx \qquad (3.3)$$

Untuk bagian terakhir dari persamaan (3.3) yaitu $\int \left(\hat{f}_h f\right)(x)dx = E_x\left[\hat{f}_h(x)\right]$ sehingga dalam hal ini $E_x\left[\hat{f}_h(x)\right]$yang dianalisis dan dihitung dalam hubungannya dengan tambahan observasi X yang independen. Untuk estimasi bagian ini, didefinisikan "$leave - one - out\ estimation$"

$$E_x\left[\hat{f}_h(x)\right] = n^{-1}\sum_{i=1}^{n}\hat{f}_{h,i}(x_i)$$

Dengan demikian, apabila digunakan estimasi ini berarti untuk menentukan suatu bandwith h yang baik yang meminimalkan ruas kanan dari persamaan (3.3) akan didefinisikan "$Least\ Squares\ Cross\ Validation$" sebagai berikut:

$$cv\ (h) \int \hat{f}_h^2\,(x)dx - \tfrac{2}{h}\textstyle\sum_{i=1}^{n}\hat{f}_{h,i}(x_i) \qquad (3.4)$$

Dimana :

$$\int \hat{f}_h^2\,(x_i) = \tfrac{1}{n-1}\textstyle\sum_{i=1} K_{n,i}(x_i\,, x_j) \qquad (3.5)$$

$(n-1)$ pada perencanaan (3.5) diperoleh dengan mengganti posisi n pada perencanaan (3.4). Pertukaran ini tidak berpengaruh pada fungsi kernel. Sedangkan bandwith yang meminimalkan perencanaan (3.4) adalah:

$$\hat{h}cv = {}^{arg}_{\ h}min\,cv(h)$$

Maka ekspetasinya adalah:

$$E\left(\hat{f}_{h,i}(x_i)\right) = E[\frac{1}{n-1}\sum_{i=1} K_{n,i}(x_i\,, x_j)\,]$$

$$= E[K_{n,i}(x_i, x_j)]$$

$$= E\int K_{n,i}\,(x, x_i)f(x)\alpha x$$

$$= \int \frac{1}{n}\sum_{i=1} K_{n,i}(x, x_i)f(x)\alpha x$$

$$= \int \hat{f}_h(x)f(x)\alpha x \qquad (3.6)$$

Berdasarkan persamaan (3.4) dan (3.6), maka untuk bandwith tertentu pada fungsi kernel secara eksak diperoleh :

$$E\big(cv(h)\big) = E[ISE(h)] + 2\left(E_x\left[\int \hat{f}_h\,(x)dx\right] - E\left[\frac{1}{n}\sum_{i=1}^{n}\hat{f}_{h,i}\,(x)\right]\right) - \|f\|_2^2$$

$$= MISE\left[\hat{f}_h\right] - \|f\|_2^2 \qquad (3.8)$$

Fraktal: Jurnal Matematika dan Pendidikan Matematika *Ekowati, Perbandingan estimator……*
Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

Dari persamaan (3.8) inilah sehingga Least Squares Cross- Validation disebut juga Unbiased Cross-Validation.

Jadi, untuk pembahasan selanjutnya (3.4) ditulis :

$$ucv(h) = \int \hat{f}_h\,(x)\alpha x - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{h,i}\,(x_i)$$

Suatu bandwith $\hat{h}_n = h(x_i, \dots, x_n)$ menjadi optimal secara asymtotik, Jika:

$$\frac{ISE(\hat{h}_n)}{Inf_{h>0} ISE(h)} \xrightarrow{AS} 1(n \to \infty)$$

Teorema 3.1

Jika SUP $\left|\frac{\mathrm{ISE(h)-ISE(h')) - [ucv(h) - ucv(h')}}{\mathrm{ISE(h)+ISE(h')}}\right| \xrightarrow{\mathrm{AS}} 1$

$(\mathrm{n} \to \infty)$ maka $\mathrm{h_{ucv}}$ optimal secara asymtotik.

Bukti:

Dari (3.1) diperoleh $\hat{\mathrm{h}}_{\mathrm{ISE}} = {}^{\mathrm{arg}}_{\mathrm{h}}\mathrm{min\ ISE(h)}$ yang berarti $0 < \mathrm{ISE}(\hat{\mathrm{h}}_{\mathrm{ucv}}) - \mathrm{ISE}(\hat{\mathrm{h}}_{\mathrm{ISE}})$, sedangkan persamaan (3.4) diperoleh $\hat{h}_{ucv} = {}^{arg}_{h} min\ ucv(h)$ yang berarti $0 < ucv(\hat{h}_{ISE})$ sehingga untuk $n \to \infty, \varepsilon > 0$ berlaku:

$$0 < \frac{ISE(\hat{h}_{ucv}) - ISE(\hat{h}_{ISE}) - [ucv(\hat{h}_{ucv}) - ucv(\hat{h}_{ISE})]}{ISE(\hat{h}_{ucv}) + ISE(\hat{h}_{ISE})} \leq \varepsilon$$

Dengan demikian diperoleh :

$$0 \geq ucv(\hat{h}_{ucv})(\hat{h}_{ISE}) \geq (1-\varepsilon)ISE(\hat{h}_{ucv}) - (1+\varepsilon)ISE(\hat{h}_{ISE})$$

Dengan menghilangkan unsur $UCV(\hat{h}_{ucv}) - (1+\varepsilon)UCV(\hat{h}_{ISE})$

Sehingga didapatkan $(1+\varepsilon)ISE(\hat{h}_{ISE}) \geq (1-\varepsilon)ISE(\hat{h}_{ucv})$.

Selanjutnya diperoleh:

$$\frac{(1+\varepsilon)}{(1-\varepsilon)} \geq \frac{ISE(\hat{h}_{ucv})}{ISE(\hat{h}_{ISE})}$$

Karena berlaku untuk setiap ε, dengan demikian mengakibatkan :

$\lim_{h\to\infty} \frac{ISE(\hat{h}_{ucv})}{ISE(\hat{h}_{ISE})} = 1$ yang berarti bahwa $\hat{h}_{ucv}$ optimal secara asymtotik.

Volume 2, No. 1, Mei 2021, Hal. 43-53 (e-ISSN 2776-0073)
Doi: https://doi.org/10.35508/fractal.v2i1.4036

KESIMPULAN

Berdasarkan pembahasan di atas, maka dapat disimpulkan sebagai berikut :

1. Estimasi densitas histogram adalah :

   $f_h(x) = (nh)\sum_{i=1}^{n}\sum I(x_i \in B_j)$ dengan $B_j = \left[xo - \frac{h}{2}.xo + \frac{h}{2}\right]$

2. Estimator densitas kernel adalah :

   $f_h(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - x_1) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-x_1}{h}\right)$

3. Estimator kernel mengatasi kelemahan- kelemahan dari estimator histogram.
4. Tingkat konvergensi dari estimator kernel lebih baik dari pada histogram.
5. Pemilihan bandwidth dengan metode unbiased cross validation (least cross validation) secara asymtotik menghasikan bandwidth yang optimum.

Pada pemilihan bandwidth dengan menggunakan metode least cross validation menghasilkan bandwidth yang optimum. Namun masih merupakan persamaan yang rumit, sehingga penulis menyarankan ada dari para penggemar statistik yang melanjutkan dengan menggunakan metode lain yang dapat menghasilkan pemilihan bandwidth yang lebih baik dan sederhana.

DAFTAR PUSTAKA

Bartlett, M. S. (1963). *Statistikal estimation of density function,* Aankhya. 25A, 245-254

Bowman, A. (1984). *An alternative method of cross-validation for density estimates*. Biometrika, 71, 353-360

Berliner, L. M. (1993). *Improving on inadmissible estimators in control problem*, Ann.Statist. 11, 814-826

Berger, J.O. (1980). *Statistical decision theory: foundation, consepts and methods,* New York: Springer-Verlag,

Ekowati, Ch.K. (1998). *Perbadingan estimator klasik dan estimator invers dalam kalibrasi linear multi-univariat,* Tesis-UGM Yogyakarta

Hall, P. (1982). *Cross-validation in density estimation,* 69, 383-390