ESTIMASI DENSITAS KERNEL PADA GABUNGAN DUA DISTRIBUSI NORMAL

(Skripsi)

Oleh

SITI RAHMATALIA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

ABSTRAK

ESTIMATION OF KERNEL DENSITY IN THE COMBINATION OF TWO NORMAL DISTRIBUTIONS

By

Siti Rahmatalia

This study aims to estimate the density of the kernel on a combination of two independent normal distribution. The optimal bandwidth is determined by *Unbiased Cross Validation* (UCV) using different the kernel method, namely the Gaussian, Epanechnikov and Uniform kernel functions. The results show that estimation of Gaussian kernel density is better in estimating the combination of independent two normal distributions.

Kata kunci : Kernel Density Estimator, Mixture of Two Normal Distributions, *Unbiased Cross-Validation* (UCV)

ABSTRAK

ESTIMASI DENSITAS KERNEL PADA GABUNGAN DUA DISTRIBUSI NORMAL

Oleh

Siti Rahmatalia

Pada penelitian ini bertujuan untuk melakukan estimasi densitas kernel pada gabungan dua distribusi normal independen. Bandwidth optimal ditentukan dengan *Unbiased Cross-Validation* (UCV) menggunakan metode kernel yang berbeda, yaitu fungsi kernel Gaussian, Epanechnikov dan Uniform. Hasil penelitian menunjukkan bahwa estimasi densitas kernel Gaussian lebih baik dalam mengestimasi terhadap gabungan dua distribusi normal independen.

Kata kunci : Estimator Densitas Kernel, Gabungan Dua Distribusi Normal, *Unbiased Cross-Validation* (UCV)

ESTIMASI DENSITAS KERNEL PADA GABUNGAN DUA DISTRIBUSI NORMAL

Oleh

SITI RAMATALIA

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA SAINS

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

Judul skripsi	: **ESTIMASI DENSITAS KERNEL PADA GABUNGAN DUA DISTRIBUSI NORMAL**
Nama Mahasiswa	: Siti Rahmatalia
Nomor Pokok Mahasiswa	: 1517031074
Jurusan	: Matematika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.	**Dr. Asmiati, S.Si., M.Si.**
NIP. 19650125 199003 2 001	NIP. 19760411 200012 2 001

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, MA., Ph.D.
NIP: 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Ir. Netti Herawati, M.Sc., Ph.D.** ....

Sekretaris : **Dr. Asmiati, S.Si., M.Si.** ....

Penguji
Bukan pembimbing : **Prof. Drs. Mustofa Usman, MA., Ph.D.** ....

2. Dekan Fakultas Matematika Dan Ilmu Pengetahuan Alam

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN UNIVERSITAS LAMPUNG FAKULTAS MATEMATIKA DAN ILMU

Drs. Suratman, M.Sc.
NIP: 19640604 199003 1 002

Tanggal Lulus Ujian Skripsi : **29 Oktober 2019**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini :

Nama	: **Siti Rahmatalia**
Nomor Pokok Mahasiswa	: **1517031074**
Judul	: **ESTIMASI DENSITAS KERNEL PADA GABUNGAN DUA DISTRIBUSI NORMAL**
Jurusan	: **Matematika**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 30 Oktober 2019
Penulis,

Siti Rahmatalia
NPM: 1517031074

RIWAYAT HIDUP

Penulis dilahirkan di Bandar Lampung pada tanggal 01 April 1997. Sebagai anak kedua Bapak Sarifuddin dan Ibu Desi Deria Herawati. Penulis menempuh pendidikan Sekolah Dasar Negeri (SDN) 1 Sukarame pada tahun 2003-2009, Madrasah Tsanawiyah Negeri (MTsN) 2 Bandar Lampung pada tahun 2009-2012, Madrasah Aliyah Negeri (MAN) 1 Bandar Lampung pada tahun 2012-2015.

Pada tahun 2018 penulis melaksanakan Kerja Praktek (KP) di KPP Pratama Teluk Betung. Pada tahun 2018 Penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Setia Agung, Kecamatan Gunung Terang, Kabupaten Tulang Bawang Barat, Provinsi Lampung. Pengalaman organisasi penulis yaitu menjadi Anggota Biro Dana dan Usaha Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) periode 2016 dan 2017.

KATA INSPIRASI

"Karena sesungguhnya sesudah kesulitan itu ada kemudahan."
(Q.S. Al-Insyirah: 5-6)

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."
(Q.S. Al-Baqarah: 286)

"Ada dua jenis orang yang susah dikalahkan di dunia ini yaitu, orang yang sabar dan orang yang tidak mudah menyerah ."
(Tere Liye)

"Setiap manusia adalah pemeran utama di kehidupannya. Jadi dirimulah yang menentukan bagaimana warna kehidupanmu."

PERSEMBAHAN

Puji Sykur kepada Allah SWT, Karena atas limpahan berkah, rahmad Allah SWT Kupersembahkan karya sederhana penuh perjuangan ini kepada:

Ayahanda Sarifuddin dan Ibu Desi Deria Herawati

Terimakasih atas limpahan kasih sayang, pengorbanan, semangat, motivasi, serta doa dan sujud yang selalu menantikan keberhasilanku dengan sabar dan penuh pengertian. Karena atas doa dan ridho kalian, Allah memudahkan setiap perjalanan hidup ini.

Terimalah bukti kecil ini sebagai kado keseriusanku untuk membalas semua pengorbanan, keikhlasan, dan jerih payah yang selama ini telah diberikan.

Almamater yang kucintai, Universitas Lampung

Terima Kasih

SANWACANA

Penulis mengucapkan puji syukur kehadirat Allah SWT, karena dengan ridho dan karunia-Nya serta atas berkah dan rahmat-Nya sehinga penulis dapat menyelesaikan skripsi yang berjudul "Estimasi Densitas Kernel pada Gabungan Dua Distribusi Normal". Selesainya penulisan skripsi ini adalah berkat motivasi, pengarahan serta bimbingan dari berbagai pihak. Dengan segala kerendahan dan ketulusan hati penulis ingin menyampaikan ucapan terima kasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D, selaku pembimbing pertama atas saran, bimbingan, arahan, motivasi, dan kesabaran dalam membimbing penulis selama penelitian hingga penyelesaian skripsi dan memberi arahan kepada penulis selama menuntut ilmu di Universitas Lampung.
2. Ibu Dr. Asmiati, S.Si., M.Si., selaku pembimbing kedua yang telah memberihan arahan, saran, serta dukungan bagi penulis.
3. Bapak Prof. Drs. Mustofa Usman, MA., Ph.D., selaku Pembahas yang telah memberikan ide, kritik, dan saran sehingga terselesaikannya skripsi ini.
4. Ibu Prof. Dra.Wamiliana, MA., Ph.D. selaku Kepala Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung.
5. Bapak Drs. Suratman, M.Sc. selaku Dekan FMIPA Universitas Lampung
6. Para Dosen dan Staff Jurusan Matematika FMIPA Universitas Lampung.
7. Mama, Papa, Kakak, Adik-adik, dan keluarga tercinta yang selalu memberikan motivasi, semangat, dan doa yang tak terhingga kepada penulis.

8. Sahabat-sahabat penulis Sekar, Luvita, Rima, Dinda, Resti, Pipin yang telah membantu, memberikan semangat, dan kecerian pada penulis.
9. Teman-temanku Wike, mba Anis, mba Novita, mba Desi, mba Resti yang telah memberikan keceriaan dan semangat bagi penulis.
10. Teman-teman Matematika 2015 yang selalu menjadi semangat bagi penulis.
11. Semua pihak yang terlibat dalam penyelesaian skripsi yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa masih ada kekurangan dari skripsi ini, akan tetapi besar harapan semoga skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung, 21 Oktober 2019
Penulis

Siti Rahmatalia

DAFTAR ISI

Halaman

DAFTAR TABEL iii

DAFTAR GAMBAR iv

I. PENDAHULUAN
1.1 Latar Belakang dan Masalah 1
1.2 Tujuan Penelitian 2
1.3 Manfaat Penelitian 2

II. TINJAUAN PUSTAKA
2.1 Distribusi Normal 3
2.2 Model Campuran 3
2.3 Campuran Distribusi Normal Independen 5
2.4 Sebaran Unimodal 7
2.5 Sebaran Bimodal 7
2.6 Estimasi Densitas Kernel 8
2.7 Funsi Kernel 10
2.7.1 Kernel Gaussian 11
2.7.2 Kernel Epanechnikov 11
2.7.3 Kernel Uniform 12
2.8 Pemilihan *Bandwidth* Optimal 13
2.8.1 *Unbiased Cross-Validation* (UCV) 13

III. METODOLOGI PENELITIAN
3.1 Waktu dan Tempat Penelitian 14
3.2 Data Penelitian 14
3.3 Metode Penelitian 15

IV. HASIL DAN PEMBAHASAN
4.1 Gabungan Dua Distribusi Normal dengan Parameter $\mu_1 = \mu_2$ dan $\sigma_1 = \sigma_2$ 16
4.2 Gabungan Dua Distribusi Normal dengan Parameter $\mu_1 \neq \mu_2$ dan $\sigma_1 = \sigma_2 = \sigma$ 17
4.3 Gabungan Dua Distribusi Normal dengan Parameter $\mu_1 \neq \mu_2$ dan $\sigma_1 \neq \sigma_2$ 18
4.4 Pemilihan *Bandwidth* Optimal 19

4.4.1 Pemilihan *Bandwidth* Optimal dan Estimasi Densitas Kernel pada Gabungan Dua Distribusi Normal dengan paremeter $\mu_1 = \mu_2$ dan $\sigma_1 = \sigma_2$ 19
4.4.2 Pemilihan *Bandwidth* Optimal dan Estimasi Densitas Kernel pada Gabungan Dua Distribusi Normal dengan paremeter $\mu_1 \neq \mu_2$ dan $\sigma_1 = \sigma_2$ 21
4.4.3 Pemilihan *Bandwidth* Optimal pada Gabungan Dua Distribusi Normal dengan paremeter $\mu_1 \neq \mu_2$ dan $\sigma_1 \neq \sigma_2$ 23

V. KESIMPULAN 26

DAFTAR PUSTAKA 27

LAMPIRAN

DAFTAR TABEL

Tebel Halaman

1. Nilai UCV beberapa kernel pada gabungan distribusi normal dengan parameter $\mu_1 = \mu_2$ dan $\sigma_1 = \sigma_2$ 20

2. Nilai UCV beberapa kernel pada gabungan distribusi normal dengan parameter $\mu_1 \neq \mu_2$ dan $\sigma_1 = \sigma_2$ 21

3. Nilai UCV beberapa kernel pada gabungan distribusi normal dengan parameter $\mu_1 \neq \mu_2$ dan $\sigma_1 \neq \sigma_2$ 23

DAFTAR GAMBAR

Gambar Halaman

1. Bentuk Sebaran Unimodal 7
2. Bentuk Sebaran Bimodal 8
3. Kurva Fungsi Gaussian 11
4. Kurva Fungsi Epanecnikov 12
5. Kurva Fungsi Uniform 12
6. Histogram $X_1 \sim N(2{,}2) + X_1 \sim N(2{,}2)$ 17
7. Histogram $X_1 \sim N(2{,}2) + X_1 \sim N(12{,}2)$ 18
8. Histogram $X_1 \sim N(2{,}3) + X_1 \sim N(10{,}5)$ 19
9. Estimasi densitas kernel $X_1 \sim N(2{,}2) + X_1 \sim N(2{,}2)$ 20
10. Estimasi densitas kernel $X_1 \sim N(2{,}2) + X_1 \sim N(12{,}2)$ 22
11. Estimasi densitas kernel $X_1 \sim N(2{,}3) + X_1 \sim N(10{,}5)$ 24

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Estimasi densitas adalah konstruksi perkiraan kerapatan dari data yang diamati. Dalam memperkirakan kerapatan data langkah awal yang dilakukan adalah menganalisis data. Analisis data dapat di lakukan melalui struktur dan pola data serta penyajian hasil dalam bentuk yang lebih ringkas dan sederhana. Penelusuran struktur data bertujuan untuk memeriksa apakah suatu data dapat membentuk suatu model tertentu, sedangkan penelusuran pola data bertujuan untuk memeriksa bagaimana sebaran data. Salah satu cara untuk memeriksa sebaran data adalah dengan menentukan fungsi kepekatan peluang (fkp) atau *probability density function*. Misalkan $\{X_i, i = 1, 2, \dots, n\}$ adalah data pengamatan yang saling bebas yang memiliki fungsi kepekatan peluang $f(x)$. Untuk menduga $f(x)$ dapat dilakukan dengan dua pendekatan, yaitu parametrik dan nonparametrik.

Gabungan dua variabel acak yang independen dan berdistribusi normal yang mana pada himpunan data tersebut dapat membentuk suatu puncak tertentu. Secara sederhana gabungan normal adalah probabilitas kontinu yang merupakan gabungan dari dua variabel acak yang independen dan distribusi normal dengan ukuran parameternya yaitu mean μ dan varian σ^2 yang telah di tentukan sehingga membentuk suatu puncak yang bergantung pada ukuran parameternya (Silesian, 2016). Dalam menggabungkan dua distribusi tersebut dilakukan dengan gabungan densitas dan menggunakan kernel Gaussian, Epanechnikov, dan Uniform. Pada penelitian ini akan dilakukan estimasi densitas pada gabungan dua distribusi normal dengan kondisi modus yang berbeda-beda menggunakan estimator kernel Gaussian, Epanechnikov, dan Uniform.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah melakukan estimasi densitas gabungan dua distribusi normal yang independen dengan modus yang berbeda.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah

1. Bagi penulis diharapkan dapat menambah pemahaman yang telah diterima mengenai estimasi densitas gabungan dua distribusi normal yang independen.
2. Bagi pembaca diharapkan dapat dijadikan sumbang saran yang akan melakukan penelitian dan dapat mengembangkan estimasi densitas pada gabungan dua distribusi normal yang independen.

II. TINJAUAN PUSTAKA

2.1 Distribusi Normal

Distribusi normal merupakan distribusi peluang kontinu yang sangat penting. Distribusi normal juga sering disebut dengan distribusi Gauss. Distribusi normal secara matematis bergantung pada parameter μ dan σ. Definisi mengenai distribusi normal sebagai berikut:

Jika x merupakan sampel acak dari distribusi normal, maka fungsi kepekatan peluang dari x adalah

$$f(x;\mu,\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{[x-\mu]^2}{2\sigma^2}\right) \quad ; -\infty < x < \infty \tag{2.1}$$

Parameter μ dan σ^2 merupakan rata-rata dan ragam dari x. Sehingga x mempunyai distribusi $N(\mu,\sigma^2)$ (Hogg, McKean, and Craig, 1995).

2.2 Model Campuran

Secara umum dapat dikatakan bahwa suatu distribusi f adalah sebuah campuran dari distribusi komponen K yaitu $f_1, f_2, \dots, f_K$ jika:

$$f(x) = \sum_{k=1}^{K} \lambda_k f_k(x) \tag{2.2}$$

Dimana λ_k adalah bobot campuran, $\lambda_k > 0, \sum_k \lambda_k = 1$. Persamaan (2.2) merupakan sebuah model stokastik sempurna, yang memberi cara untuk menghasilkan titik pada data baru, pertama pilih distribusi, dengan probabilitas yang diberikan oleh bobot pencampuran, dan kemudian menghasilkan satu pengamatan yang sesuai dengan distribusi tersebut. Secara simbolis dituliskan sebagai berikut:

$$Z \sim \text{Mult}\ (\lambda_1, \lambda_2, \dots, \lambda_k) \tag{2.3}$$

$$X|Z \sim f_z \tag{2.4}$$

dengan Z merupakan variabel acak diskrit dari komponen X. Dalam hal ini kita belum mengetahui distribusi apakah f_k tersebut. Pada prinsipnya, dapat dapat di sesuaikan dengan keinginan, dan akan tetap memiliki model campuran yang sangat baik. Namun dalam prakteknya, banyak upaya diberikan dengan menggunakan model campuran parametrik, di mana semua f_k dari keluarga parametrik yang sama, tetapi dengan parameter yang berbeda. misalnya jika semua merupakan distribusi gauss dengan mean dan varian yang berbeda, atau semua distribusi poisson dengan mean yang berbeda, atau semua hukum kekuasaan dengan eksponen yang berbeda. Dapat dituliskan parameter, atau vektor parameter, k^{th} dari komponen sebagai θ_k, sehingga model menjadi:

$$f(x) = \sum_{k=1}^{K} \lambda_k f(x; \theta_k) \tag{2.5}$$

Dengan demikian keseluruhan dari parameter vektor dari campuran model $\theta = (\lambda_1, \lambda_2, \dots, \lambda_K, \theta_1, \theta_2, \dots, \theta_K)$. Dalam pertimbangkan dua hal ekstrim. Ketika

$K = 1$, maka memiliki distribusi parametrik sederhana, dari jenis yang biasa, dan estimasi kepadatan berkurang menjadi estimasi parameter, dengan mengunakan *maximum likelihood* atau apa pun yang dinginkan. Di sisi lain ketika $K = n$ jumlah pengamatan, maka kembali ke estimasi densitas kernel yang merupakan estimasi model campuran nonparametrik (Ruzgas and Drulyte, 2013).

2.3. Campuran Distribusi Normal Independen

Menurut McLachlan (1946), Campuran terbatas distribusi normal pada pengamatan univariat diasumsikan bahwa pengamatan terhadap $x = (x_1, \dots, x_n)$ yang merupakan independen pada sebuah variabel acak X yang muncul dari campuran distribusi normal dapat di tuliskan sebagai berikut:

$$p(x|\vartheta) = \pi_1 \Phi\left(x; \mu_1, \sigma_1^2\right) + \cdots + \pi_K \Phi\left(x; \mu_K, \sigma_K^2\right) \tag{2.6}$$

dengan $f_k(x; \mu_k, \sigma_k^2)$ densitas dari sebuah distribusi normal univariat dimana rata-rata μ_k dan varians σ_k^2 serta π_k ialah bobot probabilitas. Pada campuran dua distribusi normal yang independen pada setiap variabel acak X_1 dan X_2 maka jumlah dari variabel acak tersebut akan berdistribusi normal jika:

$$X_1 \sim N(\mu_1, \sigma_1^2)$$

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

maka

$$X_1 + X_2 \sim N(\mu_1 + \mu_2\,, \sigma_1^2 + \sigma_2^2) \tag{2.7}$$

Ini berarti bahwa jumlah dua variabel acak yang terdistribusi normal adalah normal, dengan menjumlahkan dari dua rata-rata dan dua varians. Dalam kondisi ini pemisahan komponen-komponen dari campuran dua distribusi normal dapat dinyatakan dengan perbedaan antara rata-rata komponen yaitu:

$$\Delta = |\mu_2 - \mu_1| \tag{2.8}$$

Kondisi yang cukup bahwa suatu campuran adalah unimodal jika:

$$\Delta^2 < \frac{27\sigma_1^2\sigma_2^2}{4(\sigma_1^2+\sigma_2^2)} \tag{2.9}$$

Berikut beberapa kondisi yang terjadi:

1. Jika $\mu_1 = \mu_2$ akan membentuk unimodal
2. Jika $\mu_1 \neq \mu_2$ dan $\sigma_1 = \sigma_2 = \sigma$. Dua fungsi kepadatan probabilitas normal dengan standar deviasi yang sama, σ, tetapi dengan nilai μ_1 dan μ_2 yang berbeda masing-masing, adalah bimodal jika:

   $$|\mu_1 - \mu_2| > 2\sigma \tag{2.10}$$

   Bergantung pada jarak antara μ_1 dan μ_2 kepadatan campuran akan dimiliki nilai maksimum pada $x_0 = \frac{1}{2}(\mu_1 + \mu_2)$ (kasus bimodal) atau nilai *local minimum* pada $x_0 = \frac{1}{2}(\mu_1 + \mu_2)$ (kasus bimodal).

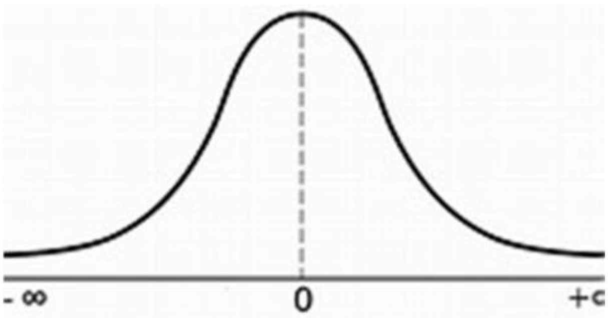
3. Jika $\mu_1 \neq \mu_2$ dan $\sigma_1 \neq \sigma_2$. Suatu kondisi yang cukup bahwa ada nilai-nilai dari $p, 0 < p < 1$ untuk itu $f(x,p)$ adalah bimodal jika:

   $$(\mu_2 - \mu_1)^2 > \frac{8\sigma_1^2\sigma_2^2}{(\sigma_1^2\sigma_2^2)} \tag{2.11}$$

Untuk setiap himpunan nilai $\mu_1,\ \mu_2,\ \sigma_1, \sigma_2$ serta $p, 0 < p < 1$

2.4 Sebaran Unimodal

Jika X adalah sebuah variabel acak bernilai nyata, dan dimisalkan μ menjadi distribusi probabilitasnya pada R, dan jika F adalah fungsi distribusi (kanan kontinu), maka variabel X (atau F atau μ) dapat dikatakan unimodal dengan mode di a dan F cembung pada $(-\infty, a)$ dan cekung pada (a, ∞). Sehingga akan memiliki puncak tunggal. Istilah "mode" dalam konteks ini mengacu pada puncak dari distribusi apa pun. Digambarkan dengan bentuk kurva simetris, berbentuk seperti lonceng, dari pola data yang diukur. Berikut ini adalah gambar sebuah sebaran unimodal (Bertin, Cuculescue, and Theodorescu, 1997).

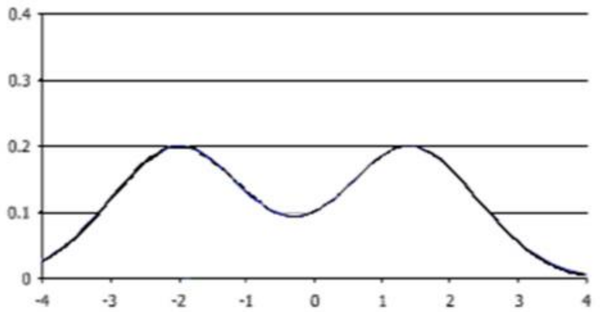


Gambar 2.1. Bentuk Sebaran unimodal

2.5 Sebaran Bimodal

Sebaran bimodal adalah distribusi probabilitas kontinu dengan dua mode yang berbeda yang membentuk puncak yang berbeda pada fungsi kepadatan probabilitas. Menurut Basset, Bremner, dan Morgan (2000), bimodal secara sederhana merupakan campuran dua distribusi normal dengan varians yang sama $(\sigma_1 = \sigma_2)$

namun nilai tengah berbeda ($\mu_1 \neq \mu_2$). Pada kedua kurva normal tersebut akan membentuk kurva yang sama namun berpusat pada posisi yang berbeda sepanjang sumbu mendatar yang membentuk dua puncak. Jika bobotnya tidak sama, distribusi yang dihasilkan masih membentuk bimodal namun dengan puncak dengan ketinggian yang berbeda seperti pada Gambar 2.2:



Gambar 2.2. Bentuk Sebaran Bimodal

2.6 Estimator Densitas Kernel

Estimator densitas kernel merupakan pengembangan dari estimator histogram. Suatu histogram disusun dengan meletakkan titik-titik data ke dalam suatu bin atau kelas. Setiap bin dinyatakan secara grafik oleh segiempat dengan lebar sama dan tinggi proposional dengan banyaknya titik-titik data yang terletak dalam bin tersebut.

Menurut Hardle (1994), estimator kernel diperkenalkan oleh Rosenblatt dan Parzen sehingga disebut estimator densitas kernel Rosenblatt-Parzen . Jika $\{X_i\}_{i=1,2,\ldots,n}$ data pengamatan independen dari suatu distribusi dengan densitas f (tak diketahui), maka estimator densitas kernel f dengan kernel K dan lebar jendela h didefinisikan sebagai berikut:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\,(x - x_i) \quad = \quad \frac{1}{nh}\sum_{i=1}^{n} K\,(\frac{x - x_i}{h}) \tag{2.12}$$

Dari persamaan di atas, dapat dilihat bahwa $\hat{f}_h$ dipengaruhi oleh fungsi kernel K dan parameter pemulus h. parameter pemulus (*bandwith*) dalam fungsi densitas kernel berfungsi untuk mengatur kehalusan kurva yang akan diestimasi. Peran *bandwith* ini diasumsikan seperti lebar interval pada histogram (Hardle, 1994). dengan meminimalkan IMSE dari $\hat{f}_h$. Berikut besar IMSE dari estimator densitas kernel.

Lema

Jika diberikan pengamatan $\{X_i\}_{i=1,2,\ldots,n}$ dari variabel random berdistribusi identik dan independen dengan densitas f, K suatu kernel order r dan f mempunyai derivatif paling sedikit tingkat r, maka $E\left(\hat{f}_h(x)\right) - f(x)) = \frac{h^r}{r!} f^{(r))}(x) \int_{-\infty}^{\infty} S^r$ untuk $h \to 0$.

Bukti:

$E\left(\hat{f}_h(x)\right) - f(x)) = E(\frac{1}{nh}\sum_{i-1}^{n} K\left(\frac{x - X_i}{h}\right)) - f(x)$

$= (\frac{1}{nh}\sum_{i-1}^{n} E(K\left(\frac{x - X_i}{h}\right)) - f(x)$

$= (\frac{1}{nh}\sum_{i-1}^{n} E(K\left(\frac{x - z}{h}\right)) - f(x)$

$= (\frac{1}{h} E(K\left(\frac{x - z}{h}\right)) - f(x)$ karena X_i iid

$= \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x-z}{h}\right) f(z)dz - f(x)$

$= \frac{1}{h}\int_{-\infty}^{\infty} K(-S)\, f(x-hs)hds - f(x)$

Sehingga menjadi

$\frac{h^r}{r!} f^{(r))}(x) \int_{-\infty}^{\infty} S^r K(s)ds$ Terbukti

Dari lema di atas, dapat disimpulkan bahwa estimator densitas kernel $\hat{f}(x)_h$ merupakan estimator yang tak bias secara asimtotis dari $f(x)$.

2.7 Fungsi Kernel

Menurut Wand and Jones (1995), secara umum kernel K dengan parameter pemulus (*bandwidth*) h didefinisikan sebagai:

$$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right) \quad \text{untuk} - \infty < x < \infty \; dan \; h > 0 \qquad (2.13)$$

Serta memenuhi:

(i) K(x) ≥ 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) $\int_{-\infty}^{\infty} xK(x)dx = 0$

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

4. Kernel Quartik : $K(x) = \frac{15}{16}(1 - x^2)^2$ $I(|x| \leq 1)$

5. Kernel Triweight : $K(x) = \frac{35}{32}(1 - x^2)^3$ $I(|x| \leq 1)$

6. Kernel Cosinus : $K(x) = \frac{\pi}{4} cos\left(\frac{\pi}{2}x\right)$ $I(|x| \leq 1)$

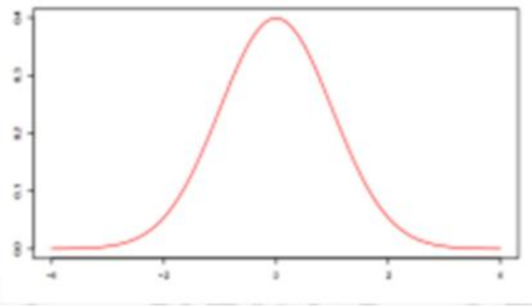
7. Kernel Gaussian : $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ $-\infty < x < \infty$

8. Kernel Tricube : $K(x) = \frac{70}{81}(1 - x^3)^3$ $I(|x| \leq 1)$

2.7.1 Kernel Gaussian

Kernel Gaussian yang dinamai oleh Carl Friedrich Gauß seorang Jerman yang cerdas ahli matematika. Menurut Wand dan Jones (1995), didefinisikan sebagai berikut:

$$K(x) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-\mu^2}{2}\right) \qquad -\infty < x < \infty \qquad (2.14)$$



Gambar 2.3. Kurva Fungsi *Gaussian*

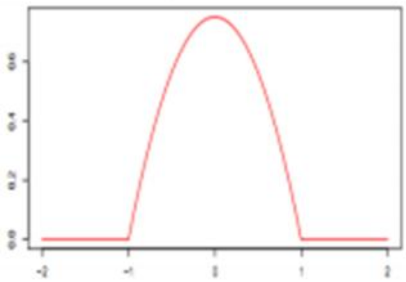
2.7.2 Kernel Epanechnikov

Menurut Hardle (1994), kernel Epanechnikov merupakan fungsi kernel yang mempunyai laju konvergensi lebih cepat menuju nilai yang diestimasi dibanding fungsi kernel yang lainnya. estimator kernel Epanechnikov dapatkan didefinisikan sebagai berikut:

$$K(x) = \frac{3}{4}(1 - x^2) \tag{2.15}$$

dengan

$$I\ (|x| \leq 1)$$



Gambar 2.3. Kurva Fungsi Epanechnikov

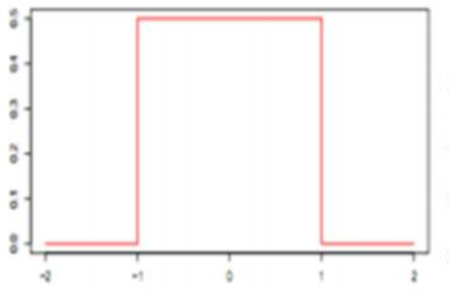
2.7.3 Kernel *Uniform*

Menurut Hardle (1994), kernel Uniform dapat didefinisikan sebagai berikut:

$$K(x) = \frac{1}{2} \tag{2.16}$$

dengan

$$I\ (|x| \leq 1)$$



Gambar 2.3. Kurva Fungsi Uniform

2.8 Pemilihan *Bandwidth* Optimal

Menurut Hardle (1994), *Bandwidth* (h) adalah parameter pemulus (*smoothing*) yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. *Bandwidth* yang terlalu kecil akan menghasilkan kurva yang *under-smoothing* yaitu sangat kasar dan sangat fluktuatif, dan sebaliknya *bandwidth* yang terlalu lebar akan menghasilkan kurva yang *over-smoothing* yaitu sangat mulus, tetapi tidak sesuai dengan pola data. Oleh karena itu perlu dipilih *bandwidth* yang optimal. Salah satu metode untuk mendapatkan *h* optimal adalah dengan menggunakan kriteria *Unbiased Cross-Validation* (UCV), yang didefinisikan sebagai berikut:

2.8.1 *Unbiased Cross-Validation* (UCV)

Menurut Bowman (1984), *Unbiased Cross-Validation (UCV)* merupakan estimator yang paling umum digunakan dalam pemilihan *bandwidth* optimal. *Unbiased Cross-Validation (UCV)* dalam penduga kepadatan kernel untuk pilihan bandwidth dengan memperkirakan h minimum pada turunan ke r dari estimator densitas kernel, didefinisikan sebagai berikut:

$$UCV(h,r) = \int \hat{f}_h^{\;2}(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{h,i}(X_i) \tag{2.17}$$

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2018-2019 dan bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi yang di bangkitkan menggunakan *sofware R* yaitu dua variabel acak yang distribusi normal yang diasumsikan independent dengan ketentuan sebagai berikut:

1. Dua distribusi normal dengan: - $X_1 \sim N\ (2,2)$

   - $X_2 \sim N\ (2,2)$

2. Dua distribusi normal dengan: - $X_1 \sim N\ (2,2)$

   - $X_2 \sim N\ (12,2)$

3. Dua distribusi normal dengan: - $X_1 \sim N\ (2,3)$

   - $X_2 \sim N\ (10,5)$

3.3 Metode Penelitian

Pada penelitian ini dilakukan estimasi dari gabungan dua variabel acak yang berdistribusi normal independent dengan fungsi kernel Gaussian, Epanechnikov dan Uniform. Program yang di gunakan dalam penelitian ini adalah R program. Adapun langkah-langkah yang dilakukan dalam penelitian ini adalah sebangai berikut:

1. Membangkitkan data yang berdistribusi normal independent dengan kondisi sebagai berikut :
   3. Dua distribusi normal dengan: 1. $X_1 \sim N\ (2,2)$
      2. $X_2 \sim N\ (2,2)$
   4. Dua distribusi normal dengan: 1. $X_1 \sim N\ (2,2)$
      2. $X_2 \sim N\ (12,2)$
   - Dua distribusi normal dengan: 1. $X_1 \sim N\ (2,3)$
      2. $X_2 \sim N\ (10,5)$
2. Menggabungkan distribusi dengan gabungan densitas
3. Menentukan bandwidth optimal dengan UCV
4. Melakukan estimasi densitas terhadap data campuran dua distribusi normal apakah membentuk bimodal atau unimodal
5. Melakukan analisa

V. KESIMPULAN

5.1 Kesimpulan

Setelah dilakukan pembahasan dalam estimasi densitas pada gabungan dua distribusi normal diperoleh kesimpulan sebagai berikut:

1. Pada parameter $\mu_1 = \mu_2 = 2$ dan $\sigma_1 = \sigma_2 = 2$ dari gabungan dua distribusi normal membentuk unimodal.
2. Pada paremeter $\mu_1 = 2,\ \mu_2 = 12$ $\sigma_1 = 2$ dan $\sigma_2 = 2$ dari gabungan dua distribusi normal membentuk bimodal.
3. Pada paremeter $\mu_1 = 2,\ \mu_2 = 10$, $\sigma_1 = 3,$ dan $\sigma_2 = 5$ dari gabungan dua distribusi normal membentuk bimodal.
4. Estimasi densitas kernel Gaussian lebih baik dalam mengestimasi terhadap gabungan dua distribusi normal independen dengan modus yang berbeda-beda.

DAFTAR PUSTAKA

Bertin, E., Cuculescue, I., and Theodorescu, R. 1997. *Unimodality of Probability Measures*. Kluwer Academic Publishers, Amsterdam.

Basset, E.E., Bremner, M.J., and Morgan, B. 2000. *Statistics Problem and Solution*. 2nd Edition. World Scientific, London.

Bowman, A.W. 1984. An alternative method of cross-validation for the smoothing of kernel density estimates. *Biometrika*. **71**(2): 353–360.

Chaterjee, S. 2007. *Regression Analysis by Example.* 4th Edition. Jhon Wiley and Sons, New Jersey.

Hardle, W. 1994. *Applied Nonparametric Regression*. Cambridge University Press, New York.

Hogg, R.V., McKean, J.W., and Craig, A.T. 1995. *Introduction to Mathematical Statistics*. Pearson, United States of America.

McLachlan, G. 1946. *Finite Mixture Models*. John Wiley and Sons, Canada.

Ruzgas, T. and Drulyte, I. 2013. Kernel Density for Gaussian Mixture Models. *Statistics Lithuania.* **52**(1): 14-21.

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

Selesian, J. 2016. The Modes Of A Mixture Of Two Normal Distrbution. *Pure Appl*. **6**(1): 59-63.

Wand, M.P. and Jones, M.C. 1995. *Kernel Smoothing*. Chapman and Hall, New York.