PENDUGAAN KEPEKATAN PELUANG DENGAN METODE KERNEL PADA DATA PEUBAH TUNGGAL DAN GANDA DUA

Oleh

BETHA LUSIANA

G 24 0668

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

INSTITUT PERTANIAN BOGOR

1992

RINGKASAN

BETHA LUSIANA. Pendugaan Kepekatan Peluang dengan Metode Kernel pada Data Peubah Tunggal dan Ganda Dua. (Di bawah bimbingan **Dr. Aunuddin** sebagai ketua dan **Dr. Khairil Anwar Notodiputro** sebagai anggota).

Pendugaan kepekatan peluang memegang peranan yang cukup penting dalam analisa data. Metode pendugaan yang umum digunakan adalah pendugaan kepekatan dengan pendekatan parametrik yang mengasumsikan data menyebar mengikuti suatu sebaran parametrik tertentu.

Tulisan ini bertujuan untuk memperkenalkan suatu metode pendugaan kepekatan yang tidak terikat dengan asumsi sebaran yaitu metode pendugaan kepekatan secara nonparametrik. Metode yang digunakan adalah metode kernel dan diterapkan pada data peubah tunggal dan ganda dua. Kajian yang dilakukan bersifat eksploratif dengan menggunakan data hasil penelitian Fisher (1936) sebagai ilustrasi.

PENDUGAAN KEPEKATAN PELUANG DENGAN METODE KERNEL PADA DATA PEUBAH TUNGGAL DAN GANDA DUA

Karya Ilmiah

sebagai Salah Satu Syarat

untuk Memperoleh Gelar Sarjana Statistika

pada

Fakultas Matematika dan Ilmu Pengetahuan Alam

di Institut Pertanian Bogor

oleh
BETHA LUSIANA
(G 24 0668)

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

INSTITUT PERTANIAN BOGOR

1992

Judul Karya Ilmiah : PENDUGAAN KEPEKATAN PELUANG DENGAN METODE KERNEL PADA DATA PEUBAH TUNGGAL DAN GANDA DUA

Nama Mahasiswa : BETHA LUSIANA

Nomor Pokok Mahasiswa : G 24 0668

Menyetujui

1. Komisi Pembimbing

(Dr. Aunuddin)
Ketua

(Dr. Khairil Anwar N.)
Anggota

Mengetahui

2. Ketua Jurusan Statistika

(Dr. Aunuddin)

Tanggal Lulus : 26 Pebruari 1992

RIWAYAT HIDUP

Penulis dilahirkan di Bogor pada tanggal 18 Desember 1968, sebagai anak kedua dari tiga bersaudara keluarga Bapak N. Supriana dan Ibu Ruspitasari.

Pada tahun 1981 penulis tamat dari Bevois Town Primary School di Southampton, Inggris. Kemudian tamat dari Sekolah Menengah Pertama Negeri 5 Bogor pada tahun 1984. Penulis melanjutkan ke Sekolah Menengah Atas Budi Mulia Pematang Siantar, dan tamat pada tahun 1987.

Melalui jalur PMDK penulis diterima di Institut Pertanian Bogor pada tahun 1987. Setahun kemudian penulis diterima di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam dengan mata kuliah penunjang Sosial Ekonomi. Semasa kuliah penulis pernah menjadi asisten mata kuliah Komputer I.

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Allah SWT, karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan penelitian dan Karya Ilmiah ini.

Pada kesempatan ini, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada Dr. Aunuddin dan Dr. Khairil Anwar Notodiputro selaku ketua dan anggota komisi pembimbing yang telah memberikan pengarahan dan bimbingan sejak awal hingga terselesaikannya tulisan ini.

Ucapan terimakasih juga penulis sampaikan kepada segenap staf pengajar Jurusan Statistika IPB yang telah membekali penulis dengan ilmu serta seluruh karyawan Jurusan Statistika yang telah banyak membantu.

Kepada bapak, ibu, Gamma dan Yusef penulis sampaikan terimakasih atas doa, dorongan serta kasihnya. Juga kepada Irma, Ika, Mila serta sahabat-sahabat lainnya yang telah banyak memberikan semangat.

Bogor, Pebruari 1992

DAFTAR ISI

DAFTAR GAMBAR

Nomor Halaman

Teks

Lampiran

DAFTAR LAMPIRAN

Nomor halaman

PENDAHULUAN

Fungsi kepekatan peluang memegang peranan penting di dalam ilmu statistika. Bila fungsi kepekatan peluang suatu peubah acak X diketahui maka dengan mudah dapat diketahui pola sebaran X serta besarnya peluang pada selang X tertentu sehingga proses penarikan kesimpulan dapat dilakukan. Cukup banyak metode analisa statistika yang menggunakan konsep fungsi kepekatan peluang.

Dalam kenyataannya seringkali fungsi kepekatan peluang ini tidak diketahui, sehingga perlu dilakukan pendugaan. Pendugaan kepekatan umumnya dilakukan dengan menggunakan pendekatan parametrik. Pendekatan ini mengasumsikan data menyebar mengikuti fungsi kepekatan tertentu yang telah diketahui bentuknya. Kemudian, nilai parameter-parameter fungsi diduga berdasarkan statistik data yang tersedia. Cara ini dikenal juga dengan nama *fitting a distribution*.

Bila ternyata data tersebut tidak menyebar mengikuti suatu sebaran parametrik tertentu maka pendugaan dapat dilakukan berdasarkan suatu pendekatan lain yang tidak terikat dengan asumsi sebaran yaitu pendekatan nonparametrik. Metode dengan pendekatan seperti ini disebut sebagai metode pendugaan kepekatan secara nonparametrik.

Literatur mengenai teori metode pendugaan kepekatan secara nonparametrik belum sebanyak metode parametrik sehingga sampai saat ini aplikasi metode pendugaan nonparametrik belum seluas metode parametrik. Meskipun demikian berbagai penelitian telah menunjukkan bahwa metode tersebut sangat efektif bila digunakan untuk kepentingan eksplorasi, presentasi data maupun analisa yang bersifat pengujian seperti uji modus ganda dan analisis diskriminan (Izenman, 1991).

Saat ini terdapat berbagai metode pendugaan kepekatan secara nonparametrik, salah satu di antaranya adalah metode **kernel**. Metode kernel pertama kali dikembangkan untuk analisis diskriminan pada tahun 1951 oleh Fix dan Hodges guna mengatasi asumsi ketat dalam penyebaran data (kenormalan). Hal yang menarik dari metode kernel ini adalah konsepnya yang sederhana dan mudah dimengerti secara matematis.

Pada data peubah tunggal metode kernel umumnya digunakan untuk kepentingan eksplorasi dan hasil pendugaannya lebih banyak disajikan secara visual (deskriptif). Sedangkan pada peubah ganda penggunaannya umumnya lebih ditekankan pada penerapan dalam analisa statistika, misalnya pada analisis diskriminan dan analisis gerombol. Penyajian hasil dalam bentuk visual kurang diminati karena sulitnya membuat grafik terutama pada data lebih dari dua peubah.

Tulisan ini bertujuan untuk memperkenalkan metode pendugaan kepekatan peluang secara nonparametrik menggunakan metode kernel. Pendugaan dilakukan pada data peubah tunggal dan ganda dua serta hasilnya disajikan dalam bentuk grafik agar pola kepekatan peluangnya dapat diamati lebih mudah.

TINJAUAN PUSTAKA

Histogram merupakan alat penduga fungsi kepekatan peluang secara nonparametrik yang paling populer (Tarter dan Kronmall, 1976). Hal ini disebabkan proses pembentukannya yang sederhana yang hanya mencakup dua tahap yaitu :

1. Pengalokasian pengamatan ke dalam salah satu kelas yang telah ditetapkan
2. Pembuatan persegi panjang yang proporsional dengan frekuensi pengamatan pada masing-masing kelas.

Namun histogram memiliki beberapa kelemahan sehingga kurang tepat bila dijadikan alat penduga kepekatan. Salah satu kelemahannya adalah sumbangan setiap pengamatan dianggap hanya mewakili nilai tengah kelas, berapapun nilai pengamatan tersebut. Selain itu belum ada aturan yang pasti mengenai penentuan jumlah kelas, lebar kelas dan lokasi nilai tengah. Padahal ketiga parameter ini sangat mempengaruhi bentuk dugaan kepekatan yang dihasilkan. Kelemahan lainnya adalah dugaan kepekatan yang dihasilkan bersifat tidak kontinu pada batas kelas, sehingga menimbulkan kesulitan bila akan digunakan untuk analisa lebih lanjut yang membutuhkan turunan dari dugaan kepekatan (Silverman, 1986 dan Izenman, 1991).

Selanjutnya dikembangkanlah metode lain, yaitu kernel, untuk mengatasi beberapa kelemahan histogram. Tahap-tahap penyusunan histogram diperbaiki sehingga sumbangan setiap pengamatan tidak lagi dipusatkan pada nilai tengah kelas, melainkan pada pengamatan itu sendiri. Selain itu sumbangan pengamatan terhadap dugaan kepekatan dapat mengambil bentuk selain persegi panjang.

Bentuk umum dari penduga kepekatan kernel pada peubah tunggal adalah :

$$f(\hat{x}) = \frac{1}{n\ h} \sum K \left(\frac{X - X_i}{h} \right)$$

dengan n : jumlah pengamatan , h : lebar kelas dan K : fungsi Kernel .

Lebar kelas (h) atau lebih sering disebut lebar jendela merupakan parameter pemulus yang sangat mempengaruhi bentuk penduga kepekatan yang dihasilkan. Semakin lebar jendela maka penduga kepekatan yang dihasilkan akan semakin mulus.

Fungsi Kernel (K) merupakan parameter yang menentukan bentuk peranan dari setiap pengamatan. Sifat penting yang dimiliki fungsi kernel adalah $\int K(t)dt = 1$ dan $K(t) \geq 0$. Karena K(t) merupakan fungsi peluang, maka dugaan kepekatan ($\hat{f}(x)$) juga akan merupakan kepekatan peluang.

Beberapa fungsi Kernel yang umum digunakan pada data peubah tunggal dapat dilihat pada Tabel 1, sedangkan gambar fungsi-fungsi tersebut dapat dilihat pada Lampiran 1.

Tabel 1. Beberapa Fungsi Kernel

Kernel	K(t)
Biweight	$15/16(1-t^2)^2$ jika $\|t\| < 1$ 0 jika lainnya
Triangular	$1 - \|t\|$ jika $\|t\| < 1$ 0 jika lainnya
Rectangular	$1/2$ jika $\|t\| < 1$ 0 jika lainnya
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-(1/2 t)^2}$

Untuk menilai kecocokan antara dugaan fungsi kepekatan dengan fungsi kepekatan yang sebenarnya biasanya digunakan Integral Kuadrat Tengah Galat (*Mean Integrated Square Error* = MISE), yaitu :

$$\text{MISE}\ (\hat{f}) = E \int \left[\hat{f}(x) - f(x) \right]^2 dx$$

Menurut Epanechnikov (<u>dalam</u> Silverman, 1986) nilai MISE yang minimum akan diperoleh jika digunakan fungsi Kernel :

$$K(t) = \begin{cases} \frac{3}{4\sqrt{5}}\left[1 - 1/5t^2\right] & ; \quad -\sqrt{5} \le t < \sqrt{5} \\ 0 & ; \quad \text{lainnya} \end{cases}$$

Fungsi kernel ini disebut juga fungsi kernel Epanechnikov-Bartlett. Oleh karena itu efisiensi fungsi-fungsi kernel lainnya dihitung dengan jalan membandingkannya dengan fungsi kernel Epanechnikov (Silverman, 1986). Pembandingan ini dilakukan pada lebar jendela yang optimum untuk masing-masing fungsi kernel. Hasil yang diperoleh menunjukkan bahwa setiap fungsi kernel tersebut memiliki efisiensi yang tidak jauh berbeda dan mendekati nilai 1. Sehingga berdasarkan integral kuadrat tengah galat sebenarnya tidak ada perbedaan yang menyolok antara fungsi kernel tersebut dan pemilihan fungsi kernel yang akan digunakan dapat didasarkan pada pertimbangan mana yang lebih sederhana.

Pemilihan parameter pemulus merupakan masalah yang penting dalam pendugaan fungsi kepekatan dengan metode kernel. Pemilihan parameter pemulus yang optimum dipengaruhi oleh tujuan pendugaan fungsi kepekatan. Bila pendugaan bertujuan untuk untuk eksplorasi atau presentasi data maka parameter pemulus dapat ditentukan secara subjektif (Silverman, 1986), yaitu dengan melakukan pendugaan pada beberapa nilai parameter pemulus dan memilih penduga kepekatan dengan pola paling sesuai berdasarkan pengetahuan dan pengalaman pemakai.

Pemilihan parameter pemulus secara subjektif kurang efektif bila pemakai sama sekali tidak memiliki gambaran akan pola kepekatan populasi yang sebenarnya. Demikian juga bila pendugaan kepekatan harus dilakukan secara rutin terhadap data yang banyak. Karena itulah dikembangkan metode pemilihan parameter pemulus secara otomatis. Seperti misalnya penentuan lebar jendela berdasarkan fungsi kenel yang digunakan (Silverman, 1986) Prinsipnya adalah menentukan lebar jendela yang akan meminimumkan MISE.

Dengan mengasumsikan data menyebar normal maka rumus lebar jendela optimum untuk kernel Gaussian adalah $1.06\sigma n^{-1/5}$. Rumus lebar jendela yang lebih resisten terhadap asumsi kenormalan adalah $0.9An^{-1/5}$ dengan A sebagai min(σ,jarak antar kuartil//1,34). Pemilihan parameter pemulus berdasarkan fungsi kernel yang digunakan ini sangat dipengaruhi pola sebaran data. Karena itu parameter pemulus yang didapatkan belum tentu menghasilkan penduga kepekatan dengan MISE minimum, terutama bila data menyimpang jauh dari asumsi kenormalan. Meskipun demikian, nilai tersebut dapat dijadikan sebagai acuan untuk pendugaan.

Selain metode di atas ada beberapa metode penentuan lebar jendela secara otomatis lainnya, di antaranya adalah metode validasi silang kuadrat terkecil (*Cross Validation Least Squares*) dan validasi silang *likelihood* (Silverman, 1986). Metode-metode ini memberikan hasil yang lebih tepat (*exact*) dibandingkan dengan metode berdasarkan fungsi kernel, tetapi memerlukan proses komputasi yang lebih kompleks.

Metode kernel juga dapat digunakan pada data peubah ganda dengan perubahan pada rumus menjadi:

$$f(\hat{X}) = \frac{1}{n\,h^d} \sum K\left(\frac{X - X_i}{h}\right) .$$

Pada kasus peubah ganda, K(t) adalah fungsi Kernel bagi t berdimensi d.

Fungsi kernel pada peubah ganda merupakan fungsi bermodus tunggal yang radial simetris. Fungsi yang paling umum digunakan adalah fungsi kernel *Gaussian* dan *Epanechnikov*.

Bentuk fungsi *Gaussian* untuk peubah ganda adalah :

$$K(t) = 2\pi^{-d/2} e^{(-1/2t^T t)}$$

dan bentuk fungsi kernel *Epanechnikov* adalah:

$$K(t) = \begin{cases} 1/2c_d^{-1}(d+2)(1-t^T t) & ; \ t^T t < 1 \\ 0 & ; \text{lainnya} \end{cases}$$

Besaran c_d menyatakan volume 1 unit dimensi d. Sebagai contoh $c_2 = 2$, $c_3 = \pi$, $c_4 = \pi/3$ dan seterusnya.

Penelitian mengenai metode pendugaan kepekatan kernel pada peubah ganda belum begitu banyak. Hal ini disebabkan karena proses pendugaan yang lebih kompleks dibandingkan pada peubah tunggal.

Kesulitan utama dalam penerapan metode kernel pada peubah ganda adalah penentuan parameter pemulus. Dengan menggunakan rumus seperti di atas berarti metode kernel pada peubah ganda menggunakan parameter pemulus tunggal. Hal ini akan berakibat buruk bila perbedaan keragamaan masing-masing peubah sangat tinggi (Silverman, 1986). Pada keadaan seperti ini sebaiknya data ditransformasi (*pre-scale*) terlebih dahulu agar mempunyai keragaman yang sama. Perlakuan seperti ini lebih mudah daripada melakukan pendugaan dengan lebih dari satu parameter pemulus.

Fukunaga dalam Silverman (1986) menyarankan untuk mentransformasi data agar memiliki matriks ragam peragam berbentuk identitas. Kemudian melakukan pemulusan menggunakan salah satu fungsi kernel dan akhirnya mentransformasi kembali. Hal ini sama dengan menggunakan metode pendugaan fungsi kepekatan dalam bentuk :

$$\hat{f}(x) = \frac{(\det S)^{-1/2}}{n\ h^{d}} \sum_{i=1}^{n} k\left\{h^{-2}(X - X_i)^{T}S^{-1}(X - X_i)\right\}$$

dengan S sebagai matriks ragam peragam dan k adalah fungsi yang memenuhi : $k(X^{T}X) = K(X)$. Jika K merupakan fungsi Gaussian maka $k(t)$ akan sama dengan $(2\pi)^{-d/2}e^{-1/2t}$.

Tukey dalam Silverman (1986) menyarankan untuk menggunakan matriks ragam peragam yang resisten. Selain matriks ragam peragam yang biasa dapat juga digunakan matriks ragam peragam gabungan, diagonal matriks ragam peragam biasa atau diagonal matriks ragam peragam gabungan (*SAS/STAT Guide for Personal Computer Rel. 6*, 1987, hal. 368).

Pada peubah ganda rumus pemilihan parameter pemulus berdasarkan fungsi kernel yang digunakan adalah $A(K)n^{-1/6}$ (Silverman, 1986 hal. 87). A(K) merupakan suatu konstanta yang besarnya 1 untuk kernel Gaussian dan 2.78 untuk kernel Epanechnikov.

Bila perbedaan keragaman masing-masing peubah cukup tinggi dan pendugaan kepekatan dilakukan tanpa transformasi terlebih dahulu, maka lebar jendela perlu dikalikan dengan simpangan baku (σ). Besarnya simpangan baku dapat dihitung berdasarkan rata-rata ragam yaitu $\sigma^2 = d^{-1} \sum s_{ii}$.

Metode kepekatan kernel khususnya pada data peubah tunggal menghasilkan penduga yang baik kecuali pada keadaan data menjulur (Silverman, 1986 dan Therrien, 1989). Hal ini karena digunakannya lebar jendela yang tetap pada setiap titik pendugaan. Sehingga ada kecenderungan munculnya gangguan berupa modus-modus acak pada daerah yang menjulur. Karena itu pada keadaan data menjulur disarankan untuk menggunakan metode lain yang tidak menggunakan lebar jendela yang tetap besarnya, seperti metode *Nearest Neighbour* dan *Variable Kernel* (Therrien, 1988 dan Silverman, 1986).

Cara lain yang dapat dilakukan adalah tetap menggunakan metode kernel tetapi dengan mentransformasi data terlebih dahulu. Penjelasan yang lebih rinci mengenai metode transformasi untuk mengatasi masalah kemenjuluran data dapat dilihat pada tulisan Wand, et al. (1991).

SUMBER DATA DAN METODE PENELITIAN

Data

Data yang digunakan dalam penelitian ini adalah data sekunder, diambil dari data hasil penelitian Fisher (1936) terhadap tiga species bunga iris yaitu *Iris setosa*, *I. versicolor*, dan *I. virginica*. (*SAS/STAT Guide for Personal Computer Rel. 6*, 1987, hal. 388). Peubah yang digunakan adalah lebar petal dan panjang petal.

Metode

Dalam penelitian ini pendugaan kepekatan peluang dilakukan menggunakan metode Kernel.

Pada peubah tunggal digunakan fungsi kernel *Triangular*, *Biweight*, *Rectangular*, *Gaussian* dan *Epanechnikov*. Pendugaan dilakukan menggunakan lebar jendela yang sama, yang dihitung menggunakan rumus lebar jendela optimum untuk kernel Gaussian yaitu $0.9An^{-1/5}$.

Pada peubah ganda dua dilakukan pendugaan dengan metode kernel biasa (tanpa transformasi) dan pendugaan dengan transformasi. Pendugaan tanpa transformasi menggunakan fungsi kernel *Gaussian* dan *Epanechnikov*, sedangkan pendugaan dengan transformasi menggunakan kernel *Gaussian*. Pada setiap pendugaan digunakan lebar jendela yang optimum untuk masing-masing fungsi kernel berdasarkan rumus $A(K)\sigma n^{-1/6}$.

Perhitungan dilakukan dengan bantuan program *Macro* pada paket *Minitab Rel. 7*. Contoh program dapat dilihat pada Lampiran 2 dan 3. Grafik plot tiga dimensi dan contour disajikan dengan bantuan paket *SASGRAPH Rel 6.03*.

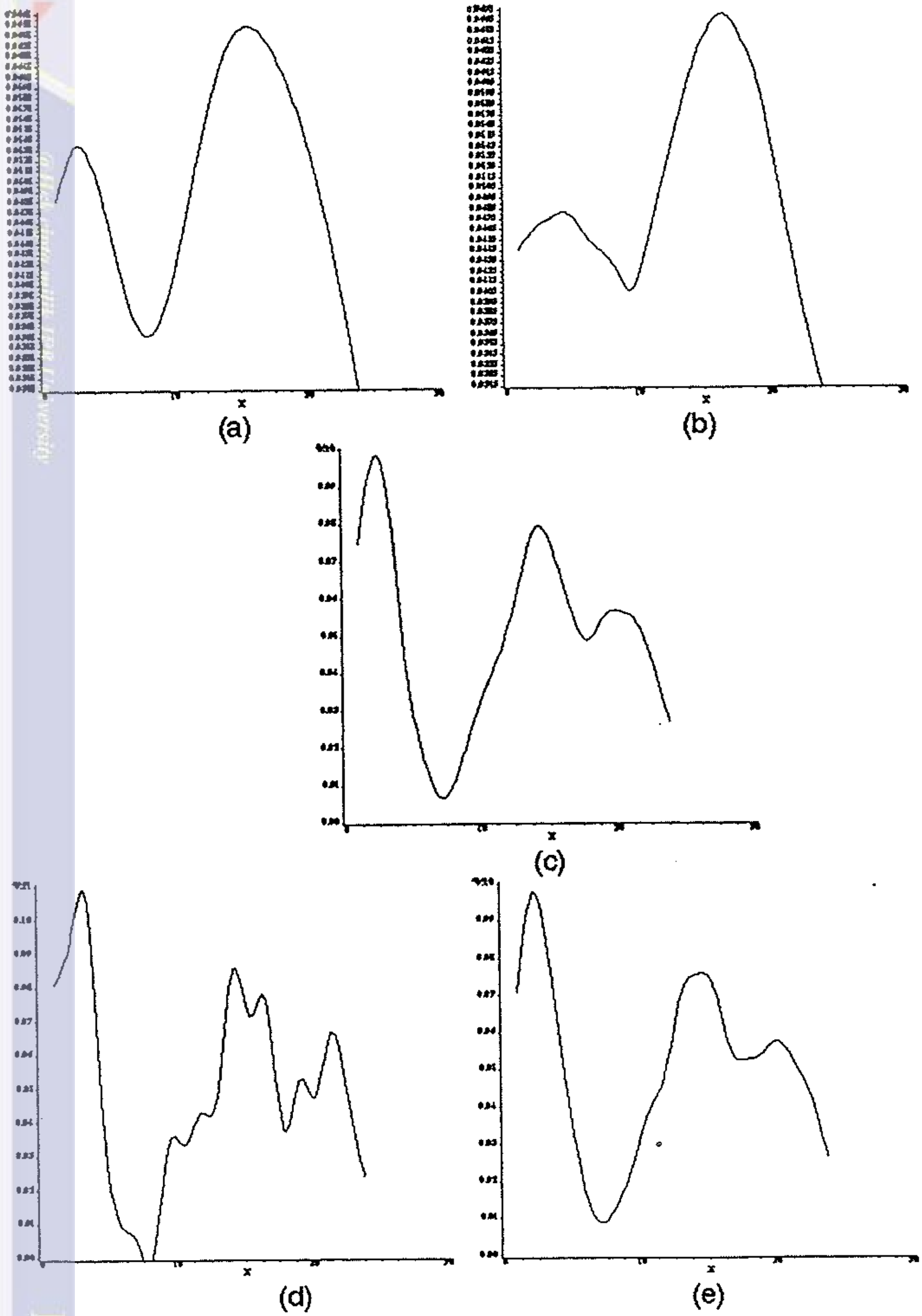
HASIL DAN PEMBAHASAN

Hasil pendugaan kepekatan peluang dengan metode kernel bagi peubah lebar petal dapat dilihat pada Peraga 1(a-e). Lebar jendela yang digunakan sebesar 3.3.

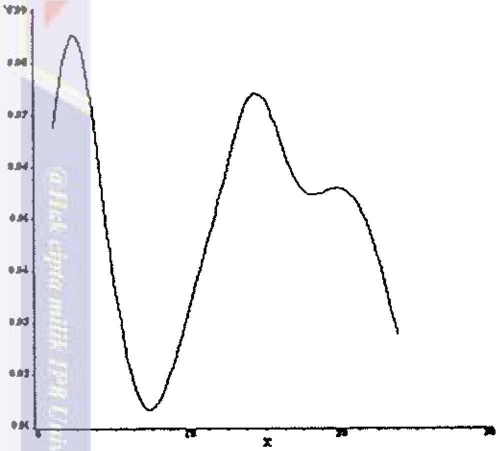
Dari kelima grafik kepekatan dugaan tersebut terlihat adanya tiga buah pola penduga kepekatan. Pada pola pertama penduga kepekatan mempunyai dua buah modus. Pola seperti ini dihasilkan kernel Gaussian dan Epanechnikov. Pola kedua adalah penduga kepekatan dengan tiga buah modus, yang dihasilkan kernel Biweight dan Triangular. Pada pola ketiga yaitu pada penduga kepekatan yang dihasilkan kernel Rectangular didapatkan tujuh buah modus.

Melalui cara uji coba ditemukan bahwa penduga kepekatan dengan kernel Gaussian akan mempunyai tiga modus pada saat lebar jendela 1.7 (Peraga 2) dan kernel Epanechnikov pada saat lebar jendela 1.8 (Peraga 3). Dengan lebar jendela 1.8 kernel Biweight menghasilkan penduga kepekatan dengan tiga buah modus, sedangkan dengan kernel Triangular dan Rectangular masing-masing lima dan tujuh buah modus (Peraga 4).

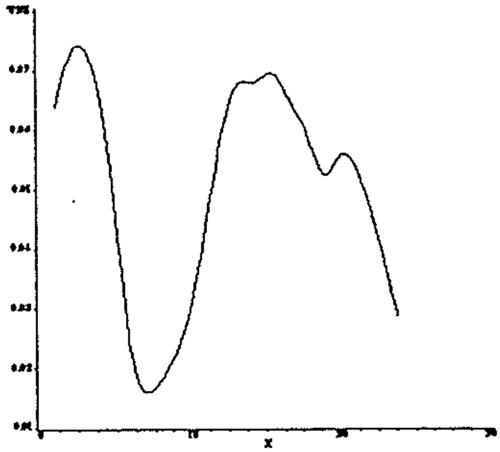
Dengan demikian pada lebar jendela yang sama fungsi kernel Gaussian dan Epanechnikov menghasilkan penduga kepekatan yang lebih mulus dibandingkan dengan yang dihasilkan ketiga kernel lainnya.



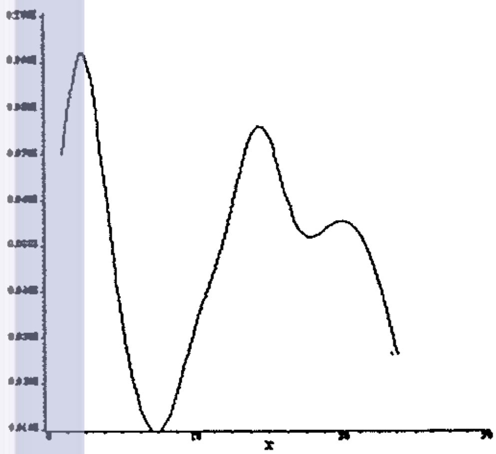
Peraga 1. Penduga Kepekatan Peluang dengan h=3.3
(a) Kernel Gaussian (b) Kernel Epanechnikov
(c) kernel Biweight (d) Kernel Rectangular
(e) Kernel Triangular



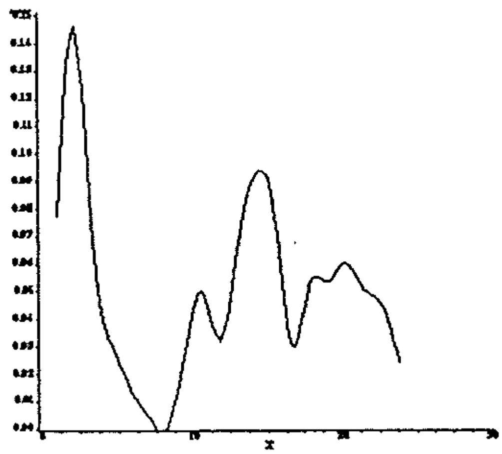
Peraga 2. Penduga Kepekatan Kernel Gaussian h=1.7



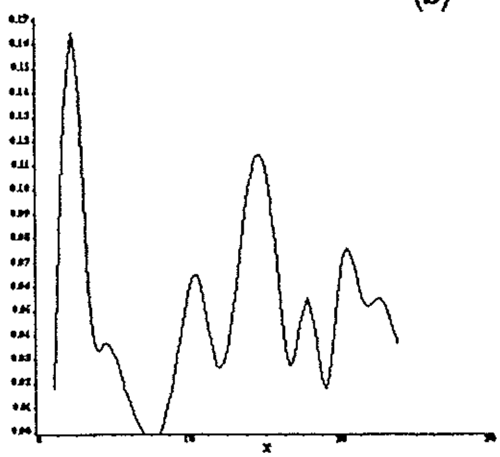
Peraga 3. Penduga Kepekatan Kernel Epanechnikov h=1.8



(a)



(b)



(c)

Peraga 4. Penduga Kepekatan dengan h=1.8
(a) Kernel Biweight (b) Kernel Triangular (c) Kernel Rectangular

Perbedaan kemulusan pola penduga kepekatan pada lebar jendela yang sama tidak menunjukkan bahwa fungsi kernel yang satu lebih baik jika dibandingkan dengan fungsi kernel lainnya. Karena lebar jendela yang digunakan bukan lebar jendela yang optimum untuk masing-masing fungsi kernel. Penyebab perbedaan tersebut sebenarnya adalah sifat dari masing-masing fungsi kernel itu sendiri.

Fungsi kernel dapat juga dianggap sebagai fungsi pembobot yang menentukan seberapa besar pengamatan (X_i) mempengaruhi besarnya kepekatan peluang pada titik pendugaan (X)

Fungsi pembobot Biweight-Tukey sering digunakan dalam pendugaan nilai tengah yang resisten (Aunuddin, 1988). Sama halnya seperti pada pendugaan nilai tengah, kernel Biweight akan menghasilkan penduga kepekatan yang resisten, artinya tidak terlalu peka terhadap pengamatan yang terletak jauh dari titik pendugaan. Pengamatan yang mempengaruhi besarnya kepekatan peluang dibatasi, yaitu X yang memenuhi $(X - X_i)/h < |1|$. Hal yang sama juga berlaku pada fungsi kernel Triangular dan Rectangular.

Pada pendugaan kepekatan dengan kernel Gaussian seluruh pengamatan mempunyai pengaruh dalam pendugaan, meskipun pengaruh tersebut semakin kecil dengan semakin jauhnya data dari titik pendugaan. Jadi, nilai dugaan tidak resisten terhadap pengamatan yang terletak jauh dari titik pendugaan.

Kernel Epanechnikov sebenarnya memiliki bentuk yang tidak jauh berbeda dengan Biweight. Tetapi fungsi ini memiliki batasan yang lebih lebar yaitu $(X-X_i)/h < |\sqrt{5}|$, sehingga penduga kepekatan yang dihasilkannya lebih peka terhadap pengamatan yang jauh.

Pada peraga 1 juga terlihat bahwa fungsi kernel Rectangular menghasilkan penduga kepekatan yang sangat kasar. Bila dilihat dari bentuk fungsinya (Tabel 1) maupun pola fungsi (Gambar Lampiran 1) maka pendugaan dengan kernel Rectangular mempunyai sifat yang sama dengan histogram, yaitu menganggap setiap pengamatan dalam lebar jendela yang sama akan memberikan sumbangan yang sama terhadap dugaan kepekatan berapapun besarnya pengamatan tersebut. Pada pendugaan dengan keempat fungsi kernel lainnya pengamatan pada lebar jendela yang sama memberikan sumbangan yang berbeda-beda, tergantung dari letaknya terhadap titik pendugaan (lihat Gambar Lampiran 1). Karena itulah pola penduga kepekatan kernel Rectangular sangat berbeda bila dibandingkan dengan yang dihasilkan keempat kernel lainnya.

Pendugaan kepekatan pada data peubah ganda dua menggunakan peubah lebar petal dan panjang petal. Matriks ragam peragam dari kedua peubah ini adalah :

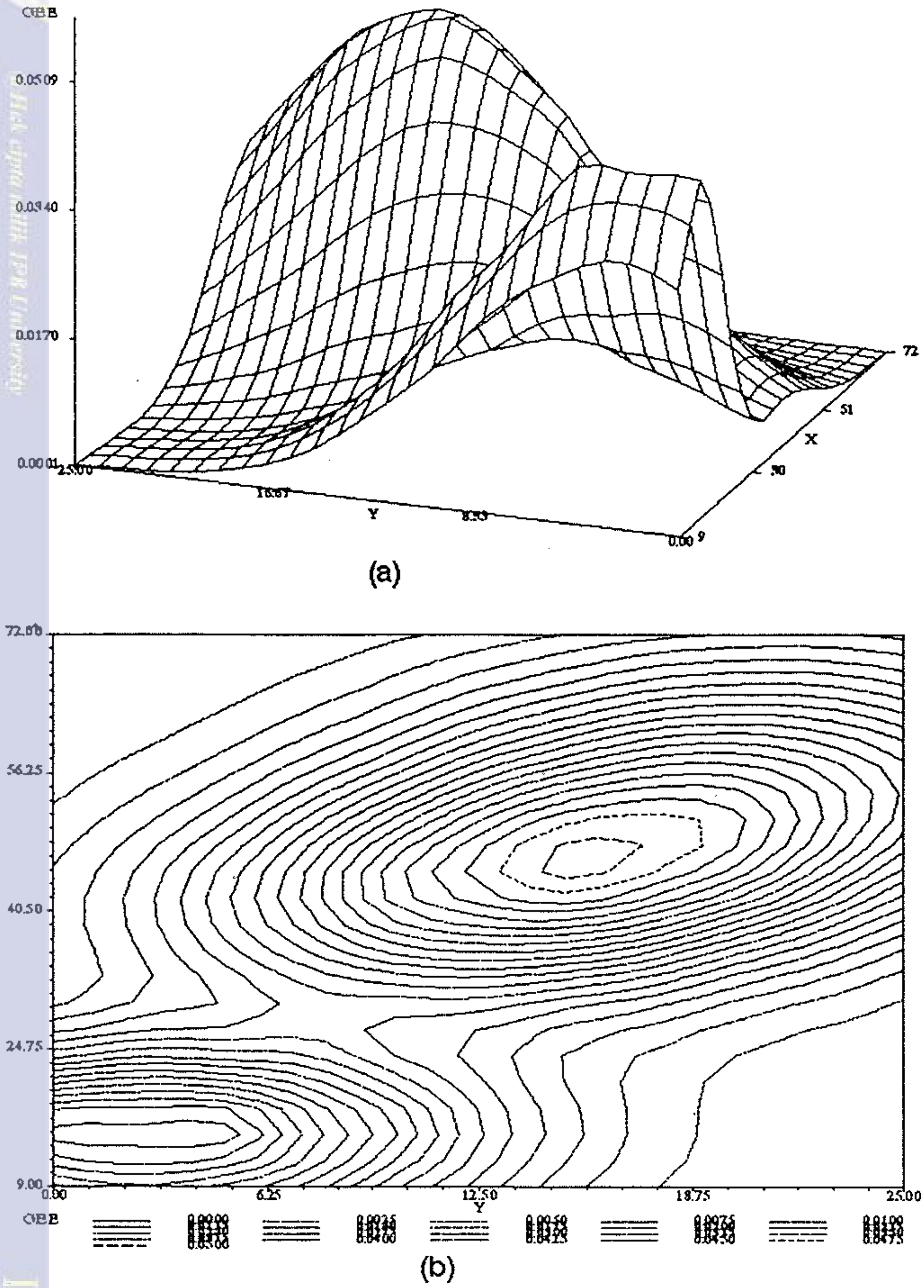
$$\begin{bmatrix} 55.98 & 129.45 \\ 129.45 & 320.42 \end{bmatrix}$$

Peraga 5 dan 6 memperlihatkan penduga kepekatan peluang menggunakan kernel Gaussian dan Epanechnikov dengan lebar jendela optimum untuk masing-masing fungsi kernel.

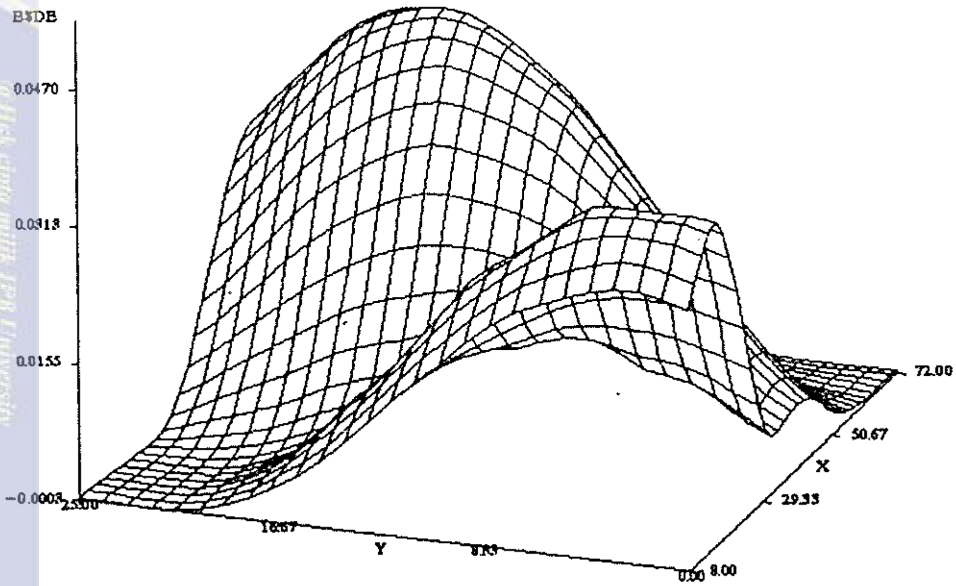
Kedua penduga kepekatan tersebut mempunyai pola yang serupa yaitu kurva mulus dengan dua buah modus. Meskipun pada penelitian ini tidak dilakukan pengukuran besarnya MISE karena keterbatasan sarana, namun pola yang serupa menunjukkan bahwa penduga kepekatan yang dihasilkan kernel Gaussian maupun Epanechnikov tidak jauh berbeda. Hal ini sesuai dengan pernyataan Silverman (1986) yang mengatakan bahwa penduga kepekatan yang dihasilkan fungsi kernel Gaussian, Epanechnikov, Biweight, Triangular dan Rectangular tidak mempunyai perbedaan yang menyolok berdasarkan MISE.

Peubah lebar dan panjang petal mempunyai perbedaan keragaman yang cukup tinggi sehingga dilakukan pendugaan kepekatan dengan transformasi agar setiap peubah mempunyai ragam yang sama yaitu satu..

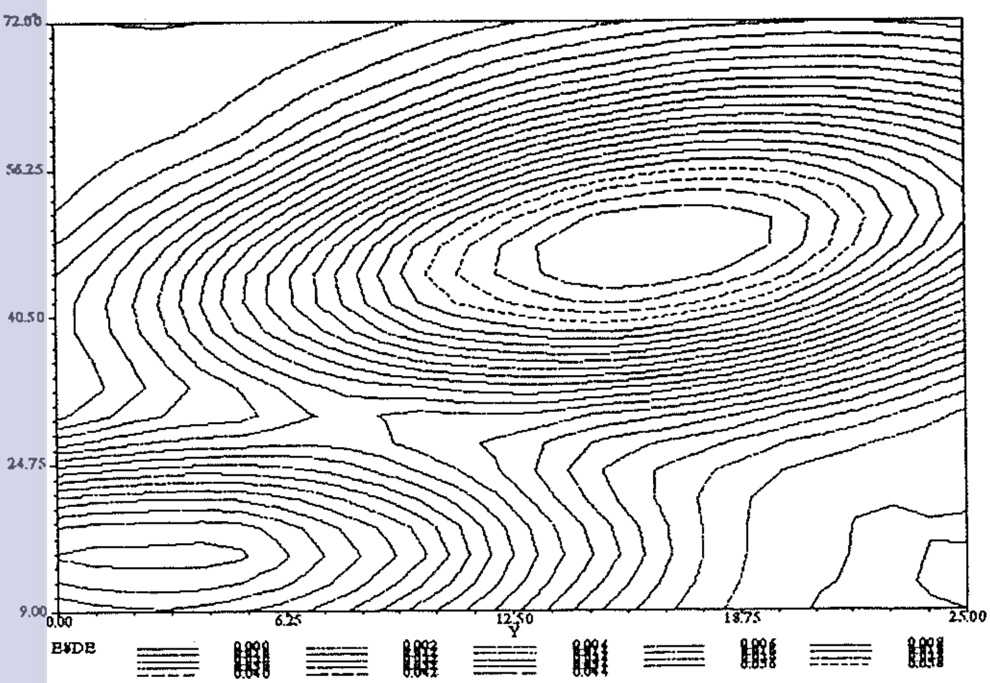
Hasil pendugaan dengan kernel Gaussian terhadap data yang telah ditransformasi ke dalam normal baku dapat dilihat pada peraga 7. Penduga kepekatan terlihat mempunyai pola yang sama dengan pendugaan tanpa transformasi, yaitu bermodus dua.



Peraga 5. Penduga Kepekatan dengan Kernel Gaussian h=6.7
(a) Plot Tiga Dimensi
(b) Plot Contour

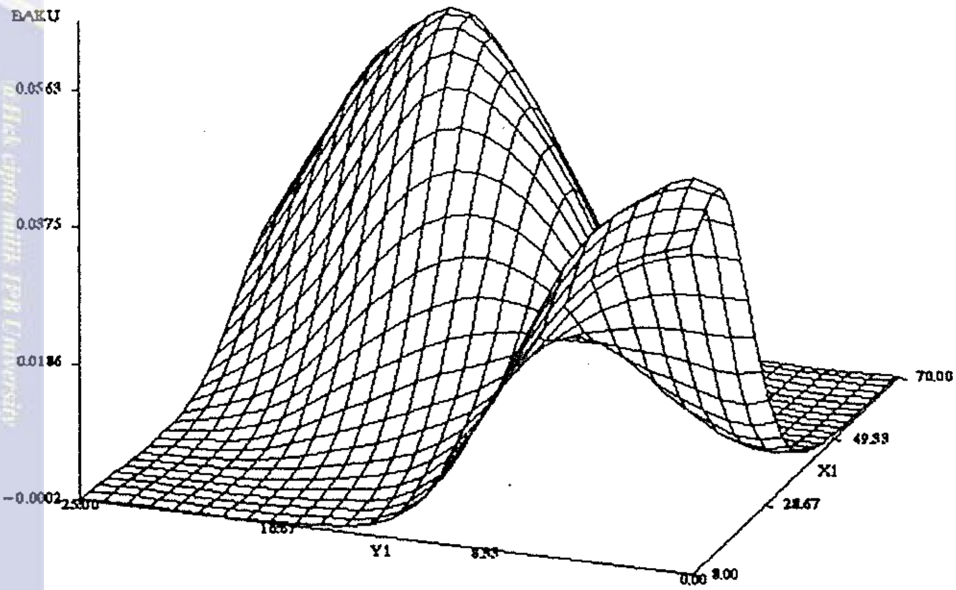


(a)

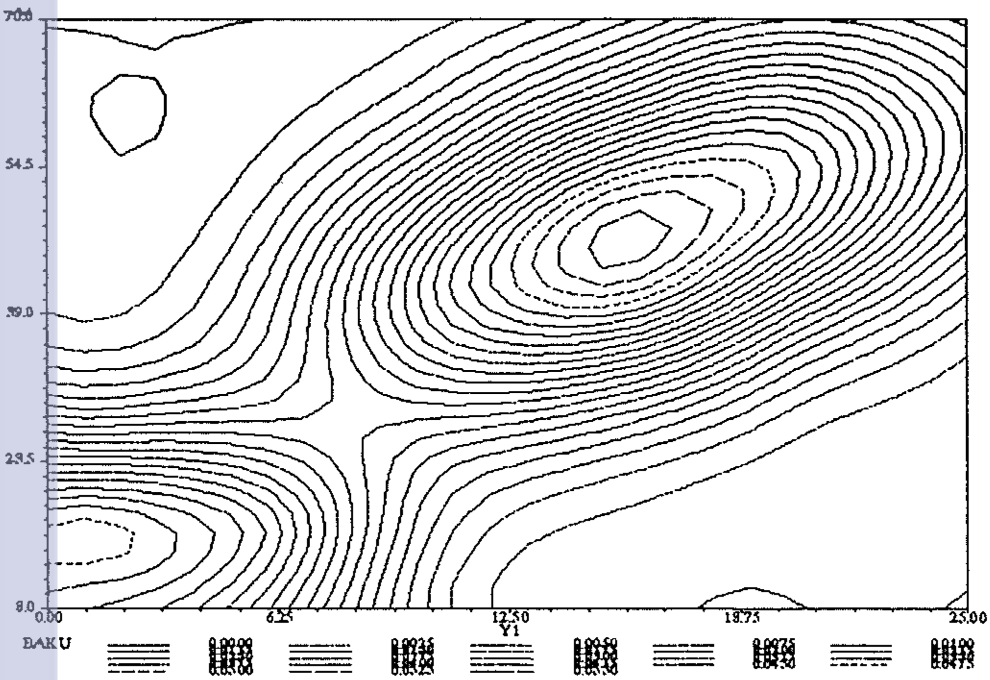


(b)

Peraga 6. Penduga Kepekatan dengan Kernel Epanechnikov h=18.7
(a) Plot Tiga Dimensi
(b) Plot Contour



(a)

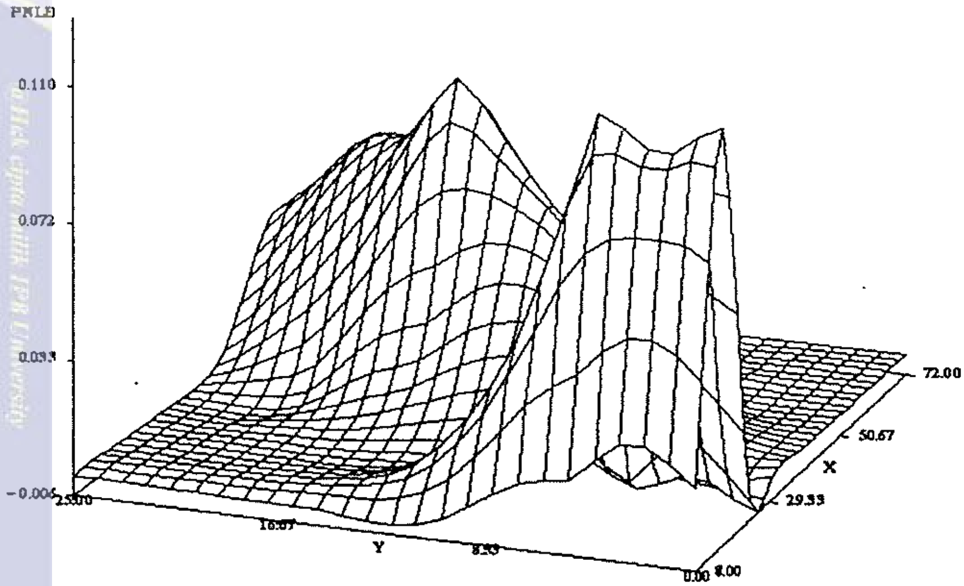


(b)

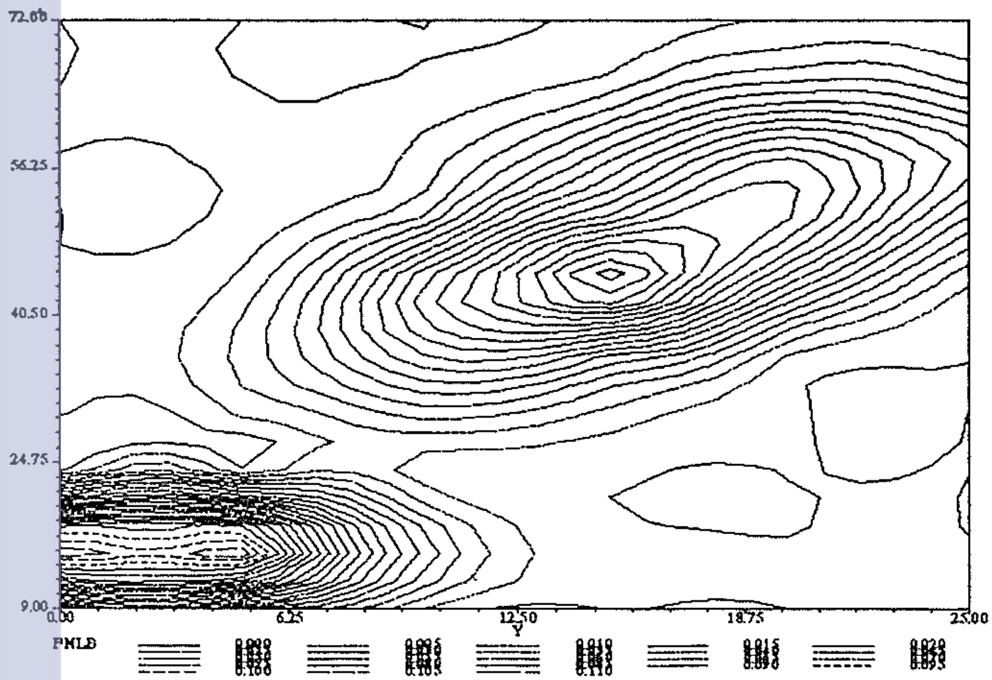
Peraga 7. Penduga Kepekatan Kernel Gaussian h=0.5
dengan Transformasi Normal Baku
(a) Plot Tiga Dimensi (b) Plot Contour

Peraga 8 memperlihatkan hasil pendugaan kepekatan dengan metode yang disarankan Fukunaga. Matriks S yang digunakan adalah matriks yang unsur-unsurnya diambil dari diagonal matriks ragam peragam. Penduga kepekatan ini mempunyai pola yang berbeda dengan ketiga penduga kepekatan sebelumnya. Pada penduga kepekatan sebelumnya didapatkan dua buah modus sedangkan pada penduga kepekatan ini diperoleh empat buah modus.

Penduga kepekatan mana yang terbaik belum dapat ditentukan pada penelitian ini karena besarnya nilai MISE tidak dihitung. Namun berdasarkan pengamatan secara visual terlihat bahwa pendugaan kepekatan dengan metode transformasi Fukunaga menghasilkan kepekatan yang lebih kasar dibandingkan dengan metode kernel biasa pada lebar jendela yang optimum untuk masing-masing metode.



(a)



(b)

Peraga 8. Penduga Kepekatan dengan Kernel Gaussian
(metode Fukunaga) h=0.5 (a) Plot Tiga Dimensi
(b) Plot Contour

KESIMPULAN

Pada penelitian ini telah dilakukan pendugaan kepekatan dengan metode kernel pada peubah tunggal dan ganda dua menggunakan data lebar dan panjang petal.

Pada peubah tunggal, kernel Gaussian dan Epanechnikov menghasilkan penduga kepekatan yang lebih mulus dibandingkan dengan kernel Triangular, Biweight dan Rectangular pada lebar jendela yang sama. Perbedaan hasil yang didapatkan disebabkan karena sifat kernel Triangular, Biweight dan Rectangular yang tidak peka atau resisten terhadap pengamatan yang terletak jauh dari titik pendugaan. Perbedaan tersebut tidak menunjukkan bahwa fungsi kernel yang satu lebih baik dari fungsi kernel lainnya.

Hasil yang diperoleh pada peubah ganda dengan lebar jendela yang optimum untuk masing-masing metode memperlihatkan bahwa pendugaan dengan metode transformasi Fukunaga menghasilkan penduga kepekatan yang lebih kasar dibandingkan dengan metode kernel biasa pada data asli maupun pada data yang telah dinormalkan.

Penduga kepekatan kernel pada peubah tunggal dan ganda dua pada lebar jendela optimum hampir seluruhnya menunjukkan bahwa bunga iris bermodus dua.

DAFTAR PUSTAKA

Aunuddin. (1988). *Analisis Data dengan Pendekatan Eksploratif.* Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.

Izenman, A. J. (1991). "Recent Developments in Nonparametric Density Estimation", *Jour. Amer. Statist. Assoc.* 86(413), 205-224

SAS Institute Inc. (1987). *SAS/STAT Guide for Personal Computers, Version 6 Edition.* Cary, NC:SAS Institute Inc.

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

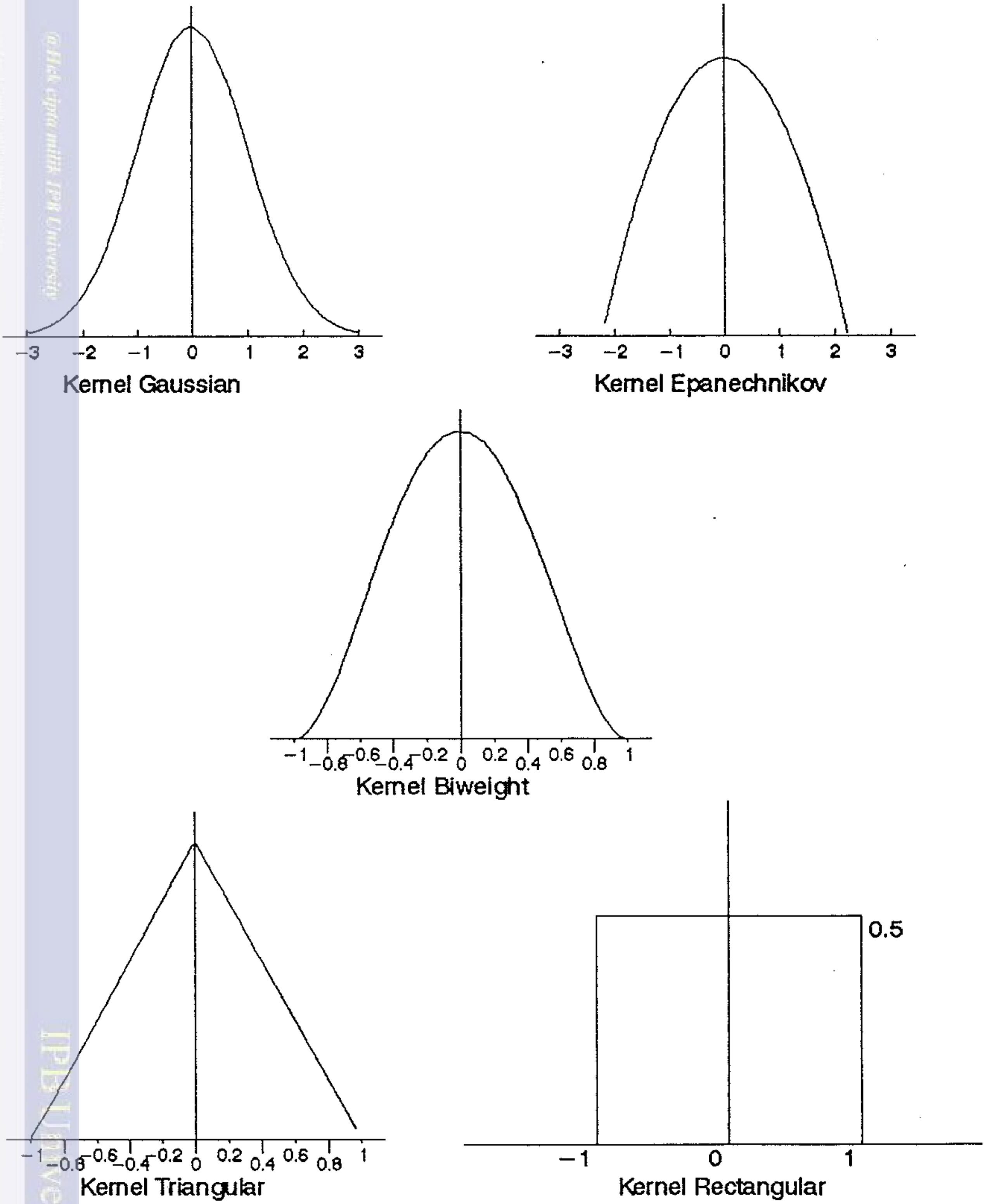
Tarter, M. E. dan R. A. Kronmall. (1976). "An Introduction to the Impementation and Theory of Nonparametrics Density Estimation", *Am. Stat.*, 30, 105-112.

Therrien, C. W. (1989). *Decision Estimation and Classification, An Introduction to Pattern Recognition and Related Topics.* John Wiley, Sons. New York.

Wand, M. P., J. S. Marron dan D Ruppert. (1991). "Transformation in Density Estimation", *Jour. Amer. Statist. Assoc.*, 86(414), 343-353.

LAMPIRAN

Gambar Lampiran 1. Pola Beberapa Fungsi Kernel



Lampiran 1. Program Macro Pendugaan Kepekatan pada Peubah Tunggal dengan Kernel Gaussian

File GAUSS.MTB

```
NOECHO
NOTE  PROGRAM INI ADALAH UNTUK PENDUGAAN FUNGSI KEPEKATAN
NOTE        DENGAN FUNGSI PEMULUS GAUSS (NORMAL)
NOTE  INFORMASI YANG DIHARAPKAN DARI PENGGUNA ADALAH
NOTE       1.  DATA YANG DISIMPAN DALAM C1
NOTE       2.  LEBAR JENDELA  (h)
NOTE  (semakin besar h kurva-nya akan semakin halus)
NOTE       3.  BANYAK GRID YANG DIPAKAI DALAM GRAFIK
NOTE          Berapa lebar jendela (h) yang diinginkan ?
set 'terminal' c40;
nobs=1.
copy c40 k99
NOTE                  BERAPA BANYAK GRID ....
SET 'TERMINAL' C40;
NOBS=1.
COPY C40 K2
let k1=count(c1)
let k3=min(c1)
let k4=max(c1)
let k5=(k4-k3)/k2
let k6=1
let c2=1
let c3=1
let c4=1
let c5=1
let k7=k3
exec 'GAUSS1' k2
let k8=k1*k99
let c4=c4/k8
LET K9=SUM(C4)     # JUMLAH FREKUENSI SELURUH GRID
LET C4=C4/K9       # RELATIF FREKUENSI PER UNIT GRID
name c4='densitas'
name c5='peubah'
NOTE SELANJUTNYA BUATLAH PLOT ANTARA C4 DAN C5
NOTE           UNTUK MELIHAT HASILNYA
NOTE GPLOT C4 C5;
NOTE LINE C4 C5.
end
```

File GAUSS1.MTB

```
let c2=(c1-k7)/k99
LET C2=C2*C2
LET C3=EXP(-0.5*C2)
let k10=sum(c3)
let c4(k6)=k10
let c5(k6)=k7
let k6=k6+1
let k7=k7+k5
end
```

Lampiran 2. Program Macro Pendugaan Kepekatan pada Peubah Ganda Dua dengan Kernel Gaussian

File GAUSS.BIV

```
noecho
note PROGRAM PENDUGAAN FUNGSI KEPEKATAN
note   DENGAN FUNGSI PEMULUS GAUSSIAN
note        UNTUK DATA BIVARIATE
note DATA YANG DIHARAPKAN DARI PEMAKAI
NOTE          PEUBAH 1 PADA  C1
NOTE          PEUBAH 2 PADA  C2
note          LEBAR JENDELA
note (banyaknya grid yang digunakan = 60)
note lebar jendela yang diinginkan (h) ?
Set 'terminal' c20;
nobs = 1.
copy c20 k1
let k3 = count(c1)
let k7 = 1
let k8 = 1
exec 'gauss1.biv'
exec 'gauss2.biv' 10
let k10 = k1*60
let c10 = c10/k10
let k11 = sum(c10)
let c10 = c10/k11
NOTE                  C10  =  DENSITAS
NOTE                  C11  =  PEUBAH PERTAMA (X1)
NOTE                  C12  =  PEUBAH KEDUA   (X2)
```

File GAUSS1.BIV

```
let k4 = min(ck7)
let k5 = max(ck7)
let k6 = (k5-k4)/9
let c4(k7) = round(k6)
let c3(k7) = round(k4)-1
let k4 = min(c2)
let k5 = max(c2)
let k6 = (k5-k4)/5
let c4(2) = round(k6)
let c3(2) = round(k4)-1
```

File GAUSS2.BIV

```
let c5 = (c3(1) - c1)/k1
exec 'gauss3.biv' 6
let c3(1) = c3(1) + c4(1)
let c3(2) = round(k4) - 1
```

File GAUSS3.BIV

```
let c6 = (c3(2) - c2)/k1
let c7 = c5*c5 + c6*c6
let c8 = exp(-0.5*c7)
let c10(k8) = sum(c8)
let c11(k8) = c3(1)
let c12(k8) = c3(2)
let c3(2) = c3(2)+c4(2)
let k8 = k8 +1
```