Jurnal Matematika UNAND
Vol. **5** No. **2** Hal. 1 – 10
ISSN : 2303–291X

BAGAN KENDALI NONPARAMETRIK DENGAN ESTIMASI FUNGSI KEPEKATAN KERNEL (STUDI KASUS: INDEKS PRESTASI MAHASISWA JURUSAN MATEMATIKA ANGKATAN 2011-2013 FMIPA UNAND PADA SEMESTER GANJIL 2015-2016)

EWI JUPIT, MAIYASTRI, HAZMIRA YOZZA
Program Studi Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email : ewijupit@yahoo.co.id

Abstrak. Bagan kendali nonparametrik dengan estimasi fungsi kepekatan kernel digunakan untuk membuat bagan kendali untuk data tidak normal. Data IP mahasiswa Jurusan Matematika angkatan 2011-2013 pada semester ganjil 2015/2016 tidak menyebar normal, oleh sebab itu, digunakan bagan kendali dengan estimasi fungsi kepekatan kernel. Hasil penelitian ini menunjukkan bahwa bagan kendali dengan estimasi fungsi kepekatan kernel Triangular paling baik digunakan untuk data IP tersebut.

Kata Kunci: Bagan kendali nonparametrik, estimasi fungsi kepekatan kernel

1. Pendahuluan

Keunggulan suatu perguruan tinggi dapat dilihat dari proses perkuliahan yang berkualitas. Salah satu ukuran penentu kualitas selama proses perkuliahan pada tiap semester dari suatu perguruan tinggi adalah Indeks Prestasi (IP). Demi menjaga konsistensi atau kestabilan nilai Indeks Prestasi (IP) khususnya Jurusan Matematika angkatan 2011-2013 FMIPA UNAND perlu dilakukan pengendalian kualitas.

Bagan kendali *Shewhart* adalah salah satu alat pengendalian kualitas yang mempunyai ketentuan bahwa data harus berdistribusi normal. Namun, untuk data yang tidak berdistribusi normal bagan kendali Shewhart tidak bisa digunakan. Sebagai solusinya dikembangkan suatu bagan kendali yang tidak membutuhkan asumsi distribusi normal yaitu bagan kendali dengan pendekatan nonparametrik yang didasarkan pada pendekatan kernel.

Bagan kendali dibangun dengan terlebih dahulu mengestimasi bentuk fungsi kepekatan peluang dari data dengan metode estimasi fungsi kepekatan kernel. Faktor yang mempengaruhi estimasi fungsi kepekatan kernel yaitu fungsi kernel dan *bandwidth (h)*, namun yang yang paling berperan dalam menentukan kemulusan dan keakuratan grafik estimasi fungsi kepekatan kernel adalah pemilihan *bandwidth (h)* yang tepat (optimal). Berdasarkan lebar batas kendali, dilakukan perbandingan bagan kendali dan menentukan bagan kendali terbaik pada data sampel.

2. Tinjauan Teori

2.1. *Estimasi Fungsi Kepekatan Kernel*

Estimasi fungsi kepekatan kernel merupakan salah satu metode nonparametrik untuk menduga fungsi kepekatan peluang dari suatu peubah acak [1]. Rumus estimasi fungsi kepekatan kernel dapat dituliskan sebagai berikut :

$$\hat{f}(x) = \frac{1}{hn} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{2.1}$$

dimana

$$\begin{aligned} n &: \text{ ukuran sampel} \\ h &: \text{ parameter pemulus (bandwidth)} \\ x &: \text{ nilai acak yang akan diestimasi} \\ X_i &: \text{ titik sampel ke-i}, i = 1, 2, \cdots, n \\ K &: \text{ fungsi kernel.} \end{aligned}$$

Untuk menentukan estimasi fungsi kepekatan kernel, dapat digunakan berbagai bentuk fungsi kernel dengan mengganti fungsi K dengan fungsi kernel lainnya.

2.2. Fungsi Kernel

Menurut [11], suatu fungsi $K(z)$ untuk $-\infty < z < \infty$ dengan $z = (x - X_i)/h$ dan $h > 0$ disebut fungsi kernel jika memenuhi sifat-sifat berikut.

(1) $K(z) \geq 0$ untuk setiap z.
(2) $K(z) : R \to R$.
(3) $\int_{-\infty}^{\infty} K(z)dz = 1$.
(4) $\int_{-\infty}^{\infty} zK(z)dz = 0$.
(5) $\int_{-\infty}^{\infty} z^2 K(z)dz < \infty$
(6) Simetri dimana $K(z) = K(-z)$ untuk setiap z.

Berikut adalah beberapa fungsi kernel.

Nama Kernel	**Bentuk Fungsi** $K(z)$
Epanechnikov	$K(z) = \{\frac{3}{4}\frac{1-0,2z^2}{\sqrt{5}}$ untuk $\mid z \mid< \sqrt{5}$ dan 0 lainnya
Biweight	$K(z) = \{\frac{15}{6}(1 - z^2)^2$ untuk $\mid z \mid< 1$ dan 0 lainnya
Triangular	$K(z) = \{1- \mid z \mid$ untuk $\mid z \mid< 1$ dan 0 lainnya
Gaussian	$K(z) = \{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}}z^2$
Rectangular	$K(z) = \{\frac{1}{2}$ untuk $\mid z \mid< 1$ dan 0 lainnya

2.3. *Kebaikan Estimasi Fungsi Kepekatan Kernel*

Menurut [6], salah satu metode pemilihan *bandwidth (h)* optimal adalah *Cross Validation* (CV). Metode ini digunakan untuk menduga kesalahan estimasi. *Bandwidth* dapat didekati secara numerik dengan memilih h yang meminimalkan

$$CV(h) = \int_{-\infty}^{\infty} \hat{f}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(X_i). \tag{2.2}$$

2.4. *Bagan Kendali dengan Estimasi Fungsi Kepekatan Kernel*

Batas-batas kendali untuk bagan kendali dengan estimasi fungsi kepekatan kernel, dengan $q = 0,0027$ dapat ditentukan sebagai berikut. UCL *(Upper Control Limit)* atau batas kendali atas adalah nilai x, sehingga

$$\int_{-\infty}^{UCL} \hat{f}(x)dx = \left(1 - \frac{q}{2}\right) 100. \tag{2.3}$$

Sebagai garis tengah *(Center Line)* diambil nilai median, yaitu nilai x, sehingga

$$\int_{-\infty}^{m} \hat{f}(x)dx = (\frac{1}{2})100. \tag{2.4}$$

LCL *(Lower Control Limit)* atau batas kendali bawah adalah nilai x, sehingga

$$\int_{-\infty}^{LCL} \hat{f}(x)dx = (\frac{q}{2})100. \tag{2.5}$$

3. Metode Penelitian

Berikut adalah langkah-Langkah penelitian yang dilakukan.

(1) Melakukan pengujian distribusi normal pada data sampel.
(2) Mendefinisikan masing-masing fungsi kernel dan rumus estimasi fungsi kepekatan kernel dalam bahasa pemrograman R.3.1.0.
(3) Mencari nilai bandwidth (h) optimal.
(4) Menghitung estimasi fungsi kepekatan masing-masing kernel.
(5) Menggambarkan grafik estimasi fungsi kepekatan kernel.
(6) Menentukan nilai UCL, CL dan LCL.
(7) Menggambarkan bagan kendali dengan estimasi fungsi kepekatan kernel.
(8) Membandingkan bagan kendali dengan estimasi fungsi kepekatan kernel Epanechnikov, Biweight, Triangular, Gaussian, Rectangular.

4. Hasil dan Pembahasan

4.1. *Pengujian Distribusi Normal pada Data Sampel*

Untuk mengetahui apakah data sampel berdistribusi normal atau tidak, maka dilakukan uji *Kolmogorov-Smirnov*. Perhitungan dilakukan dengan bantuan SPSS 16 dan diperoleh nilai D_{hitung} adalah sebesar 0,089. Kemudian dengan taraf nyata $\alpha = 0,05$ dan banyak data $n = 252$ diperoleh nilai Dtabel adalah sebesar 0,0857. Oleh karena $0,089 > 0,0857$ atau $_{hitung} > D_{tabel}$ maka data tolak H_0 sehingga dapat disimpulkan bahwa data sampel tidak berdistribusi normal.

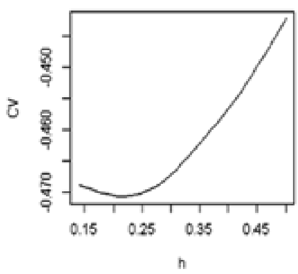
4.2. *Estimasi Fungsi Kepekatan Kernel*

4.2.1. *Pemilihan Bandwidth (h) Optimal*

Dalam menentukan h yang optimal berdasarkan CV yang minimum pada kernel Epanechnikov, ditentukan nilai awal untuk h adalah 0,14 dengan selang kenaikan nilai h sebesar 0,01. Dengan menggunakan *software* R.3.1.0 diperoleh nilai h dan CV seperti pada Tabel 1, serta grafik pada Gambar 1.

h	CV		h	CV		h	CV
0,14	-0,4689357		0,24	-0,4704044		0,34	-0,4633146
0,15	-0,4691660		0,25	-0,4701645		0,35	-0,4622591
0,16	-0,4694936		0,26	-0,4697528		0,36	-0,4611643
0,17	-0,4699507		0,27	-0,4693528		0,37	-0,4600577
0,18	-0,4701812		0,28	-0,4688009		0,38	-0,4590051
0,19	-0,4703521		0,29	-0,4680793		0,39	-0,4578552
0,20	-0,4706058		0,30	-0,4672745		0,40	-0,4566676
0,21	-0,4708020		0,31	-0,4663724		0,41	-0,4554363
0,22	-0,4707802		0,32	-0,4653836		0,42	-0,4541155
0,23	-0,4706457		0,33	-0,4643626		0,43	-0,4527601

Tabel 1. Nilai *Bandwidth (h)* dan CV dengan Kernel Epanechnikov



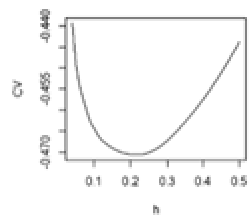
Gambar 1. Grafik CV dengan Kernel Epanechnikov

Berdasarkan Tabel 1 dan Gambar 1 terlihat bahwa nilai CV pada *bandwidth* $h = 0,21$ yaitu $-0,4708020$ benilai paling minimum diantara nilai CV pada *bandwidth* $h = 0,20$ yaitu $-0,4706058$ dan nilai CV pada *bandwidth* $h = 0,22$ yaitu $-0,4707802$. Berarti *bandwidth (h)* yang optimal sebenarnya berada disekitar $h = 0,21$. Oleh karena itu, dilakukan perhitungan nilai CV kembali disekitar $h = 0,21$. Dengan bantuan *software* R.3.1.0 diperoleh nilai CV yang paling minimum adalah $-0,4708974$ yang terletak pada nilai *bandwidth* $h = 0,2124484$, sehingga *bandwidth (h)* yang optimal untuk kernel Epanechnikov adalah $0,2124484$.

Dengan cara yang sama dapat dilakukan untuk fungsi kernel lainnya, sehingga diperoleh hasil untuk kernel Biweight pada Tabel 2 dan Gambar 2. Diperoleh h yang optimal untuk kernel Biweight adalah $0,2143558$ dengan nilai CV yang paling minimum adalah $-0,4705497$.

h	CV		h	CV		h	CV
0,04	-0,4394583		0,14	-0,4683944		0,24	-0,4702433
0,05	-0,4489806		0,15	-0,4689297		0,25	-0,4699573
0,06	-0,4541702		0,16	-0,4693677		0,26	-0,4695677
0,07	-0,4572847		0,17	-0,4697212		0,27	-0,4690785
0,08	-0,4602818		0,18	-0,4700637		0,28	-0,4685022
0,09	-0,4628003		0,19	-0,4703197		0,29	-0,4678451
0,10	-0,4642603		0,20	-0,4704712		0,30	-0,4671265
0,11	-0,4657478		0,21	-0,4705437		0,31	-0,4663497
0,12	-0,4669732		0,22	-0,4705252		0,32	-0,4655306
0,13	-0,4677357		0,23	-0,4704293		0,33	-0,4646793

Tabel 2. Nilai *Bandwidth* dan CV dengan Kernel Biweight



Gambar 2. Grafik CV dengan Kernel Biweight

Selanjutnya untuk kernel Triangular diperoleh hasil pada Tabel 3 dan Gambar 3. Diperoleh h yang optimal untuk kernel Triangular adalah $0,2132032$ dengan nilai CV yang paling minimum adalah $-0,4701527$.

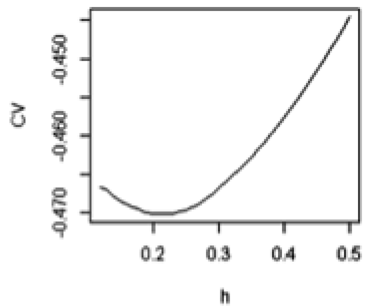
Selanjutnya untuk kernel Gaussian diperoleh hasil pada Tabel 4 dan Gambar 4. Diperoleh *bandwidth (h)* yang optimal untuk kernel Gaussian adalah $0,2150014$ dengan nilai CV yang paling minimum adalah $-0,469936$.

Selanjutnya untuk kernel Rectangular diperoleh hasil pada Tabel 5 dan Gambar 5. Diperoleh *bandwidth (h)* yang optimal untuk kernel Rectangular adalah $0,3350483$ dengan nilai CV yang paling minimum adalah $-0,4645963$.

Selanjutnya dapat digambarkan grafik estimasi fungsi kepekatan kernel dengan menggunakan *software* R.3.1.0, sehingga diperoleh grafik estimasi pada Gambar 6. Berdasarkan Gambar 6 terlihat bahwa grafik estimasi fungsi kepekatan kernel Epanechnikov dan Gaussian memiliki grafik yang paling mulus dibandingkan Biweight, Triangular dan Rectangular.

h	CV		h	CV		h	CV
0,12	-0,4666712		0,22	-0,4700965		0,32	-0,4651684
0,13	-0,4669482		0,23	-0,4700402		0,33	-0,4643475
0,14	-0,4677816		0,24	-0,4698552		0,34	-0,4635066
0,15	-0,4683465		0,25	-0,4696061		0,35	-0,4626586
0,16	-0,4687474		0,26	-0,4692215		0,36	-0,4617569
0,17	-0,4692466		0,27	-0,4687414		0,37	-0,4608056
0,18	-0,4695037		0,28	-0,4681277		0,38	-0,4597713
0,19	-0,4699226		0,29	-0,4674680		0,39	-0,4587278
0,20	-0,4700786		0,30	-0,4667333		0,40	-0,4576125
0,21	-0,470130		0,31	-0,4659624		0,41	-0,4564528

Tabel 3. Nilai *Bandwidth* dan CV dengan Kernel Triangular



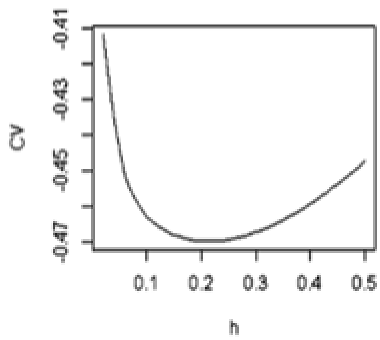
Gambar 3. Grafik CV dengan Kernel Triangular

h	CV		h	CV		h	CV
0,02	-0,4117334		0,12	-0,4656579		0,22	-0,4699228
0,03	-0,4248289		0,13	-0,4666520		0,23	-0,4698436
0,04	-0,4365760		0,14	-0,4674633		0,24	-0,469688
0,05	-0,4453423		0,15	-0,4681296		0,25	-0,4694611
0,06	-0,4513208		0,16	-0,4686688		0,26	-0,4691693
0,07	-0,4555303		0,17	-0,4691081		0,27	-0,4688092
0,08	-0,4586382		0,18	-0,4694454		0,28	-0,4683871
0,09	-0,4610422		0,19	-0,4696938		0,29	-0,4679058
0,10	-0,4629310		0,20	-0,4698519		0,30	-0,4673687
0,11	-0,4644422		0,21	-0,4699284		0,31	-0,4667769

Tabel 4. Nilai *Bandwidth* dan CV dengan Kernel Gaussian

4.3. *Batas-Batas Kendali dengan Estimasi Fungsi Kepekatan Kernel*

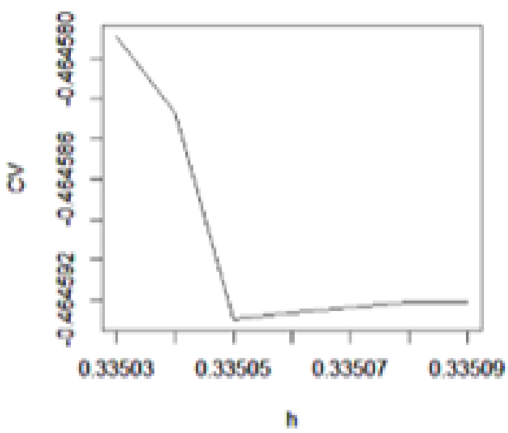
Berdasarkan persamaan (2.3), (2.4), (2.5) dan $q = 0,0027$ dengan bantuan R.3.1.0 diperoleh batas-batas kendali untuk masing-masing kernel seperti pada Tabel 6.



Gambar 4. Grafik CV dengan Kernel Gaussian

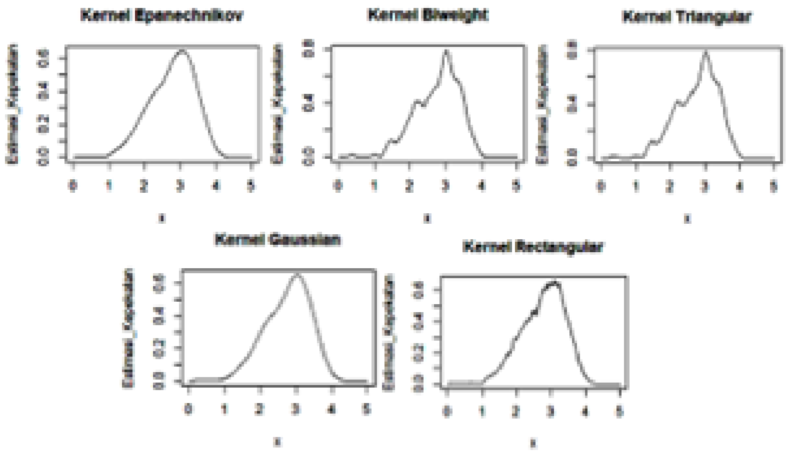
h	CV
0,33503	-0,4645789
0,33504	-0,4645827
0,33505	-0,4645930
0,33506	-0,4645926
0,33507	-0,4645924
0,33508	-0,4645921
0,33509	-0,4645921

Tabel 5. Nilai *Bandwidth* dan CV dengan Kernel Rectangular



Gambar 5. Grafik CV dengan Kernel Rectangular

Selanjutnya dapat dibangun bagan kendali berdasarkan estimasi fungsi kepekatan kernel pada Gambar 7. Dari Gambar 7 terlihat bahwa bagan kendali



Gambar 6. Grafik Estimasi Fungsi Kepekatan Kernel

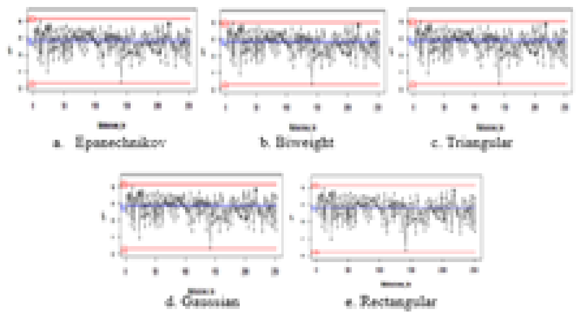
Tabel 6. Batas-Batas Kendali untuk Masing-Masing Fungsi Kernel

No	Kernel	UCL	CL	LCL
1	Epanechnikov	4,165333	2,826314	0,276667
2	Biweight	3,977407	2,851568	0,317500
3	Triangular	3,967576	2,849984	0,317500
4	Gaussian	4,201667	2,828605	0,291429
5	Ractangular	4,129167	2,827590	0,248333

yang dihasilkan dengan estimasi fungsi kepekatan kernel Epanechnikov, Biweight, Triangular, Gaussian dan Rectangular menunjukkan semua titik-titik sampel berada di dalam batas-batas kendali, sehingga dapat disimpulkan bahwa data Indeks Prestasi mahasiswa Jurusan Matematika angkatan 2011-2013 FMIPA UNAND pada semester ganjil 2015/2016 berada dalam keadaan terkendali atau terkendali secara statistik.

4.4. *Membandingkan Bagan Kendali Nonparametrik dengan Estimasi Fungsi Kepekatan Kernel yang Berbeda*

Melalui Tabel 6 dapat dicari lebar batas kendali pada bagan kendali dengan mencari selisih dari UCL dan LCL, sehingga diperoleh lebar batas kendali untuk kernel Epanechnikov sebesar $3,888666$, kernel Biweight sebesar $3,659907$, kernel Triangular sebesar $3,650076$, kernel Gaussian sebesar $3,910238$ dan kernel Rectangular sebesar $3,880834$. Oleh karena itu, bagan kendali dengan estimasi fungsi kepekatan kernel Triangular lebih baik karena memiliki lebar batas kendali yang lebih sempit sehingga lebih sensitif dalam mengidentifikasi titik-titik sampel pada kasus ini.



Gambar 7. Bagan Kendali dengan Estimasi Fungsi Kepekatan Kernel

5. Kesimpulan

Dari hasil pembahasan yang telah dilakukan dapat diperoleh beberapa kesimpulan sebagai berikut.

(1) Dari bagan kendali dengan estimasi fungsi kepekatan kernel diketahui bahwa tidak terdapat titik sampel yang berada di luar batas kendali, sehingga dapat disimpulkan bahwa proses perkuliahan mahasiswa Jurusan Matematika angkatan 2011-2013 FMIPA UNAND yang dilihat dari Indeks Prestasi semester ganjil periode 2015/2016 berada dalam keadaan terkendali atau terkendali secara statistik.

(2) Dari perbandingan kelima bagan kendali tersebut, jika dilihat dari lebar batas kendalinya maka bagan kendali dengan estimasi fungsi kepekatan kernel Triangular lebih baik jika dibandingkan dengan bagan kendali Epanechnikov, Biweight, Gaussian dan Rectangular karena lebar batas kendali yang diperoleh lebih sempit, sehingga lebih sensitif dalam mengidentifikasi titik-titik sampel pada kasus ini.

Daftar Pustaka

[1] Anonymous. 2016. Kernel Density Estimation. http://en.wikipedia.org/wiki/Kernel-density-estimation. diakses 9 Maret 2016.
[2] Bain, L.J and M. Engelhardt. 1992. *Introduction to Probability and Mathematical Statistics. Second Edition.* Duxbury Press, United States of America.
[3] Hansen, B.E. 2009. *Lecture Notes on Nonparametrics.* University Wisconsin, Spring.
[4] Montgomery, D.C.1990. *Pengantar Pengendalian Kualitas Statistik.* Gadjah Mada University Press, Yogyakarta.
[5] Oaklan, J.S. 2003. *Statistical Process Control Fifth Edition.* MPG Books Limited, Great Britian.
[6] Santoso, Rukun. 2008. Grafik pengendali nonparametrik empirik. *Media Statistika.* **1**(2): 83 – 90
[7] Steel, R. G. D., Torrie, J. H., and Dickey, D. A. (1996). *Principles and Procedure of Statistics : A Biometrical Approach,* 3rd ed. New York : McGraw-Hill.

[8] Taungke, Novriyanthi., Adi, Setiawan. dan Hanna, A.P. 2011. Pengendalian Kualitas Produk Untuk Karakteristik pH dengan Menggunakan Grafik Pengendali Berdasarkan Densitas Kernel. http://adisetiawan26.files.wordpress.com.2012/08/Pengendalian-kualitas.pdf. diakses 12 September 2015.

[9] Universitas Trisakti. BAB II Proses Belajar Mengajar. http://www.trisakti.ac.id/ftke/files/03BAB2BELAJARMENGAJAR2012-2013.pdf. diakses 1 Maret 2016.

[10] Vermaat, M.B., R.A. Ion., R.J.M.M. Does., C.A.J. Klaassen. 2003. A comparison of Shewhart individuals control chart based on normal, *Non-Parametric and Extreme-Value Theory* **19**(4) : 337 − 353

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