Estimation de densités par noyaux

Cuvelier F.

6 juin 2009

Table des matières

1 Rappels

Soit $X = (X_1, \ldots, X_N)$ un N-échantillon de variables aléatoires indépendantes et identiquement distribuées selon une loi μ de densité f inconnue.

On veut illustrer les techniques d'estimation par noyaux introduitent par Rosenblatt (1956) et Parzen (1962).

Les estimateurs $\hat{f}_h(x)$ sont définis par

$$\hat{f}_{N,h}(x) = \frac{1}{N}\sum_{i=1}^{N} K_h(x - X_i) \tag{0.1}$$

où h est la largeur de fenêtre et K le noyau, $K_h(u) = K(u/h)/h$.

Habituellement, on impose les conditions suivantes sur K :

Propriété 1.1. *On appelle noyau d'ordre k, l'application K de $\mathbb{R}$ dans $\mathbb{R}$ vérifiant*

1. *K est symétrique, $K(u) = K(-u)$,*
2. $\int_{\mathbb{R}} K(u)du = 1$
3. $\int_{\mathbb{R}} u^j K(u)du = 0, \ \forall j \in [\![1, k-1]\!]$
4. $\int_{\mathbb{R}} u^k K(u)du \neq 0$

Pour $k = 2$, on peut imposer à K d'être positif, i.e. K est alors une densité de probabilité. Pour $k > 2$, il est necessaire que K puisse prendre des valeurs négatives (Berwin A. Turlach).

Voici quelques noyaux usuels d'ordre 2, où l'on note $I(\bullet)$ la fonction indicatrice :

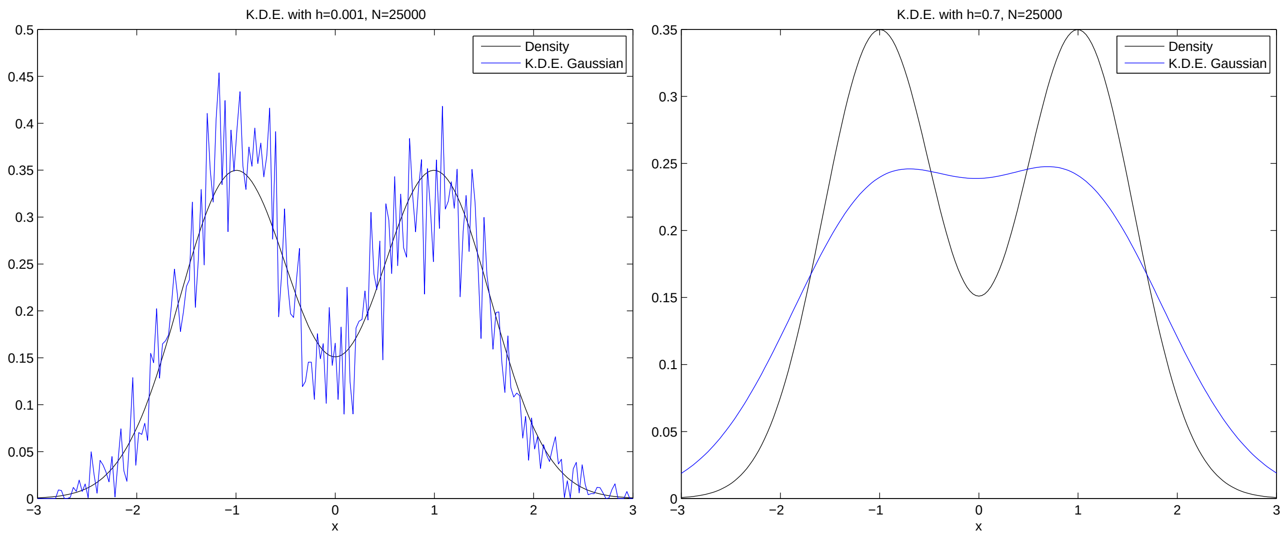


FIG. 1 – Mauvais choix du paramètre h : trop petit (gauche), trop grand (droite)

Noyau d'ordre 2	$K(u)$
Uniform	$\frac{1}{2}I(\|u\| \leqslant 1)$
Triangle	$(1-\|u\|)I(\|u\| \leqslant 1)$
Epanechnikov	$\frac{3}{4}(1-u^2)I(\|u\| \leqslant 1)$
Quartic	$\frac{15}{16}(1-u^2)^2 I(\|u\| \leqslant 1)$
Triweight	$\frac{35}{32}(1-u^2)^3 I(\|u\| \leqslant 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)$

Il ressort de Silverman (1986), le fait que le choix de h est plus important pour le comportement de $\hat{f}_{N,h}$ que le choix du noyau K. De petites valeurs de h font que l'estimation est fortement <bruitée> et de grandes valeurs régularisent de trop l'estimation (voir figure 1).

Avec un paramètre h optimisé, on obtient la figure 1

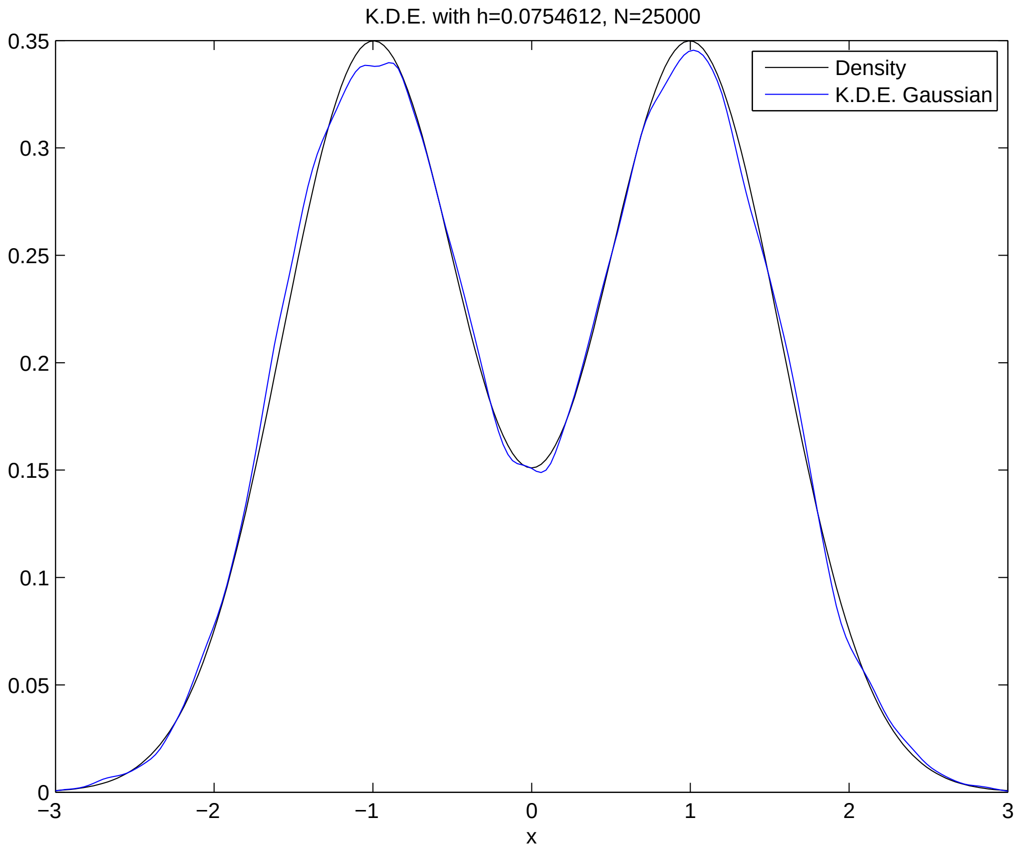
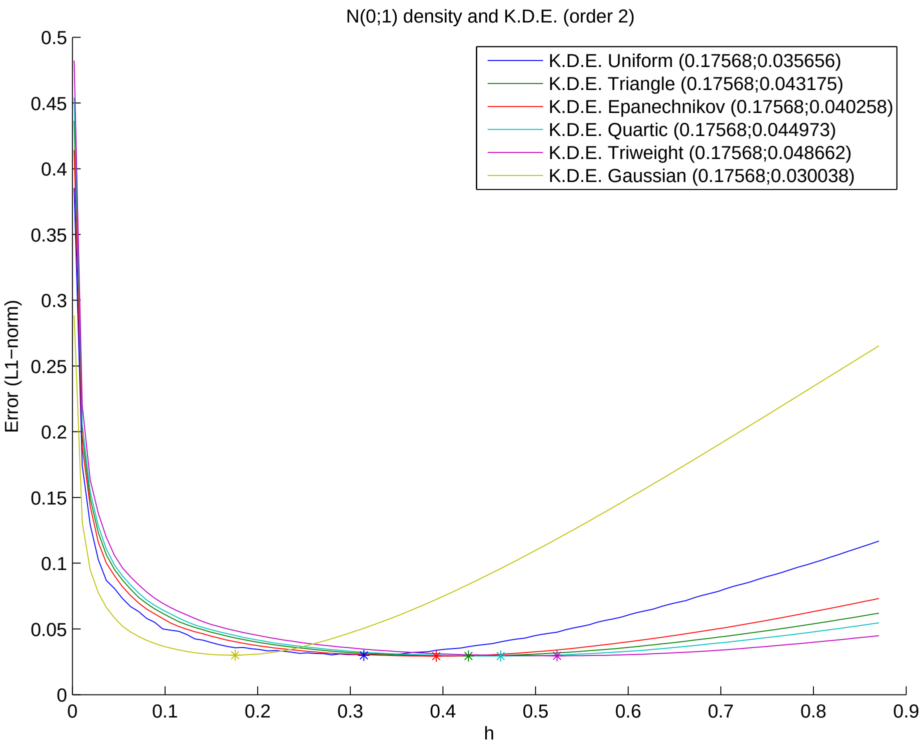
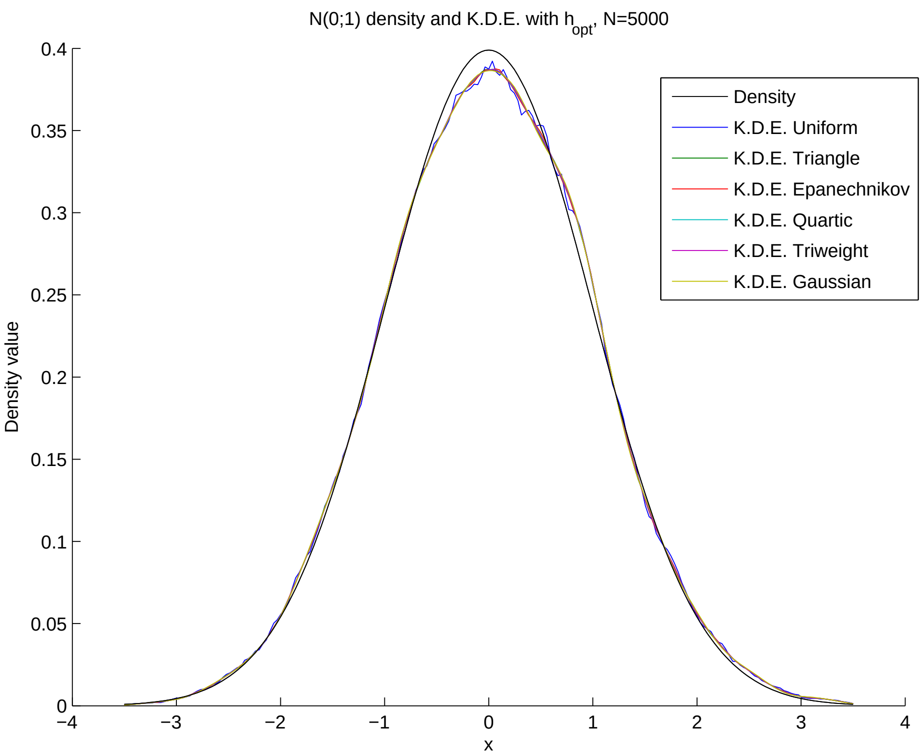


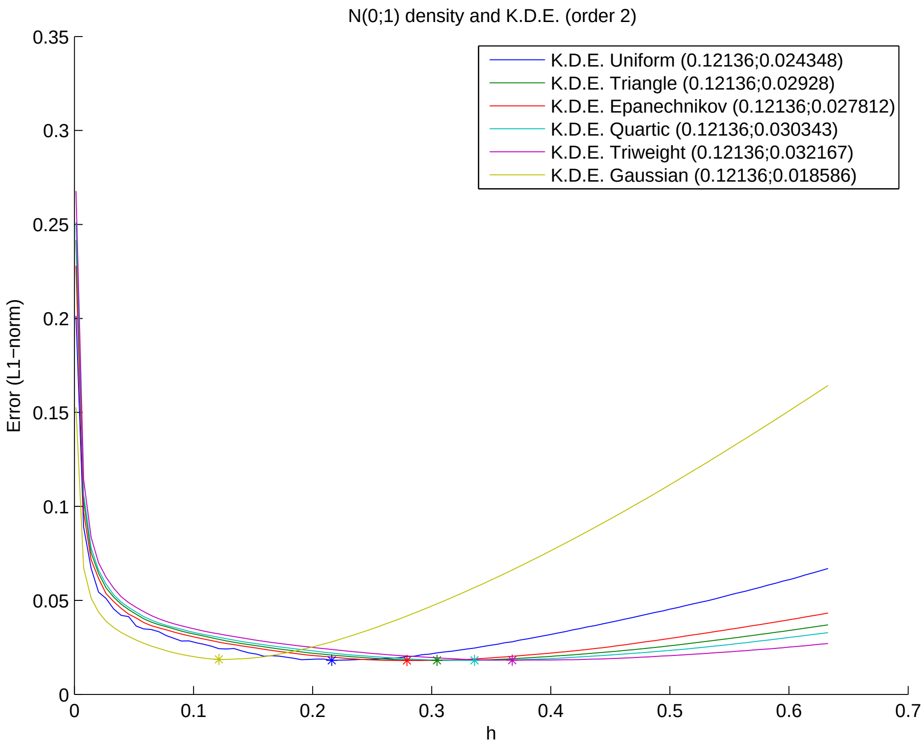
FIG. 2 – choix du paramètre h optimé : $h_{opt} \approx 0.075$

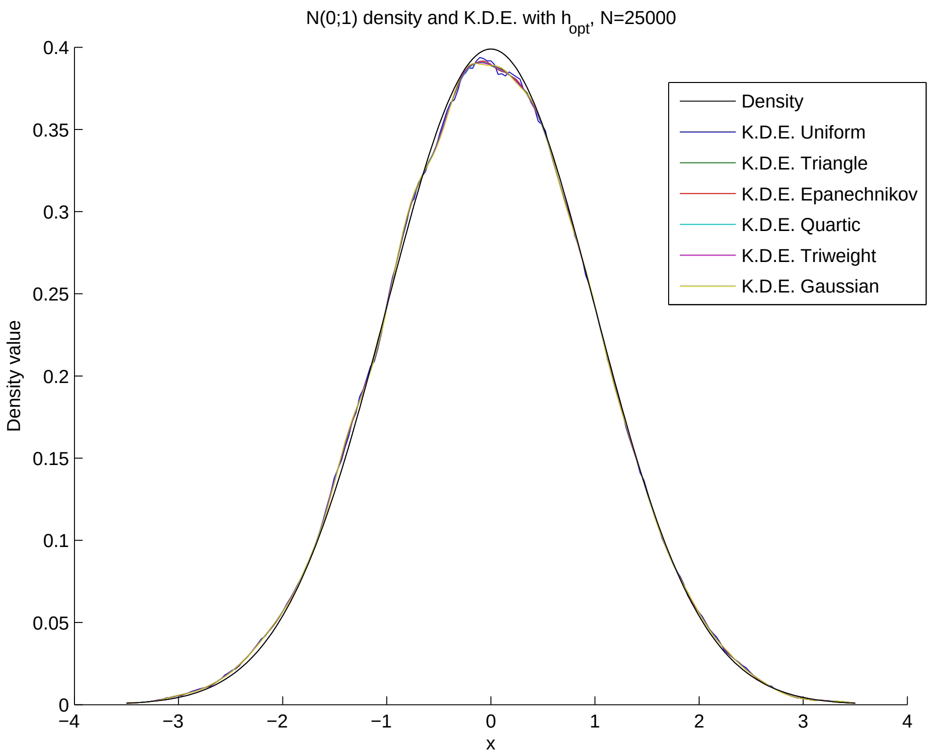
2 Comparaison de différents noyaux

2.1 Noyaux d'ordre 2 et $X \sim \mathcal{N}(0;1)$

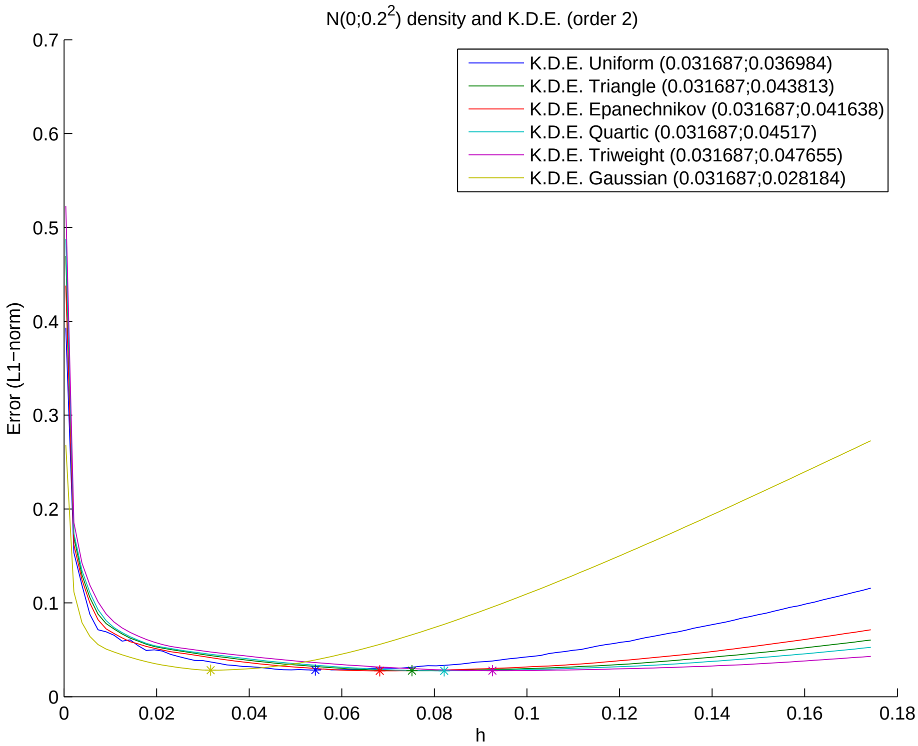


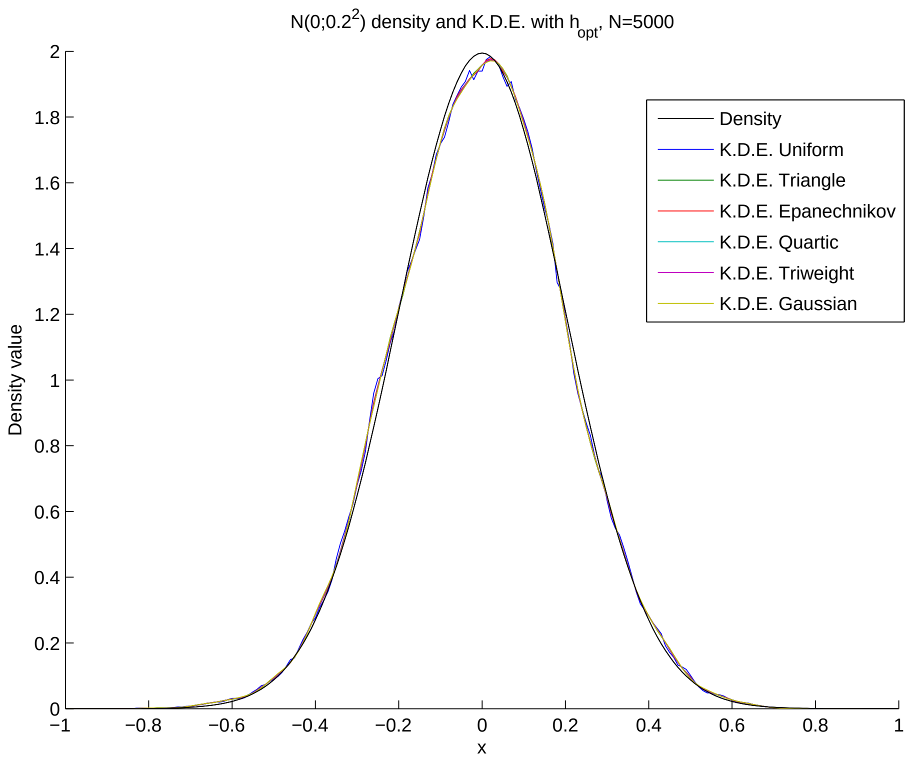


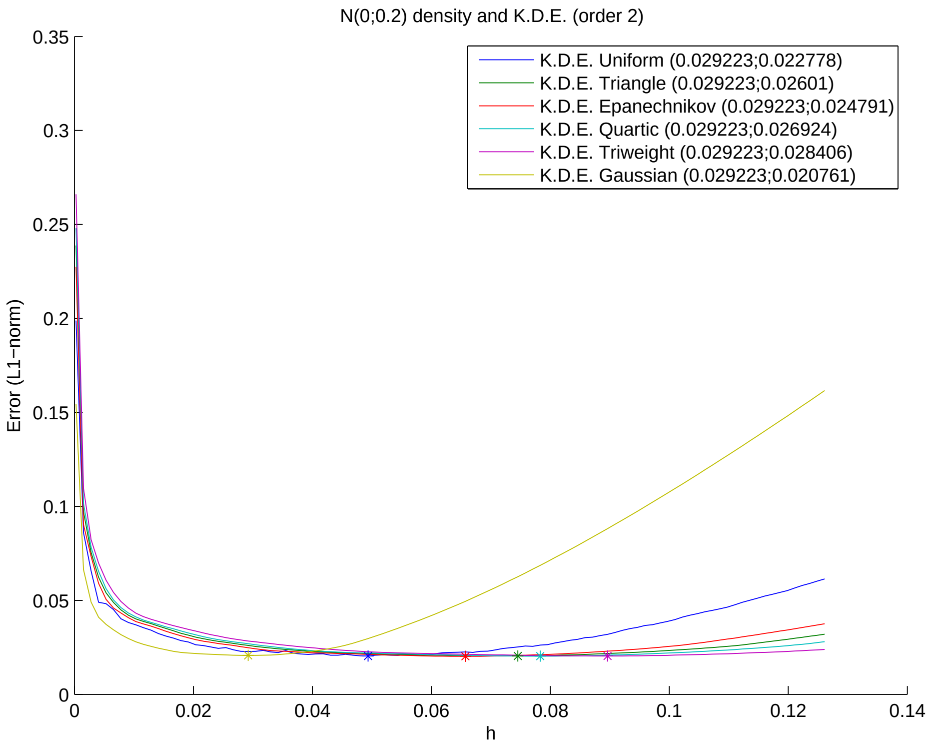

N(0;1) density and K.D.E. (order 2)
K.D.E. Uniform (0.12136;0.024348)
K.D.E. Triangle (0.12136;0.02928)
K.D.E. Epanechnikov (0.12136;0.027812)
K.D.E. Quartic (0.12136;0.030343)
K.D.E. Triweight (0.12136;0.032167)
K.D.E. Gaussian (0.12136;0.018586)
Error (L1−norm)
h

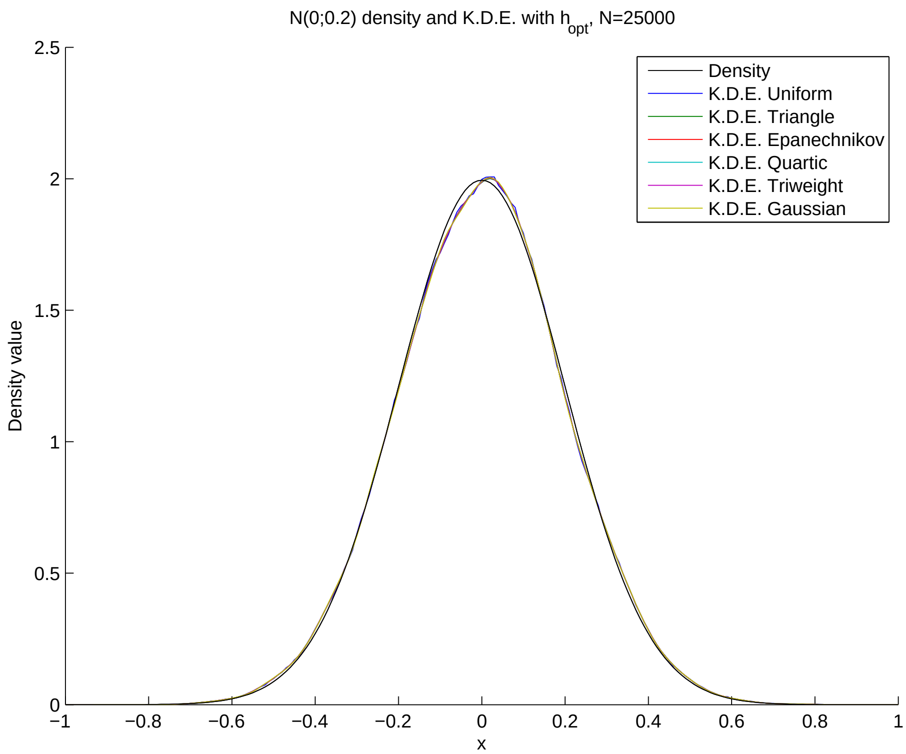

N(0;1) density and K.D.E. with h_{opt}, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x

2.2 Noyaux d'ordre 2 et $X \sim \mathcal{N}(0; 0.2^2)$

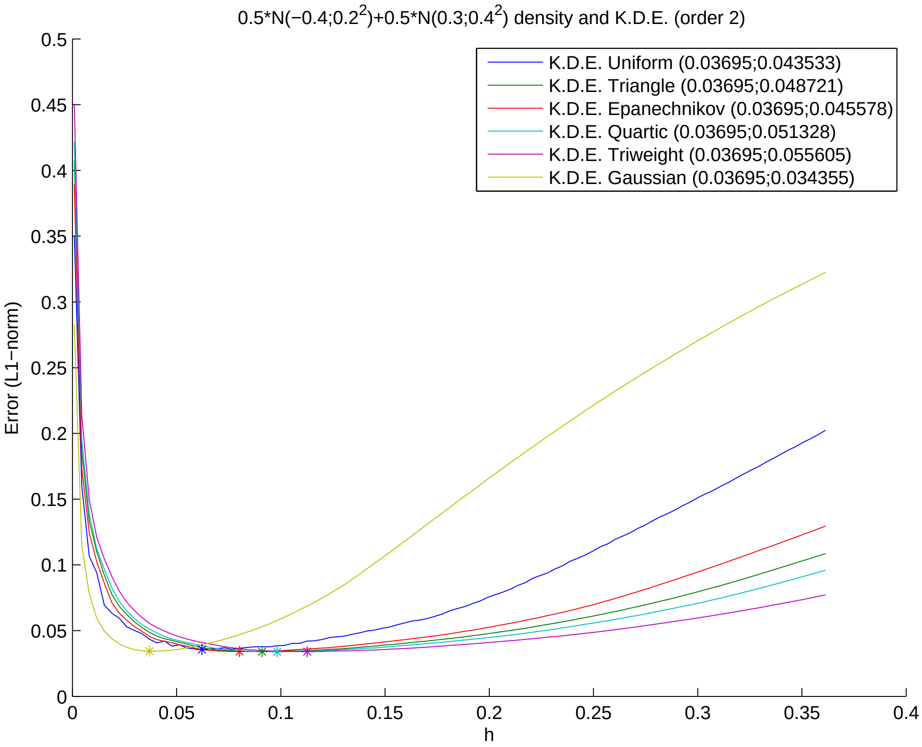


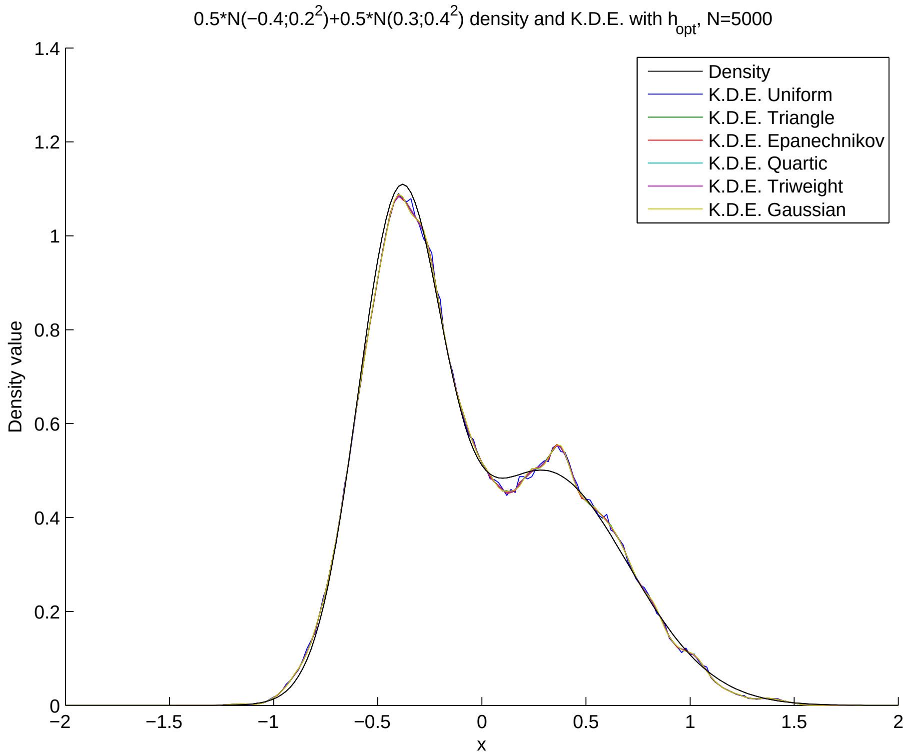


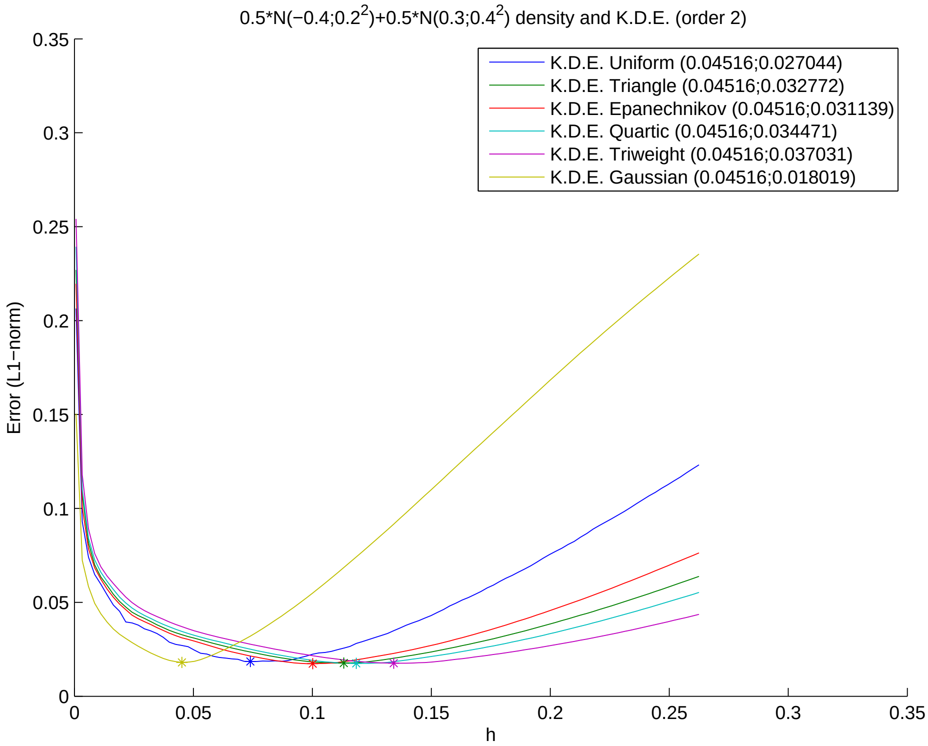

N(0;0.2) density and K.D.E. (order 2)
K.D.E. Uniform (0.029223;0.022778)
K.D.E. Triangle (0.029223;0.02601)
K.D.E. Epanechnikov (0.029223;0.024791)
K.D.E. Quartic (0.029223;0.026924)
K.D.E. Triweight (0.029223;0.028406)
K.D.E. Gaussian (0.029223;0.020761)
Error (L1−norm)
h
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.02
0.04
0.06
0.08
0.1
0.12
0.14

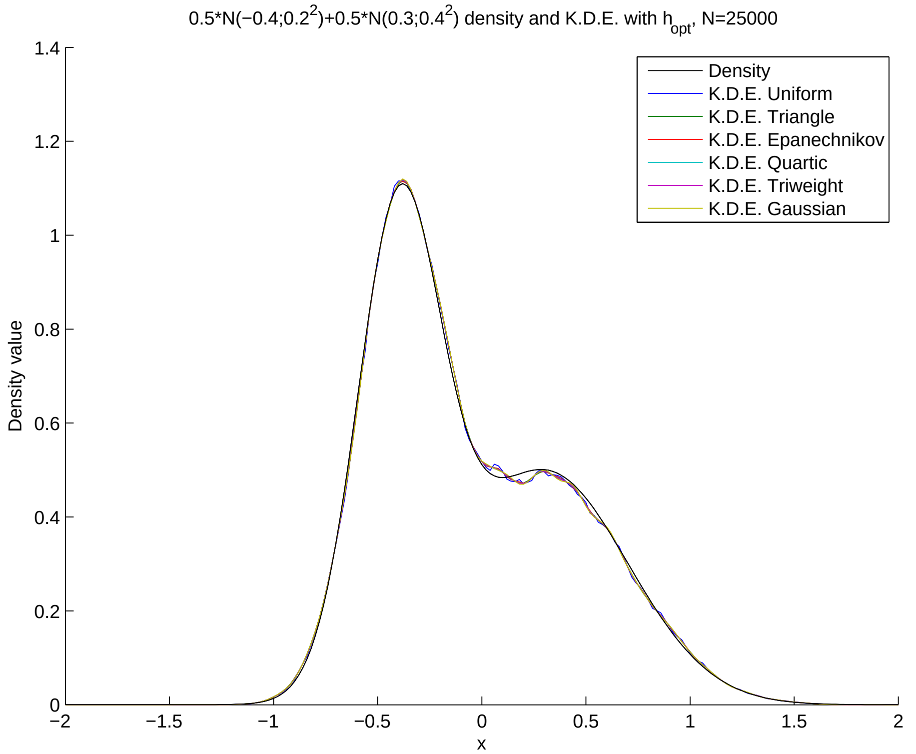

N(0;0.2) density and K.D.E. with h_{opt}, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x
0
0.5
1
1.5
2
2.5
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

2.3 Noyaux d'ordre 2 et $X \sim \frac{1}{2}\mathcal{N}(-0.4; 0.2^2) + \frac{1}{2}\mathcal{N}(0.3; 0.4^2)$

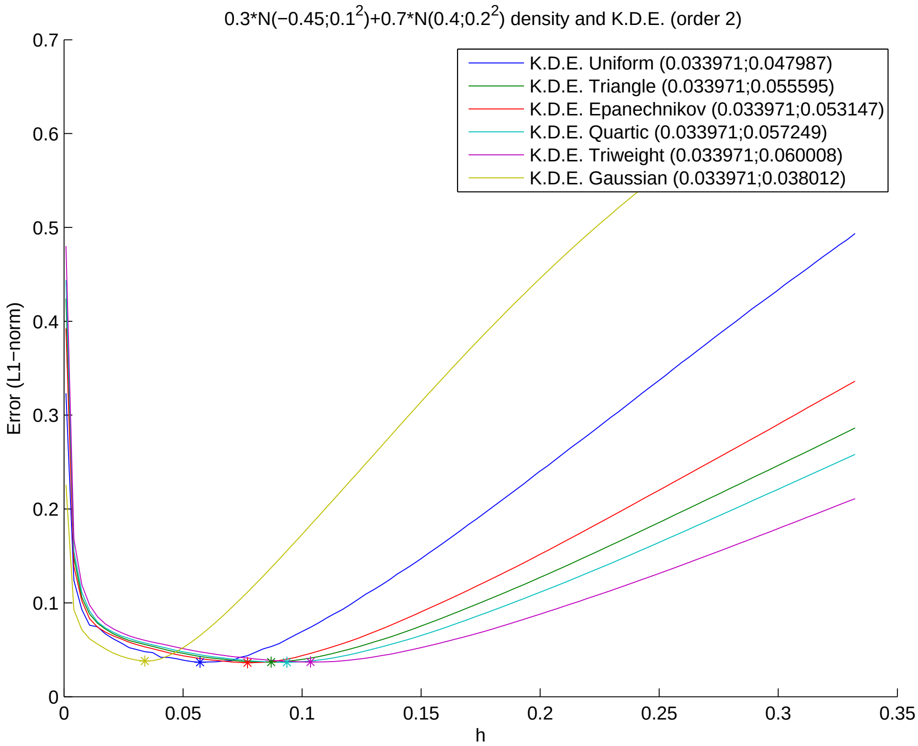


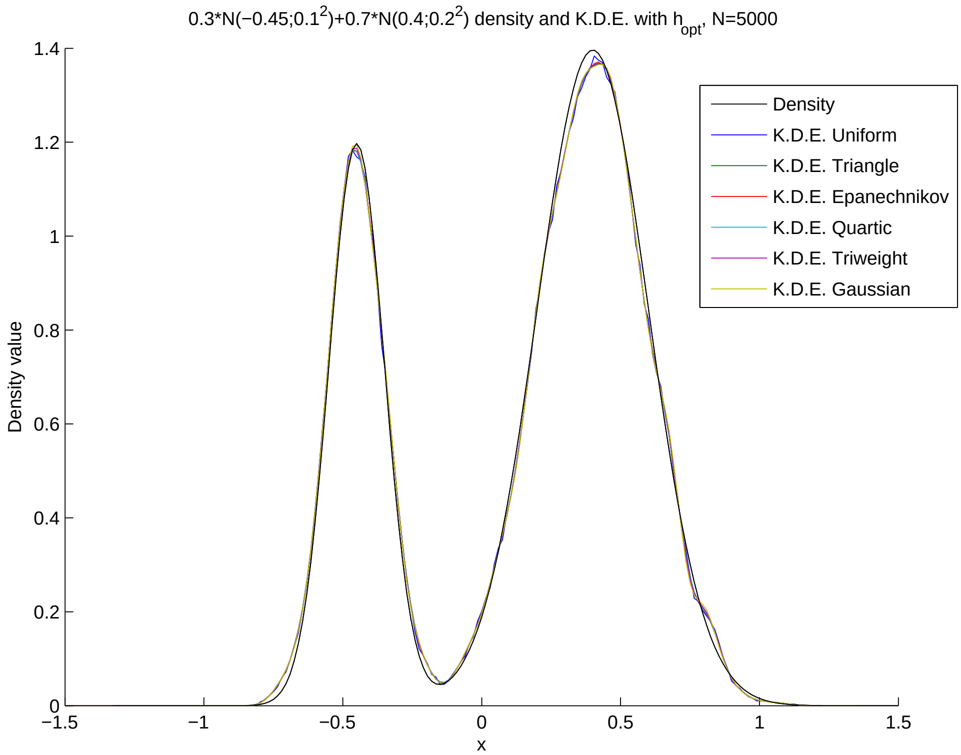


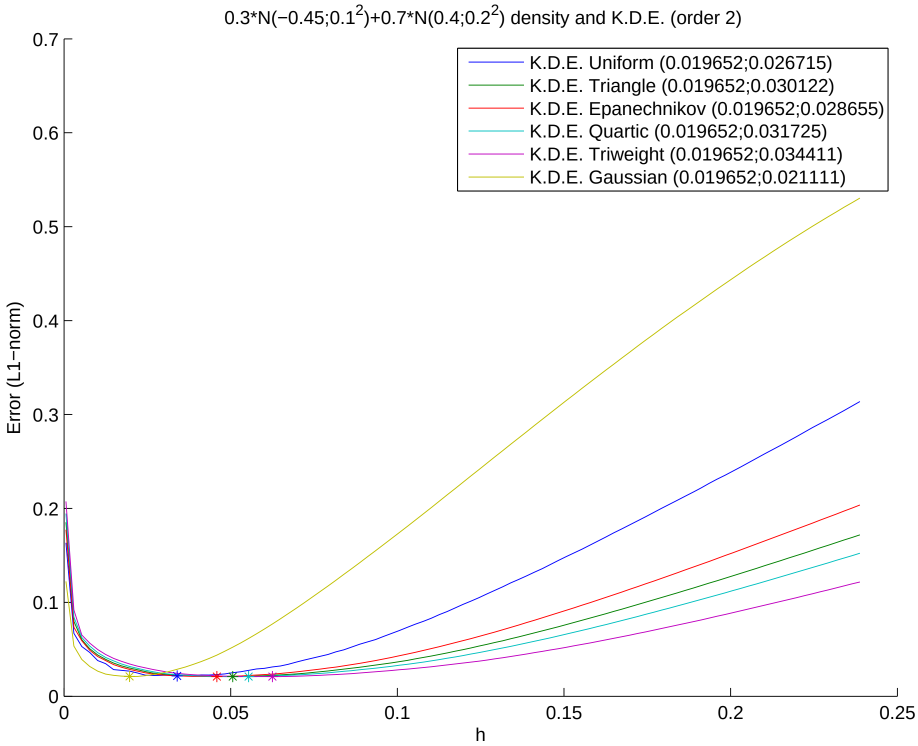

0.5*N(−0.4;0.2²)+0.5*N(0.3;0.4²) density and K.D.E. (order 2)
K.D.E. Uniform (0.04516;0.027044)
K.D.E. Triangle (0.04516;0.032772)
K.D.E. Epanechnikov (0.04516;0.031139)
K.D.E. Quartic (0.04516;0.034471)
K.D.E. Triweight (0.04516;0.037031)
K.D.E. Gaussian (0.04516;0.018019)
Error (L1−norm)
h
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

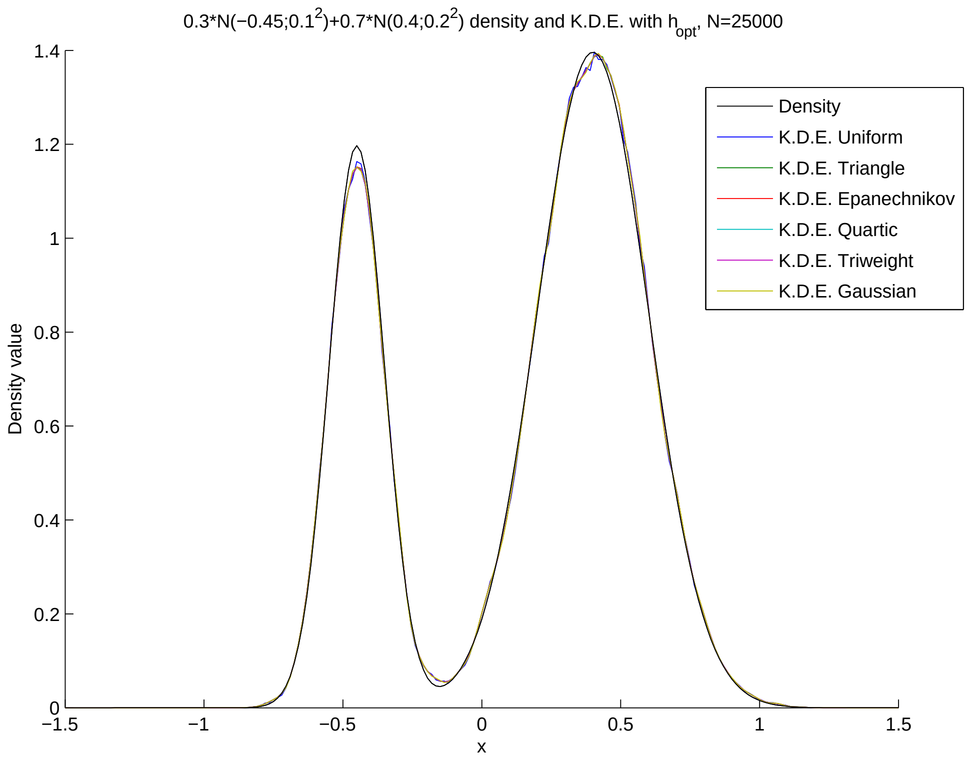

0.5*N(−0.4;0.2²)+0.5*N(0.3;0.4²) density and K.D.E. with h_opt, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
0
0.2
0.4
0.6
0.8
1
1.2
1.4

2.4 Noyaux d'ordre 2 et $X \sim 0.3\,\mathcal{N}(-0.45; 0.1^2) + 0.7\,\mathcal{N}(0.4; 0.2^2)$

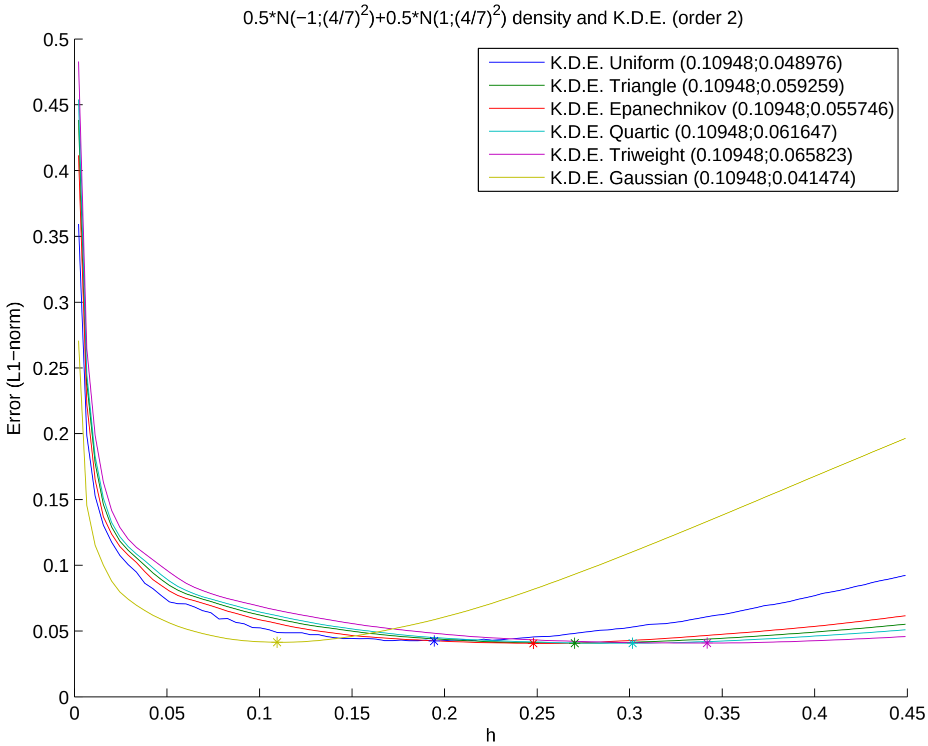


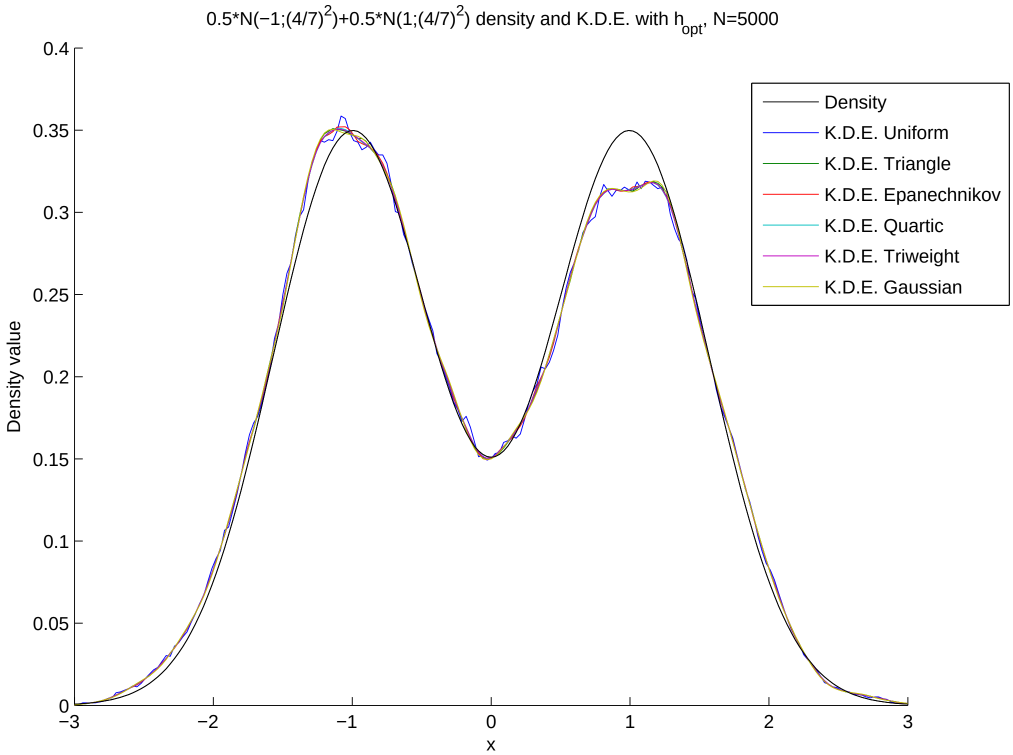


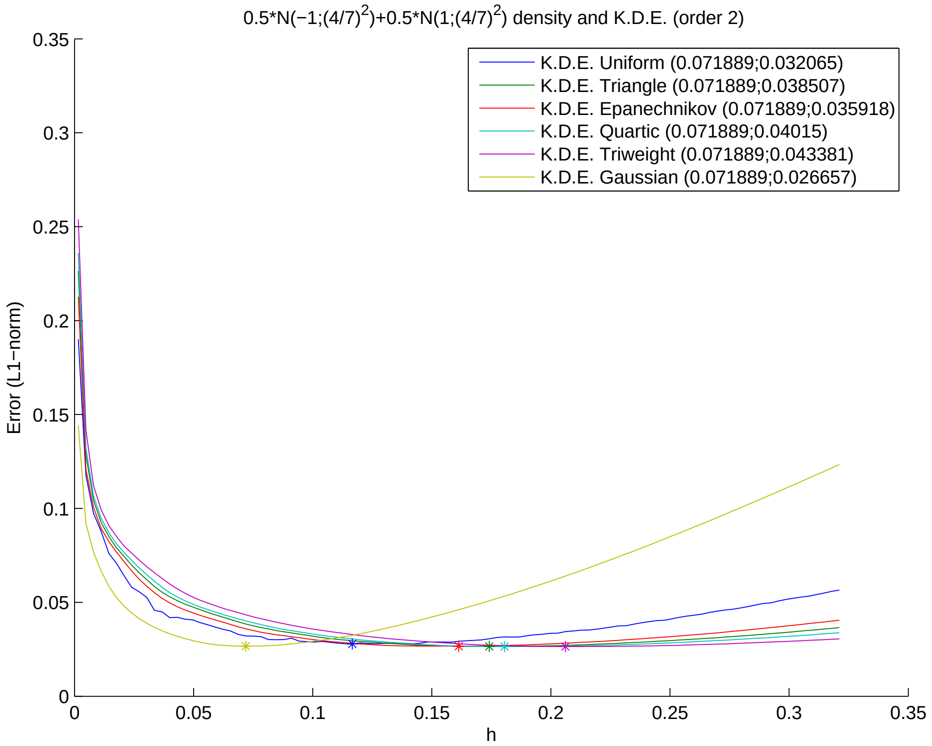

0.3*N(−0.45;0.1^2)+0.7*N(0.4;0.2^2) density and K.D.E. (order 2)
K.D.E. Uniform (0.019652;0.026715)
K.D.E. Triangle (0.019652;0.030122)
K.D.E. Epanechnikov (0.019652;0.028655)
K.D.E. Quartic (0.019652;0.031725)
K.D.E. Triweight (0.019652;0.034411)
K.D.E. Gaussian (0.019652;0.021111)
Error (L1−norm)
h

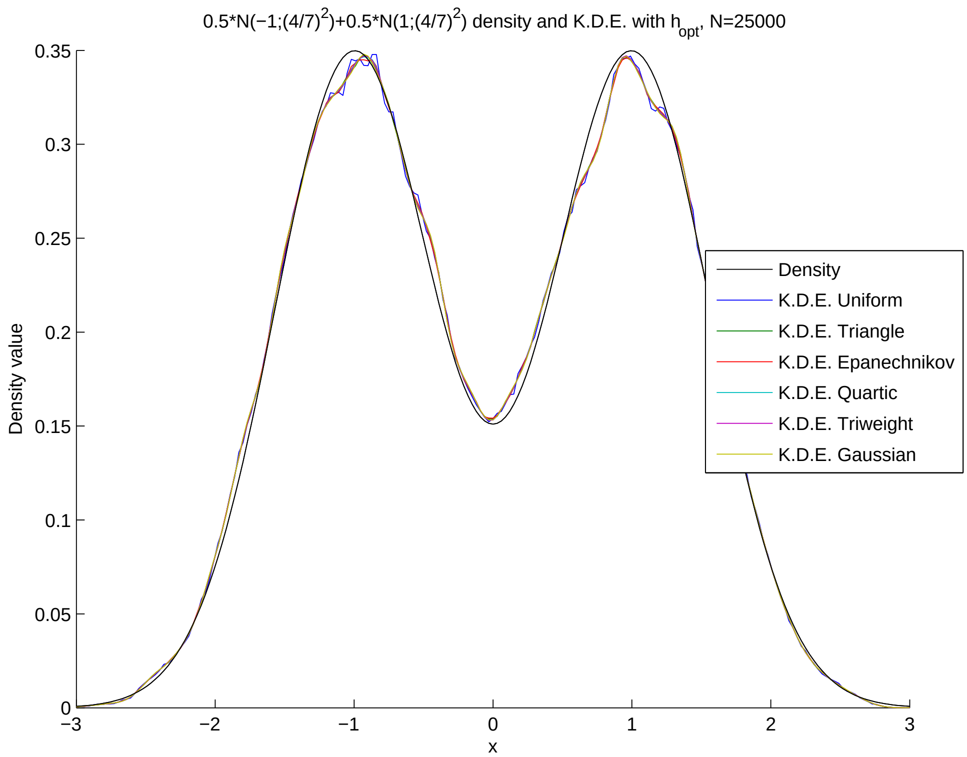

0.3*N(−0.45;0.1^2)+0.7*N(0.4;0.2^2) density and K.D.E. with h_opt, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x

2.5 Noyaux d'ordre 2 et $X \sim \frac{1}{2}\mathcal{N}\left(1;(4/7)^2\right) + \frac{1}{2}\mathcal{N}\left(1;(4/7)^2\right)$

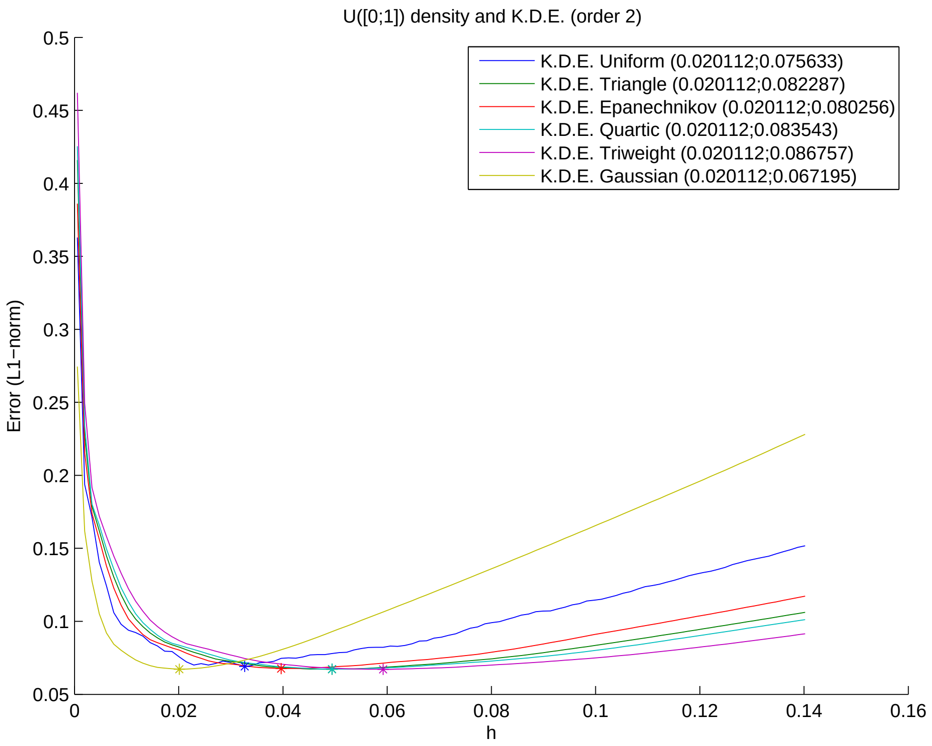


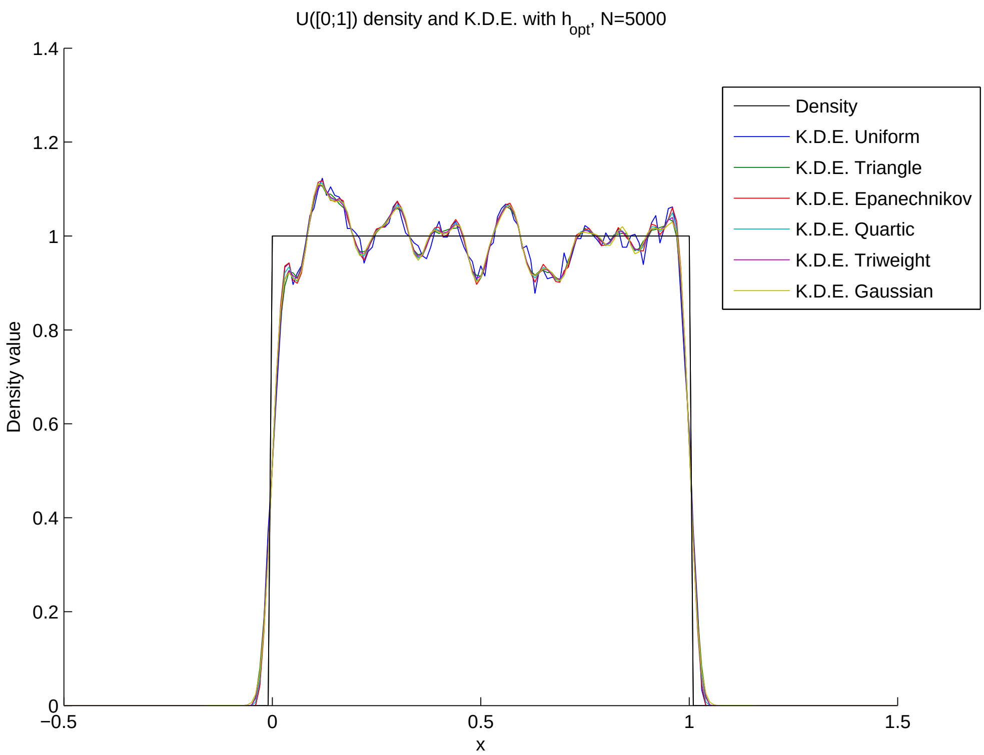


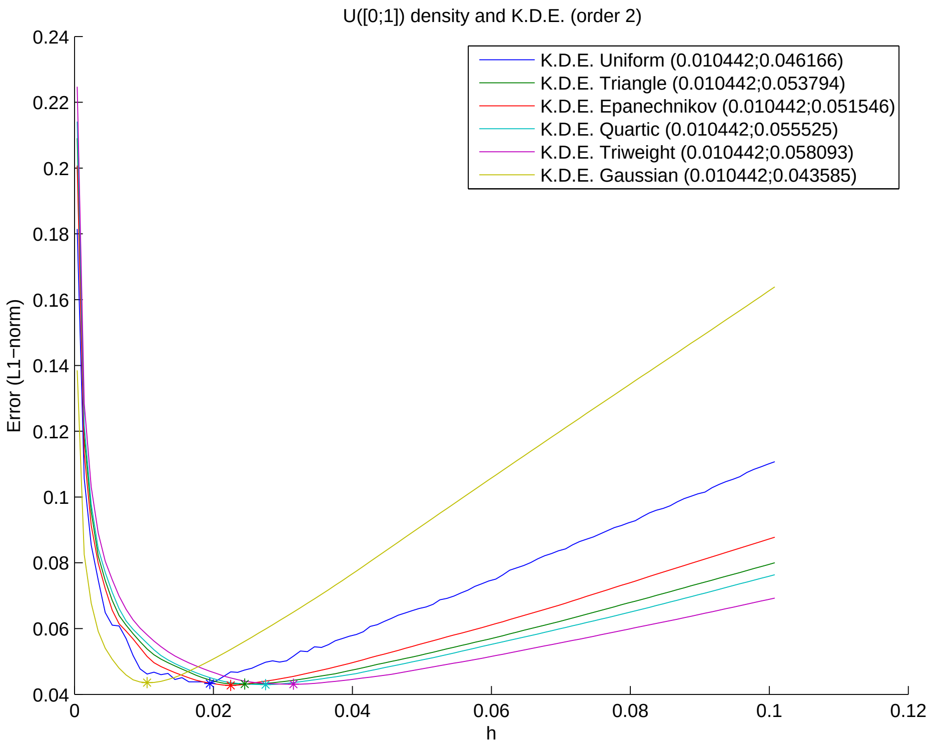

0.5*N(−1;(4/7)^2)+0.5*N(1;(4/7)^2) density and K.D.E. (order 2)
K.D.E. Uniform (0.071889;0.032065)
K.D.E. Triangle (0.071889;0.038507)
K.D.E. Epanechnikov (0.071889;0.035918)
K.D.E. Quartic (0.071889;0.04015)
K.D.E. Triweight (0.071889;0.043381)
K.D.E. Gaussian (0.071889;0.026657)
Error (L1−norm)
h

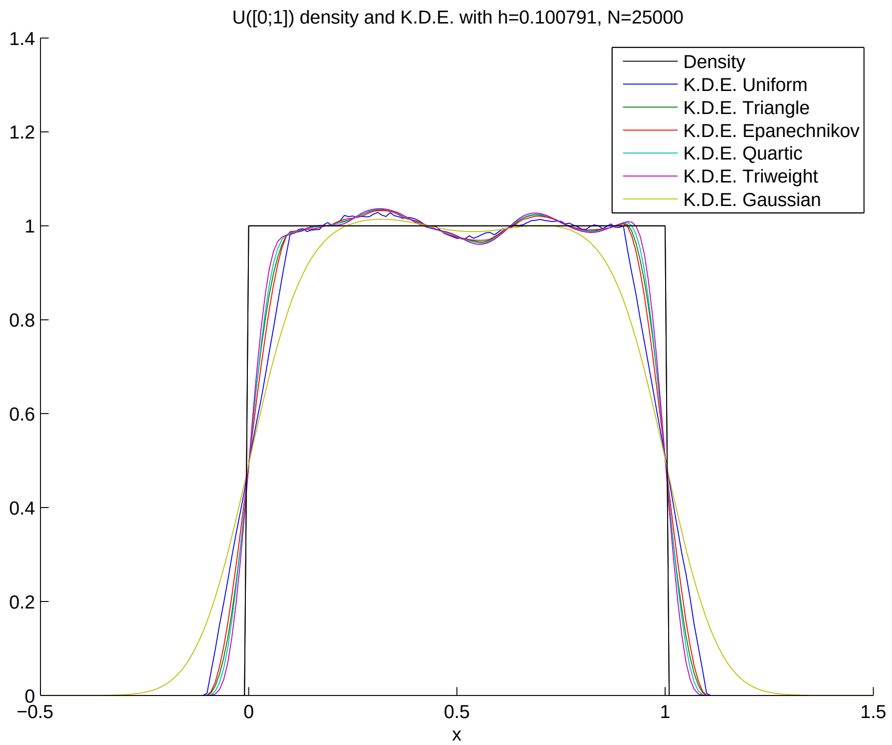

0.5*N(−1;(4/7)^2)+0.5*N(1;(4/7)^2) density and K.D.E. with h_opt, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x

2.6 Noyaux d'ordre 2 et $X \sim \mathcal{U}([0;1])$

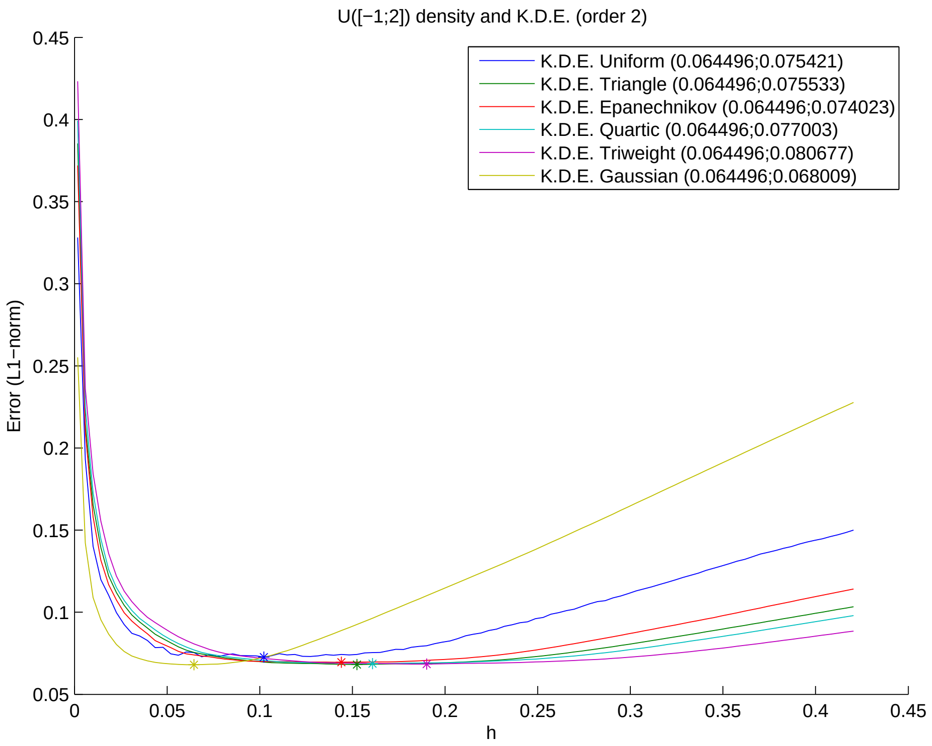


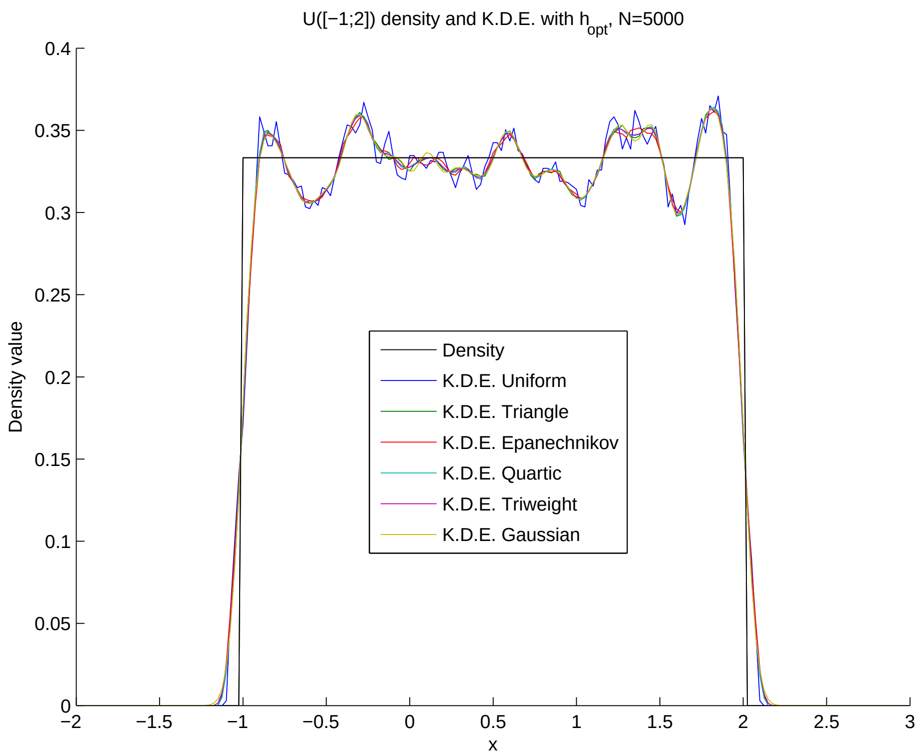


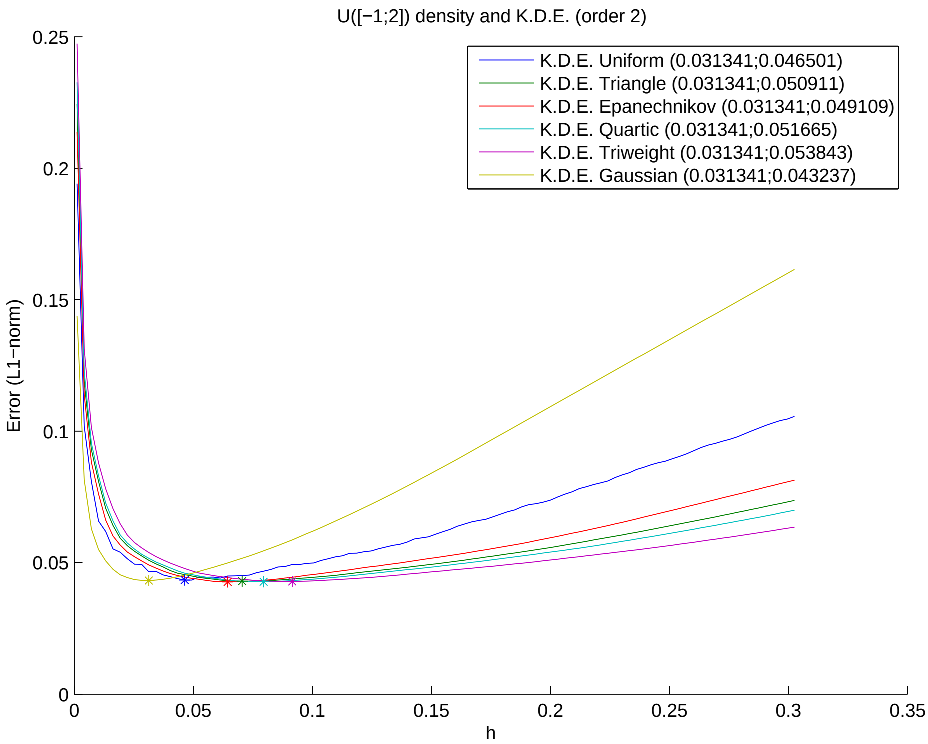

U([0;1]) density and K.D.E. (order 2)
K.D.E. Uniform (0.010442;0.046166)
K.D.E. Triangle (0.010442;0.053794)
K.D.E. Epanechnikov (0.010442;0.051546)
K.D.E. Quartic (0.010442;0.055525)
K.D.E. Triweight (0.010442;0.058093)
K.D.E. Gaussian (0.010442;0.043585)
Error (L1−norm)
h

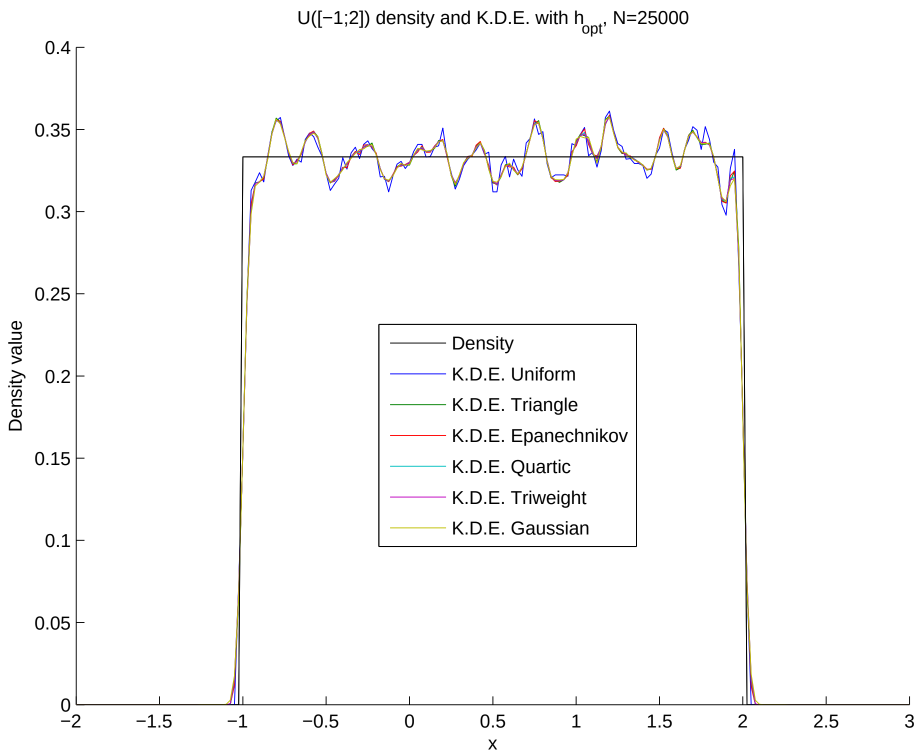

U([0;1]) density and K.D.E. with h=0.100791, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
x

2.7 Noyaux d'ordre 2 et $X \sim \mathcal{U}([-1;2])$

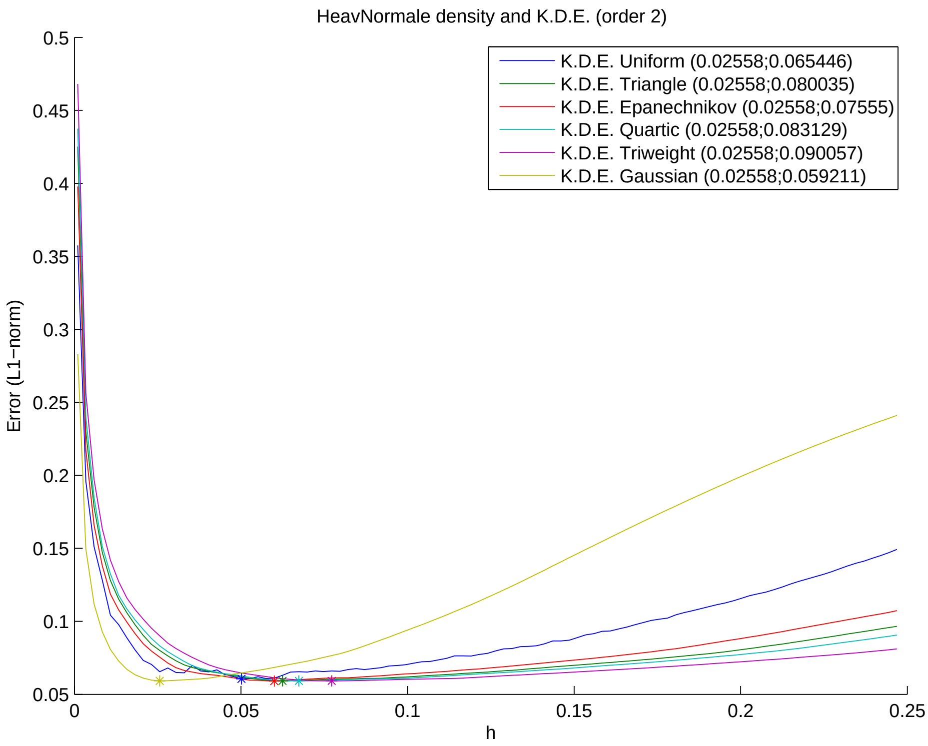


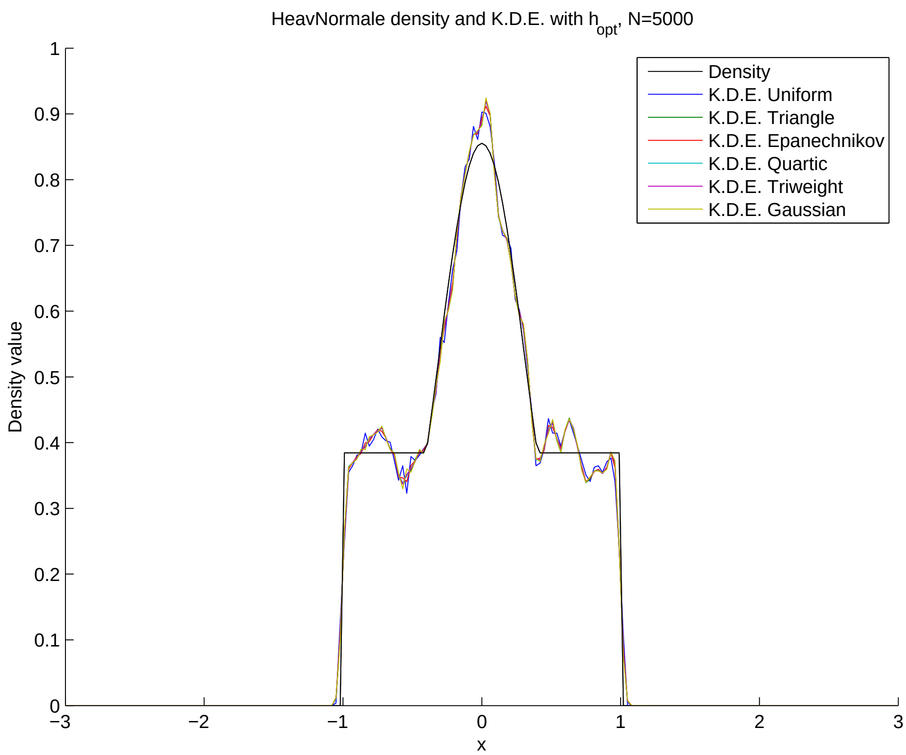


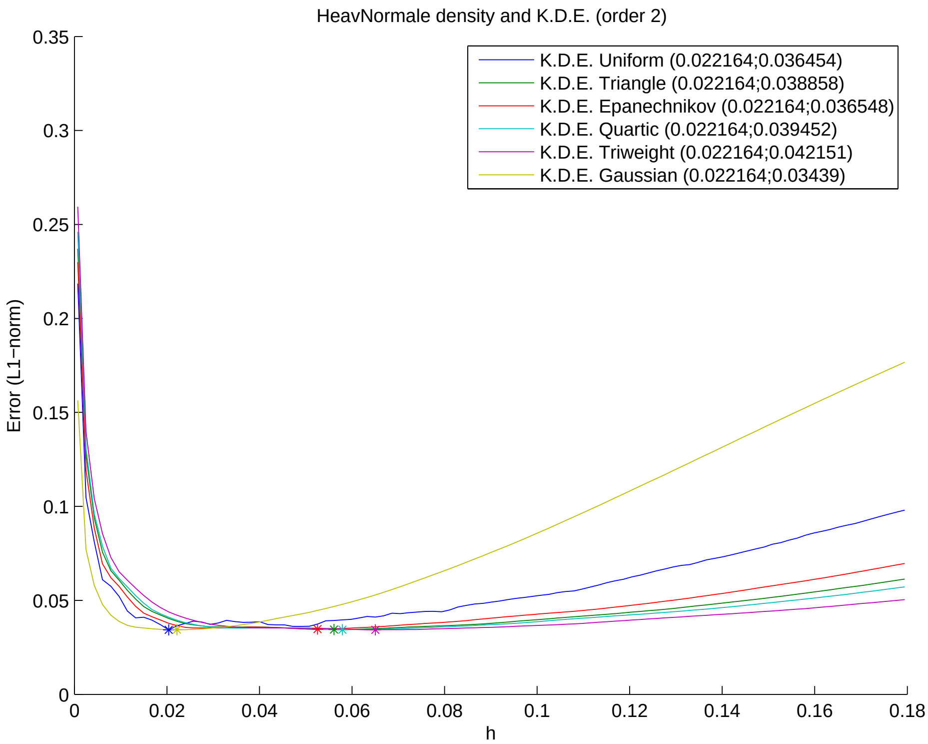

U([−1;2]) density and K.D.E. (order 2)
K.D.E. Uniform (0.031341;0.046501)
K.D.E. Triangle (0.031341;0.050911)
K.D.E. Epanechnikov (0.031341;0.049109)
K.D.E. Quartic (0.031341;0.051665)
K.D.E. Triweight (0.031341;0.053843)
K.D.E. Gaussian (0.031341;0.043237)
Error (L1−norm)
h
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

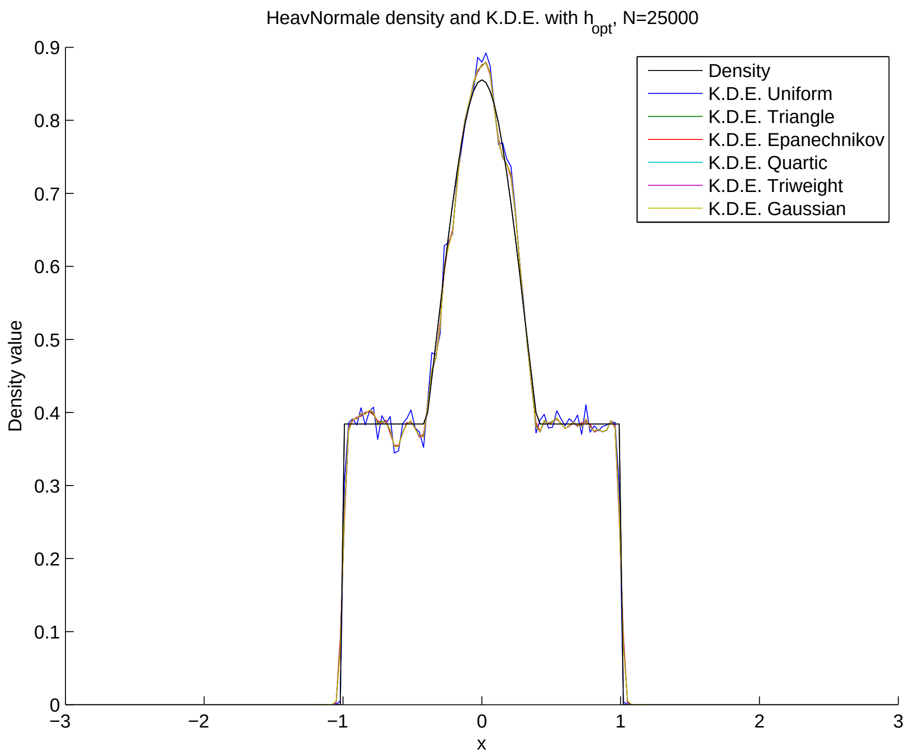

U([−1;2]) density and K.D.E. with h_{opt}, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
3
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4

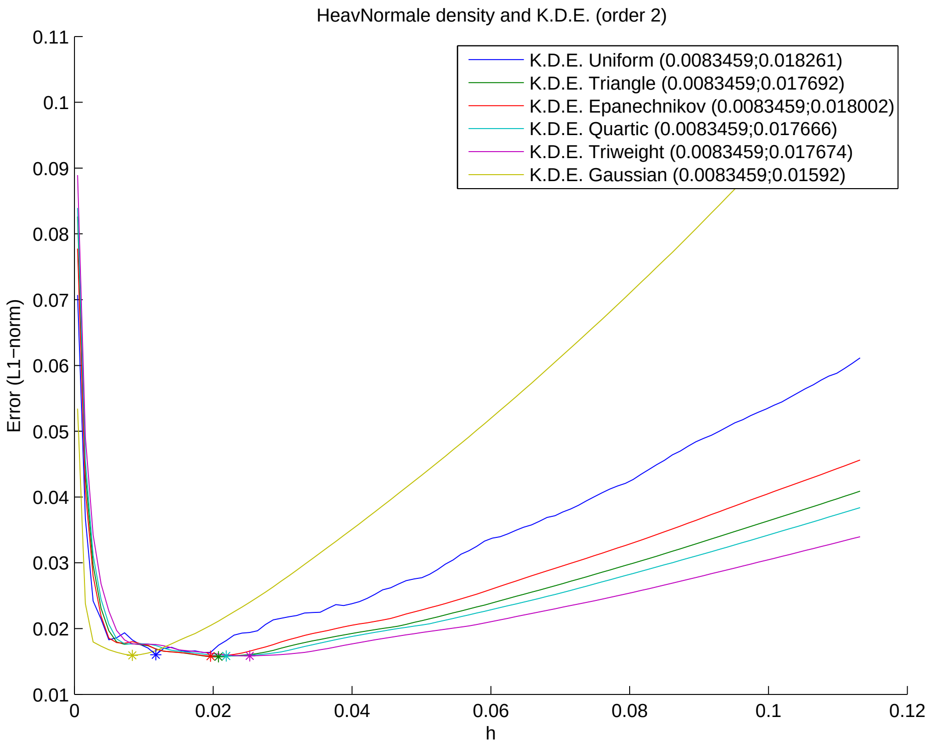
2.8 Noyaux d'ordre 2 et Densité discontinue (Density1.m)

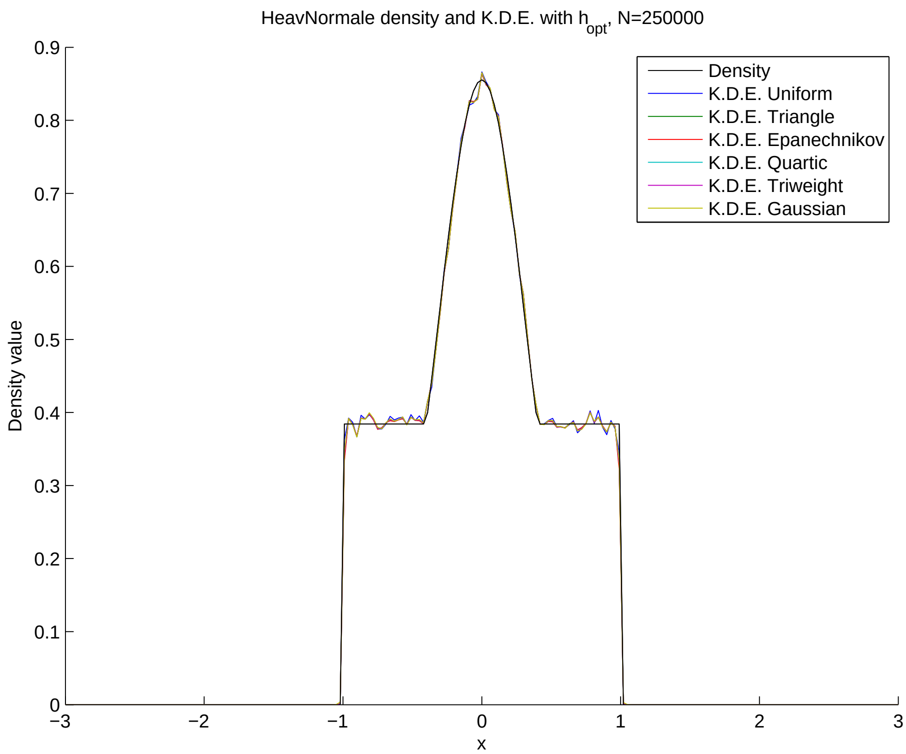




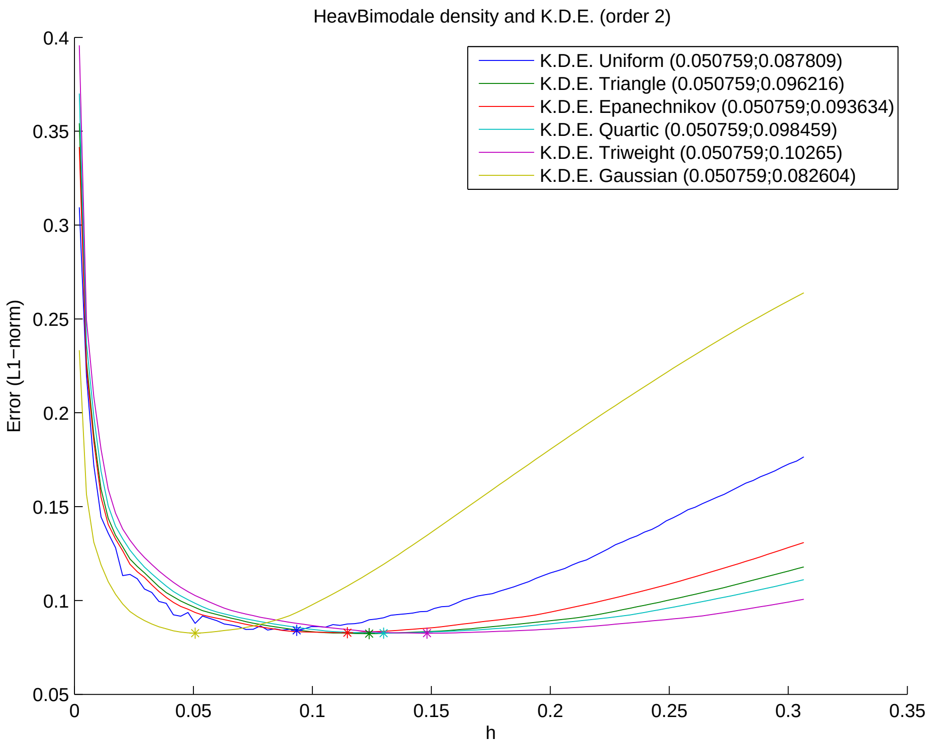

HeavNormale density and K.D.E. (order 2)
K.D.E. Uniform (0.022164;0.036454)
K.D.E. Triangle (0.022164;0.038858)
K.D.E. Epanechnikov (0.022164;0.036548)
K.D.E. Quartic (0.022164;0.039452)
K.D.E. Triweight (0.022164;0.042151)
K.D.E. Gaussian (0.022164;0.03439)
Error (L1−norm)
h

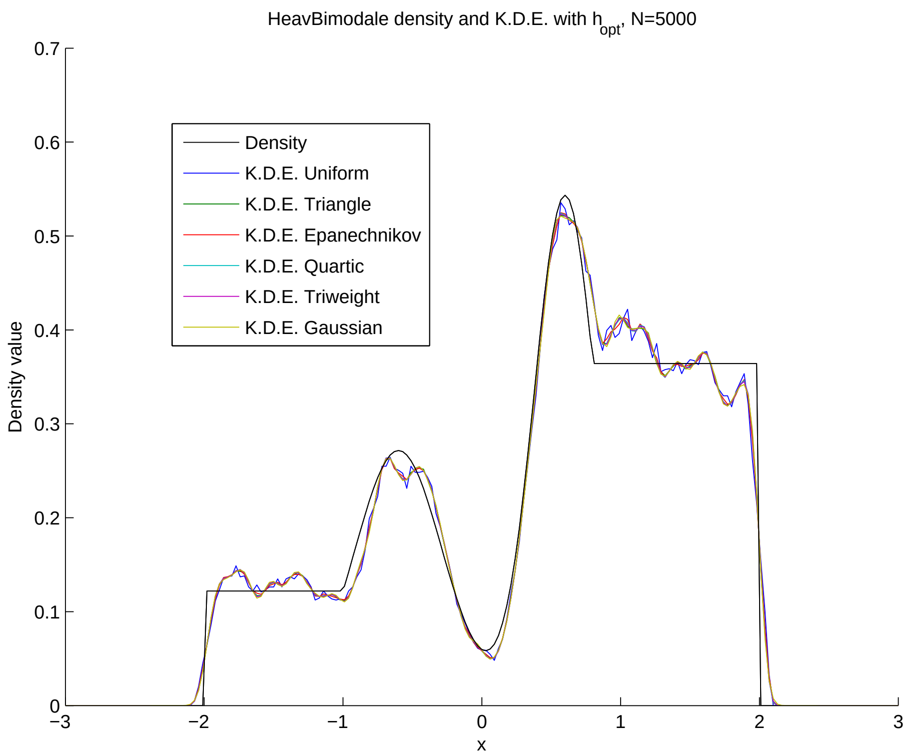

HeavNormale density and K.D.E. with h_opt, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x

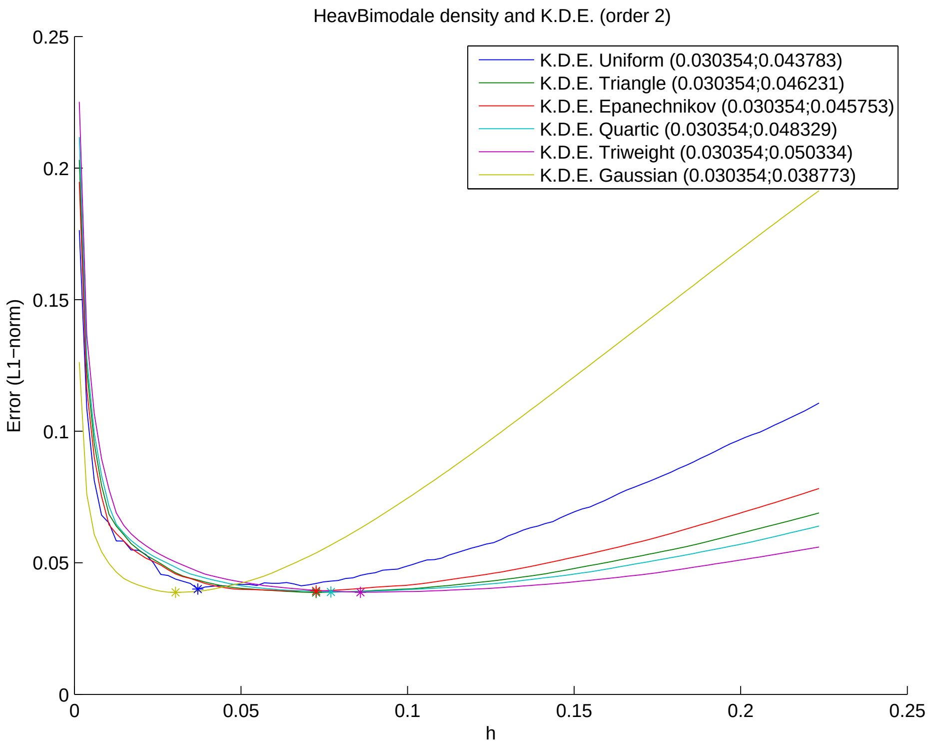

HeavNormale density and K.D.E. (order 2)
K.D.E. Uniform (0.0083459;0.018261)
K.D.E. Triangle (0.0083459;0.017692)
K.D.E. Epanechnikov (0.0083459;0.018002)
K.D.E. Quartic (0.0083459;0.017666)
K.D.E. Triweight (0.0083459;0.017674)
K.D.E. Gaussian (0.0083459;0.01592)
Error (L1−norm)
h
0.11
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0.02
0.04
0.06
0.08
0.1
0.12

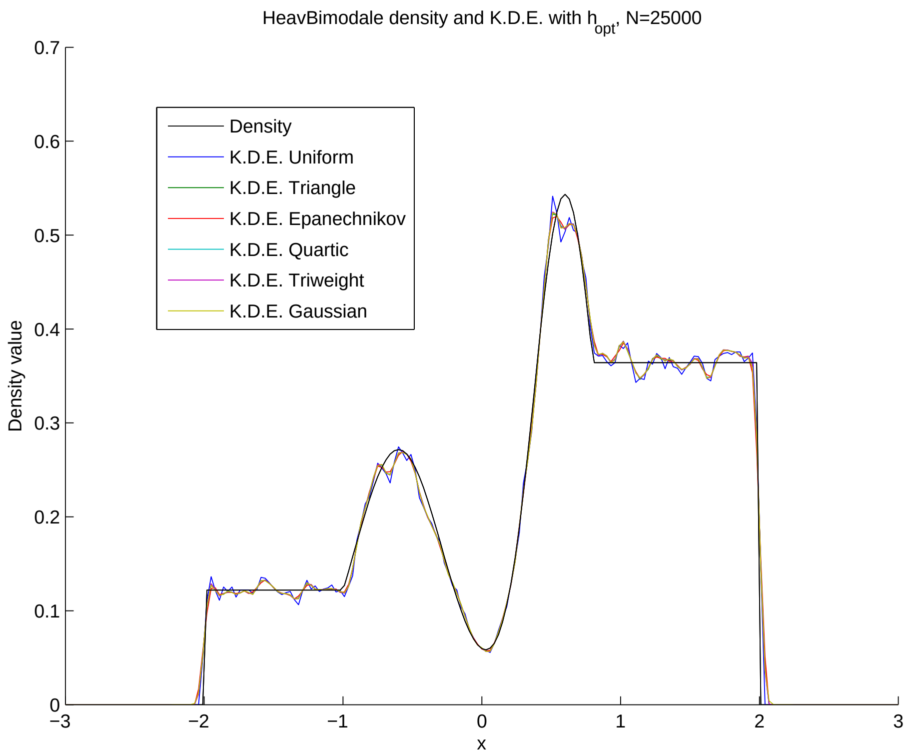

HeavNormale density and K.D.E. with h_{opt}, N=250000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−3
−2
−1
0
1
2
3

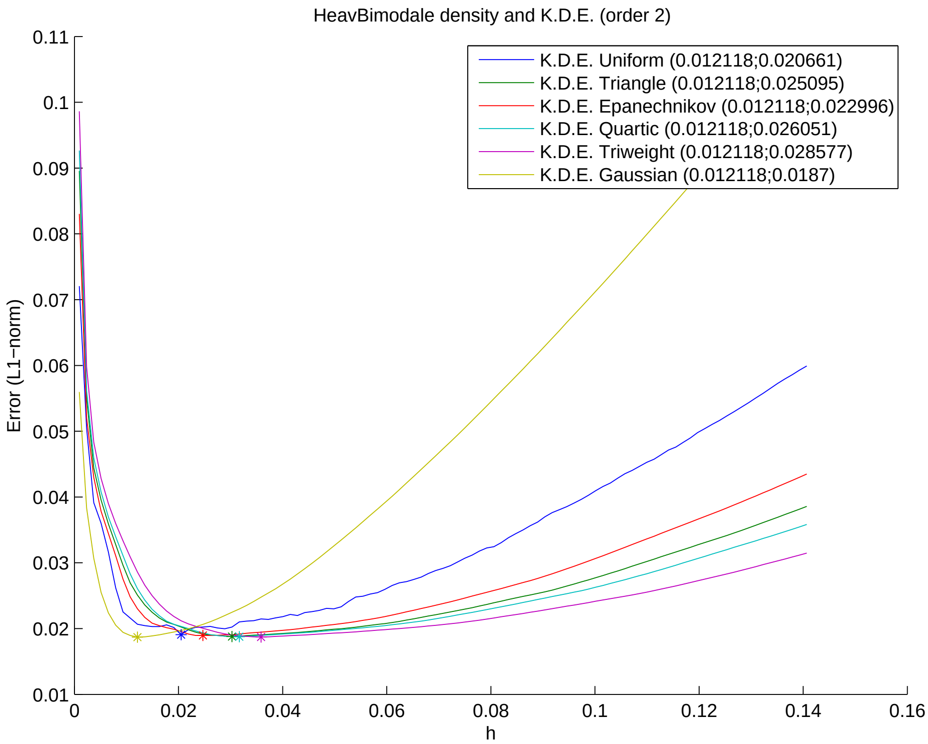
2.9 Noyaux d'ordre 2 et Densité discontinue (Density2.m)

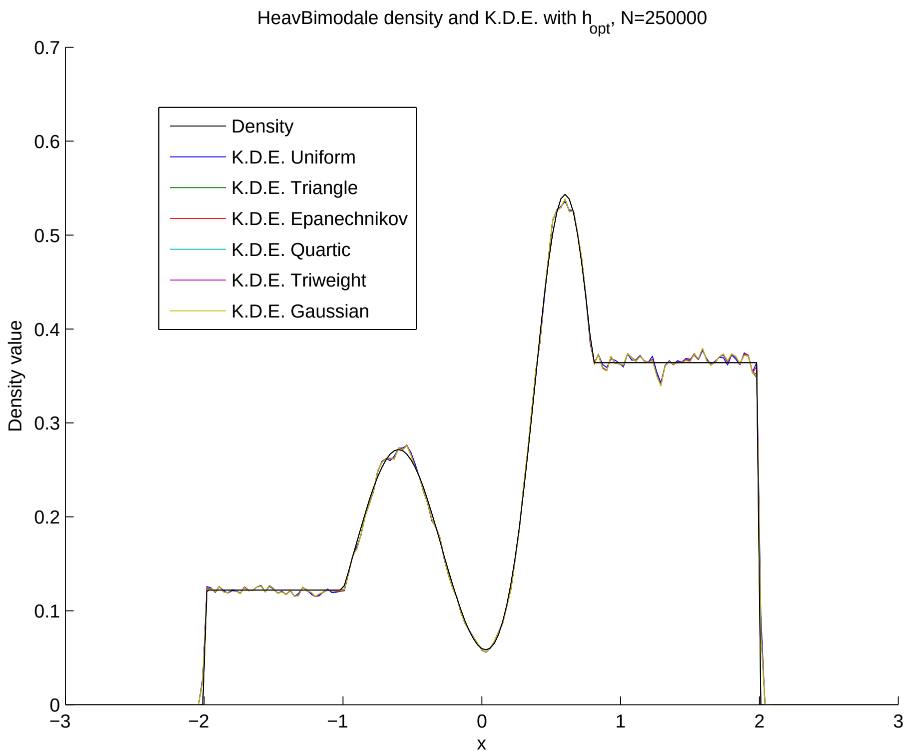





HeavBimodale density and K.D.E. (order 2)
K.D.E. Uniform (0.030354;0.043783)
K.D.E. Triangle (0.030354;0.046231)
K.D.E. Epanechnikov (0.030354;0.045753)
K.D.E. Quartic (0.030354;0.048329)
K.D.E. Triweight (0.030354;0.050334)
K.D.E. Gaussian (0.030354;0.038773)
Error (L1−norm)
h
0.25
0.2
0.15
0.1
0.05
0
0.05
0.1
0.15
0.2
0.25


HeavBimodale density and K.D.E. with h_opt, N=25000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−3
−2
−1
0
1
2
3


HeavBimodale density and K.D.E. (order 2)
K.D.E. Uniform (0.012118;0.020661)
K.D.E. Triangle (0.012118;0.025095)
K.D.E. Epanechnikov (0.012118;0.022996)
K.D.E. Quartic (0.012118;0.026051)
K.D.E. Triweight (0.012118;0.028577)
K.D.E. Gaussian (0.012118;0.0187)
Error (L1−norm)
h


HeavBimodale density and K.D.E. with h_opt, N=250000
Density
K.D.E. Uniform
K.D.E. Triangle
K.D.E. Epanechnikov
K.D.E. Quartic
K.D.E. Triweight
K.D.E. Gaussian
Density value
x