UNIVERSITEIT VAN AMSTERDAM

MASTER THESIS

Estimation of multivariate decreasing densities

Author:
CHENCHEN WANG
10226389
diudiu922@gmail.com

Supervisor:
DR.A.J.VAN ES
Korteweg-de Vries instituut
voor Wiskunde

A thesis submitted in fulfilment of the requirements for the degree of Master of Science

October 2013

UNIVERSITEIT VAN AMSTERDAM

UNIVERSITEIT VAN AMSTERDAM

Abstract

Faculty of Science

Master of Science

Estimation of multivariate decreasing densities

by Chenchen WANG

In this work we will explore the theoretical and practical aspects of nonparametric exponential deconvolution in the two-dimensional setting. After a transformation, exponential deconvolution can be used to estimate a decreasing density from direct observations. First, we will rigorously derive an inversion formula that can be implemented in an actual software implementation. We will then proceed with summarizing some of the known statistical properties of nonparametric kernel density estimators. Finally, we will combine the obtained results and construct, implement and test an exponential deconvolution method based on kernel estimators. An interesting application for estimation of decreasing densities will be pointed out and elaborated.
Keywords: deconvolution, decreasing densities, kernel estimation.

Acknowledgements

I would like to thank my supervisor Dr.A.J.van Es for providing me with a constant guidance and professional assistance. I would also like express my thanks to everyone who supported me during my final year of studies.

Contents

Chapter 1

Introduction

In this thesis we consider a deconvolution problem for estimating the unknown probability density function of a two-dimensional random vector, that we observe after a certain random noise has been added. We assume that the noise can be decomposed into a number of factors, each following an exponential distribution. The distributions of the noise factors are known beforehand and they are mutually independent, as well as independent from the original source. This gives us several instruments to capture the unknown density of the source. In the simple case of one source of exponential noise, we have

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

where $\mathbf{X} = (X_1, X_2)$ is the observed vector, $\mathbf{Y} = (Y_1, Y_2)$ is the vector with unknown distribution and Z_1, Z_2 are independent random variables with an exponential distribution (not necessarily with the same mean). While obtaining a formula for the probability density function of X is simple, since it is a standard convolution, in case we need the probability density function of Y, we need to invert the equation

$$f_{\mathbf{X}} = f_{\mathbf{Y}} * f_{\mathbf{Z}}.$$

where * denotes convolution, in order to obtain a formula for expressing $f_{\mathbf{Y}}$. Such formula is called an inversion formula.

A nonparametric approach using kernel estimators and an inversion formula has been introduced by Van Es and Kok (1998) in the one dimensional case.

Assuming in the one dimensional setting that $X = Y + Z$, where Y and Z are independent and Z has an exp(1) distribution, they derive the inversion formula

$$f_Y = f_X + f'_X. \tag{1.1}$$

We can estimate the density f_X for the data $X_1, ..., X_n$ by a so called kernel estimator

$$g_{nh}(x) = \frac{1}{nh} \sum_{i=1}^{n} w(\frac{x - Xi}{h}) \tag{1.2}$$

where w is a probability density function, the kernel, and $h > 0$ is a bandwidth.
This leads to an estimator $f_{nh}(x)$ of the density of y. We get

$$f_{nh}(x) = \frac{1}{nh} \sum_{i=1}^{n} w(\frac{x - Xi}{h}) + \frac{1}{nh^2} \sum_{i=1}^{n} w'(\frac{x - Xi}{h}) \tag{1.3}$$

Our main goal in this paper is to estimate decreasing densities. We will outline our approach for dimension one.

If X has a decreasing density f_X on $[0, \infty)$ then X can be written as $Y \cdot U$ where Y is a density on $[0, \infty)$ and U is uniform. The random variable Y and U are assumed to be independent. By taking logarithm we now get

$$-\log X = -\log Y - \log U \tag{1.4}$$

Since $-\log U$ has an exp(1) distribution estimating the density of $-\log Y$, for observations, X_i, or $-\log X_i$, is then an exponential deconvolution problem where we can apply the estimator (1.2). Let us write B for $-\log Y$, and f_B for its density. One can then show that we have

$$f_X(y) = \int_y^{\infty} f_B(-\log x) \frac{1}{x^2} dx, \; y > 0 \tag{1.5}$$

Substituting a kernel estimator $f_B^{(n)}$ for f_B we get an estimator of the decreasing density f_X.

This approach will be extended to the two dimensional decreasing density problem.Based on the results of Krycha(2011) for two dimensional exponential deconvolution, we will derive an estimator of two dimensional decreasing densities.

Some methods of estimating decreasing densities have been elaborated. One of these methods is the method of isotonic estimators. This method was proposed by Groeneboom and Lopuha in 1993. However this method only works in one dimension.
Let F_n be the empirical distribution function of a sample $X_1, ..., X_n$ and suppose that the distribution function F of the X_i's is continuous and concentrated on $[0, \infty)$. By the concave majorant $\hat{F_n}$ of F_n we mean the smallest concave function on $[0, \infty)$ such that $\hat{F_n}(x) \geqslant F_n(x)$. Since F_n is piecewise constant, $\hat{F_n}$ is zero at the origin, is piecewise linear and has a finite number of vertices on the interval $[0, \infty)$. One can think of $\hat{F_n}$ as of pinning down a rubber band at the origin and streching it over the points of the so-called "cumulative sum diagram", consisting of points $(X_{(i)}, \frac{i}{n})$, where $X_{(i)}$ are the order statistics of the sample
Statistics based on either concave majorants or convex minorants arise in different context.Behnen(1975) and Scholz(1983) proposed different statistics which are expressible in terms of the L_2 norm of the density of the concave majorant of different empirical distribution functions. Other examples occur in censoring problems. If one wants to estimate the distribution of interval censored observation problems or estimate a distribution which is part of a convolution, the nonparametric maximum likelihood estimator can be characterized in terms of convex minorants. (see for instance Groeneboom and Wellner, 1992)
The connection with nonparametric maximum likelihood estimators of the density of Groeneboom was already known from Grenander (1956), who showed the following classical result, which might seem suprising at first sight.

Theorem 1.1: *Let :* $\mathbb{F}$ *be the class of nonincreasing densities on* $[0, \infty)$ *and let* $X_1, ..., X_n$ *be a sample from a density* $f \in \mathbb{F}$. *Let* $\hat{f_n}$ *be the left continuous derivative of the concave majorant* $\hat{F_n}$ *of the empirical distribution function of* $X_1, ..., X_n$. *Then* $\hat{f_n}$ *maximizes the function*

$$f \rightarrow \prod_{i=1}^{n} f(X_i)$$

among all $f \in \mathbb{F}$.
Similarly the nonparametric maximum likelihood estimator of an increasing density is the left derivative of the convex minorant of the empirical distribution.

While the problem of estimating a decreasing density by nonparametric maximum likelihood seemd to be solved by the Grenander estimator, the two dimensional problem is much more difficult. In that case there is no closed solution for the maximizer of the likelihood though some theory have been derived, see Polonik(1995,1998) and Biau and Devroye(2003).

Moreover, we will present our method for one and two dimensional data. While for one dimensional data smooth vectors of the Grenander estimator will probability give good results, the one dimensional case. The two dimensional estimator is of higher interest since no easy competitor exists.

The main purpose of Chapter 2 is to introduce the kernel estimator. We will illustrate the effect of changing the bandwidth, bin size and kernels on both simulated data and data sets adopted from the literature. Finally, we will derive statistical properties and extend these observations to second dimension problems.

In Chapter 3 we will introduce Gamma and exponential deconvolution in one and two dimensions, and derive the key inversion formulas. We will combine the preceding results and introduce a two-dimensional nonparametric deconvolution estimator. The key objective of the chapter is then to derive the statistical properties such as asymptotic expectation and variance.

The aim of Chapter 4 is to develop a method for nonparametric estimation of decreasing densities. While using the main results of the preceding chapters, we will concentrate on obtaining a closed-form formula that would be practical for later implementation purposes. The resulting estimator will prove to be more effective for estimating decreasing densities in comparison to standard non-parametric estimators. We will also give interesting examples to apply this method to.

Finally, in Chapter 5 we will implement the introduced two-dimensional density estimation methods in Matlab and verify that the obtained formulas are of practical use. We will focus on estimating densities from simulated samples while observing the right choice of bandwidth and required computational complexity. We will use a range of distributions and visualization methods in order to present the potential use of the proposed methods. The source code will be included in the appendix so that the results can be verified and so that the implemented algorithm can be further used.

Chapter 2

The kernel estimator

Assume that we have a random variable X (or similarly a random vector X) and we observe its independent realizations $X_1, ...X_n$. We want to estimate its unknown density based on those observations. In this chapter we will discuss and illustrate one of the nonparametric estimation methods - the kernel estimation method, which can be applied both to one-dimensional and multivariate data. The results that are reviewed are standard and can be found for instance in Sliverman(1988) and Matt P.Wand(1995)

2.1 One-dimensional kernel estimation

Kernel density estimation is a method which assigns weighted positive functions (kernels) to the observations. The density estimate is calculated as a sum of these weight functions ($Härdle$, 1991).This approach boasts several advantages: we can avoid jumps in our estimate by appropriate selection of the kernels and the resulting estimate function will then be smooth (up to the class of the kernel) (Silverman,1998). Let us first introduce the one-dimensional kernel estimator.

2.1.1 One-dimensional kernel estimators

For one dimensional observations the kernel density estimator introduced by Matt P.Wand(1995) is defined as follows:

Definition 2.1 (Kernel): *Let K(x) be a nonnegative bounded integrable function with real values. We say that K(x) is a kernel if and only if the following holds:*
(1) $\forall x \in \mathrm{R} : K(-x) = K(x)$,
(2)$\int_{-\infty}^{\infty} K(y)dy = 1$.

Definition 2.2 (Kernel density estimator): *Let $X_1, ..., X_n$ be independent identically distributed random variables with probability density function f(x), K(x) a kernel and $h > 0$. Then the kernel density estimator $f_{nh}(x)$ is defined as*

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

The positive number h is called a bandwidth of the estimator.

Usually K is chosen to be a unimodal probability density function that is symmetric about zero. This ensures that $f_{nh}(x)$ is itself also a density. However, kernels that are not densities are also sometimes used. These are the kernels we mainly use in this paper:

kernel	
Uniform	$\frac{1}{2}\mathbb{I}\{\lvert x\rvert \leq 1\}$
Triangular	$(1 - \lvert x\rvert)\mathbb{I}\{\lvert x\rvert \leq 1\}$
Epanechnikov	$\frac{3}{4}(1 - x^2)\mathbb{I}\{\lvert x\rvert \leq 1\}$
Biweight	$\frac{15}{16}(1 - x^2)^2\mathbb{I}\{\lvert x\rvert \leq 1\}$
Triweight	$\frac{35}{32}(1 - x^2)^3\mathbb{I}\{\lvert x\rvert \leq 1\}$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$
Cosinus	$\frac{\pi}{4}\cos(\frac{\pi}{2}x)\mathbb{I}\{\lvert x\rvert \leq 1\}$

Figure 2.2 shows graphs of the mentioned kernels.

In order to understand kernel estimators, we will give an example of their use.We will use all the kernels listed above to estimate the density of a random sample of size 500 from an exponential distribution with parameter one. We will choose the bandwidth equal to 2.5. In Figure 2.3 we can see not only the resulting estimates, but also the weighted kernels centered at the observations. The estimate is then constructed as a sum of the kernel weights. The source code is included in the Appendix for more detailed examination of the graphs.

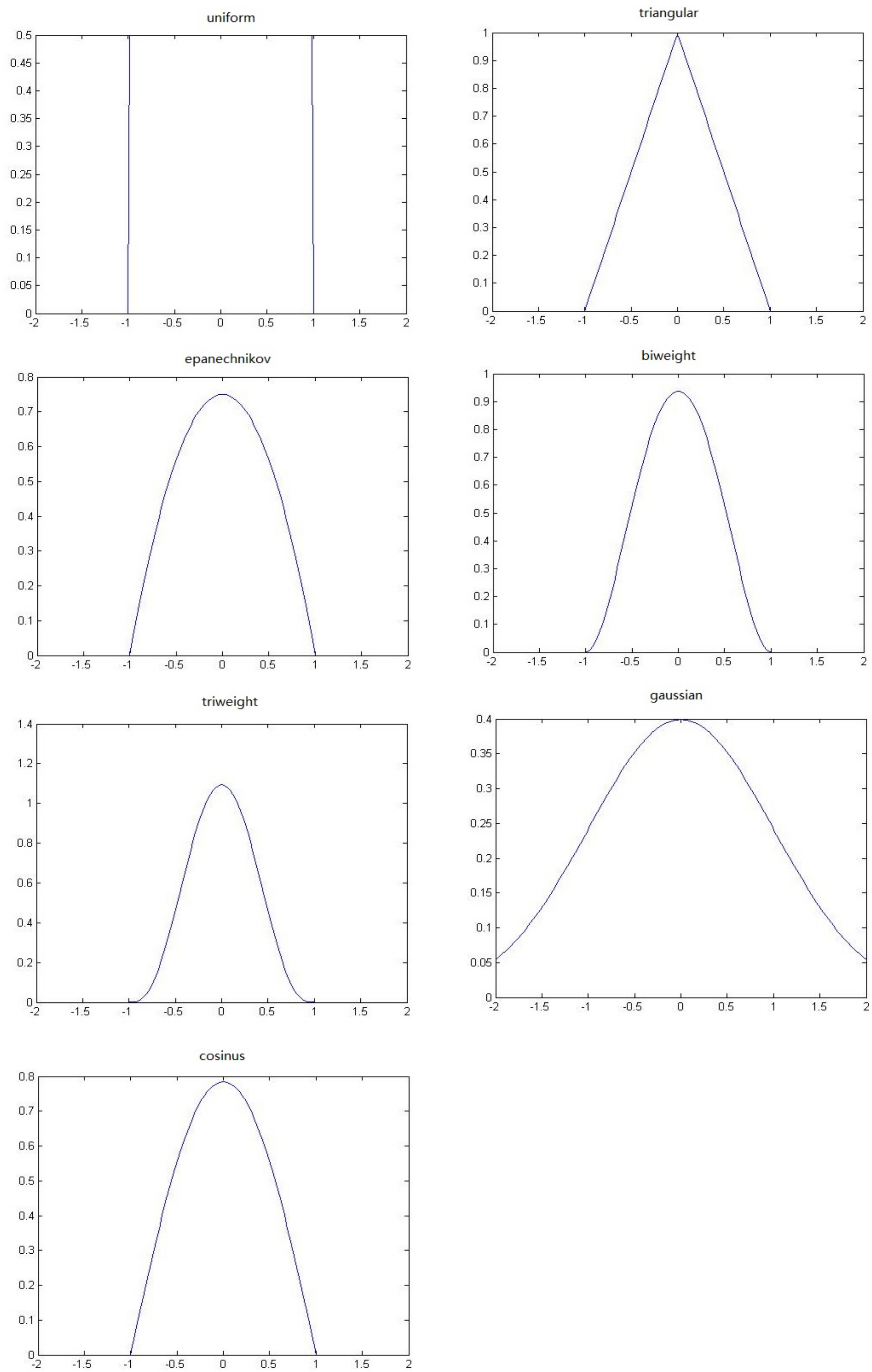

Figure 2.2: Graphical comparison of frequently used univariate kernels.

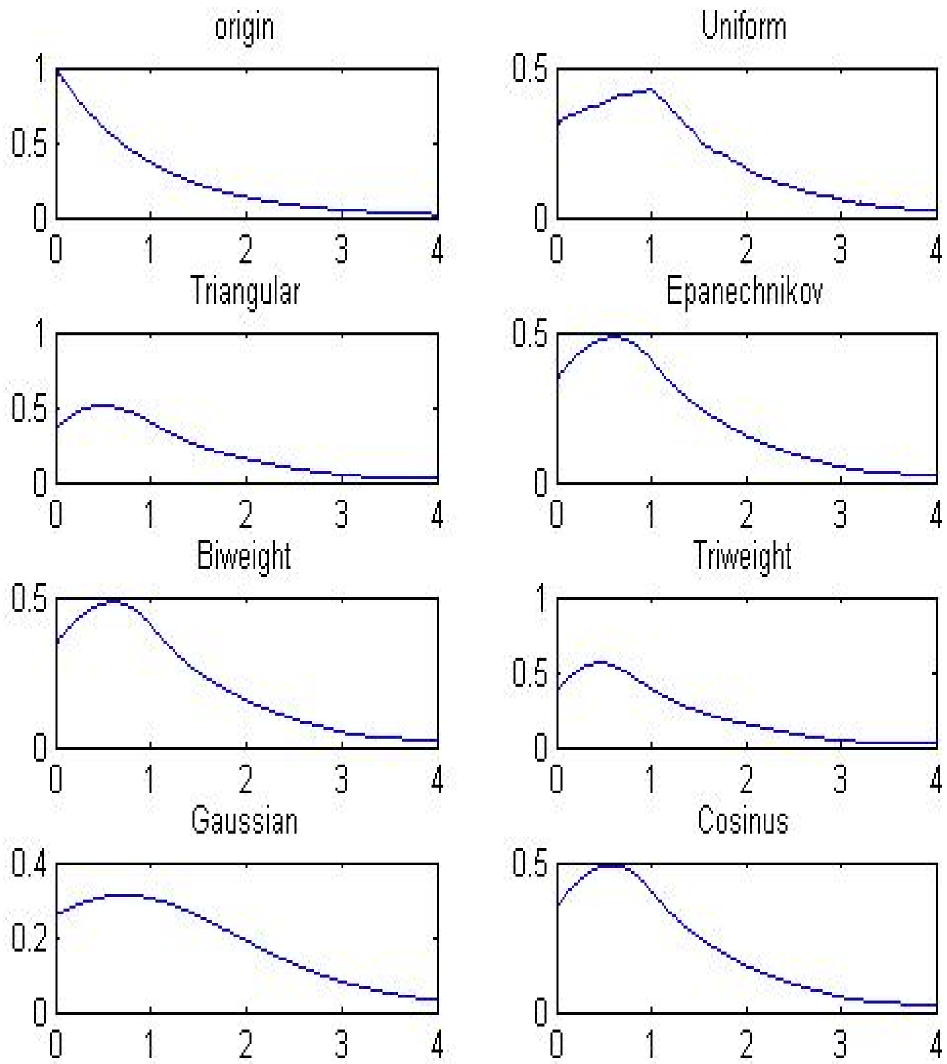

Figure 2.3: Construction of kernel density estimates.

Obviously the kernel estimators of the decreasing density have a bias at zero because of the boundary effect. That shows the need of special estimators for decreasing densities.

2.1.2 The MSE and MISE criteria

Our analysis of the performance of the kernel density estimator will require the specification of appropriate error criteria for measuring the error when estimating the density at a single point as well as the error when estimating the density over the whole real line.

In classical parametric statistics it is common to measure the closeness of an estimator $\hat{\theta}$ to its target parameter θ by the size of the mean squared error(MSE), given by

$$MSE(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta)^2$$

One appealing feature of MSE is its simple decomposition into variance and squared bias. We have

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (\mathbb{E}\hat{\theta} - \theta)^2$$

This error criterion is often preferred to other criteria such as mean absolute error $MAE(\hat{\theta}) = \mathbb{E}|\hat{\theta} - \theta|$ since it is mathematically simpler to work with.The variance-bias decomposition allows easier analysis and interpretation of the performance of the kernel density estimator as we will now show.

Consider $f_{nh}(x)$ as an estimator of the density function $f(x)$ at some point $x \in \mathbb{R}$. To compute MSE($f_{nh}(x)$) we will require expressions for the mean and variance of $f_{nh}(x)$.These can be derived directly from Definition 2.2.Let X be a random variable having density f.Firstly, we have

$$\mathbb{E}f_{nh}(x) = \mathbb{E}K_h(x - X) = \int K_h(x - y)f(y)dy, \tag{2.2}$$

where ,$K_h(x) = \left(\frac{1}{h}\right)K\left(\frac{x}{h}\right)$.
At this stage it is convennient to introduce the convolution notation,

$$(f * g)(x) = \int f(x - y)g(y)dy,$$

since this allows us to write the bias of $f_{nh}(x)$ as

$$\mathbb{E}f_{nh}(x) - f(x) = (K_h * f)(x) - f(x).$$

Convolution is usually thought of as a smoothing operation, so the bias is the difference between a smooth version of f and f itself. Similar calculations lead to

$$Var(f_{nh}(x)) = n^{-1}\{(K_h^2 * f)(x) - (K_h * f)^2(x)\} \tag{2.3}$$

and these may be combined to give

$$Var(f_{nh}(x)) = n^{-1}\{(K_h^2 * f)(x) - (K_h * f)^2(x)\} + \{(K_h * f)(x) - f(x)\}^2. \qquad (2.4)$$

Rather than simply estimating f at a fixed point, it is usually desirable,especially from a data analytic viewpoint, to estimate f over the entire real line. In this case our estimate is the function f_{nh} so we need to consider an error criterion that globally measures the distance between the functions f_{nh} and f. One such error criterion is the mean integrated squared error (MISE), given by

$$MISE(f_{nh}) = \mathbb{E}\int \{f_{nh}(x) - f(x)\}^2 dx$$

Note that by changing the order of integration we have

$$MISE(f_{nh}) = \mathbb{E}\int \{f_{nh}(x) - f(x)\}^2 dx = \int MSE(f_{nh}(x))dx.$$

This leads to the following representation of the mean integrated error,

$$MISE(f_{nh}) = n^{-1}\int \{(K_h^2 * f)(x) - (K_h^f)^2(x)\}dx + \int \{(K_h * f)(x) - f(x)\}^2 dx \quad (2.5)$$

Some straightforward manipulations lead to the more manageable expression

$$\begin{aligned} MISE(f_{nh}) =& (nh)^{-1}\int K^2(x)dx + (1-n^{-1})\int (K_h * f)^2(x)dx \\ & - 2\int (K_h * f)(x)f(x)dx \int f(x)^2 dx. \end{aligned}$$

While this is a relatively compact expression for MISE, it has the misfortune of depending on the bandwidth h in a fairly complicated fashion.This makes it difficult to interpret the influence of the bandwidth on the performance of the kernel density estimator.We will investigate a way of overcoming this problem that involves the derivation of large sample approximations for the leading variance and bias terms.These approximations have very simple expressions that allow a deeper appreciation of the role of the bandwidth. They can also be used to obtain the rate of convergence of the kernel density estimator and the MISE-optimal bandwidth.

Throughout this section we will make the following assumptions.

(i)The density f is such that its second derivative f'' is continuous, square integrable

and ultimately monotone.
(ii)The bandwidth $h = h_n$ is a non-random sequence of positive numbers. To keep the notation less cumbersome the dependence of h on n will be suppresses in our calculations. We also assume that h satisfies

$$\lim_{n\to\infty} h = 0, \lim_{n\to\infty} nh = \infty$$

which is equivalent to saying that h approaches zero, but at a rate slower than n^{-1}.
(iii)The kernel K is a bounded probability density function having finite fourth moment and symmetry about the origin.
An ultimately monotone function is one that is monotone over both $(-\infty, -M)$ and $(M, +\infty)$ for some $M > 0$. Note that conditions (i) and (iii) can be replaced by numerous other combinations of conditions on f and K so that the results presented here remain valid.

We first consider the estimation of $f(x)$ at $x \in \mathbb{R}$. We have, from (2.3) and a change of variables,

$$\mathbb{E}f_{nh}(x) = \int K(z)f(x - hz)dz.$$

Expanding $f(x - hz)$ in a Taylor series about x we obtain

$$f(x - hz) = f(x) - hzf'(x) + \frac{1}{2}h^2z^2f''(x) + o(h^2).$$

This leads to

$$\mathbb{E}f_{nh}(x) = f(x) + \frac{1}{2}h^2f''(x)\int z^2K(z)dz + o(h^2),$$

where we have used

$$\begin{cases} \int K(z)dz & = 1, \\ \int zK(z)dz & = 0, \\ \int z^2K(z)dz & < \infty. \end{cases}$$

each of which follow from assumption (iii). We now introduce the notation $\mu_2(K) = \int z^2K(z)dz$, which leads to the bias expansion

$$\mathbb{E}f_{nh}(x) - f(x) = \frac{1}{2}h^2\mu_2(K)f''(x) + o(h). \tag{2.6}$$

Notice that the bias is of order h^2 which implies that $f_{nh}(x)$ is asymptotically unbiased. Also noteworthy is the way that the bias depends on the true f. The bias is large whenever the absolute value of the second derivative of f is large, and this happens in regions where the curvature of the density is high. For many densities this occurs in peaks where the bias is negative, and valleys where thebias is positive. This shows that f_{nh} has a tendency to smooth out such features on average.
For the variance, note that from (2.3), we get

$$\begin{aligned} Var(f_{nh}(x)) &= \frac{1}{nh}\int K(z)^2 f(x-hz)dz - \frac{1}{n}\{\mathbb{E}f_{nh}(x)\}^2 \\ &= \frac{1}{nh}\int K(z)^2\{f(x)+o(1)\}dz - \frac{1}{n}\{f(x)+o(1)\}^2 \\ &= \frac{1}{nh}\int K(z)^2 dz f(x) + o(\frac{1}{nh}) \end{aligned}$$

Another useful notation is $R(g) = \int g(x)^2 dx$ for any square integrable function g. This allows us to write the variance as

$$Var(f_{nh}(x)) = \frac{1}{nh}R(K)f(x) + o\{\frac{1}{nh}\}. \tag{2.7}$$

Since the variance is of order $\frac{1}{nh}$ assumption (ii) ensures that $Var\big(f_{nh}(x)\big)$ converges to zero as $n \to \infty$.
Adding (2.7) and the square of (2.6) we obtain

$$MSE(f_{nh}(x)) = \frac{1}{nh}R(K)f(x) + \frac{1}{4}h^4\mu_4(K)^2 f''(x)^2 + o(\frac{1}{nh} + h^4).$$

If we integrate this expression then, under our integrability assumption on f, we obtain

$$MSE(f_{nh}(x)) = AMISE(f_{nh}(x)) + o\{\frac{1}{nh} + h^4\}$$

where

$$AMSE(f_{nh}(x)) = \frac{1}{nh}R(K)f(x) + \frac{1}{4}h^4\mu_4(K)^2 f''(x)^2.$$

We call this asymptotic MISE since it provides a useful large sample approximation to MISE.Notice that the integrated squared bias is asymptotically proportional to h^4, so for this quantity to decrease one needs to take h to be small. However, taking h small means an increase in the leading term of the integrated variance since this quantity is proportional to $\frac{1}{nh}$. This indicates the existence of an asymptotic optimal bandwidth.

2.2 Two-dimensional kernel estimation

2.2.1 Two-dimensional kernel estimators

Now we will investigate the extension of the univariate kernel density estimator to the bivariate setting.The bivariate kernel density estimator that we study in this section is a direct extension of the univariate estimators.

Definition 2.3(two-dimensional kernel):*Let K(x) be a nonnegative bounded integrable function with real values defined on $\mathbb{R}^2$. We say that K(x) is a two-dimensional kernel if and only if the following holds:*
(1) $\forall x \in \mathrm{R}^2 : K(-x) = K(x)$
(2)$\int_{\mathrm{R}^2} K(y)dy = 1$

Definition 2.4(two- dimensional Kernel density estimator): *Let $X_1, ..., X_n$ be independent identically distributed two-dimensional random vectors with probability density function f(x), K(x) be a two-dimensional kernel and $h > 0$. Then the kernel density estimator $f_{nh}(x)$ is defined as*

$$f_{nh}(\mathbf{x}) = \frac{1}{nh^2} \sum_{i=1}^{n} K\big(\frac{\mathbf{x} - \mathbf{X_i}}{h}\big) \tag{2.8}$$

In that case h is called a bandwidth of the estimator.

In this paper, we choose the two dimensional kernel as the product of two univariate kernels.

As in the previous section,Figure 2.3 shows comparisons of the kernels which are used in Section 2.1.
Also, we will give an example of use. We will using all the kernels listed above to estimate a random sample of size 500 from a two dimensional exponential distribution with parameter 1. So $\mathbf{X} = (E_1, E_2)$ with E_1 and E_2 independent and exp(1) distributed. We will choose bandwidth equal to 2.5.

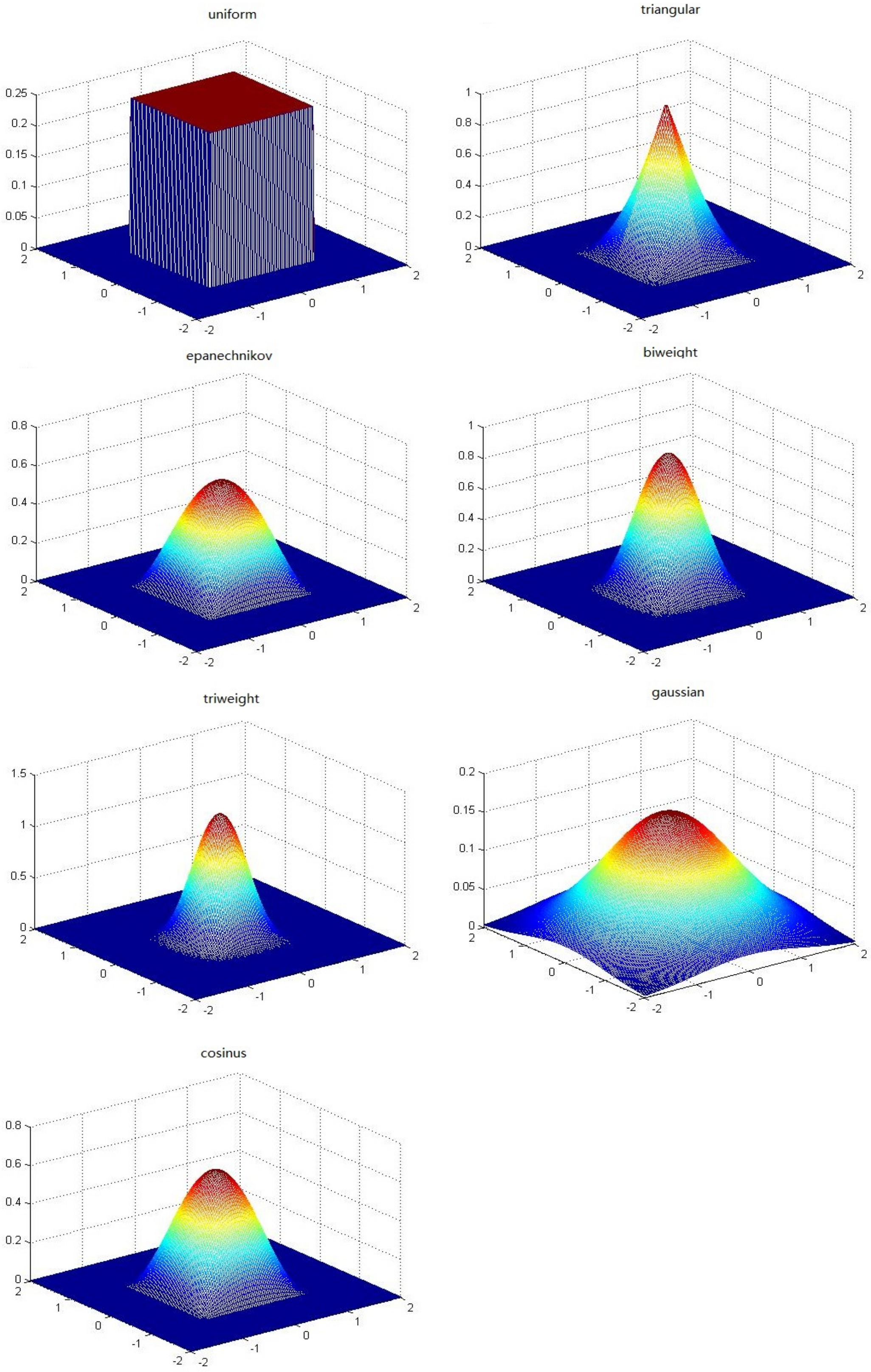

Figure 2.4: Graphical comparison of frequently used multivariate kernels.

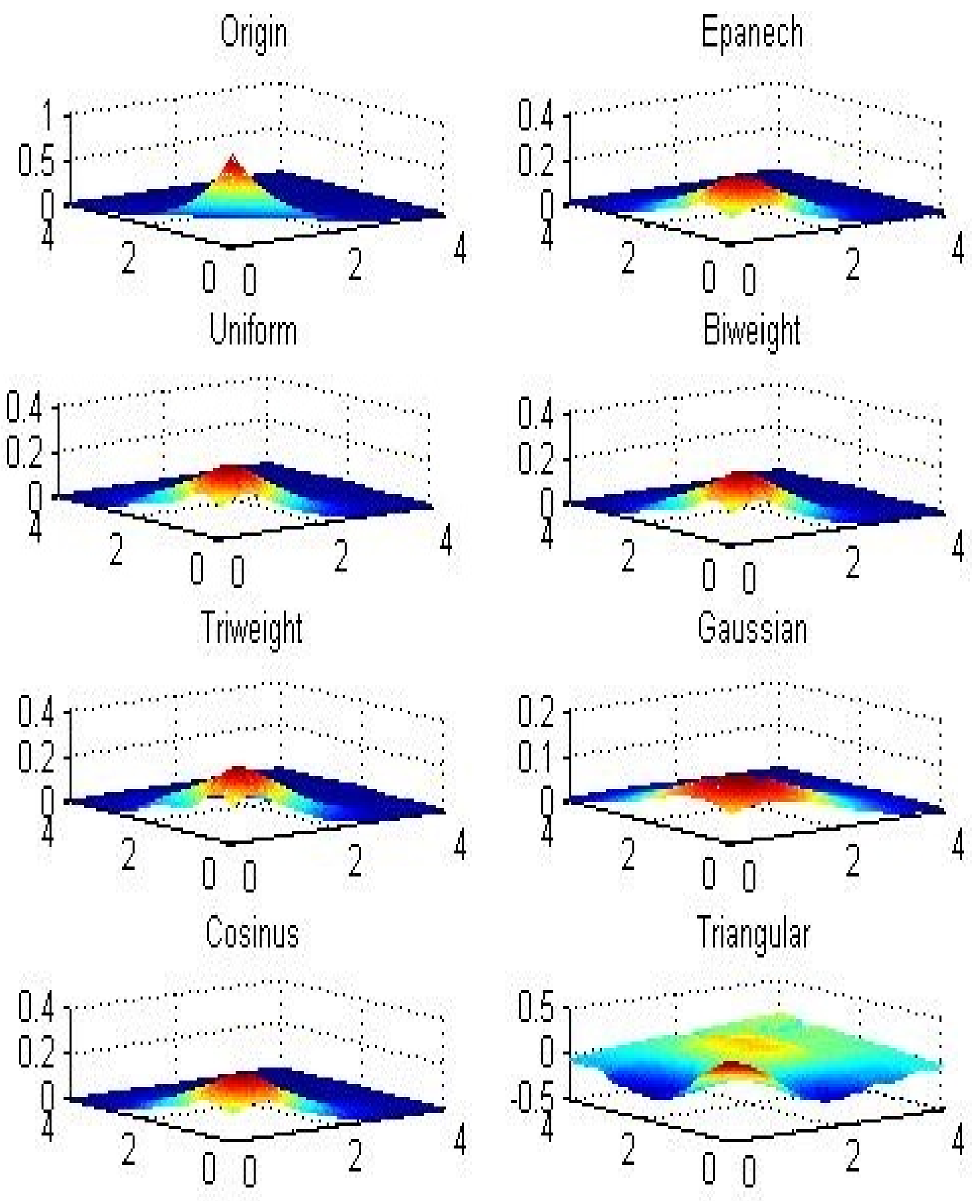

Figure 2.5: Construction of 2-dimensional kernel density estimates.

Similar to the previous section, the example also shows the boundary problem of the decreasing density at the zero point. Therefore, we need to find a better estimator for two dimensional decreasing densities.

Chapter 3

The Gamma Deconvolution kernel estimator

In the previous chapter, we have introduced the concept of convolution. Now we are going to deal with the deconvolution problem.

3.1 One-dimensional Gamma deconvolution kernel estimator

In many instances in statistics we have to deal with observations X_i which are equal in distribution to the independent sum of random variables of interest Y_i and random variables Z_i, where the distribution of Z_i can be assumed to be known. For instance consider a value Y_i which is measured with measurement error Z_i. The deconvolution problem can be described as follows. Let X_i denote an observation which is equal to the sum of two independent random variables Z_i and Y_i. Assume that Z_i has distribution function K which is known, and that Y_i has distribution function F which is unknown. Now assume that the observation X_i has distribution function G, say, which is equal to the convolution of the functions K and F. So, we have $G(x) = (K * dF)(x) := \int_{\mathbb{R}} K(x-t)dF(t)$. If K has density k, then G also has a density $g(x) = (k * dF)(x) = \int_{\mathbb{R}} k(x-t)dF(t)$. Our aim is to estimate F or its density f, based on a sample of $X_1, ..., X_n$ of independent identically distributed observations from G. The following theorem can be found in Van Es and Kok(1995)

Theorem 3.1(one-dimensional Gamma deconvolution): *Let :*

$$X = Y + Z,$$
$$Z = \lambda_1 E_1 + ... + \lambda_m E_m$$

where the λ_i are positive constants, $E_1, ...E_m$ and Y are independent, that the E_i are standard exponential random variables and that Y has a distribution function F. Let G denote the distribution function of X.Then

$$F = \sum_{j=0}^{m} s_m^{(j)}(\lambda_1, ..., \lambda_m) G^{(j)}, \tag{3.1}$$

where $s_m^{(j)}(\lambda_1, ..., \lambda_m)$ denotes the jth-order elementary symmetric polynomial of the m variables $\lambda_1, ..., \lambda_m$ given by $s_m^{(0)} = 1$ and

$$s_m^{(j)}(\lambda_1, ..., \lambda_m) = \sum_{1 \le i_1 < ... < i_j \le m} \lambda_{i_1} ... \lambda_{i_j}.$$

Proof. We prove the theorem by an induction argument. For $m = 1$, the inversion formula equals to $F = G + \lambda g$, which is straightforwardly derived as follows:

$$G(x) + \lambda g(x) = \int_{-\infty}^{x} (1 - e^{-\frac{1}{\lambda}(x-t)}) dF(t) + \lambda \int_{-\infty}^{x} \frac{1}{\lambda} e^{-\frac{1}{\lambda}(x-t)} dF(t) = F(x)$$

Note that the equality holds for all x, and λ an arbitrary positive number.
Now we assume that the inversion formula holds for $m = m_0$. Let $\tilde{G}$ denote the distribution function of $\lambda_1 E_1 + ... + \lambda_{m0} E_{m_0} + Y$.
For $X = \lambda_1 E_1 + ... + \lambda_{m0} E_{m_0} + Y + \lambda_{m_0+1} E_{m_0+1}$, we have by the equality for m=1

$$\tilde{G}(x) = G(x) + \lambda_{m_0+1} G(x)^{(1)}$$

Note that $\xi_{m+1} s_m^{(j-1)}(\xi_1, ..., \xi_m) + s_m^{(j)}(\xi_1, ...\xi_m) = s_{(m+1)}^{(j)}(\xi_1, ..., \xi_{m+1})$. We will denote $s_m^{(j)}(\lambda_1, ..., \lambda_m)$ by $s_m^{(j)}$. $i = 0, ..., m$. For $i = m + 1$ the equality holds under the convention $s_m^{(m+1)}(\xi_1, .., \xi_m) = 0$. Suppose that (3.1) holds for $m = m_0$. Then

$$
\begin{aligned}
F &= \sum_{j=0}^{m_0} s_{m_0}^{(j)} \tilde{G}^{(j)} \\
&= \sum_{j=0}^{m_0} s_{m_0}^{(j)} (G + \lambda_{m_0+1} g)^{(j)} \\
&= \sum_{j=0}^{m_0} s_{m_0}^{(j)} G^{(j)} + \lambda_{m_0+1} \sum_{j=0}^{m_0} s_{m_0}^{(j)} G^{(j+1)} \\
&= \sum_{j=0}^{m_0} s_{m_0}^{(j)} G^{(j)} + \sum_{j=0}^{m_0} s_{m_0+1}^{(j+1)} G^{(j+1)} - \sum_{j=0}^{m_0} s_{m_0}^{(j+1)} G^{j+1} \\
&= G + \sum_{j=1}^{m_0+1} s_{m_0+1}^{(j)} G^{(j)} - s_{m_0}^{(m_0+1)} G^{(m_0+1)} \\
&= \sum_{j=0}^{m_0+1} s_{m_0+1}^{(j)} G^{(j)}
\end{aligned}
$$

since $s_{m_0}^{(m_0+1)} = 0$.
So (3.1) holds for $m = m_0 + 1$ and by induction it is valid for all m.

□

By plugging in kernel estimators of G and the derivatives of G, the inversion formulas lead to estimators of F and, when it exists, its density f. For Gamma deconvolution we get

$$
F_{nh}^{mE}(x) = G_{nh}(x) + \sum_{j=1}^{m} s_m^{(j)}(\lambda_1, ..., \lambda_m) g_{nh}^{(j-1)}(x)
$$

as estimator of $F(x)$ and

$$
f_{nh}^{mE}(x) = g_{nh}(x) + \sum_{j=1}^{m} s_m^{(j)}(\lambda_1, ..., \lambda_m) g_{nh}^{(j)}(x)
$$

as estimator of $f(x)$.

Theorem 3.2:*Assume that F is twice differentiable with a bounded continuous second derivative and that w is a symmetric bounded support probability density with continuous $(m-1)th$ derivative $w^{(m-1)}$.Then, with$\sigma^2 = \int u^2 w(u) du$, we have*

$$\mathbb{E}F_{nh}^{mE}(x) = F(x) + \frac{1}{2}h^2\sigma^2 \sum_{j=0}^{m} s_m^{(j)}(\lambda_1, ..., \lambda_m)g^{(j+1)}(x) + o(h^2), \tag{3.2}$$

$$VarF_{nh}^{mE}(x) = \frac{(\lambda_1, ..., \lambda_m)^2}{nh^{2m-1}}g(x)\int_{-\infty}^{\infty} w^{(m-1)}(u)^2du + o(\frac{1}{nh^{2m-1}}). \tag{3.3}$$

Theorem 3.3:*Assume that F is three times differentiable with a bounded continuous second derivative and that w is a symmetric bounded support probability density with continuous (m)th derivative $w^{(m)}$.Then, with $\sigma^2 = \int u^2w(u)du$, we have*

$$\mathbb{E}f_{nh}^{mE}(x) = f(x) + \frac{1}{2}h^2\sigma^2 \sum_{j=0}^{m} s_m^{(j)}(\lambda_1, ..., \lambda_m)g^{(j+2)}(x) + o(h^2), \tag{3.4}$$

$$Varf_{nh}^{mE}(x) = \frac{(\lambda_1, ..., \lambda_m)^2}{nh^{2m+1}}g(x)\int_{-\infty}^{\infty} w^{(m)}(u)^2du + o(\frac{1}{nh^{2m+1}}). \tag{3.5}$$

The proofs of Theorem 3.2 and Theorem 3.3 are very similar, and can be found in Van Es and Kok(1995)

3.2 Two dimensional Gamma deconvolution kernel estimator

Now, we extend X, Y, Z to two-dimensional versions.Similarly, as in the previous section we have

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

where $\mathbf{X} = (X_1, X_2)$ is the observed vector, $\mathbf{Y} = (Y_1, Y_2)$ is the vector with unknown distribution and Z_1, Z_2 are independent random variables equal to weighted sum of independent exponential random variables.

The following results can be found in Krycha(2011).

Theorem 3.4(2-dimensional Gamma deconvolution): *Let :*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$
$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 E_{11} + ... + \lambda_m E_{1m} \\ \tau_1 E_{21} + ... + \tau_n E_{2n} \end{pmatrix}$$

$n, m \in \mathbb{N}$,assume that Z_i are mutually independent, $(Y_1, Y_2)^T$ is independent from $\mathbf{Z}$ *and the E_{1i} and E_{2j} are standard exponential random variables, $\lambda_i \geqslant 0, \tau_i \geqslant 0$, $i = 1, ..., m; j = 1, , ..., n$. Let g be the density of $X = (X_1, X_2)^T$ and f the density of $Y = (Y_1, Y_2)^T$, where f is of class C^{m+n} on $\mathbb{R}^2$ with integrable partial derivatives up to order $m+n$. Then*

$$f(x_1, x_2) = \sum_{j=0}^{n} \sum_{i=0}^{m} s_m^{(i)}(\lambda_1, ...\lambda_m) s_n^{(j)}(\tau_1, ...\tau_n) \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g(x_1, x_2).$$

Corollary 3.5: *Let:*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} \lambda E_1 \\ \tau E_2 \end{pmatrix}$$

Assume that $(Y_1, Y_2)^T$, E_1, E_2 are independent. Let g be the density of $X = (X_1, X_2)$ and f the density of $Y = (Y_1, Y_2)$. Assume that E_1, E_2 are standard exponential random variables, $\lambda, \tau \geq 0$ and f is of class $\mathbb{C}^2$ on $\mathbb{R}^2$ with integrable partial derivatives up to order 2.Then

$$f(x_1, x_2) = g(x_1, x_2) + \lambda \frac{\partial}{\partial x_1} g(x_1, x_2) + \tau \frac{\partial}{\partial x_2} g(x_1, x_2) + \lambda\tau \frac{\partial^2}{\partial x_1 \partial x_2} g(x_1, x_2). \quad (3.6)$$

Now, we are going to combine the findings established above in order to define a kernel estimator in the deconvolution problem. We will want to derive the corresponding statistical properties of the estimator just as we described them in Chapter2 in relation to multivariate estimations, i.e.dominating parts of the bias and variance. We have already known the result of the Gamma deconvolution. If we know the observe realization of independent random vectors $X_1, ...X_N$, we can substitute the unknown density g with its kernel estimator counterpart. For the derivatives of the unknown density g we will simply use corresponding derivatives of the kernel estimator, which will yield desirable properties. The resulting the deconvolution kernel estimator is:

$$f_{Nh}(x_1, x_2) = \sum_{j=0}^{n} \sum_{i=0}^{m} s_m^{(i)}(\lambda_1, ..., \lambda_m) s_n^{(j)}(\tau_1, ..., \tau_n) \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g_{Nh}(x_1, x_2)$$

In order to obtain the bias and variance of this estimator, we must first expand the properties derived in the previous chapter to the used derivative estimators. We will do so in the proof of the following theorem, which can be found in Krycha.

Theorem 3.6: *Let the assumption of theorem 3.4 hold. More over, assume that :*
1.*$g(x)$ is of class C^{m+n+2} and its partial derivatives up to order $m+n+2$ are bounded*

2.*For the kernel K satisfying the Definition in Chapter 2 and $1 \leq i \leq m$,$1 \leq j \leq n$, the following holds:*

$$\int (\frac{\partial^{i+j} K(\mathbf{y})}{\partial y_1^i \partial y_2^j})^2 d\mathbf{y} < \infty$$

3.*For the kernel K and some constant $\mu_2(K)$ the following holds:*

$$\int \mathbf{y}^T \mathbf{y} K(\mathbf{y}) d\mathbf{y} = \begin{pmatrix} \mu_2(K) & 0 \\ 0 & \mu_2(K) \end{pmatrix}$$

Then the bias and the variance of the two-dimensional gamma deconvolution kernel estimator are as follows:

$$\mathbb{E}[f_{Nh}(\mathbf{x}) - f(\mathbf{x})] = \frac{1}{2} h^2 \mu_2(K) \nabla^2 f(\mathbf{x}) + o(h^2) \tag{3.7}$$

$$Var(f_{Nh}(\mathbf{x}) = \int ((\frac{\partial^{m+n} K(\mathbf{y})}{\partial y_1^n \partial y_2^m})^2) \frac{(\lambda_1 ... \lambda_m \tau_1 ... \tau_n)^2 g(\mathbf{x})}{N h^{2n+2m+1}} \tag{3.8}$$

Proof. Analogically to (2.6) and (2.7), $Chac\acute{o}n et al.$ (2011) show that under the stated assumptions we can write

$$\mathbb{E}[\frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g_{Nh}(\mathbf{x})] = \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g(\mathbf{x}) + \frac{1}{2} h^2 \mu_2(K) \nabla^2 \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g(\mathbf{x}) + o(h^2) \tag{3.9}$$

$$Var(\frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g_{Nh}(\mathbf{x})) = \int (\frac{\partial^{i+j} K(\mathbf{x})}{\partial y_1^i \partial y_2^j})^2 d\mathbf{y} \frac{g(\mathbf{x})}{N h^{2i+2j+1}} + o(\frac{1}{N h^{2i+2j+1}}) \tag{3.10}$$

We can calculate the bias of the deconvolution kernel estimator as

$$\begin{aligned}
&\mathbb{E}[f_{Nh}(\mathbf{x})] = \sum_{j=0}^{n} \sum_{i=0}^{m} s_m^{(i)}(\lambda_1, ... \lambda_m) s_n^{(j)}(\tau_1, ..., \tau_n) \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g(\mathbf{x}) \\
&+ \frac{1}{2} h^2 \mu_2(K) \sum_{j=0}^{n} \sum_{i=0}^{m} s_m^{(i)}(\lambda_1, ... \lambda_m) s_n^{(j)}(\tau_1, ..., \tau_n) \nabla^2 \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g(\mathbf{x}) + o(h^2) \\
&= f(\mathbf{x}) + \nabla^2 (\frac{1}{2} h^2 \mu_2(K) \sum_{j=0}^{n} \sum_{i=0}^{m} s_m^{(i)}(\lambda_1, ... \lambda_m) s_n^{(j)}(\tau_1, ..., \tau_n) \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g(\mathbf{x}) + o(h^2)) + o(h^2) \\
&= f(\mathbf{x}) + \frac{1}{2} h^2 \mu_2(K) \nabla^2 f(\mathbf{x}) + o(h^2)
\end{aligned}$$

For the variance, one has to realize that $f_{Nh}(\mathbf{x})$ has the same asymptotic variance as the last summand of the double sum, while the variance of the other summands is negligible (van Es and Kok, 1998). First we have

$$Var(f_{Nh}(\mathbf{x})) = \sum_{i=0}^{m}\sum_{j=0}^{n}\sum_{k=0}^{m}\sum_{l=0}^{n} c_{i,j,k,l}Cov(\frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g_{Nh})(x_1,x_2), \frac{\partial^{k+l}}{\partial x_1^k \partial x_2^l} g_{Nh}(x_1,x_2)),$$

where

$$c_{i,j,k,l} = s_m^{(i)}(\lambda_1,...,\lambda_m)s_n^{(j)}(\tau_1,...,\tau_n)s_m^{(k)}(\lambda_1,...,\lambda_m)s_n^{(l)}(\tau_1,...,\tau_n)$$

Using the Cauchy-Schwarz inequality, we can write

$$\begin{aligned}
&|Cov(\frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g_{Nh}(x_1,x_2), \frac{\partial^{k+l}}{\partial x_1^k \partial x_2^l} g_{Nh}(x_1,x_2))| \\
&\leq \sqrt{Var(\frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} g_{Nh}(x_1,x_2)}\sqrt{\frac{\partial^{k+l}}{\partial x_1^k \partial x_2^l} g_{Nh}(x_1,x_2)} \\
&= \begin{cases} \int(\frac{\partial^{m+n}K((y))}{\partial y_1^n \partial y_2^m})^2 d\mathbf{y}\frac{g(\mathbf{x})}{Nh^{2n+2m+1}} + o(\frac{1}{Nh^{2n+2m+1}}) & \text{if } i=k=m, j=l=n, \\ O(\frac{1}{Nh^{i+j+k+l+1}}) = o(\frac{1}{Nh^{2n+2m+1}}) & \text{otherwise.} \end{cases}
\end{aligned}$$

Finally

$$\begin{aligned}
Var(f_{Nh}(\mathbf{x})) &= c_{m,n,m,n}\int(\frac{\partial^{m+n}K((y))}{\partial y_1^n \partial y_2^m})^2 d\mathbf{y}\frac{g(\mathbf{x})}{Nh^{2n+2m+1}} + o(\frac{1}{Nh^{2n+2m+1}}) \\
&= \int(\frac{\partial^{m+n}K((y))}{\partial y_1^n \partial y_2^m})^2 d\mathbf{y}\frac{(_1...\lambda_m\tau_1...\tau_n)^2 g(\mathbf{x})}{Nh^{2n+2m+1}} + o(\frac{1}{Nh^{2n+2m+1}})
\end{aligned}$$

which finishes the proof.

□

This provides us with a rigorous setting for the implementation part. We can end this chapter with stating the AMISE-optimal bandwidth for the covered estimators. Let us start with the gamma deconvolution estimator. The mean square error is equal to the sum of the squared bias and the variance We integrate the mean square error and get

$$\begin{aligned}
AMISE(f_{Nh}(\mathbf{x})) &= \frac{h}{4}\mu_2^2(K)\int_{\mathbf{R}^2}(\nabla^2 f(\mathbf{x}))^2 d\mathbf{x} \\
&+ \int(\frac{\partial^{m+n}K((y))}{\partial y_1^n \partial y_2^m})^2 d\mathbf{y}\frac{(_1...\lambda_m\tau_1...\tau_n)^2}{Nh^{2n+2m+1}}.
\end{aligned}$$

Minimizing AMISE in a standard way leads us to an expression for the optimal bandwidth:

$$\hat{h} = (\int (\frac{\partial^{m+n} K((y))}{\partial y_1^n \partial y_2^m})^2 d\mathbf{y} \frac{(2n+2m+1)({}_1 ... \lambda_m \tau_1 ... \tau_n)^2}{N\mu_2^2(K) \int_{\mathbb{R}^2} (\nabla^2 f(\mathbf{x}))^2 d\mathbf{x}})^{\frac{1}{2n+2m+3}}$$

Tn this thesis we will not consider bandwidth selection. But it is in principle possible to estimate this optimal bandwidth from the data.

Chapter 4

Estimation of decreasing densities

We will apply the results for exponential deconcolution to the problem of estimating decreasing densities.

4.1 One dimension

4.1.1 Estimation of decreasing densities

Using one dimensional exponential deconvolution, we are going to develop a method to estimate decreasing densities. Consider a positive random variable X. Assume that X has a positive density function f(x) such that for all $x \in [0, \infty), \forall \delta > 0, f(x + \delta) \geq f(x)$. Assume that there are random variables U and Y, such that we can write $X = UY$, where U follows the uniform distribution on (0,1), U and Y are independent. In fact this presentation holds for every decreasing density, see Feller(1971), page 158. If we apply the logarithm to both sides of the equation and define random $A = -\log X$, $B = -\log Y$, then we can write $-\log X = -\log Y - \log U$, which is equivalent to $A = B - \log U$.
If we denote $t(x) = -\log x$ and use the theorem for transformed random variables, we get that the probability of the density function of $-\log U$, equals

$$f_{-\log U}(u) = f_U(t^{-1}(u))\dot{|}[t^{-1}(u)']| = \mathbb{I}_{(0,1)}(\exp(-u))\exp(-u) = \mathbb{I}_{(0,\infty)}(u)\exp(-u),$$

i.e. $-\log U$ is an independent random variable with exponential distribution with parameter equal to 1. Assume that we have a random sample of observations $X_1, ...X_n$ from the density f_X. Denote by

$$A_i = -\log X_i, i = 1, ..., n,$$

the observations of the transformed random vector. Note that recovering the density of B, and hence of Y, from the observed A_i, is a one dimensional exponential deconvolution problem. For any selected kernel and bandwidth h we can obtain a non-parametric kernel density estimator f_B^n of the unknown density of the random variable B. The estimator is equal to

$$f_B^{(n)}(x) = f_A^{(n)}(x) + \frac{\partial}{\partial x} f_A^{(n)}(x),$$

where

$$f_A^{(n)}(x) = \frac{1}{nh} \sum_{i=1}^{n} w\big(\frac{x - A_i}{h}\big). \tag{4.1}$$

Having obtained an estimator of f_B, we can easily proceed to estimating the density of $Y = \exp(-B)$. Again using the random transformation theorem we get

$$f_Y(y) = f_B(-\log(y))\frac{1}{y}, y > 0 \tag{4.2}$$

and the estimator is defined as

$$f_Y^{(n)}(y) = f_B^{(n)}(-\log(y))\frac{1}{y}, y > 0 \tag{4.3}$$

Next we establish the relation between the densities of X and B.

Theorem 4.1:*Given the context described in this section, for $y > 0$, we can write*

$$f_X(y) = \int_0^\infty f_B(-\log x)\mathbb{I}_{(0,x)}(y)\frac{1}{x^2}dx$$

Proof. To derive the density of X, we to consider the function:

$$t(Y, U) = t(\mathbf{W}) = (YU, U)$$

Now the density of the transformed variable $\mathbf{W}$ is

$$f_{\mathbf{W}}(y, u) = f_{Y,U}(t^{-1}(y, u)) \cdot |D_{\frac{\partial}{\partial \mathbf{W}} t^{-1}(y,u)}|$$

where

$$D_{\frac{\partial}{\partial \mathbf{W}} t^{-1}(y,u)} = \begin{vmatrix} \frac{1}{u} & -\frac{y}{u^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{u}.$$

Then we can write

$$f_{\mathbf{W}}(y,u) = f_Y(\frac{y}{u})\mathbb{I}_{(0,1)}(u)\frac{1}{u}, y > 0,$$

and we obtain the density of X by integrating. We get

$$f_X(y) = \int_0^\infty f_Y(\frac{y}{u})\mathbb{I}_{(0,1)}(u)\frac{1}{u}du,$$

which can be rewritten by substitution $x = \frac{y}{u}$ as

$$f_X(y) = \int_0^\infty f_Y(x)\mathbb{I}_{(0,x)}(y)\frac{1}{x}dx,$$

Substitution of (4.1) yields the formula for the density of X, i.e.

$$f_X(y) = \int_0^\infty f_B(-\log x)\mathbb{I}_{(0,x)}(y)\frac{1}{x^2}dx. \tag{4.4}$$

□

If we substitute an estimator $f_B^{(n)}$ for f_B in this formula then the result will not automatically be decreasing since the estimators used here are not automatically positive. Hence we have to truncate $f_B^{(n)}$ at zero and renormalize.
Finally, we can write the estimator of the decreasing density of X as

$$\tilde{f_X}^{(n)}(y) = \frac{1}{C_n}\int_0^\infty (f_B^{(n)}(-\log x) \wedge 0)\mathbb{I}_{(o,x)}(y)\frac{1}{x^2}dx, \tag{4.5}$$

where the normalizing constant C_n equals $\int_{-\infty}^{\infty}(f_B^{(n)}(u) \wedge 0)du$. The nontruncated estimator will be denoted by $f_X^{(n)}$.

4.1.2 Expectation and Variance

Recall the asymptotics for one dimensional exponential deconvolution. We have the following theorem.

Theorem 4.2: *Let the assumption of the Theorem 3.2 hold. Moreover, we assume* *that $m = 1$, $\lambda_1 = 1$. Then the bias and the variance of the one-dimensional exponential deconvolution kernel estimator defined by (4.5) are as follows, we have*

$$\begin{aligned} \mathbb{E}f_X^{(n)}(y) &= f_X(y) + \frac{1}{2}h^2\sigma^2[\int_y^\infty g^{(2)}(-\log x)dx + \int_y^\infty g^{(3)}(-\log x)dx] + o(h^2), \\ Varf_X^{(n)}(y) &= O(\frac{1}{nh^2}). \end{aligned}$$

Proof. In Chapter 3, by theorem 3.3, for $m = 1$, $\lambda_1 = 1$, we have

$$\begin{aligned} \mathbb{E}f_{nh}^{1E}(x) &= f(x) + \frac{1}{2}h^2\sigma^2(g^{(2)}(x) + g^{(3)}(x)) + o(h^2), \\ Varf_{nh}^{1E}(x) &= \frac{1}{nh^3}g(x)\int_{-\infty}^\infty w^{(1)}(u)^2du + o(\frac{1}{nh^3}). \end{aligned}$$

Therefore, using the fact that $f_B^{(n)}$ equals an estimator like f_{nh}^{1E} based on the sample $-X_1, ..., -\log X_n$, we have

$$\begin{aligned} \mathbb{E}f_X^{(n)}(y) &= \mathbb{E}\int_0^\infty f_B^{(n)}(-\log x)\mathbb{I}_{(0,x)}(y)\frac{1}{x^2}dx \\ &= \mathbb{E}\int_y^\infty f_B^{(n)}(-\log x)\frac{1}{x^2}dx \\ &= \int_y^\infty \mathbb{E}f_B^{(n)}(-\log x)\frac{1}{x^2}dx \\ &= \int_y^\infty \mathbb{E}(f(-\log x) + \frac{1}{2}h^2\sigma^2(g^{(2)}(-\log x)dx + g^{(3)}(-\log x)dx) + o(h^2))\frac{1}{x^2}dx \\ &= f_X(y) + \frac{1}{2}h^2\sigma^2\big[\int_y^\infty g^{(2)}(-\log x)dx + \int_y^\infty g^{(3)}(-\log x)dx\big] + o(h^2) \end{aligned}$$

Now, we are going to focus on the variance. We have for standard exponential deconvolution $f = g + g$.
Hence, by (4.1) and (4.4) we get

$$\begin{aligned} f_X^{(n)}(y) &= \int_0^\infty f_B^{(n)}(-\log x)\mathbb{I}_{(0,x)}(y)\frac{1}{x^2}dx \\ &= \int_0^\infty [\frac{1}{nh}\sum_{i=1}^n w(\frac{x+\log X_i}{h}) + \frac{1}{nh^2}\sum_{i=1}^n w'(\frac{x+\log X_i}{h})]\mathbb{I}_{(0,x)}(y)\frac{1}{x^2}dx \\ &= \int_y^\infty [\frac{1}{nh}\sum_{i=1}^n w(\frac{x+\log X_i}{h}) + \frac{1}{nh^2}\sum_{i=1}^n w'(\frac{x+\log X_i}{h})]\frac{1}{x^2}dx \end{aligned}$$

where w is the kernel function. Note that in this case, $f_X^{(n)}(y)$ is calculated without cutting off at zero and renormalizing.
Let

$$\begin{aligned} Q_1 &= \int_y^\infty \frac{1}{nh}\sum_{i=1}^n w\big(\frac{x+\log X_i}{h}\big)\frac{1}{x^2}dx \\ &= \frac{1}{nh}\sum_{i=1}^n \int_y^\infty w\big(\frac{x+\log X_i}{h}\big)\frac{1}{x^2}dx \end{aligned}$$

and

$$\begin{aligned} Q_2 &= \int_y^\infty \frac{1}{nh^2}\sum_{i=1}^n w'\big(\frac{x+\log X_i}{h}\big)]\frac{1}{x^2}dx \\ &= \frac{1}{nh^2}\sum_{i=1}^n \int_y^\infty w'\big(\frac{x+\log X_i}{h}\big)\frac{1}{x^2}dx. \end{aligned}$$

Then, we have

$$\begin{aligned} VarQ_1 &= Var\big(\frac{1}{nh}\sum_{i=1}^n \int_y^\infty w\big(\frac{x+\log X_i}{h}\big)\frac{1}{x^2}dx\big) \\ &= \frac{1}{n}Var\big(\frac{1}{h}\int_y^\infty w\big(\frac{x+\log X_1}{h}\big)\frac{1}{x^2}dx\big) \\ &\leq \frac{1}{n}\mathbb{E}\big(\frac{1}{h}\int_y^\infty w\big(\frac{x+\log X_1}{h}\big)\frac{1}{x^2}dx\big)^2 \end{aligned}$$

Note that

$$\begin{aligned} \big(\frac{1}{h}\int_y^\infty w\big(\frac{x+\log X_1}{h}\big)\frac{1}{x^2}dx\big)^2 &\leq \frac{1}{h^2}\big(\int_y^\infty w\big(\frac{x+\log X_1}{h}\big)\frac{1}{y^2}dx\big)^2 \\ &= \frac{1}{h^2y^4}\big(\int_y^\infty w\big(\frac{x+\log X_1}{h}\big)dx\big)^2. \end{aligned}$$

Let $t = \frac{x+\log X_1}{h}$, then this is equal to

$$\frac{1}{y^4}\big(\int_{\frac{y+\log X_1}{h}}^\infty w(t)dt\big)^2 \leq \frac{1}{y^4}\big(\int_{-\infty}^\infty w(t)dt\big)^2 = \frac{1}{y^4}.$$

On the other hand, we have

$$
\begin{aligned}
VarQ_2 &= Var(\frac{1}{nh^2}\sum_{i=1}^{n}\int_y^\infty w'(\frac{x+\log X_i}{h})\frac{1}{x^2}dx) \\
&= \frac{1}{n}Var(\frac{1}{h^2}\int_y^\infty w'(\frac{x+\log X_1}{h})\frac{1}{x^2}dx) \\
&\leq \frac{1}{n}\mathbb{E}(\frac{1}{h^2}\int_y^\infty w'(\frac{x+\log X_1}{h})\frac{1}{x^2}dx)^2.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&(\frac{1}{h^2}\int_y^\infty w'(\frac{x+\log X_1}{h})\frac{1}{x^2}dx)^2 \leq (\frac{1}{h^2}\int_y^\infty |w'|(\frac{x+\log X_1}{h})\frac{1}{y^2}dx)^2 \\
\leq &\frac{1}{h^4}(\int_y^\infty |w'|(\frac{x+\log X_1}{h})\frac{1}{y^2}dx)^2 = \frac{1}{h^4y^4}(\int_y^\infty w'(\frac{x+\log X_1}{h})dx)^2
\end{aligned}
$$

Let $t = \frac{x+\log X_1}{h}$, then this is equal to

$$
\frac{1}{h^2y^4}(\int_{\frac{y+\log X_1}{h}}^\infty |w'|(t)dt)^2 \leq \frac{1}{h^2y^4}(\int_{-\infty}^\infty |w'|(t)dt)^2
$$

Now, use

$$
\begin{aligned}
Var(Q_1+Q_2) &\leq \mathbb{E}(Q_1+Q_2)^2 \\
&\leq 2(\mathbb{E}Q_1^2+\mathbb{E}Q_2^2),
\end{aligned}
$$

since $(a+b)^2 \leq 2(a^2+b^2)$.
Therefore we get the result.

$$
\begin{aligned}
Varf_X^{(n)}(y) &\leq 2(\mathbb{E}Q_1^2+\mathbb{E}Q_2^2) \\
&\leq 2(\mathbb{E}(\frac{1}{ny^2})+\mathbb{E}(\frac{1}{nh^2y^4})) \\
&= O(\frac{1}{nh^2})
\end{aligned}
$$

□

The theorem shows that for pointwise consistency we need $n \to \infty$, $h \to 0$, $nh^2 \to \infty$. This theorem establishes the asymptotics for the nontruncated estimator $f_X^{(n)}(x)$. We expect that the same asymptotics hold for $\tilde{f_X}^{(n)}(x)$.

4.2 Two dimensions

4.2.1 Estimating decreasing densities

The approach we have developed so far can provide us with an algorithm for estimating two-dimensional decreasing densities. Consider a positive two-dimensional random vector $X = (X_1, X_2)$ with support $S = [0, \infty) \times [0, \infty)$. Assume that X has a probability density function $f(x_1, x_2)$ such that for all $(x_1, x_2) \in S$ the following holds:

$$\forall \delta > 0 : f(x_1 + \delta, x_2) \leq f(x_1, x_2),$$
$$\forall \delta > 0 : f(x_1, x_2 + \delta) \leq f(x_1, x_2),$$

i.e. it is decreasing in both directions. In this case, the standard non-parametric estimation approach with kernel density estimation proves ineffective as the constructed estimate is generally a function that is positive outside the support and increasing. Building a different algorithm might diminish this disadvantage, while keeping the desired properties of a kernel estimator. Assume that there are random vectors $\mathbf{U}$,$\mathbf{Y}$ such that we can write

$$\mathbf{X} \stackrel{\text{i.d.}}{=} \mathbf{Y} \bigodot \mathbf{U} = (Y_1 U_1, Y_2 U_2)$$

where $U1, U2$ are independent random variables with uniform distribution on $(0, 1)$, $\mathbf{U}$ and Y are independent, $=^{\text{i.d.}}$stands for equal in distribution and $\bigodot$

denotes the Hadamard product, i.e. product of the corresponding elements of the matrix. In fact all two dimensional decreasing densities can be represented this way, but a reference is hard to be found.

If we apply the logarithm to the both sides of the equation and define random vectors $\mathbf{A} = -\log(\mathbf{X})$ and $\mathbf{B} = -\log(\mathbf{Y})$, then we can write

$$\begin{aligned}
&\mathbf{X} \overset{\text{id}}{=} \mathbf{Y} \bigodot \mathbf{U} \\
\Leftrightarrow & -\log(\mathbf{X}) = -\log(\mathbf{Y}) - \log(\mathbf{U}) \\
\Leftrightarrow & \mathbf{A} = \mathbf{B} - \log(\mathbf{U})
\end{aligned}$$

If we denote $t(x) = -\log x$ and consider the transformation theorem for random vectors, then we get that the probability density function of $-\log(U_i)$, $i = 1, 2$ is equal to

$$f_{-\log U_i}(u) = f_{U_i}(t^{-1}(u)) \cdot |(t^{-1}(u))'| = \mathbb{I}_{(0,1)}(\exp(-u))\exp(-u) = \mathbb{I}_{(0,\infty)}(u)\exp(-u)$$

i.e. the $\log(U_i)$ are independent random variables with an exponential distribution with parameters equal to 1.
Assume that we have a random sample of observations $\mathbf{X_1}, ..., \mathbf{X_n}$ from the density $f_{\mathbf{X}}$. Denote

$$\mathbf{A_i} = (A_{1i}, A_{2i}) = (-\log(X_{1i}), -\log(X_{2i})) = -\log(\mathbf{X_i}), i = 1, ..., n$$

the observations of the transformed random vector. Provided that the necessary assumptions are fulfilled, we are now in the situation specified by the Corollary. For any selected two-dimensional kernel and bandwidth h we can construct a two dimensional non-parametric density estimator $f_{\mathbf{B}}^{(n)}$ of the unknown density of the random vector $\mathbf{B}$. The estimator is defined as

$$f_{\mathbf{B}}^{(n)}(x_1, x_2) = f_{\mathbf{A}}^{(n)}(x_1, x_2) + \frac{\partial}{\partial x_1} f_{\mathbf{A}}^{(n)}(x_1, x_2) + \frac{\partial}{\partial x_2} f_{\mathbf{A}}^{(n)}(x_1, x_2) + \frac{\partial^2}{\partial x_{12}} f_{\mathbf{A}}^{(n)}(x_1, x_2)$$

where $f_{\mathbf{A}}^{(n)}$ equals a two dimensional kernel estimator,

$$f_{\mathbf{A}}^{(n)}(x_1, x_2) = \frac{1}{nh^2} \sum_{i=1}^{n} w(\frac{x1 - X_{i1}}{h}) w(\frac{x2 - X_{i2}}{h})$$

Having obtained the estimator $f_{\mathbf{B}}$, we can easily proceed to estimate $\mathbf{Y} = \exp(-\mathbf{B})$. Since

$$\begin{aligned}
&\mathbb{P}(-\log Y_1 \leq y_1, -\log Y_2 \leq y_2) = \mathbb{P}(Y_1 \geq e^{-y_1}, Y_2 \geq e^{-y_2}) \\
&= 1 - \mathbb{P}(Y_1 \leq e^{-y_1}) - \mathbb{P}(Y_2 \leq e^{-y_2}) + \mathbb{P}(Y_1 \leq e^{-y_1}, Y_2 \leq e^{-y_2})
\end{aligned}$$

We get

$$
\begin{aligned}
f_{\mathbf{B}}(y_1, y_2) &= \frac{\partial^2}{\partial y_1 \partial y_2} \mathbb{P}(-\log Y_1 \leq y_1, -\log Y_2 \leq y_2) \\
&= \frac{\partial^2}{\partial y_1 \partial y_2} (\mathbb{P}(Y_1 \leq e^{-y_1}, Y_2 \leq e^{-y_2})) \\
&= \frac{\partial}{\partial y_1} (\frac{\partial}{\partial y_2} \mathbb{P}(Y_1 \leq e^{-y_1}, Y_2 \leq e^{-y_2})) \\
&= \frac{\partial}{\partial y_1} (-e^{-y_2} \frac{\partial}{\partial y_2} F_{\mathbf{Y}}(e^{-y_1}, e^{-y_2})) \\
&= e^{-y_1} e^{-y_2} \frac{\partial^2}{\partial y_1 \partial y_2} F(e^{-y_1}, e^{-y_2}) \\
&= e^{-(y_1+y_2)} f_{\mathbf{Y}}(e^{-y_1}, e^{-y_2}).
\end{aligned}
$$

Therefore we have

$$
f_{\mathbf{Y}}(y_1, y_2) = f_{\mathbf{B}}(-\log(y_1), -\log(y_2)) \frac{1}{y_1 y_2}, y_1, y_2 > 0
$$

and the estimator is equal to

$$
f_{\mathbf{Y}}^{(n)}(y_1, y_2) = f_{\mathbf{B}}^{(n)}(-\log(y_1), -\log(y_2)) \frac{1}{y_1 y_2}, y_1, y_2 > 0.
$$

As a last step will now establish the relation between the density of $\mathbf{X}$ and the density of $\mathbf{B}$.

Theorem 4.3: *Given the context described in this chapter, for* $y_1, y_2 > 0$ *we can write*:

$$
f_{\mathbf{X}}(y_1, y_2) = \int_0^\infty \int_0^\infty f_{\mathbf{B}}(-\log(x_1), -\log(x_2)) \mathbb{I}_{(0,x_1)}(y_1) \mathbb{I}_{(0,x_2)}(y_2) \frac{1}{x_1^2 x_2^2} dx_1 dx_2
$$

Proof. To derive the density of X, consider the function

$$
t(Y_1, Y_2, U_1, U_2) = t(\mathbf{W}) = (Y_1 U_1, Y_2 U_2, U_1, U_2)
$$

Now the density of the transformed variable W is equal to

$$
f_{\mathbf{W}}(y_1, y_2, u_1, u_2) = f_{\mathbf{Y},\mathbf{U}}(t^{-1}(y_1, y_2, u_1, u_2)) \cdot |D_{\frac{\partial}{\partial \mathbf{W}} t^{-1}(y_1, y_2, u_1, u_2)}|
$$

where

$$D_{\frac{\partial}{\partial \mathbf{W}}t^{-1}(y_1,y_2,u_1,u_2)} = \begin{vmatrix} \frac{1}{u_1} & 0 & -\frac{y_1}{u_1^2} & 0 \\ 0 & \frac{1}{u_2} & 0 & -\frac{y_1}{u_2^2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = \frac{1}{u_1 \cdot u_2}$$

Then we can write:

$$f_{\mathbf{W}}(y_1,y_2,u_1,u_2) = f_{\mathbf{Y}}(\frac{y_1}{u_1},\frac{y_2}{u_2})\mathbb{I}_{(0,1)}(u_1)\mathbb{I}_{(0,1)}(u_2)\frac{1}{u_1 \cdot u_2}, y_1, y_2 > 0$$

and we obtain the density of X by integrating

$$f_{\mathbf{X}}(y_1,y_2) = \int_0^\infty \int_0^\infty f_{\mathbf{Y}}(\frac{y_1}{u_1},\frac{y_2}{u_2})\mathbb{I}_{(0,1)}(u_1)\mathbb{I}_{(0,1)}(u_2)\frac{1}{u_1 \cdot u_2}du_1du_2,$$

which can be rewritten by substitution $(x_1,x_2) = (\frac{y_1}{u_1},\frac{y_2}{u_2})$ as

$$f_{\mathbf{X}}(y_1,y_2) = \int_0^\infty \int_0^\infty f_{\mathbf{Y}}(x_1,x_2)\mathbb{I}_{(0,x_1)}(y_1)\mathbb{I}_{(0,x_2)}(y_2)\frac{1}{x_1 \cdot x_2}dx_1dx_2.$$

Substituting the previous estimator results, we get

$$f_{\mathbf{X}}(y_1,y_2) = \int_0^\infty \int_0^\infty f_{\mathbf{B}}(-\log(x_1),-\log(x_2))\mathbb{I}_{(0,x_1)}(y_1)\mathbb{I}_{(0,x_2)}(y_2)\frac{1}{x_1^2x_2^2}dx_1dx_2$$

□

To get an estimator which is itself decreasing we estimate $f_{\mathbf{B}}$ by

$$f_{\mathbf{B}}^{(n)}(x_1,x_2) = \frac{1}{D_n}(f_{\mathbf{B}}^{(n)}(x_1,x_2) \wedge 0),$$

where $D_n = \int_{-\infty}^\infty \int_{-\infty}^\infty f_{\mathbf{B}}^{(n)}(u_1,u_2)du_1du_2$. We have

$$\begin{aligned} f_{\mathbf{B}}^{(n)}(x_1,x_2) =& \frac{1}{nh^2}\sum_{i=1}^n w(\frac{x+\log X_1i}{h})w(\frac{x+\log X_2i}{h}) \\ &+ \frac{1}{nh^3}\sum_{i=1}^n w'(\frac{x+\log X_1i}{h})w(\frac{x+\log X_2i}{h}) \\ &+ \frac{1}{nh^3}\sum_{i=1}^n w(\frac{x+\log X_1i}{h})w'(\frac{x+\log X_2i}{h}) \\ &+ \frac{1}{nh^4}\sum_{i=1}^n w'(\frac{x+\log X_1i}{h})w'(\frac{x+\log X_2i}{h}) \end{aligned}$$

The final estimator is then equal to

$$\tilde{f_{\mathbf{X}}}^{(n)}(y_1, y_2) = \frac{1}{D_n} \int_{y_1}^{\infty} \int_{y_2}^{\infty} \big(f_{\mathbf{B}}(-\log x_1, -\log x_2) \wedge 0\big) \frac{1}{x_1^2 x_2^2} dx_1 dx_2 \tag{4.6}$$

We will denote the nontruncated estimator of $f_{\mathbf{X}}(\mathbf{x})$ by $f_{\mathbf{X}}^{(n)}(\mathbf{x})$.

4.2.2 Expectation and Variance

Similar to one dimensional case, we can device asymptotics for the expectation and variance in the two dimensional case. We have the following theorem

Theorem 4.4: *Let the assumption of the theorem 3.6 hold. Moreover, we assume that* $m = 1$,$n = 1$,$\lambda_1 = 1$,$\tau_1 = 1$. *Then the bias and the variance of the two-dimensional exponential deconvolution kernel estimator* $f_{\mathbf{X}}^{(n)}$ *are as follows*:

$$\begin{aligned} &\mathbb{E} f_{\mathbf{X}}^{(n)}(y_1, y_2) = f_{\mathbf{X}}(y_1, y_2) + \frac{1}{2} h^2 \mu_2(K) \int_{y_1}^{\infty} \int_{y_2}^{\infty} \nabla^2 f(\mathbf{x}) dx_1 dx_2 + o(h^2), \\ &Var f_X^{(n)}(y) = O(\frac{1}{nh^4}). \end{aligned}$$

Proof. Since Theorem 3.6 tell us that

$$\mathbb{E}[f_h^{(n)}(\mathbf{x}) - f(\mathbf{x})] = \frac{1}{2} h^2 \mu_2(K) \nabla^2 f(\mathbf{x}) + o(h^2)$$

We have

$$\mathbb{E}[f_{\mathbf{B}}^{(N)}(-\log \mathbf{x}) - f(-\log \mathbf{x})] = \frac{1}{2} h^2 \mu_2(K) \nabla^2 f(-\log \mathbf{x}) + o(h^2)$$

We need to calculate $\mathbb{E}[f_{\mathbf{X}}^{(n)}(x)]$,which is equal to

$$
\begin{aligned}
&\mathbb{E}\int_0^\infty\int_0^\infty f_{\mathbf{B}^{(n)}}(-\log(x_1),-\log(x_2))\mathbb{I}_{(0,x_1]}(y_1)\mathbb{I}_{(0,x_2]}(y_2)\cdot\frac{1}{x_1^2x_2^2}dx_1dx_2\\
=&\mathbb{E}\int_{y_1}^\infty\int_{y_2}^\infty f_{\mathbf{B}}^{(n)}(-\log(x_1),-\log(x_2))\cdot\frac{1}{x_1^2x_2^2}dx_1dx_2.
\end{aligned}
$$

Therefore we have the following://

$$
\begin{aligned}
\mathbb{E}f_{\mathbf{X}}^{(n)}(y_1,y_2)=&\int_{y_1}^\infty\int_{y_2}^\infty[f(-\log(x_1),-\log(x_2))\frac{1}{x_1^2x_2^2}dx_1dx_2\\
&+\int_{y_1}^\infty\int_{y_2}^\infty\frac{1}{2}h^2\mu_2(K)\bigtriangledown^2 f(-\log(x_1),-\log(x_2))]\cdot\frac{1}{x_1^2x_2^2}dx_1dx_2\\
=&f_{\mathbf{X}}(y_1,y_2)+\frac{1}{2}h^2\mu_2(K)\int_{y_1}^\infty\int_{y_2}^\infty\nabla^2f(\mathbf{x})dx_1dx_2+o(h^2)
\end{aligned}
$$

Now, we are going to focus on the variance of $f_{\mathbf{X}}^{(n)}(y_1,y_2)$. We have

$$
\begin{aligned}
f_{\mathbf{X}}^{(n)}(y_1,y_2)=&\int_{y_1}^\infty\int_{y_2}^\infty[\frac{1}{nh^2}\sum_{i=1}^n w(\frac{x+\log X_1i}{h})w(\frac{x+\log X_2i}{h})\\
&+\frac{1}{nh^3}\sum_{i=1}^n w'(\frac{x+\log X_1i}{h})w(\frac{x+\log X_2i}{h})\\
&+\frac{1}{nh^3}\sum_{i=1}^n w(\frac{x+\log X_1i}{h})w'(\frac{x+\log X_2i}{h})\\
&+\frac{1}{nh^4}\sum_{i=1}^n w'(\frac{x+\log X_1i}{h})w'(\frac{x+\log X_2i}{h})]\frac{1}{x_1^2x_2^2}dx_1dx_2
\end{aligned}
$$

where w is the a one dimensional kernel function. Note that in this case, $f_{\mathbf{X}}(y_1,y_2)$ is calculated without cutting off at zero and renormalizing.
Let

$$
\begin{aligned}
Q_1=&\int_{y_1}^\infty\int_{y_2}^\infty\frac{1}{nh^2}\sum_{i=1}^n w(\frac{x_1+\log X_{1i}}{h})w(\frac{x_2+\log X_{2i}}{h})\frac{1}{x_1^2x_2^2}dx_1dx_2\\
=&\frac{1}{n}\sum_{i=1}^n\frac{1}{h^2}\int_{y_1}^\infty\int_{y_2}^\infty w(\frac{x_1+\log X_{1i}}{h})w(\frac{x_2+\log X_{2j}}{h})\frac{1}{x_1^2x_2^2}dx_1dx_2
\end{aligned}
$$

$$Q_2 = \int_{y_1}^{\infty}\int_{y_2}^{\infty} \frac{1}{nh^3}\sum_{i=1}^{n} w'\big(\frac{x_1+\log X_{1i}}{h}\big)w\big(\frac{x_2+\log X_{2i}}{h}\big)]\frac{1}{x_1^2x_2^2}dx_1dx_2$$
$$= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^3}\int_{y_1}^{\infty}\int_{y_2}^{\infty} w'\big(\frac{x_1+\log X_{1i}}{h}\big)w\big(\frac{x_2+\log X_{2i}}{h}\big)\frac{1}{x_1^2x_2^2}dx_1dx_2.$$

$$Q_3 = \int_{y_1}^{\infty}\int_{y_2}^{\infty} \frac{1}{nh^3}\sum_{i=1}^{n} w\big(\frac{x_1+\log X_{1i}}{h}\big)w'\big(\frac{x_2+\log X_{2i}}{h}\big)]\frac{1}{x_1^2x_2^2}dx_1dx_2$$
$$= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^3}\int_{y_1}^{\infty}\int_{y_2}^{\infty} w\big(\frac{x_1+\log X_{1i}}{h}\big)w'\big(\frac{x_2+\log X_{2i}}{h}\big)\frac{1}{x_1^2x_2^2}dx_1dx_2.$$

and

$$Q_4 = \int_{y_1}^{\infty}\int_{y_2}^{\infty} \frac{1}{nh^4}\sum_{i=1}^{n} w'\big(\frac{x_1+\log X_{1i}}{h}\big)w'\big(\frac{x_2+\log X_{2i}}{h}\big)]\frac{1}{x_1^2x_2^2}dx_1dx_2$$
$$= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^4}\int_{y_1}^{\infty}\int_{y_2}^{\infty} w'\big(\frac{x_1+\log X_{1i}}{h}\big)w'\big(\frac{x_2+\log X_{2i}}{h}\big)\frac{1}{x_1^2x_2^2}dx_1dx_2.$$

It turns out the variance of Q_4 dominates the other three variances. Hence we will only derive a bound for the variance of Q_4.

$$VarQ_4 = Var\big(\frac{1}{nh^4}\sum_{i=1}^{n}\int_{y_1}^{\infty}\int_{y_2}^{\infty} w'\big(\frac{x_1+\log X_{1i}}{h}\big)w'\big(\frac{x_2+\log X_{2i}}{h}\big)\frac{1}{x_1^2x_2^2}dx_1dx_2\big)$$
$$= \frac{1}{n}Var\big(\frac{1}{h^4}\int_{y_1}^{\infty}\int_{y_2}^{\infty} w'\big(\frac{x_1+\log X_{11}}{h}\big)w'\big(\frac{x_2+\log X_{21}}{h}\big)\frac{1}{x_1^2x_2^2}dx_1dx_2\big)$$
$$\leq \frac{1}{n}\mathbb{E}\big(\frac{1}{h^4}\int_{y_1}^{\infty}\int_{y_2}^{\infty} w'\big(\frac{x_1+\log X_{11}}{h}\big)w'\big(\frac{x_2+\log X_{21}}{h}\big)\frac{1}{x_1^2x_2^2}dx_1dx_2\big)^2$$

Note that

$$\big(\frac{1}{h^4}\int_{y_1}^{\infty}\int_{y_2}^{\infty} w'\big(\frac{x_1+\log X_{11}}{h}\big)w'\big(\frac{x_2+\log X_{21}}{h}\big)\frac{1}{x_1^2x_2^2}dx_1dx_2\big)^2$$
$$\leq \frac{1}{h^8}\big(\int_{y_1}^{\infty}\int_{y_2}^{\infty} |w'|\big(\frac{x_1+\log X_{11}}{h}\big)|w'|\big(\frac{x_2+\log X_{21}}{h}\big)\frac{1}{y_1^2y_2^2}dx_1dx_2\big)^2$$
$$= \frac{1}{h^8y_1^4y_2^4}\big(\int_{y_1}^{\infty}\int_{y_2}^{\infty} |w'|\big(\frac{x_1+\log X_{11}}{h}\big)|w'|\big(\frac{x_2+\log X_{21}}{h}\big)dx_1dx_2\big)^2.$$

Let $t_1 = \frac{x_1+\log X_{11}}{h}$ and $t_2 = \frac{x_2+\log X_{21}}{h}$, then it equals to

$$\begin{aligned}
&\frac{1}{h^4}\Big(\int_{\frac{y_1+\log X_{11}}{h}}^{\infty}\int_{\frac{y_2+\log X_{21}}{h}}^{\infty}|w'|(t_1)|w'|(t_2)dt_1dt_2\Big)^2\\
&\le \frac{1}{h^4y_1^4y_2^4}\Big(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|w'|(t_1)|w'|(t_2)dt_1dt_2\Big)^2\\
&= \frac{1}{h^4y_1^4y_2^4}\Big(\int_{-\infty}^{\infty}|w'|(t)dt\Big)^4
\end{aligned}$$

Now, use

$$\begin{aligned}
Var(Q_1+Q_2+Q_3+Q_4) &\le \mathbb{E}(Q_1+Q_2+Q_3+Q_4)^2\\
&\le 16(\mathbb{E}Q_1^2+\mathbb{E}Q_2^2+\mathbb{E}Q_3^2+\mathbb{E}Q_4^2),
\end{aligned}$$

since $(a+b+c+d)^2 \le 16(a^2+b^2+c^2+d^2)$.
Therefore, we get the result

$$\begin{aligned}
Var f_{\mathbf{X}}^{(n)}(y_1,y_2) &\le 2(\mathbb{E}Q_1^2+\mathbb{E}Q_2^2+\mathbb{E}Q_3^2+\mathbb{E}Q_4^2)\\
&\le 2(\mathbb{E}(\frac{1}{ny_1^4y_2^4})+2\mathbb{E}(\frac{1}{nh^2y_1^4y_2^4})+\mathbb{E}(\frac{1}{nh^2y_1^4y_2^4})\mathbb{E}(\frac{1}{nh^4y_1^4y_2^4}))\\
&= O(\frac{1}{nh^4}),
\end{aligned}$$

which is claimed in the theorem. □

4.3 An application

4.3.1 One-dimension

Now we will specify a problem to illustrate applications of this method. Let X be our observation, and denote $D = -\log X$. Let $D = N + E$, where N denote the random variables with normal distribution, and E denote the random variables with exponential distribution. The density of X is decreasing on $[0,\infty)$, since $X = e^{-D} = e^{-N}\cdot e^{-E}$. As we know $U = e^{-E}$ obey an Uniform distribution.
Now we are going to use the method above to implement this problem.
Let $N \sim N(0,1)$ and $E \sim \exp(1)$. We are going to calculate the real density of X in order to get the absolute error of our method when we do simulations.
The density of X is given by

$$
\begin{aligned}
f_X(x) &= \frac{d}{dx}\mathbb{P}(X \le x) = \frac{d}{dx}\mathbb{P}(-\log X \ge -\log x) \\
&= \frac{d}{dx}(1 - F_{-\log X}(-\log x)) = \frac{1}{x} f_{-\log X}(-\log x)
\end{aligned}
$$

Taking the convolution of $f_E(x) = e^{-x}\mathbb{I}_{[0,\infty)}(x)$ and $f_N(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, we get

$$
\begin{aligned}
f_D(x) = f_{-\log X}(x) &= \int_{-\infty}^{\infty} f_E(u) f_N(x-u)du \\
&= \int_0^{\infty} e^{-u}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-u)^2}{2}}du \\
&= \frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-\frac{x^2+2u(1-x)+u^2}{2}}du \\
&= \frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-\frac{(1-x)^2+2u(1-x)+u^2+(2x-1)}{2}}du \\
&= \frac{1}{\sqrt{2\pi}}e^{-x+\frac{1}{2}}\int_0^{\infty} e^{-\frac{(u+(1-x))^2}{2}}du \\
&= (1-\Phi(1-x))e^{-x+\frac{1}{2}} = e^{-x+\frac{1}{2}}\Phi(x-1)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
f_X(x) &= \frac{1}{x}(1-\Phi(1+\log x))e^{\log x+\frac{1}{2}} \\
&= \Phi(-\log x - 1)\sqrt{e}.
\end{aligned}
$$

When $x \to 0$, it is easy to see that $-\log x - 1 \to +\infty$, so $\Phi(-\log x - 1) \to 1$.Then we get that f_X converges to $\sqrt{e}$ as $x \to 0$.

Next, we are going to consider the situation that N does not follow a standard normal distribution , but $N \sim N(\mu, \sigma^2)$. Repeating the steps above, we have

$$
\begin{aligned}
f_D(x) = f_{-\log X}(x) &= \int_{-\infty}^{\infty} f_E(u) f_N(x-u)du \\
&= \int_0^{\infty} e^{-u} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu-u)^2}{2\sigma^2}} du \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\infty} e^{-\frac{2\mu\sigma^2+x^2-2(\mu+u)x+(\mu+u)^2}{2\sigma^2}} du \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2-2+\mu^2}{2\sigma^2}} \int_0^{\infty} e^{-\frac{u^2+2u(\mu+\sigma^2-x)}{2\sigma^2}} du \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mu-x)^2}{2\sigma^2}} \int_0^{\infty} e^{-\frac{[u+(\mu+\sigma^2-x)]^2}{2\sigma^2}} du \cdot e^{\frac{(\mu+\sigma^2-x)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{2\mu+\sigma^2-2x}{2}} \sigma \cdot \int_0^{\infty} e^{-\frac{[u+(\mu+\sigma^2-x)]^2}{2\sigma^2}} d(\frac{u+(\mu+\sigma^2-x)}{\sigma}) \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{2\mu+\sigma^2-2x}{2}} \cdot \int_0^{\infty} e^{-\frac{[\frac{u+(\mu+\sigma^2-x)}{\sigma}]^2}{2}} d(\frac{u+(\mu+\sigma^2-x)}{\sigma})
\end{aligned}
$$

Let $q = \frac{u+(\mu+\sigma^2-x)}{\sigma}$, then we get

$$
\begin{aligned}
f_{-\log X}(x) =& \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{2\mu+\sigma^2-2x}{2}} \int_{\frac{\mu+\sigma^2-x}{\sigma}}^{\infty} e^{-\frac{q^2}{2}} dq \\
=& (1-\Phi(\frac{\mu+\sigma^2-x}{\sigma})) e^{\frac{2\mu+\sigma^2-2x}{2}}
\end{aligned}
$$

Hence

$$
\begin{aligned}
f_X(x) &= \frac{1}{x} e^{\frac{2\mu+\sigma^2+2\log x}{2}} (1-\Phi(\frac{\mu+\sigma^2+\log x}{\sigma})) \\
&= e^{\frac{2\mu+\sigma^2}{2}} (1-\Phi(\frac{\mu+\sigma^2+\log x}{\sigma})) \\
&= e^{\frac{2\mu+\sigma^2}{2}} \Phi(\frac{-\mu-\sigma^2-\log x}{\sigma}).
\end{aligned}
$$

Therefore $f_X(x) \to e^{\frac{2\mu+\sigma^2}{2}}$ as $x \to 0$.

4.3.2 Two dimensions

We will approach the construction of two dimensional decreasing densities as in the previous paragraph. Write

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \mathbf{E} = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix}, \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \mathbf{e} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$
$$\mathbf{D} = -\log \mathbf{X}, \mathbf{N} \sim N(0, \Sigma), E_1 \sim \exp(1), E_2 \sim \exp(1).$$

We will compute $f_{-\log \mathbf{X}}$, we get

$$\begin{aligned}
f_{\mathbf{D}}(\mathbf{x}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{E}}(u_1,u_2) f_{\mathbf{N}}(x_1-u_1, x_2-u_2)du_1du_2 \\
&= \int_0^{\infty}\int_0^{\infty} e^{-u_1}e^{-u_2}\frac{1}{2\pi}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{u})^T\Sigma^{-1}(\mathbf{x}-\mathbf{u})}du_1du_2 \\
&= \frac{1}{2\pi|\Sigma|^{1/2}}\int_0^{\infty}\int_0^{\infty} e^{-\mathbf{u}^T\mathbf{e}-\frac{1}{2}(\mathbf{x}^T\Sigma^{-1}\mathbf{x}-\mathbf{x}^T\Sigma^{-1}\mathbf{u}-\mathbf{u}^T\Sigma^{-1}\mathbf{x}+\mathbf{u}^T\Sigma^{-1}\mathbf{u})}du_1du_2 \\
&= \frac{1}{2\pi|\Sigma|^{1/2}}\int_0^{\infty}\int_0^{\infty} e^{-\frac{1}{2}(\mathbf{x}^T\Sigma^{-1}\mathbf{x}+2\mathbf{u}^T(\Sigma^{-1}\Sigma\mathbf{e}-\Sigma^{-1}\mathbf{x})+\mathbf{u}^T\Sigma^{-1}\mathbf{u})}du_1du_2 \\
&= \frac{1}{2\pi|\Sigma|^{1/2}}\int_0^{\infty}\int_0^{\infty} e^{-\frac{1}{2}(\mathbf{x}^T\Sigma^{-1}\mathbf{x}+2\mathbf{u}^T\Sigma^{-1}(\Sigma\mathbf{e}-\mathbf{x})+\mathbf{u}^T\Sigma^{-1}\mathbf{u})}du_1du_2 \\
&= \frac{1}{2\pi|\Sigma|^{1/2}}\int_0^{\infty}\int_0^{\infty} e^{-\frac{1}{2}(\mathbf{x}^T\Sigma^{-1}\mathbf{x}+(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}))^T\Sigma^{-1}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}))-(\Sigma\mathbf{e}-\mathbf{x})^T\Sigma^{-1}(\Sigma\mathbf{e}-\mathbf{x})}du_1du_2 \\
&= \frac{1}{2\pi|\Sigma|^{1/2}}e^{-\frac{1}{2}(\mathbf{x}^T\Sigma^{-1}\mathbf{x}-(\Sigma\mathbf{e}-\mathbf{x})^T\Sigma^{-1}(\Sigma\mathbf{e}-\mathbf{x}))}\int_0^{\infty}\int_0^{\infty} e^{-\frac{1}{2}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}))^T\Sigma^{-1}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}))}du_1du_2 \\
&= \frac{1}{2\pi|\Sigma|^{1/2}}e^{-\frac{1}{2}(-\mathbf{e}^T\Sigma\mathbf{e}-2\mathbf{e}\mathbf{x})}\int_0^{\infty}\int_0^{\infty} e^{-\frac{1}{2}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}))^T\Sigma^{-1}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}))}du_1du_2
\end{aligned}$$

Note that $\frac{1}{2\pi|\Sigma|^{1/2}}e^{-\frac{1}{2}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}))^T\Sigma^{-1}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}))}$ is the bivariate density of $N(\mathbf{x}-\Sigma\mathbf{e}, \Sigma)$ distribution. Since the previous section has shown

$$f_{\mathbf{Y}}(y_1,y_2) = f_{\mathbf{B}}(-\log(y_1), -\log(y_2))\frac{1}{y_1 \cdot y_2}, y_1, y_2 > 0,$$

we see that

$$f_{\mathbf{X}}(x_1,x_2) = e^{\frac{1}{2}\mathbf{e}^T\Sigma\mathbf{e}} \cdot Q(-\log(\mathbf{x}) - \Sigma\mathbf{e}),$$

where Q is the Cumulative Distribution Function of a normal distribution, i.e. $N(\mathbf{0}, \Sigma)$

Let us now consider the situation that $\mathbf{N}$ has a normal distribution $N(\mu, \Sigma)$, where $\mu = (\mu_1, \mu_2)^T$. Then we have

$$f_{-\log \mathbf{X}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{E}}(u_1,u_2) f_{\mathbf{N}}(x_1-u_1,x_2-u_2)du_1du_2$$

If $\mathbf{Z}$ has a $N(0,\Sigma)$ distribution then $\mu+\mathbf{Z}$ has a $N(\mu,\Sigma)$ distribution. So we can just replace $\mathbf{x}$ by $\mathbf{x}-\mu$ in the previous result. Therefore, we get

$$f_{\mathbf{D}} = f_{-\log \mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi|\Sigma|^{1/2}} e^{-\frac{1}{2}(-\mathbf{e}^T\Sigma\mathbf{e}-2\mathbf{e}(\mathbf{x}-\mu))} \int_0^{\infty}\int_0^{\infty} e^{-\frac{1}{2}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}+\mu))^T\Sigma^{-1}(\mathbf{u}+(\Sigma\mathbf{e}-\mathbf{x}+\mu))}du_1du_2$$

The integration part of the equation has a normal $N(\mathbf{x}-\Sigma\mathbf{e}-\mu,\Sigma)$ distribution, thus we can get the true density of this example.

4.3.3 Calculation

Now, we are going to implement this by Matlab7.1. we are going to give a way to calculate the integration faster.

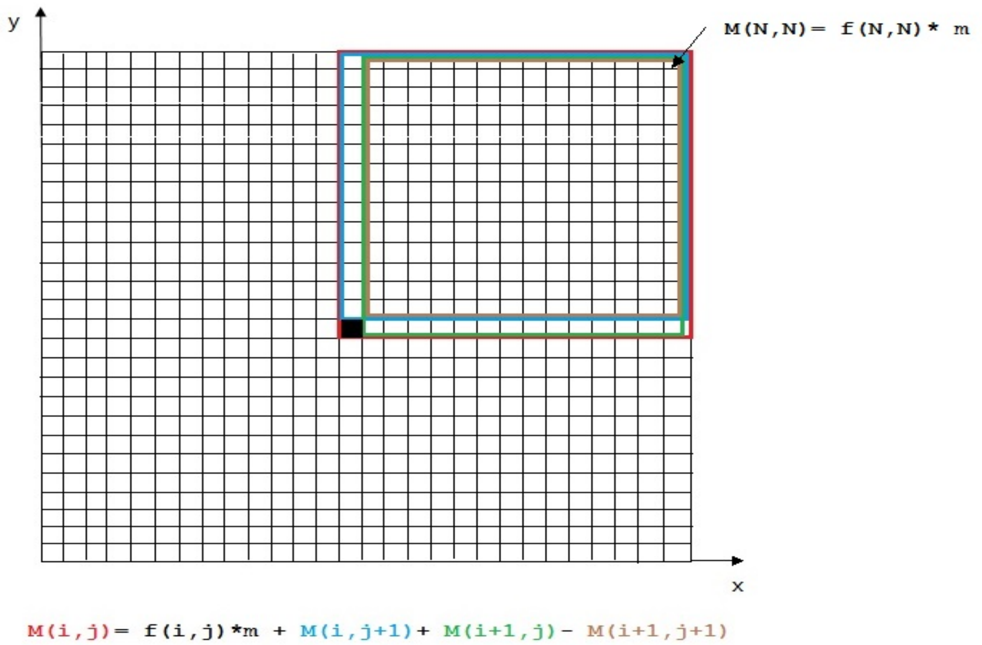

Figure 4.1: Calculation of the integration

We divide x,y into N, then we have $N\times N$ squares. The area $m=\frac{x}{N}\times\frac{y}{N}$.
$f(i,j)$ is the average of the four points of the square which are $f_B^{(n)}(-\log x_1,-\log x_2)$

calculated by the method we introduced in previous Chapter. As it is showed in the picture, $M(i,j)$ is the intgration of the whole area of the red square.Figure4.1 gives the way of calculating $M(i,j)$. which is exactly what we need.

Compared to the way that calculating the integration point by point, our way is much more efficient.

Chapter 5

Simulation

In this chapter we will apply the described algorithms of nonparametric deconvolution to several two-dimensional problems. The implementation is carried out in the software Matlab7.1 and all corresponding parts of the source code can be found in the appendix.

Let us start with a standard bivariate normal distribution, i.e. with mean equal to (0,0), no correlation and variances equal to 1. We will make 1000 simulations from this distribution in convolution with a two-dimensional exponential distribution. Each coordinate random variable of the two-dimensional exponential distribution has the mean value parameter equal to 0.5 and they are mutually independent. The setting corresponds to the one in Theorem 3.6. The simulated random sample is showed in the scatter plot in Figure 6.1. The density function of the original bivariate distribution is plotted in Figure 6.2. We can see the original distribution has been shifted by the convolution.

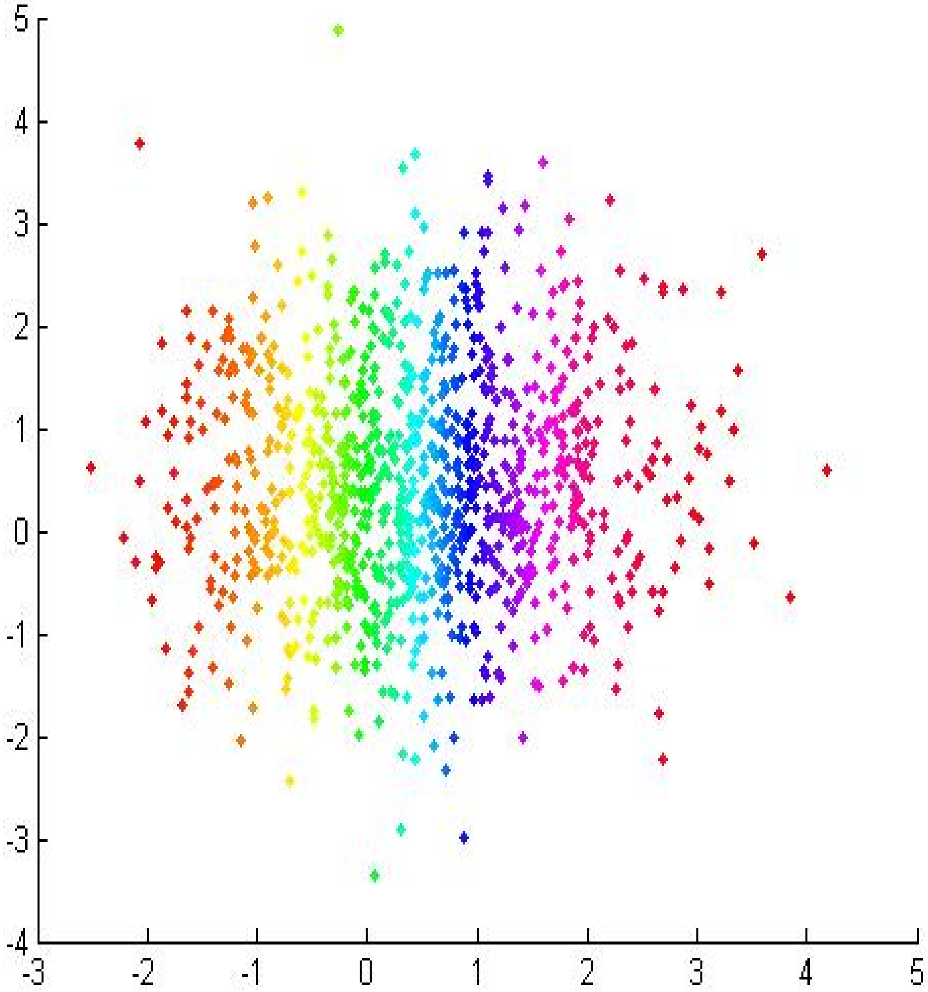

Figure 6.1: Convolution of bivariate normal ((0,0),((0,1),(1,0))) and two dimensional exponential (0.5) distribution, 1000 simulations.

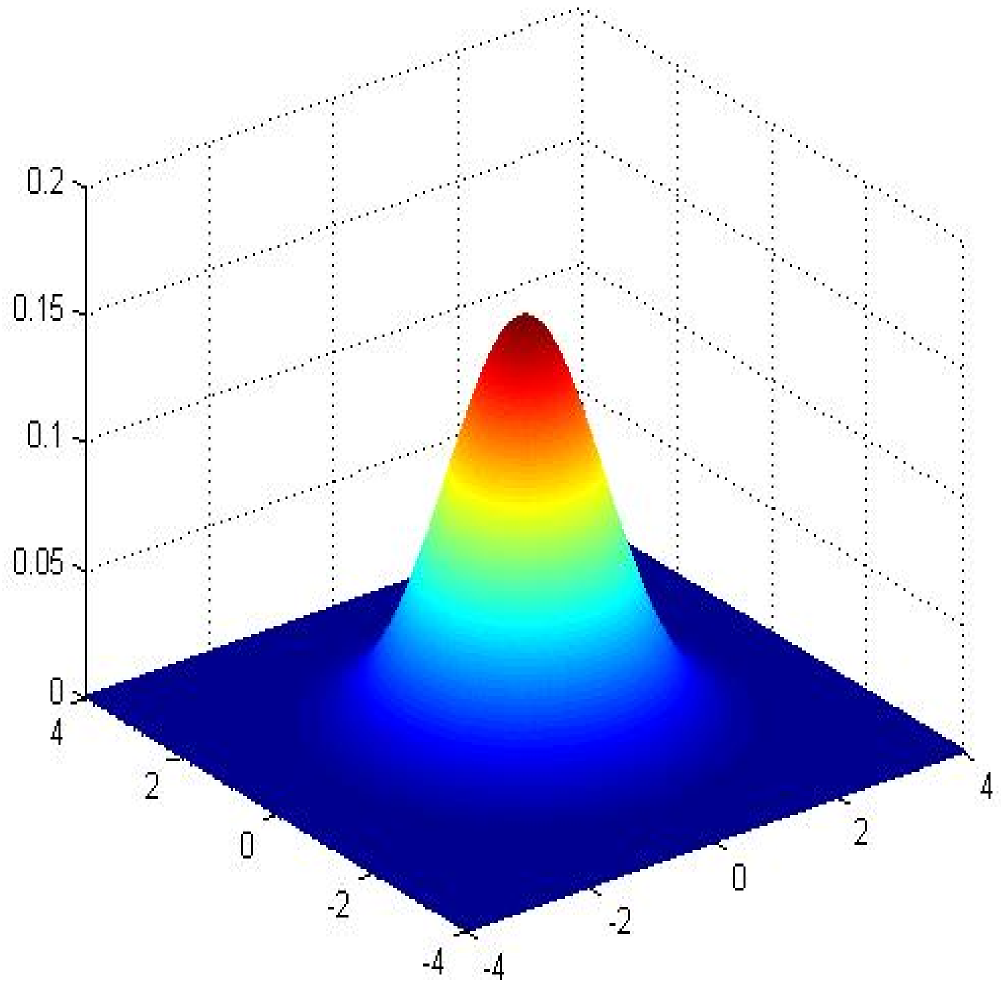

Figure 6.2: Density function of bivariate normal ((0,0),((0,1),(1,0))) 1000 simulations.

In Chapter 3, we have introduced the deconvolution kernel estimator, we are going to give an example of this.

We will apply the gamma deconvolution theorem and obtain a nonparametric kernel density estimator $f_{2.5}^{(500)}$. We used the product of the univariate kernels with bandwidth 2.5. The resulting estimate can be seen in Figure 6.3. The absolute error is then pictured in Figure 6.4. We can see that the estimation is quite close to the original two-dimensional density. Different kernels have different results. Some are good and soe are not good enough. The triweight kernel is the best,even though the absolute error reaches 0.06 around the high-densed center area, which might suggest we oversmoothed the estimate. However, in this case taking a lower bandwidth results in an undersmoothed estimate since the number of observations is not adequate.

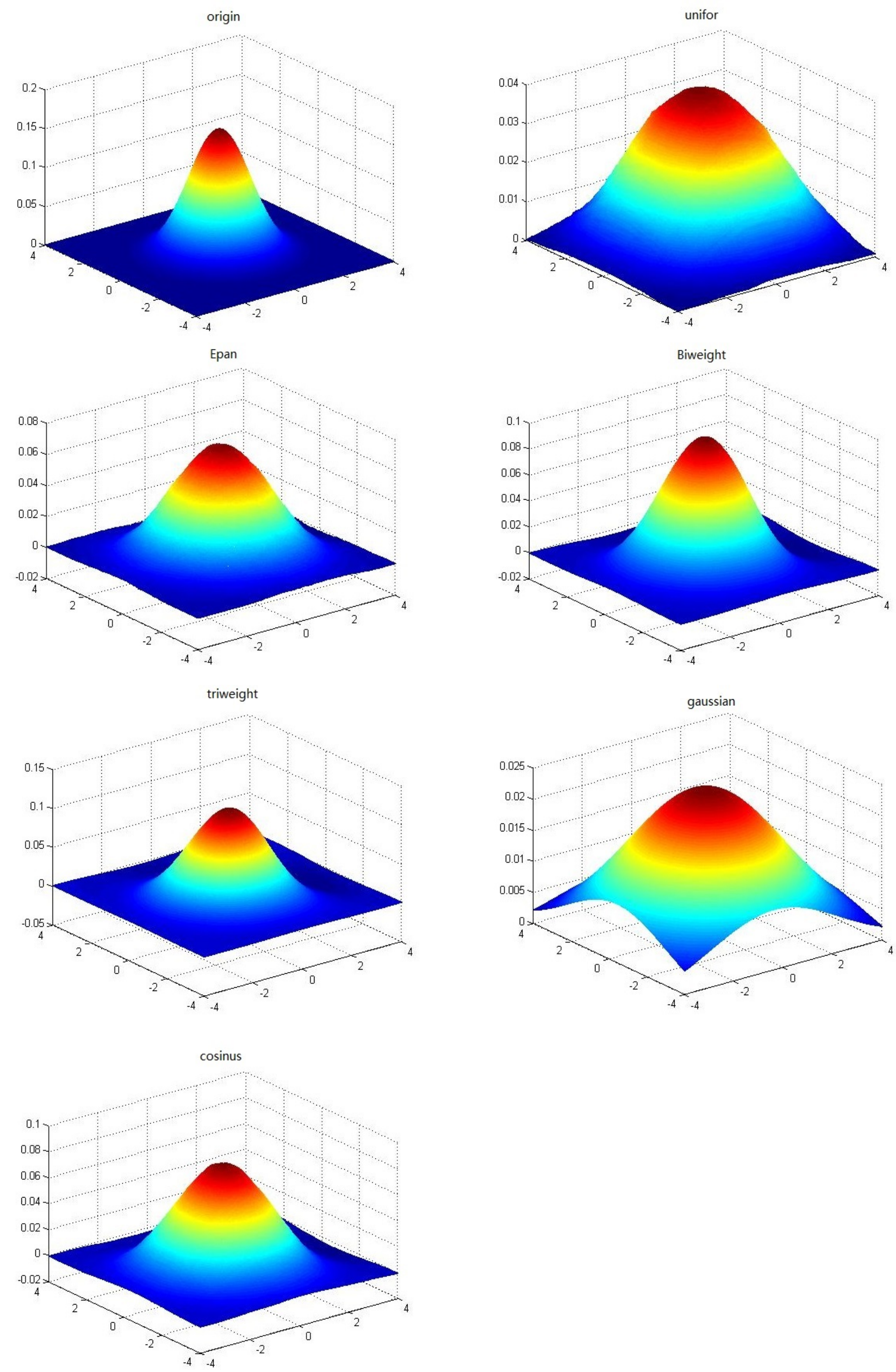

Figure 6.3: Deconvolution of bivariate normal ((0,0),((0,1),(1,0))) and two dimensional exponential (0.5) distribution with bandwidth 2.5, 500 simulations.

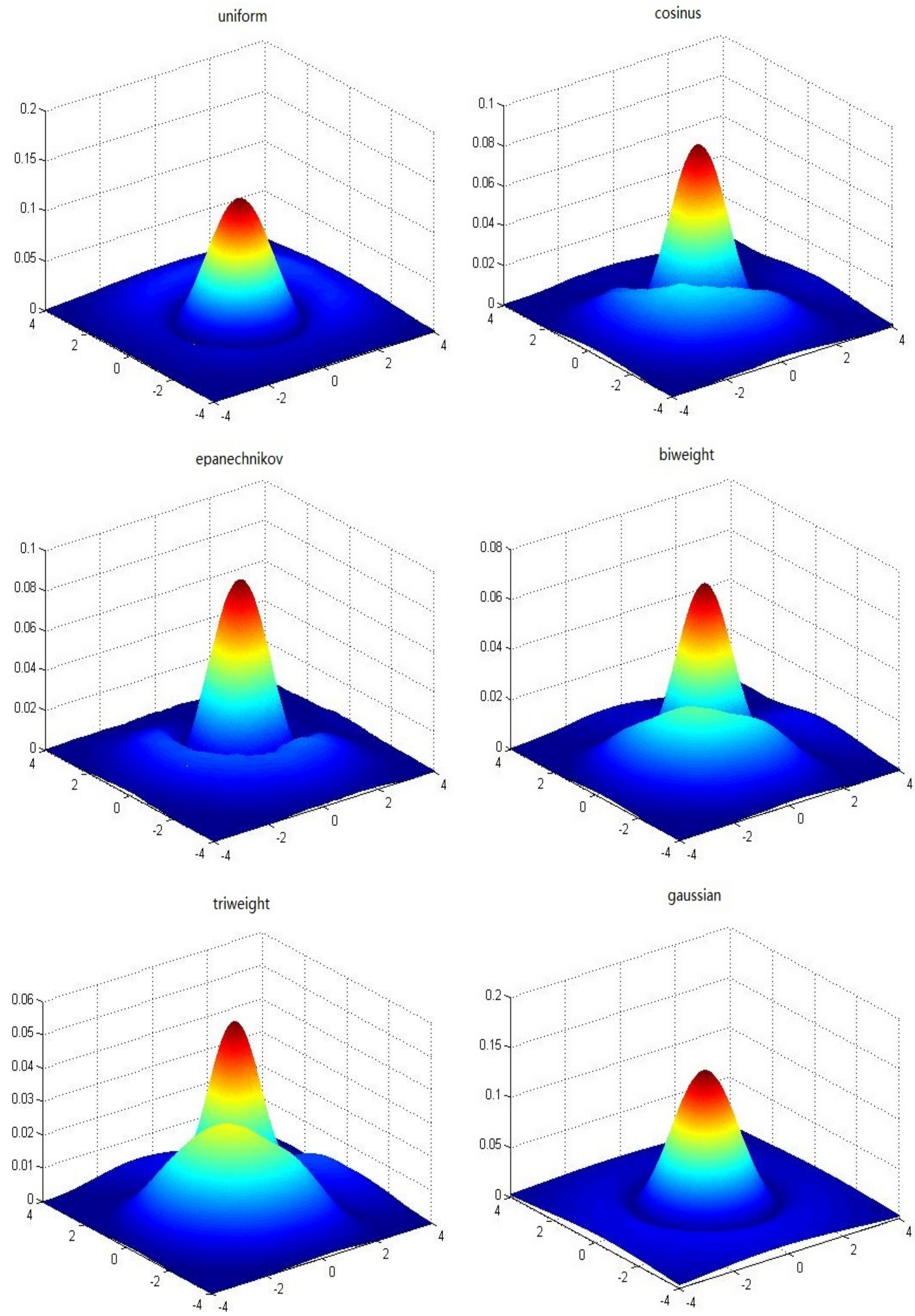

Figure 6.4: Absolute error of deconvolution of bivariate normal((0,0),((0,1),(1,0))) and two-dimensional exponential(0.5) 500 simulations,bandwidth 2.5.

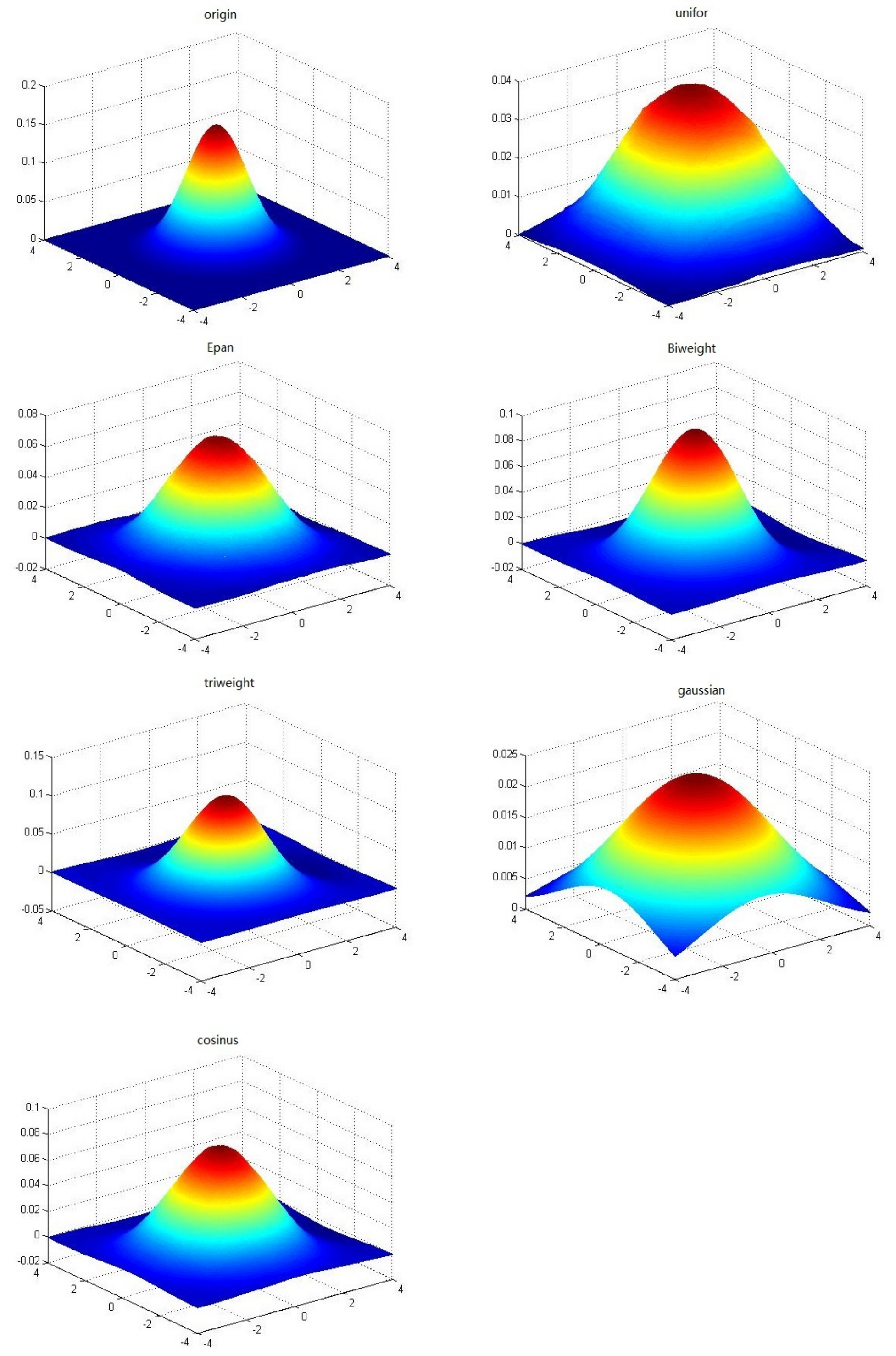

Figure 6.5: Deconvolution of bivariate normal ((0,0),((0,1),(1,0))) and two dimensional exponential (0.5) distribution with bandwidth 2.5, 2500 simulations.

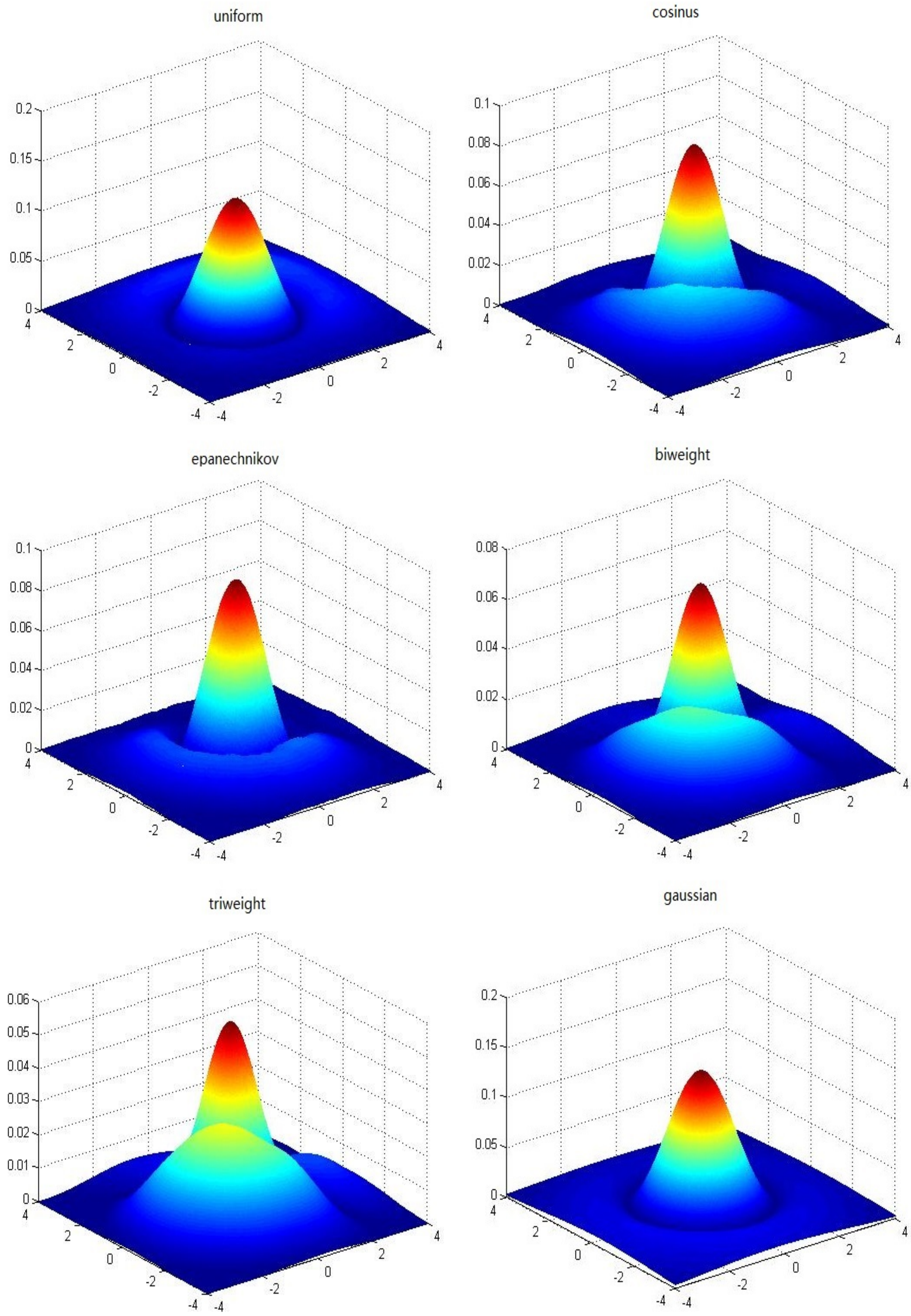

Figure 6.6: Absolute error of deconvolution of bivariate normal((0,0),((0,1),(1,0))) and two-dimensional exponential(0.5) 2500 simulations,bandwidth 2.5..

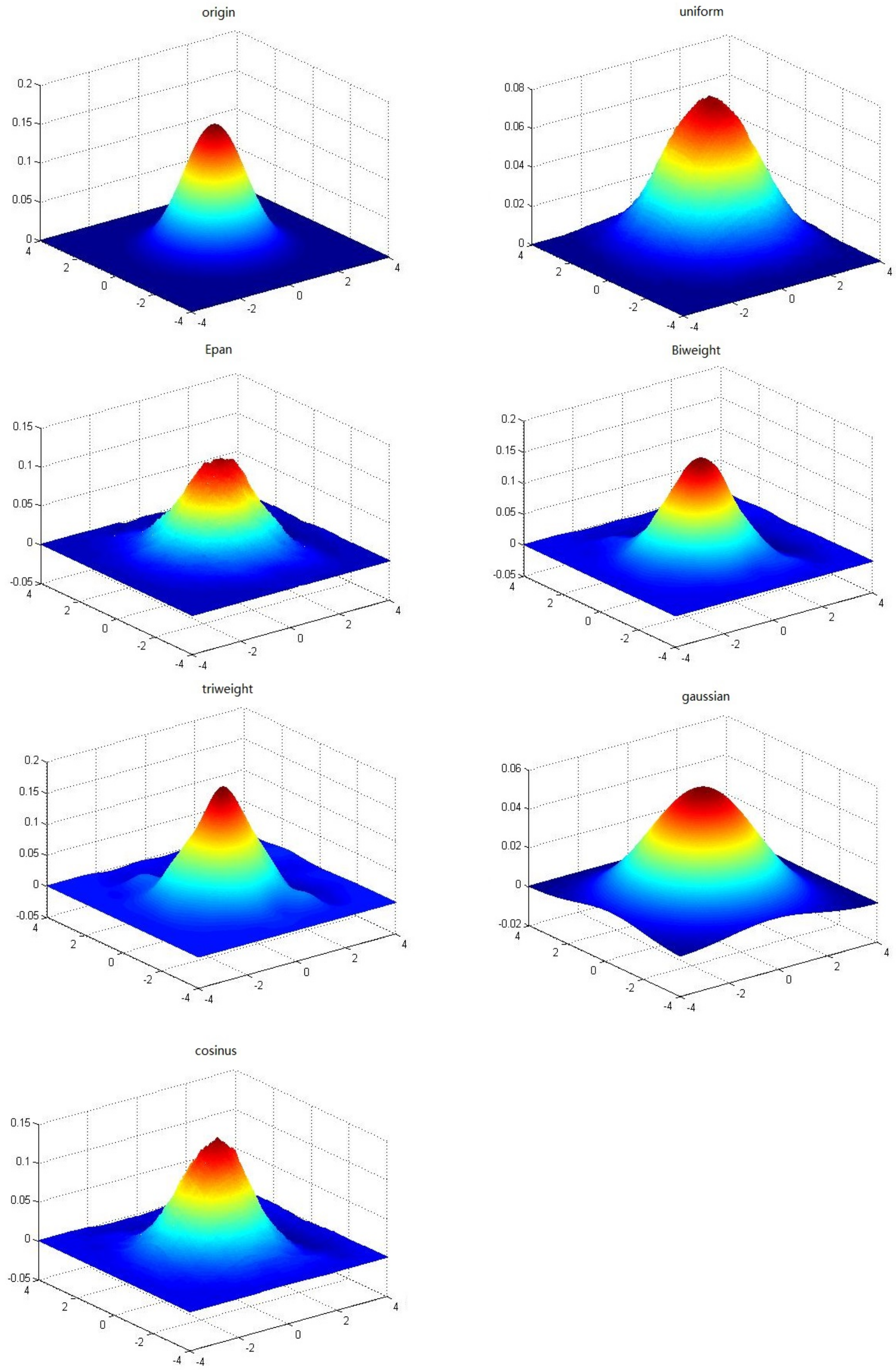

Figure 6.7: Deconvolution of bivariate normal ((0,0),((0,1),(1,0))) and two dimensional exponential (0.5) distribution with bandwidth 1.5, 500 simulations.

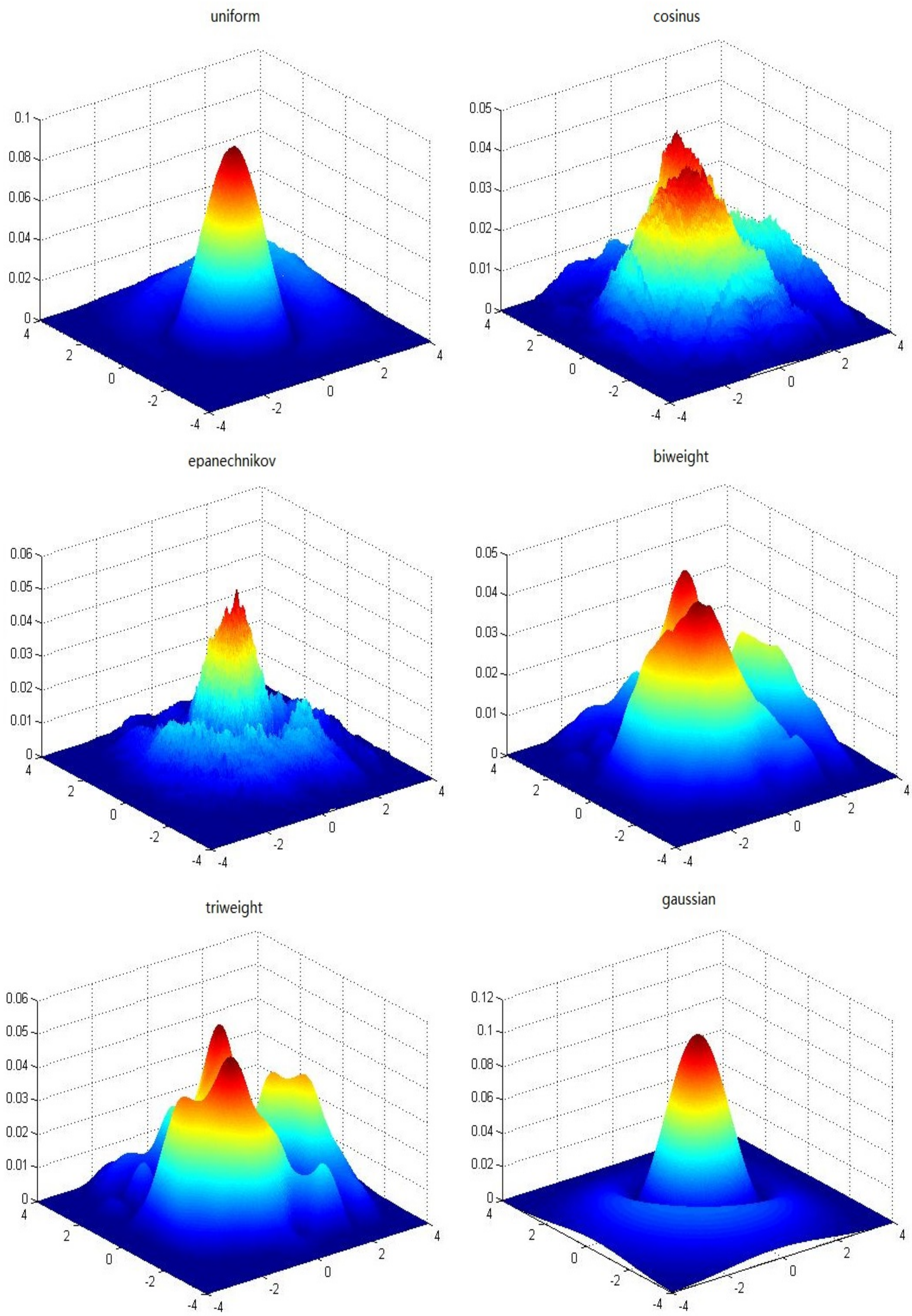

Figure 6.8: Absolute error of deconvolution of bivariate normal((0,0),((0,1),(1,0))) and two-dimensional exponential(0.5) 500 simulations,bandwidth 1.5.

To be able to improve our estimate in the area around $(0, 0)$ with high original density, we will now consider a new sample of a size 2500. All the distributions and the described setting remains the same. As we can see in Figure 6.5, the deconvolution density estimator $f_{2.5}^{(2500)}$, using the product of the univariate kernels with bandwidth 2 did not significantly change. The nature of the original density is captured by the estimator, free of any shifts previously caused by the exponential noise. In Figure 6.6 we also see that the absolute error is again highest around the central area because of oversmoothing. If we now use the deconvolution density estimator $f_{1.5}^{(500)}$ with the product of the univariate of kernels with bandwidth equal to 1.5, the estimate remains sufficiently smooth and represents a more solid indicator of the central density, as we can see in Figure 6.7. If we look at Figure 6.8, we can see that the maximum of the absolute error is approximately reduced differently from different kernels. The estimating procudure in the case of the bigger sample size already has higher demands on the program. On an average laptop the calculation runs for 3-5 minutes.

We will then show the application case in Chapter 4 in both one and two dimension.
In one dimensional case, we will first implement the standard normal distribution case referred to section 4.3.1. Figure 6.9 shows the result of estimation ,and Figure 6.10 shows the absolute error of this method. It is easy to see that this method does not convergence at the point (0,0), and the result is also different from different kernels. From the picture, it is obvious to see, that the uniform kernel is the worst, and the triweight kernel is the best.

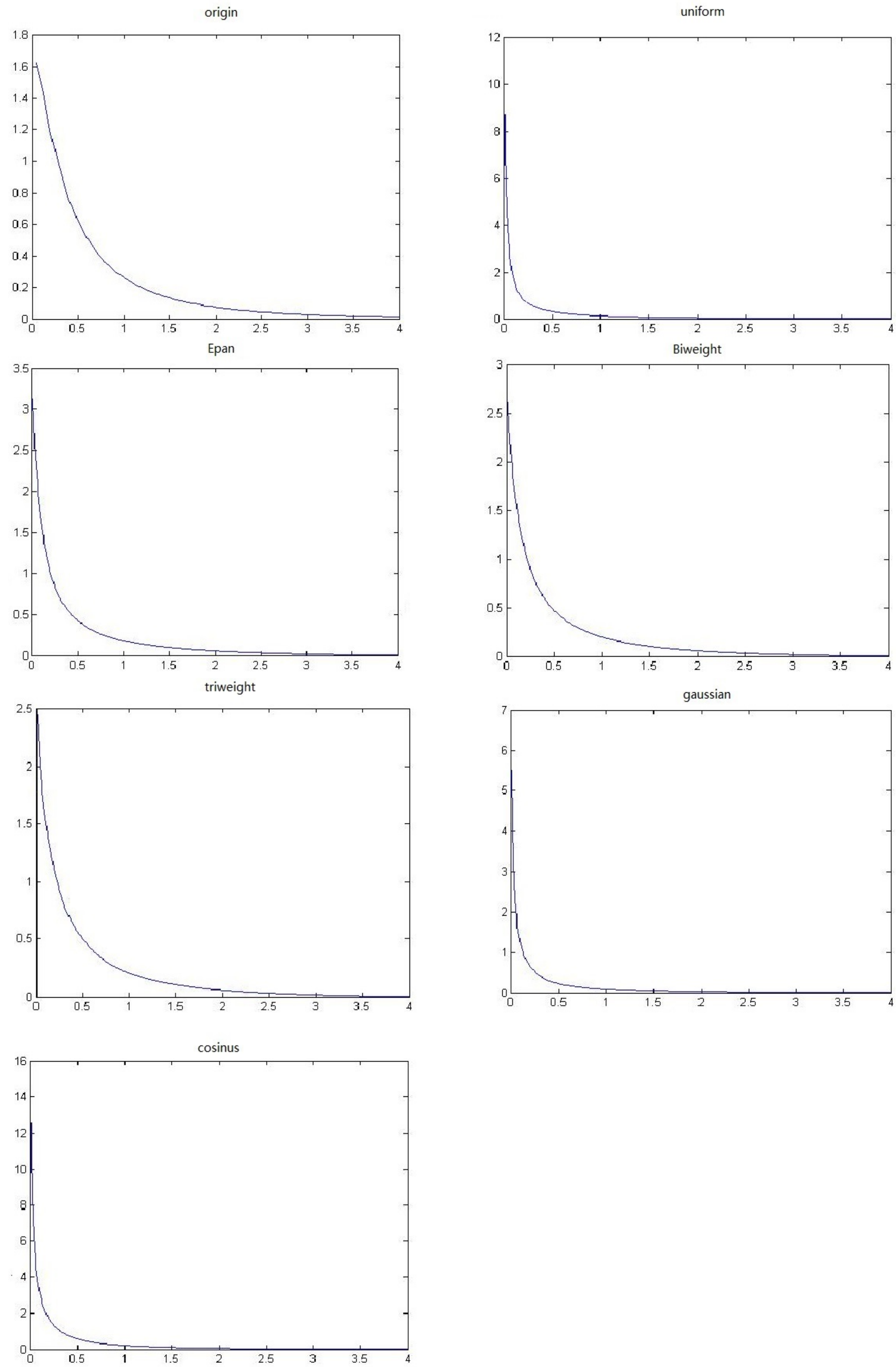

Figure 6.9: Application in one dimensional with standard normal distribution and exponential(1) 500 simulations,bandwidth 2.5.

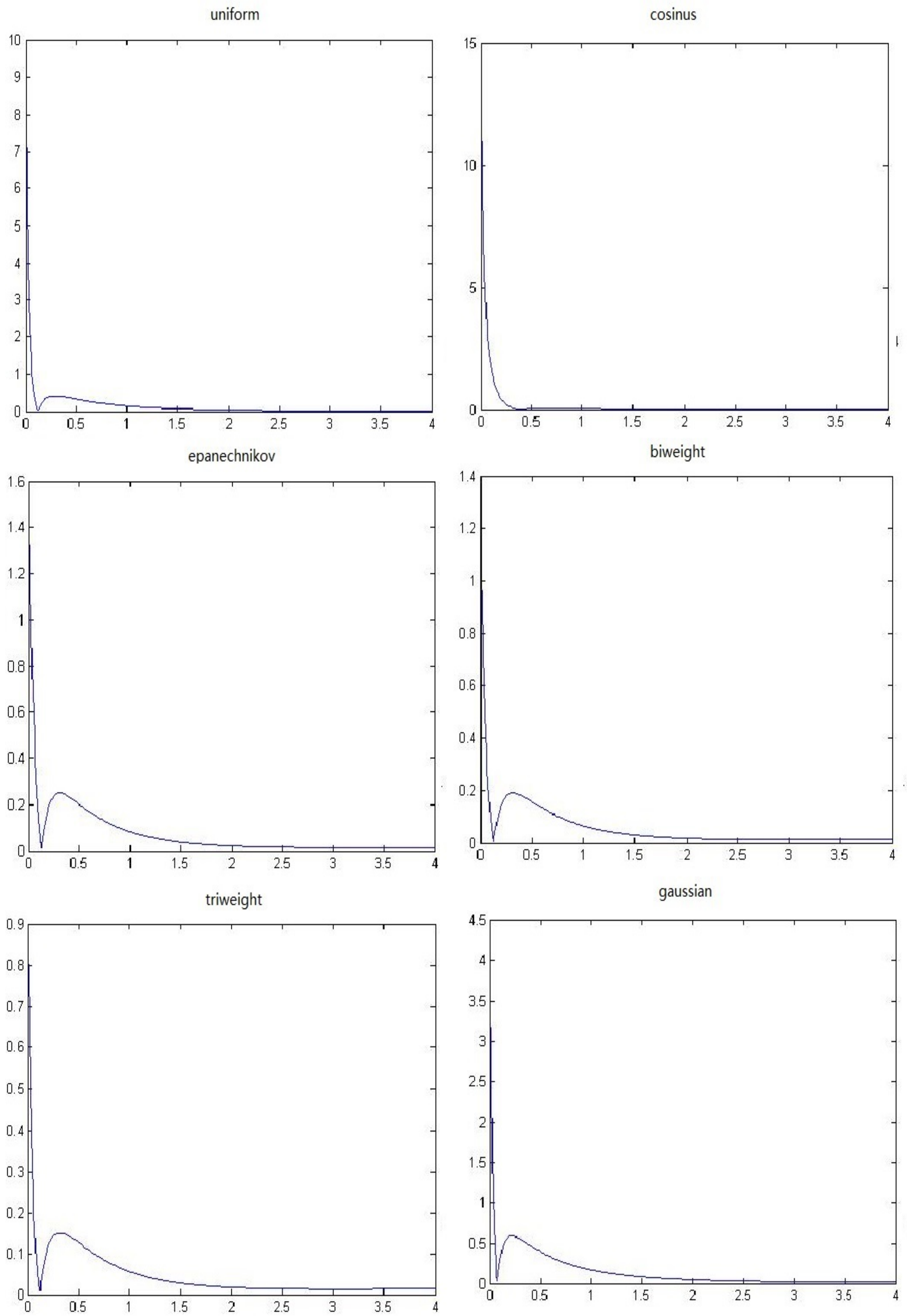

Figure 6.10: Absolute error of the Application in one dimensional with standard normal distribution and exponential(1) 500 simulations,bandwidth 2.5.

In two dimensional case, we will first implement the standard normal distribution case referred to Section 4.3.2. Figure 6.11 shows the result of estimation ,and Figure 6.12 shows the absolute error of this method. It is easy to see that this method is not convergence at the point (0,0), and the result is also different from different kernels.

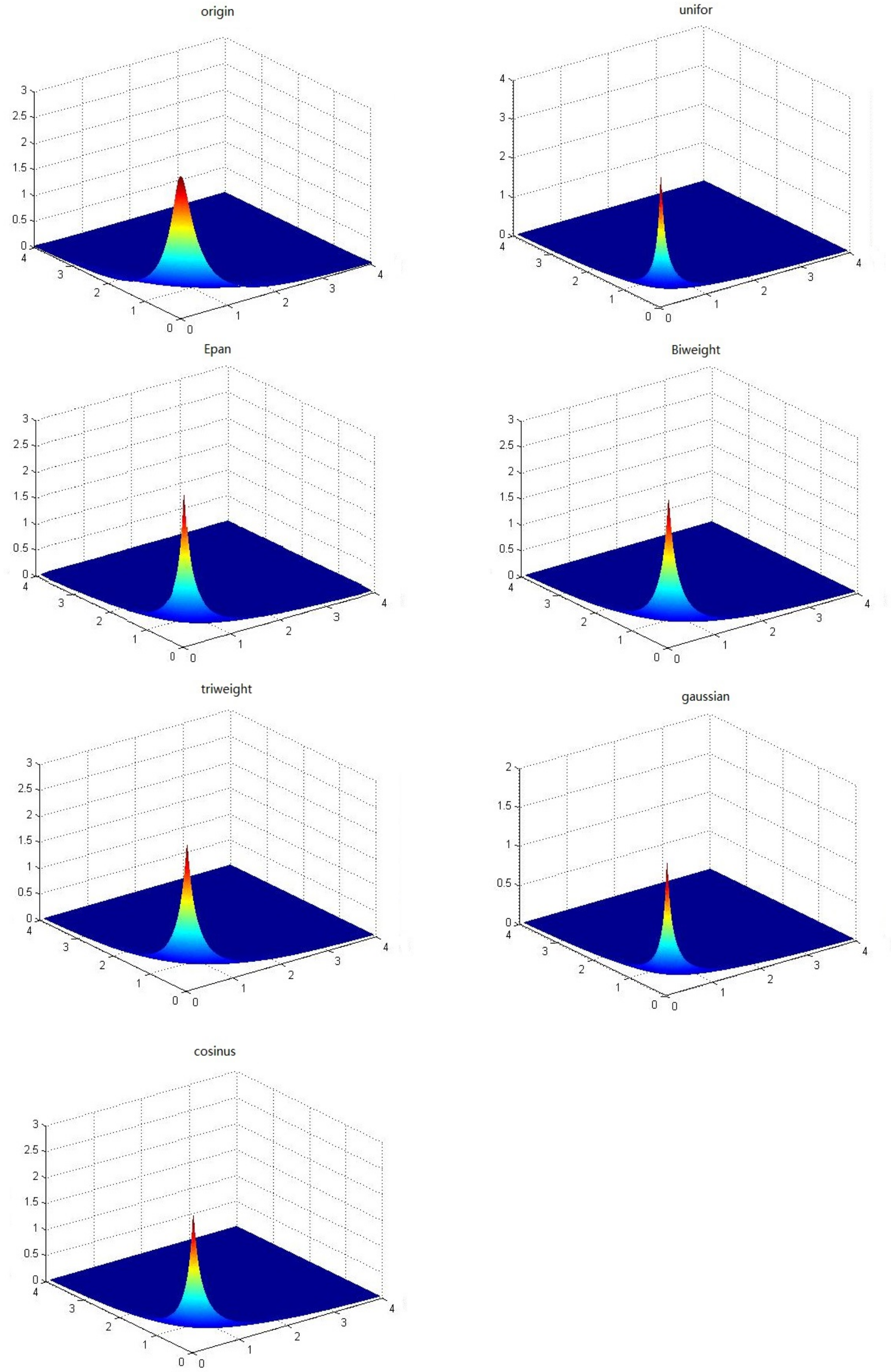

Figure 6.11: Application in two-dimensional with standard normal distribution and exponential(1) 500 simulations,bandwidth 2.5.

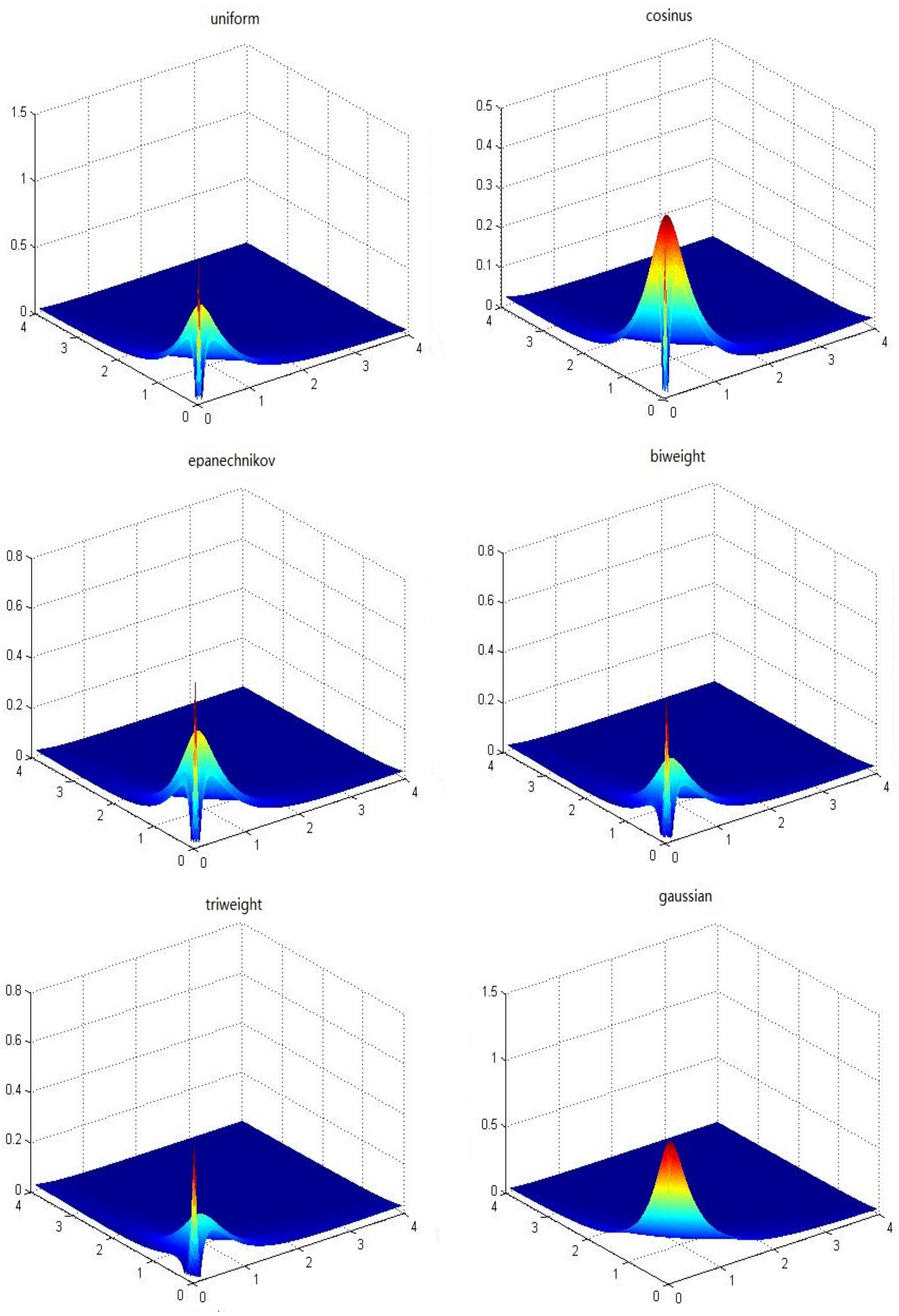

Figure 6.12: Absolute error of the Application in two-dimensional with standard normal distribution and exponential(1) 500 simulations,bandwidth 2.5.

We can see from the picture that Gaussian kernel and Uniform kernel are the worst. Since the limit of our method that X can not be zero, because we need to use -log X. If X is very small, then $-\log X$ will come to the infinity.In our case we will ask X to be calculated from 0.1. Let begin to calculate with X=0.00001, we will have the following picture: Figure 6.13.

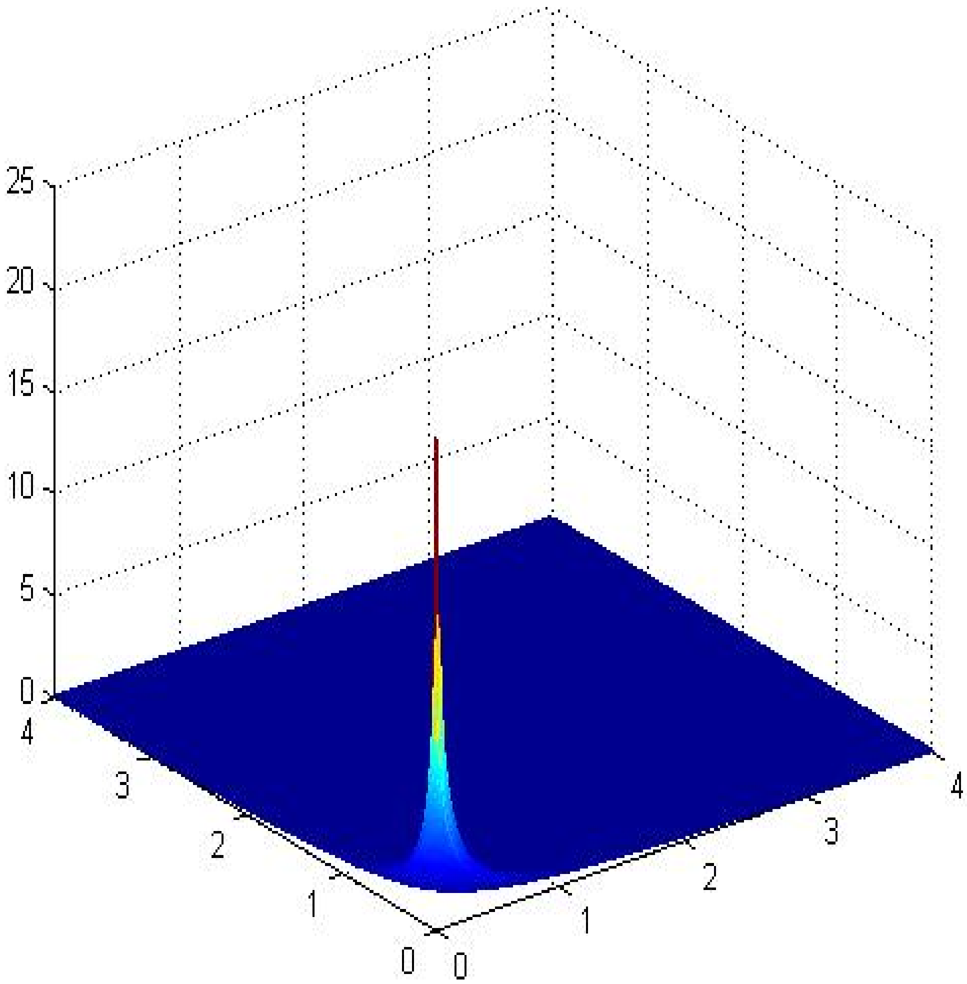

Figure 6.13: Application using Epan kernel in two-dimensional with standard normal distribution and exponential(1) 500 simulations,bandwidth 2.5, X from 0.00001.

From the picture we can see that this method is not convergence when $x \to 0$.

Picture 6.14 shows the result of Isotonic method. Compared to picture 6.12 we can see that this method has better result around the point (0,0), but it can only be applied to one dimension estimation.

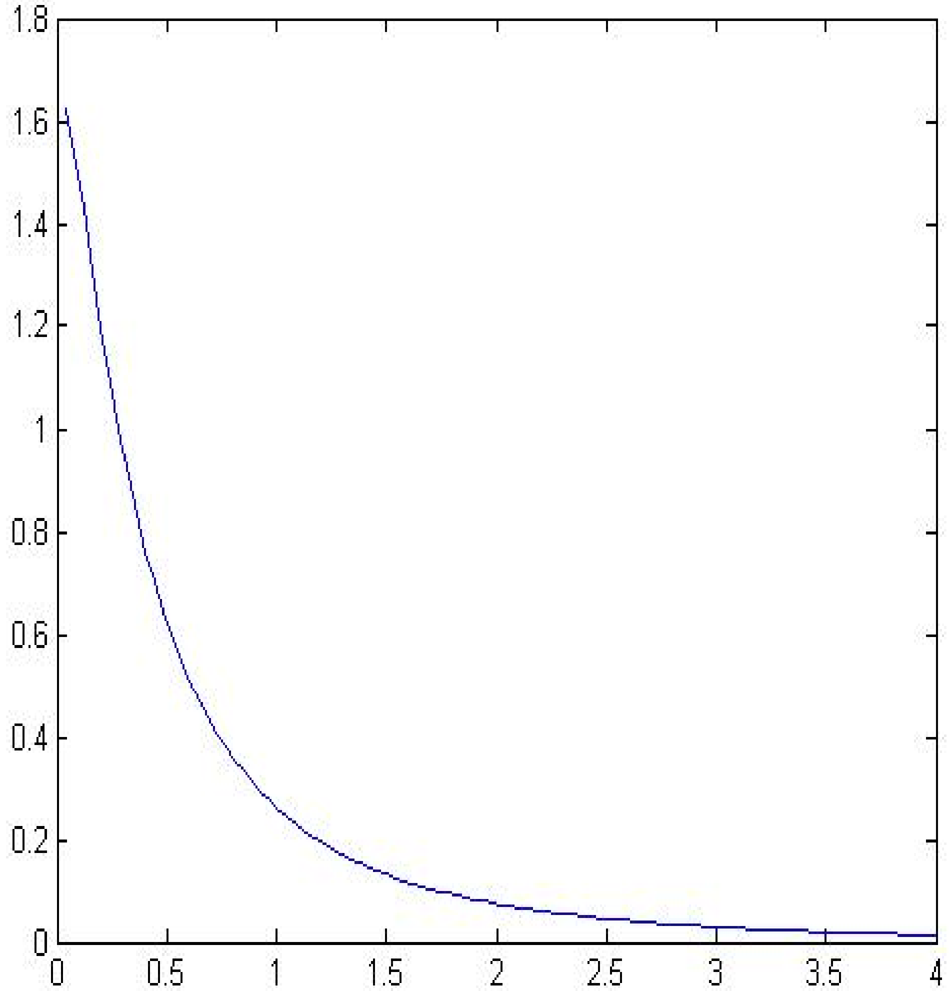

Figure 6.14: Application using method of Isotonic estimators of decreasing densities with normal(0,1)and exponential(0.5) .

Picture 6.15 shows the result of the method in Appendix B. Compared to picture 6.3, this method is much faster.But the result is not as accurate as the result of picture 6.3.And we can see from the picture that the result is not smoothing.

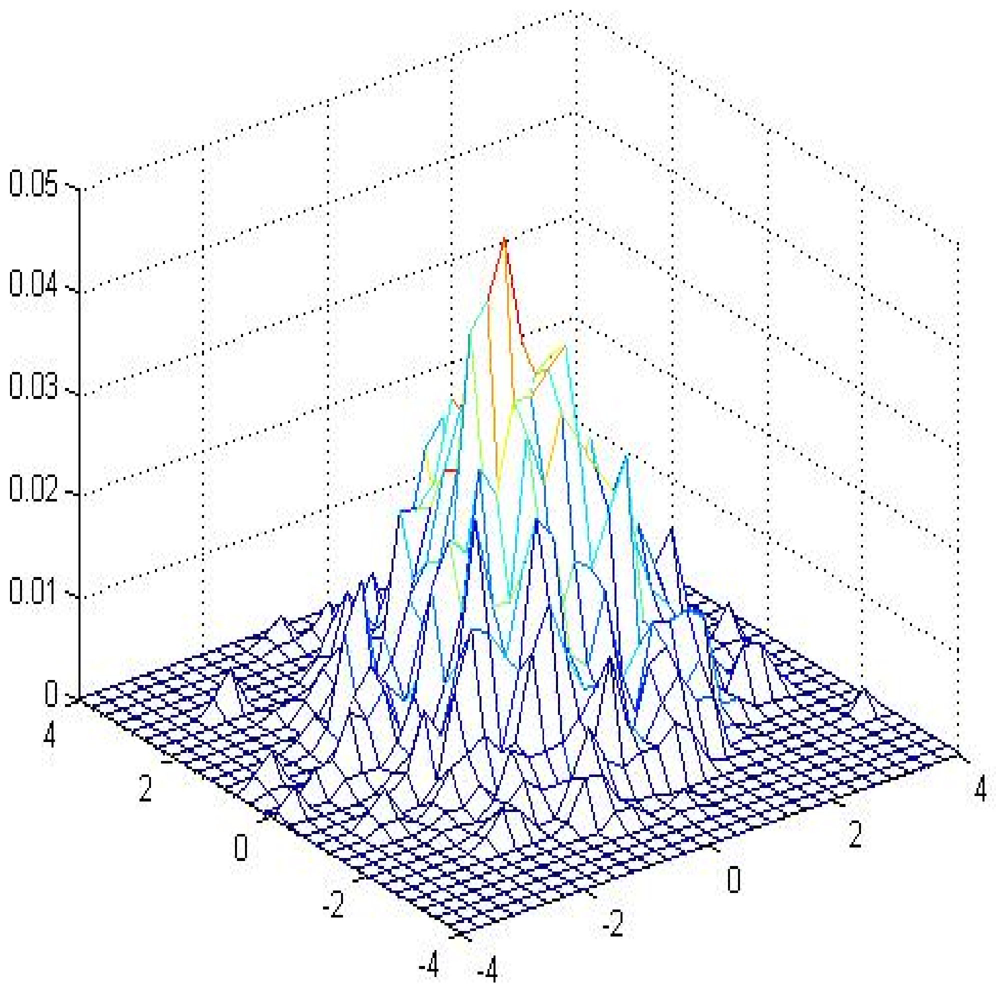

Figure 6.15: Application using method of Fast computation of multivariate kernel density estimator .

Chapter 6

Summary

In this work we represent elaborated statistical methods for solving the two-dimensional deconvolution decreasing density estimation problem in a nonparametric way. The core of the thesis is based on an existing solution of this problem in a one-dimensional setting and expands the theory to the second dimension. We construct the required theory with original proofs, deriving inversion formulas for exponential deconvolution and its generalizations. We present a practical overview of nonparametric estimation methods, that are later built iton deconvolution algorithms and implemented in computational software. The processed methods are accompanied with corresponding illustrations, allowing an immediate insight in the covered theory.

A significant focus is aimed at deriving statistical properties of the derived estimators. A particularly interesting application of the described approach is the problem of estimating decreasing densities. As we were able to derive an explicit expression for the transformed nonparametric estimator, this result can enjoy a wide range of applications.

All developed methods were implemented in Matlab and tested on simulated data. During the software implementation, the estimators described in this thesis met our expectations and produced solid and accurate density estimates.The results were obtained within a reasonable time frame while using samples large enough to produce smooth estimates.

There are two significant contributions that this work has brought. We use the extension of the nonparametric deconvolution, based on an inversion, to a higher dimension of Krycha(2001).We derive an elaborated application for estimating decreasing densities in a nonparametric way. And secondly, the actual implementation of the described methods, establishing the practical relevance of the assembled theory.

Appendix A

Appendix Proofs

A.1 Proofs of Chapter 3

A.1.1 Proof of Theorem 3.4

Before the proof, we will introduce a definition and a lemma first.

Definition A1.1(Smooth function): *Let :Let $U \subset \mathbb{R}^d$ be an open set and f be a real valued function on U. We say that f is of class $\mathbb{C}^k (k \in \mathbb{N})$if for all $1 \leq k$, $l_1, ..., l_d \geq 0$,$l_1 + ... + l_d = l$ the partial derivative*

$$\frac{\delta^l}{\delta x_1^{l_1} ... \delta x_d^{l_d}} f(x_1, ..., x_d)$$

exists and is continuous on U. We define $\mathbb{C}^0$ as a set of all continuous functions on U. We say that f is of class $\mathbb{C}^\infty$ if f is of class $\mathbb{C}^k$ for all $k \geq 0$.

Lemma A1.2: *Let :Let f and h be a densities of some random vectors, denote * as the presentation of convolution. If f is of class k($k \in \mathbb{N} \cup \{\infty\}$)if for all $1 \leq k$, then $g = f * h$ is at least of class k . Moreover, if h is bounded and f has integrable partial derivatives up to order k, then*

$$\lim_{x_i \to \pm\infty} h(x_1, x_2, ..., x_n) = 0 \Rightarrow \lim_{x_i \to \pm\infty} \frac{\delta^l}{\delta x_1^{l_1} ... \delta x_d^{l_d}} g(x_1, ..., x_d) = 0$$

for all $0 \leq l \leq k, l_1, ..., l_d \geq 0, l_1 + ... l_d = l$ This proof is referred to Krycha(2011)

Proof of theorem 3.4. We will prove this theorem by induction. First, Let

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \lambda E_{11} \\ \tau E_{21} \end{pmatrix}$$

we compute the characteristic function of the random factor X.Then we have:

$$\begin{aligned}
\phi_X(t_1, t_2) &= \mathbb{E}e^{i(t_1,t_2)(X_1,X_2)} \\
&= \mathbb{E}e^{i(t_1,t_2)(X_1,X_2)} \\
&= \mathbb{E}e^{i(t_1Y_1,t_2Y_2)}\mathbb{E}e^{it_1Z_1}\mathbb{E}e^{it_2Z_2} \\
&= \phi_Y(t_1, t_2)\phi_{Z1}(t_1)\phi_{Z2}(t_2)
\end{aligned}$$

For exponentially distributed random variables we have

$$\begin{aligned}
\phi_{Z1}(t_1) &= \mathbb{E}e^{it_1Z_1} \\
&= \int_0^\infty \frac{1}{\lambda} e^{it_1z} e^{-\frac{z}{\lambda}} dz \\
&= \int_0^\infty \frac{1}{\lambda} e^{z(it_1 - \frac{1}{\lambda})} dz \\
&= \frac{1}{\lambda} \frac{1}{it_1 - \frac{1}{\lambda}} [e^{z(it_1 - \frac{1}{\lambda})}]_0^\infty \\
&= \frac{1}{1 - it_1\lambda}
\end{aligned}$$

Since for $t_1 \in \mathbb{R}$ we have $|e^{z(it_1-1)}| = e^{-z}$ and $\lim_{z\to\infty} e^{it_1 - \frac{1}{\lambda}} = 0$,then, we have

$$\phi_X(t_1, t_2) = \phi_Y(t_1, t_2) \frac{1}{1 - it_1\lambda} \frac{1}{1 - it_2}$$

and

$$\phi_Y(t_1, t_2) = \phi_X(t_1, t_2) - it_1\lambda\phi_X(t_1, t_2) - it_2\tau\phi_X(t_1, t_2)$$

The density of Y is of class $\mathbb{C}^{m+n}$, which, using Lemma3.8 and the fact that the partial derivatives are integrable, ensures the same property for g. Now both $g(x_1, x_2)$ and $e^i(t_1x_1, t_2x_2)$ are continuously differentiable as functions of x_1,x_2, so we can use integration by parts and get

$$\begin{aligned}
&\int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} e^{i(t_1x_1+t_2x_2)}\frac{\partial}{\partial x_1}g(x_1,x_2)dx_1)dx_2 \\
&= \int_{-\infty}^{\infty}([e^{i(t_1x_1+t_2x_2)}g(x_1,x_2)]_{x_1\to\infty}^{x_1\to\infty} \\
&- \int_{-\infty}^{\infty} it_1 e^{i(t_1x_1+t_2x_2)}g(x_1,x_2)dx_1)dx_2
\end{aligned}$$

Notice the density of Z_1,i.e.

$$h(x_1,x_2) = -\frac{1}{\lambda\tau}e^{-\frac{x_1}{\lambda}-\frac{x_2}{\tau}}\mathbb{I}_{[0,\infty)}(x_1)\mathbb{I}_{[0,\infty)}(x_2)$$

is a bounded function and that

$$\lim_{x_1\to\pm\infty} h(x_1,x_2) = 0$$

Hence by A1.2, we have

$$\lim_{x_1\to\pm\infty} g(x_1,x_2) = 0$$

Now, since $|e^{i(t_1x_1+t_2x_2)}| \leqslant 1$, we see that

$$[e^{i(t_1x_1+t_2x_2)}g(x_1,x_2)]_{x_1\to-\infty}^{x_1\to+\infty} = \lim_{x_1\to+\infty} e^{i(t_1x_1+t_2x_2)}g(x_1,x_2) - \lim_{x_1\to-\infty} e^{i(t_1x_1+t_2x_2)}g(x_1,x_2) = 0$$

This results in the equalities:

$$\begin{aligned}
-it_1\phi_X(t_1,t_2) &= \int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} e^{i(t_1x_1+t_2x_2)}\frac{\partial}{\partial x_1}g(x_1,x_2)dx_1)dx_2 \\
-it_2\phi_X(t_1,t_2) &= \int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} e^{i(t_1x_1+t_2x_2)}\frac{\partial}{\partial x_2}g(x_1,x_2)dx_1)dx_2
\end{aligned}$$

With the same argument, and using the Lemma A1.2, we can show that :

$$\begin{aligned}
&\int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} e^{i(t_1x_1+t_2x_2)}\frac{\partial^2}{\partial x_1\partial x_2}g(x_1,x_2)dx_1)dx_2\\
&=\int_{-\infty}^{\infty}([\frac{\partial}{\partial x_2}g(x_1,x_2)dx_2)]_{x_1\to-\infty}^{x_1\to+\infty}-\int_{-\infty}^{\infty} it_1e^{i(t_1x_1+t_2x_2)}\frac{\partial}{\partial x_2}g(x_1,x_2)dx_1)dx_2\\
&=0-it_1\int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} e^{i(t_1x_1+t_2x_2)}\frac{\partial}{\partial x_2}g(x_1,x_2)dx_1)dx_2\\
&=-t_1t_2\phi_X(t_1,t_2)
\end{aligned}$$

Note that since g is of class $\mathbb{C}^{m+n}\supset\mathbb{C}^2$, the second-order mixed derivative are interchangeable. Substituting into

$$\phi_Y(t_1,t_2)=\phi_X(t_1,t_2)-it_1\lambda\phi_X(t_1,t_2)-it_2\tau\phi_X(t_1,t_2)$$

then we get

$$\begin{aligned}
\int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} e^{i(t_1x_1+t_2x_2)}f(x_1,x_2)dx_1)dx_2 &= \int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} e^{i(t_1x_1+t_2x_2)}[g(x_1,x_2)\\
&+\lambda\frac{\partial}{\partial x_1}g(x_1,x_2)+\tau\frac{\partial}{\partial x_2}g(x_1,x_2)+\lambda\tau\frac{\partial^2}{\partial x_1\partial x_2}g(x_1,x_2)]dx_1)dx_2
\end{aligned}$$

Since f as a convolution of densities is absolutely integrable and in this case even continuous, we can use the inverse Fourier transform to yield the result(Stein and Weiss,1971):

$$f(x_1,x_2)=g(x_1,x_2)+\lambda\frac{\partial}{\partial x_1}g(x_1,x_2)+\tau\frac{\partial}{\partial x_2}g(x_1,x_2)+\lambda\tau\frac{\partial^2}{\partial x_1\partial x_2}g(x_1,x_2)$$

Proving the assertion for $m=1$.
Let us now assume that the claim holds for $m\geq 1$ and denote

$$\begin{aligned}
\tilde{X}&=\begin{pmatrix}Y_1\\Y_2\end{pmatrix}+\begin{pmatrix}\lambda_1E_{11}+...+\lambda_mE_{1m}\\ \tau_1E_{21}+...+\tau_mE_{2m}\end{pmatrix}\\
X&=\begin{pmatrix}Y_1\\Y_2\end{pmatrix}+\begin{pmatrix}\lambda_1E_{11}+...+\lambda_{m+1}E_{1m+1}\\ \tau_1E_{21}+...+\tau_m+1E_{2m+1}\end{pmatrix}=\tilde{X}+\begin{pmatrix}\lambda_{m+1}E_{1m+1}\\ \tau_{m+1}E_{2m+1}\end{pmatrix}\\
s_m^{(i)}&=s_m^{(i)}(\lambda_1,...,\lambda_m),\\
t_m^{(j)}&=s_m^{(j)}(\tau_1,...,\tau_m)
\end{aligned}$$

f being the density of Y, g the density of X and $\tilde{g}$ the density of $\tilde{X}$. The induction hypothesis gives us that

$$f(x_1,x_2)=\sum_{j=0}^{m}\sum_{i=0}^{m}s_m^{(i)}t_m^{(j)}\frac{\partial^{i+j}}{\partial x_1^i\partial x_2^j}\tilde{g}(x_1,x_2). \tag{A.1}$$

From the proved case for $m=1$ we get expression for

$$\begin{aligned}\tilde{g}(x_1,x_2)&=g(x_1,x_2)+\lambda_{m+1}\frac{\partial}{\partial x_1}g(x_1,x_2)+\tau_{m+1}\frac{\partial}{\partial x_2}g(x_1,x_2)\\&+\lambda_{m+1}\tau_{m+1}\frac{\partial^2}{\partial x_1\partial x_2}g(x_1,x_2.)\end{aligned}$$

Substituting into the equation (3.2) leads to

$$\begin{aligned}f(x_1,x_2)=&\sum_{j=0}^{m}\sum_{i=0}^{m}(s_m^{(i)}t_m^{(j)}\frac{\partial^{i+j}}{\partial x_1^i\partial x_2^j}g(x_1,x_2)+\lambda_{m+1}s_m^{(i)}t_m^{(j)}\frac{\partial^{i+j+1}}{\partial x_1^{i+1}\partial x_2^j}g(x_1,x_2)+\\&+\tau_{m+1}s_m^{(i)}t_m^{(j)}\frac{\partial^k i+j+1}{\partial x_1^i x_2^{j+1}}g(x_1,x_2)+\lambda_{m+1}\tau_{m+1}s_m^{(i)}t_m^{(j)}\frac{\partial^{i+j+2}}{\partial x_1^{i+1}\partial x_2^{j+1}})\\&=\sum_{j=0}^{m+1}\sum_{i=0}^{m+1}k_{m+1}^{(i,j)}\frac{\partial^{i+j}}{\partial x_1^i\partial x_2^j}g(x_1,x_2)\end{aligned}$$

where the coefficients are $(1\le i,j\le m)$

$$\begin{aligned}k_{m+1}^{(0,0)}&=s_m^{(0)}t_m^{(0)}=1=s_{m+1}^{(0)}t_{m+1}^{(0)}\\k_{m+1}^{(0,j)}&=s_m^{(0)}t_m^{(j)}+s_m^{(0)}\tau_{m+1}t_m^{(j-1)}=t_{m+1}^{(j)}=s_{m+1}^{(0)}t_{m+1}^{(j)}\\k_{m+1}^{(i,0)}&=s_{m+1}^{(i)}t_{m+1}^{(0)}\\k_{m+1}^{(i,j)}&=s_m^{(i)}t_m^{(j)}+\lambda_{m+1}s_m^{(i-1)}t_m^{(j)}+\tau_{m+1}s_m^{(i)}t_m^{(j-1)}\\&=t_m^{(j)}(s_m(i)+\lambda_{m+1}s_m^{(i-1)})+\tau_{m+1}t_m^{(j-1)}(s_m^{(i)}+\lambda_{m-1}s_m^{(i-1)})\\&=t_m^{(j)}s_{m+1}^{(i)}+\tau_{m+1}t_m^{(j-1)}s_{m+1}^{(i)}=s_{m+1}^{(i)}t_{m+1}^{(j)}\\k_{m+1}^{(i,m+1)}&=\tau_{m+1}s_m^{(i)}t_m^{(m)}+\lambda_{m+1}\tau_{m+1}s_m^{(i-1)}t_m^{(m)}=s_{m+1}^{(i)}\tau_{m+1}t_m^{(m)}\\&=s_{m+1}^{(i)}\tau_1...\tau_m\tau_{m+1}=s_{m+1}^{(i)}t_{m+1}^{(m+1)}\\k_{m+1}^{(m+1,j)}&=s_{m+1}^{(m+1)}t_{m+1}^{(j)}\\k_{m+1}^{(m+1,m+1)}&=\lambda_{m+1}\tau_{m+1}s_m^{(m)}t_m^{(m)}=s_{m+1}^{(m+1)}t_{m+1}^{(m+1)},\end{aligned}$$

which finishes the proof.

□

A.2 The decreasing density estimator with the product normal kernel

In order to bring the result of Theorem 4.3 even further and obtain a closed-form expression that would be usable for implementation purposes, let us make an additional assumption. And this is referred to Krycha(2011). Assume that the two-dimensional kernel is a product of two one-dimensional normal(Gaussian) kernels,i.e.

$$\mathbf{K}(x_1, x_2) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x_1^2}{2})\frac{1}{\sqrt{2\pi}}\exp(-\frac{x_2^2}{2}) = \phi(x_1)\phi(x_2)$$

Then we can write:

$$\begin{aligned}
&f_{\mathbf{X}}^{(n)}(y_1, y_2) = \\
&\frac{1}{nh^2}\sum_{i=1}^{N}\int_{y_1}^{\infty}\frac{1}{x_1^2}\phi(\frac{-\log(x_1)-a_{1i}}{h})dx_1\int_{y_2}^{\infty}\frac{1}{x_2^2}\phi(\frac{-\log(x_2)-a_{2i}}{h})dx_2 + \\
&\frac{1}{nh^3}\sum_{i=1}^{N}\int_{y_1}^{\infty}\frac{1}{x_1^2}\phi'(\frac{-\log(x_1)-a_{1i}}{h})dx_1\int_{y_2}^{\infty}\frac{1}{x_2^2}\phi(\frac{-\log(x_2)-a_{2i}}{h})dx_2 + \\
&\frac{1}{nh^3}\sum_{i=1}^{N}\int_{y_1}^{\infty}\frac{1}{x_1^2}\phi(\frac{-\log(x_1)-a_{1i}}{h})dx_1\int_{y_2}^{\infty}\frac{1}{x_2^2}\phi'(\frac{-\log(x_2)-a_{2i}}{h})dx_2 + \\
&\frac{1}{nh^4}\sum_{i=1}^{N}\int_{y_1}^{\infty}\frac{1}{x_1^2}\phi'(\frac{-\log(x_1)-a_{1i}}{h})dx_1\int_{y_2}^{\infty}\frac{1}{x_2^2}\phi'(\frac{-\log(x_2)-a_{2i}}{h})dx_2 +
\end{aligned}$$

Let us see how we can progress to obtain a closed form of the above integrals.
Firstly,

$$\int_{y}^{\infty}\frac{1}{x_2^2}\phi(\frac{-\log(x)-A_i}{h})dx = \int_{y}^{\infty}\frac{1}{x^2}\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{\log(x)+A_i}{h})^2)dx$$

then substitute

$$\begin{aligned}
z &= \frac{\log(x)+A_i}{h} = \frac{\log x - \log(X_i)}{h} \\
x &= X_i\exp(hz)
\end{aligned}$$

obtaining

$$
\begin{aligned}
&\int_{\frac{\log(y)-\log(X_i)}{h}} \exp(-2hz)\frac{1}{X_i^2}\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})hX_i\exp(hz)dz = \\
&\frac{h}{X_i}\int_{\frac{\log(y)-\log(X_i)}{h}} \frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2+2hz}{2})dz \\
&= \frac{h}{X_i}\exp(\frac{h^2}{2})\int_{\frac{\log(y)-\log(X_i)}{h}} \frac{1}{\sqrt{2\pi}}\exp(-\frac{(z+h)^2}{2})dz \\
&= \frac{h}{X_i}\exp(\frac{h^2}{2})(1-\Phi(\frac{\log(y)-\log(X_i)}{h}))
\end{aligned}
$$

The intervals containing a derivative of ϕ can be transformed into a closed-form formula using a similar procedure,

$$
\begin{aligned}
&\int_y^\infty \frac{1}{x^2}\phi'(\frac{-\log(x)-A_i}{h})dx \\
&= \frac{h}{X_i}\int_{\frac{\log(y)-\log(X_i)}{h}}^\infty \exp(-hz)\phi'(-z)dz \\
&= -\frac{h}{X_i}\int_{\frac{\log(y)-\log(X_i)}{h}}^\infty \exp(-hz)\phi'(z)dz \\
&= -\frac{h}{X_i}[\exp(-hz)\phi(z)]_{\frac{\log(y)-\log(X_i)}{h}} - \frac{h^2}{X_i}\int_{\frac{\log(y)-\log(X_i)}{h}}^\infty \exp(-hz)\phi'(z)dz \\
&= \frac{h}{X_i}\exp(\log(X_i)-\log(y))\phi(\frac{\log(y)-\log(X_i)}{h}) \\
&-\frac{h^2}{X_i}\exp(\frac{h^2}{2})(1-\Phi(\frac{\log(y)-\log(X_i)}{h}))
\end{aligned}
$$

Plugging these result into $f_{\mathbf{X}}^{(n)}(y_1, y_2)$, will result in the desires closed-form formula.We can still simplify the result if we denote

$$
\begin{aligned}
H(X_i, y) &= \frac{h}{X_i}\exp(\frac{h^2}{2})(1-\Phi(\frac{\log(y)-\log(X_i)}{h})) \\
F(X_i, y) &= \frac{h}{X_i}\exp((\log X_i)-\log(y))\phi(\frac{\log(y)-\log(X_i)}{h}) \\
&= \frac{h}{y}\phi(\frac{\log(y)-\log(X_i)}{h})
\end{aligned}
$$

Then we can write

$$
\begin{aligned}
&f_{\mathbf{X}}^{(n)}(y_1, y_2) \\
&= \frac{1}{nh^2} \sum_{i=1}^{N} (H(X_{1i}, y_1)H(X_{2i}, y_2) + \frac{1}{h}(F(X_{1i}, y_1) - hH(X_{1i}, y_1))H(X_{2i}y_2) \\
&+ \frac{1}{h}H(X_{1i}, y_1)(F(X_{2i}, y_2) - hH(X_{2i}, y_2)) \\
&+ \frac{1}{h^2}(F(X_{1i}, y_1) - hH(X_{1i}, y_1))(F(X_{2i}, y_1) - hH(X_{2i}, y_2))) \\
&= \frac{1}{nh^4} \sum_{i=1}^{N} F(X_1 i, y_1)F(X_{2i}, y_2) \\
&= \frac{1}{nh^2} \sum_{i=1}^{N} \frac{1}{y_1 y_2} \phi(\frac{\log(y_1) - \log(X_{1i})}{h})\phi(\frac{\log(y_2) - \log(X_{2i})}{h})
\end{aligned}
$$

We see that we obtained an estimator of the kernel type, applied on the logarithmic transform of the original data. An estimator in this form can be easily implemented and produce effective estimations of decreasing densities, as we will see in the chapter5.
A disadvantage of this approach can be seen if we look at the formula.If the estimator $f_{\mathbf{B}}^{(n)}$ is nonnegative on the support, then the resulting integral indeed produces a decreasing density. However, in the general case, the kernel estimator can attain also negative values and we lose the desirable property of a decreasing estimator. One way to deal with that is to treat the areas with negative density estimator as if they were not part of the support. Then we get a new estimator that we have to rescale to make sure it has the fundamental property of the density. The modified kernel estimator of $((B))$ is then

$$
\hat{f}_{\mathbf{B}}^{(n)}(b_1, b_2) = \frac{max\{f_{\mathbf{B}}^{(n)}, 0\}}{\int_{\{c_1, c_2 : f_{\mathbf{B}}^{(n)}(c_1, c_2) > 0\}} f_{(B)}^{(n)}(c_1, c_2)dc_1 dc_2}
$$

However, this alteration prevents us from simplifying the formula as we did originally. The implementation of the algorithm with the modified kernel estimator is then problematic.

Appendix B

Appendix Fast computation of the multivariate kernel density estimator

In this section, we are going to introduce a method which will give a faster way to esitimate the density function by using kernel estimators.This method is given by M.P.Wang in 1994.
Binned kernel estimates are usually computed over an equally-spaced mesh of grid points. The same ideas can be applied to obtain quickly computable approximations to kernel functional estimates, which arise in many common automatic bandwidth selection algorithms. Their calculation requires three distinct steps:

1. Bin the data by assigning the raw data to neighboring grid points to obtain grid counts. A grid count can be thought of as representing the amount of data in the neighborhood of its corresponding grid point.
2. Compute the required kernel weights. The fact that the grid points are equally spaced means that the number of distinct kernel weights is comparatively small.
3. Combine the grid counts and the kernel weights to obtain the approximation to the kernel estimate.

First we will introduce binned multivariate kernel estimators.
Let $\mathbf{X}_1, ..., \mathbf{X}_n$ are $\mathbb{R}^d$-valued random variables having a d-variate density f.The kernel density estimate of $f(x)$ and statistics based on this estimate depend on quantities of the form

$$\hat{s}_{\mathbf{k}}(\mathbf{x}) = \sum_{i=1}^{n} (\mathbf{X_i} - \mathbf{x})^{\mathbf{k}} K_{\mathbf{h}}^{P}(\mathbf{X_i} - \mathbf{x})$$

where $\mathbf{k} = (k_1, ..., k_d)$, and for a d-vector $\mathbf{u} = (u_1, ..., u_d)$, the convention $\mathbf{u}^{\mathbf{k}} = u_1^{k_1}...u_d^{k_d}$ is used.
The notation $K_{\mathbf{h}}^{P}$ applies to the rescaling of the product kernel $K^P(\mathbf{x})$ by the vector of bandwidths $\mathbf{h} = (h_1, ..., h_d)$:

$$K_{\mathbf{h}}^{P}(\mathbf{x}) = K_{h_1}(x_1)...K_{h_d}(x_d),$$

where K is a symmetric probability density function and $K_h(x) = K(x/h)/h$ is a rescaling of K by the bandwidth $h > 0$. If K has compact support then we will let $[-\tau, \tau]$ denote the interval outside which K is 0. If K has infinite support, then one could replace K by $K\mathbb{I}_{[-\tau,\tau]}$, where τ is chosen so that K is effectively 0 outside of $[\tau, \tau]$. The choice of τ should be such that the truncation to $[-\tau, \tau]$ has negligible effect on the final estimation. For example, if K is the standard normal density, then $\tau = 4$ is a safe choice.
Note that $K_{\mathbf{h}}^{P}$ does not include all multivariate kernels of interest, such as those based on rotations of a univariate kernel. However, for computational purposes great savings in the number of kernel evaluations are possible if the product structure of $K_{h_i}(xi)$ is available. Also not included is the possibility of smoothing in orientations different to those coordinate axes. This can be done by rescaling K^P by a bandwidth matrix which was demonstrated by Wand and Jones(1993) to be an important extension in certain circumstances. If smoothing in different orientations is desired, then it is recommended that the data be prorated so that the product scaling is adequate. The computation can be done on the rotated data, and then the data can be rotated back to correspond to the coordinates of the original data.
For $i = 1, ..., d$,let $g_{i1} < ... < g_{iM_i}$ be an equally spaced grid in the ith coordinate directions such that $[g_{i1}, g_{i,M_i}]$ contains the ith coordinate values of the X's. Here M_i is a positive integer representing the grid size in direction i. Let

$$\mathbf{g_j} = (g_{1j_1}, ..., g_{dj_d}), 1 \leq j_i \leq M_i,\ i = 1, ..., d$$

denote the grid point indexed by $\mathbf{j} = (j_1, ..., j_d)$ and the ith binwidth be denoted by $_i = \frac{g_{iM_i} - g_{i1}}{M_i - 1}$. Fast binned approximations of kernel estimators involve binning the original data to obtain grid counts c_j that represent the amount of $\mathbf{X}$ data near each grid point. Strategies for obtaining grid counts are described later. The binned approximation to $\hat{s}_{\mathbf{k}}(\mathbf{g_j})$ is

$$\tilde{s}_{\mathbf{k}}(\mathbf{g_j}) = \sum_{l_1=1}^{M_1} ... \sum_{l_d=1}^{M_d} (\mathbf{g_j} - \mathbf{g_l})^{\mathbf{k}} K_{\mathbf{h}}(\mathbf{g_j} - \mathbf{g_l}) c_{\mathbf{l}}.$$

It is easy to show that

$$\tilde{s}_{\mathbf{k}}(\mathbf{g_j}) = \sum_{l_1=-L_1}^{L_1} ... \sum_{l_d=-L_d}^{L_d} c_{\mathbf{j}-\mathbf{l}} \kappa_{\mathbf{k},\mathbf{l}}$$

where

$$\kappa_{\mathbf{k},\mathbf{l}} = \prod_{i=1}^{d} K_{h_i}(l_i \delta_i)(l_i \delta_i)^{k_i}$$
$$L_i = min(\tau h_i / \delta_i \rfloor, M_i - 1)$$

and $x\rfloor$ denotes the greatest integer less than or equal to x.If the $\tilde{s}_{\mathbf{k}}(\mathbf{g_j})$ is substituted for$\hat{s}_{\mathbf{k}}(\mathbf{g_j})$ in the formulas for $\hat{f}(\mathbf{g_j})$ then its binned approximation $\tilde{f}(\mathbf{g_j})$ results.
The binned approximations $\tilde{s}_{\mathbf{k}}$ represent enormous computational savings, because only $\sum_{i=1}^{d} L_i$ kernel evaluations are required to obtain the $\kappa_{\mathbf{k},\mathbf{l}}$ regardless of the value of n.

Second, we will introduce multivariate binning rules.

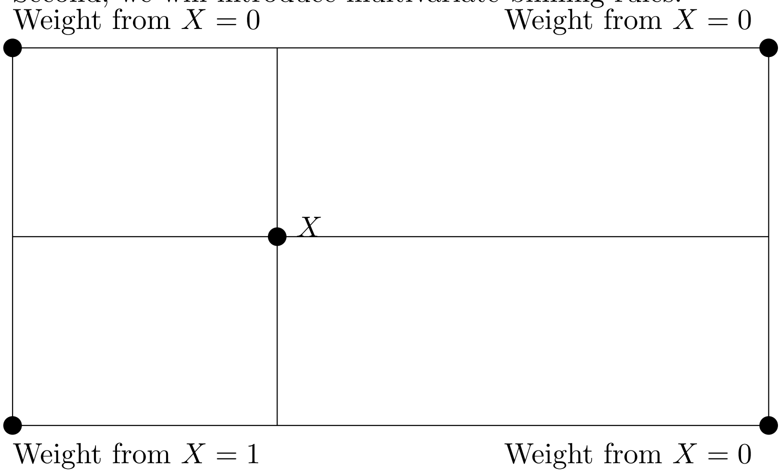

Figure 5.1 Simple bining

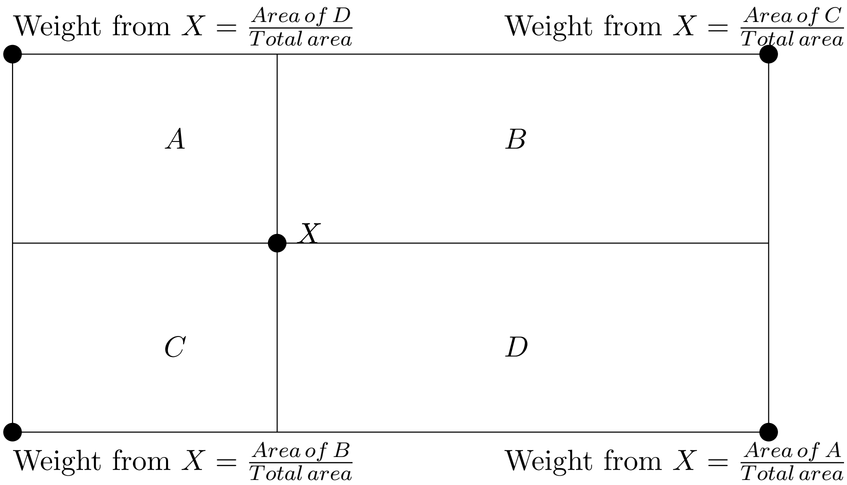

Figure 5.2 Linear bining

The two most common univariate binning rules are simple binning and linear binning. If a data point at y has surrounding grid points at x and z, then simple binning involves assigning a unit mass to the grid point closest to y. Linear binning assigns a mass of (z - y)/(z - x) to the grid point at x, and (y - x)/(z - x) to the grid point at z (see Jones and Lotwick (1983). Multivariate binning rules may be defined by taking the product of univariate rules. Figure 1 gives a graphical description of how a data point $\mathbf{X}$ distributes its weight to neighboring grid points for the bivariate extension of simple and linear binning. For simple binning, the point $\mathbf{X}$ gives all of its weight to its nearest grid point, in this case the grid point at the lower left vertex of the rectangle formed by joining the four grid points neighboring $\mathbf{X}$. In the case of linear binning, the contribution from $\mathbf{X}$ is distributed among each of the four surrounding grid points according to areas of the opposite subrectangles induced by the position of the data point. Higher-dimensional extensions of simple binning and linear binning, where areas are replaced by volumes, are obvious. Let $w_{\mathbf{l}}(\mathbf{X})$ be the weight that $\mathbf{X}$ assigns to $\mathbf{g}_{\mathbf{l}}$, by one of the described binning rules. Then the $c_{\mathbf{l}}$ is given by:

$$c_{\mathbf{l}} = \sum_{i=1}^{n} w_{\mathbf{l}}(\mathbf{X}_i)$$

For both simple and linear binning the $c_{\mathbf{l}}$ can be computed using a fast $O(n)$algorithm by extending the "integer division" idea of Fan and Marron (1994).

Two obvious questions that arise are:

1. How do simple and linear binning compare?
2. How many bins should one use in each direction?

Question 1 is partially answered by asymptotic results of Hall and Wand (1993). A straightforward extension of their arguments leads to, for constants A_i and B_i,

$$\mathbb{E}[\tilde{s}_k(x) - \hat{s}_k(x)]^2 = \sum_{i=1}^{d} A_i\delta_i^2 + o(\sum_{i=1^d} \delta_i^2) \text{ , for simple bining,}$$
$$= \sum_{i=1}^{d} B_i\delta_i^4 + o(\sum_{i=1^d} \delta_i^4) \text{ , for linear bining,}$$

,

as $\delta_i \to 0$,$i = 1, ..., d$.Therefore, in terms of how well the $\tilde{s}_k$ approximate the $\hat{s}_k$, linear binning is an order of magnitude better than simple binning..

It is impossible to give an absolute answer to Question 2 because functions with finer structure require more grid points to achieve a given level of accuracy. Insight into the effects of binning on accuracy can only be realized through examples.

Appendix C

Appendix Matlab Code

```
%Epanech
n=500;
sample=zeros(n,2);
h=2.5;
rand( seed ,1);
randn( state ,1);
s1=exprnd(1,n,2);
s2=normrnd(0,1,n,2);
sample=s1+s2;
x1=zeros(n,1);
y1=zeros(n,1);
x=linspace(0.000001,4,n);
y=linspace(0.000001,4,n);
f=zeros(n,n,2);
f1=zeros(n,n);
f3=zeros(n,n);
for i=1:n
    for j=1:n
        for m=1:n
    x1(m)=(-log(x(i))-sample(m,1))/h;
    y1(m)=(-log(y(j))-sample(m,2))/h;
    if x1(m)^2+y1(m)^2<=1
    f(i,j,2)=f(i,j,1)+9/(16*n*h^2)*(1-x1(m)^2)*(1-y1(m)^2)+(-9)/(8*n*h^3)*x1(m)*(1-y1(m)^2)
    f(i,j,1)=f(i,j,2);
    end
    end
    f1(i,j)=f(i,j,1);
    end
end
for i=1:n
    for j=1:n
```

```
        f1(i,j)=f1(i,j)/(x(i)^2*y(j)^2);
    end
end
for i=1:n
    for j=1:n
        if f1(i,j)<0
            f1(i,j)=0;
        end
    end
end
m=((4-0.000001)/n)^2;
f3(n,n)=f1(n,n)*m;
f3(n,n-1)=(f1(n,n)+f1(n,n-1))*m;
f3(n-1,n)=(f1(n,n)+f1(n-1,n))*m;
for i= n-1:-1:1
    for j= n-1:-1:1
        f3(i,j)=f1(i,j)*m + f3(i+1,j) + f3(i,j+1) - f3(i+1,j+1);
    end
end
fn=zeros(n,n);
zn=zeros(n,n);
mu = [1,1]; SIGMA = [1 0; 0 1];
[X1,X2] = meshgrid(x ,y );
X = [-log(X1(:)) -log(X2(:))];
p = mvncdf(X,mu,SIGMA);
fn=reshape(p,n,n)*exp(1);
for i=1:n
    for j=1:n
        zn(i,j)=abs(fn(i,j)-f3(i,j));
    end
end
mesh(x,y,zn)
rotate3d on
%--------------------------------------------------------------------------
%Uniform
n=500;
sample=zeros(n,2);
h=2.5;
rand( seed ,1);
randn( state ,1);
s1=normrnd(0,1,n,2);
s2=exprnd(1,n,2);
sample=s1+s2;
x1=zeros(n,1);
y1=zeros(n,1);
x=linspace(0.1,4,n);
y=linspace(0.1,4,n);
f=zeros(n,n,2);
```

```
f1=zeros(n,n);
for i=1:n
    for j=1:n
        for m=1:n
    x1(m)=(-log(x(i))-sample(m,1))/h;
    y1(m)=(-log(y(j))-sample(m,2))/h;
    if abs(x1(m))<=1 & abs(y1(m))<=1
    f(i,j,2)=f(i,j,1)+1/(4*n*h^2);
    f(i,j,1)=f(i,j,2);
    else f(i,j,1)=f(i,j,1);
    end
    end
    f1(i,j)=f(i,j,1);
    end
end
for i=1:n
    for j=1:n
        f1(i,j)=f1(i,j)/(x(i)^2*y(j)^2);
    end
end
m=((4-0.0000001)/n)^2;
f3(n,n)=f1(n,n)*m;
f3(n,n-1)=(f1(n,n)+f1(n,n-1))*m;
f3(n-1,n)=(f1(n,n)+f1(n-1,n))*m;
for i= n-1:-1:1
    for j= n-1:-1:1
        f3(i,j)=f1(i,j)*m + f3(i+1,j) + f3(i,j+1) - f3(i+1,j+1);
    end
end
mesh(x,y,f3)
rotate3d on
%--------------------------------------------------------------------------
%Biweight
n=500;
sample=zeros(n,2);
h=2.5;
rand( seed ,1);
randn( state ,1);
s2=exprnd(1,n,2);
sample=s2;
x1=zeros(n,1);
y1=zeros(n,1);
x=linspace(0.1,4,n);
y=linspace(0.1,4,n);
f=zeros(n,n,2);
f1=zeros(n,n);
for i=1:n
    for j=1:n
```

```
        for m=1:n
    x1(m)=(-log(x(i))-sample(m,1))/h;
    y1(m)=(-log(y(j))-sample(m,2))/h;
    if abs(x1(m))<=1 && abs(y1(m))<=1
    f(i,j,2)=f(i,j,1)+(15)^2/(16^2*n*h^2)*(1-x1(m)^2)^2*(1-y1(m)^2)^2-(15)^2/(16*4*n*h^3)*(1-y1(m)
    f(i,j,1)=f(i,j,2);
    else f(i,j,1)=f(i,j,1);
    end
    end
    f1(i,j)=f(i,j,1);
    end
end
for i=1:n
    for j=1:n
        if f1(i,j)>=0
        f1(i,j)=f1(i,j)/(x(i)^2*y(j)^2);
        else
            f1(i,j)=0;
        end
    end
end
m=((4-0.000001)/n)^2;
f3(n,n)=f1(n,n)*m;
f3(n,n-1)=(f1(n,n)+f1(n,n-1))*m;
f3(n-1,n)=(f1(n,n)+f1(n-1,n))*m;
for i= n-1:-1:1
    for j= n-1:-1:1
        f3(i,j)=f1(i,j)*m + f3(i+1,j) + f3(i,j+1) - f3(i+1,j+1);
    end
end
mesh(x,y,f3)
%--------------------------------------------------------------------------
%Triweight
n=500;
sample=zeros(n,2);
h=2.5;
rand( seed ,1);
randn( state ,1);
s2=exprnd(1,n,2);
sample=s2;
x1=zeros(n,1);
y1=zeros(n,1);
x=linspace(0.1,4,n);
y=linspace(0.1,4,n);
f=zeros(n,n,2);
f1=zeros(n,n);
for i=1:n
    for j=1:n
```

```
        for m=1:n
    x1(m)=(-log(x(i))-sample(m,1))/h;
    y1(m)=(-log(y(j))-sample(m,2))/h;
    if abs(x1(m))<=1 && abs(y1(m))<=1
    f(i,j,2)=f(i,j,1)+(35)^2/(32^2*n*h^2)*(1-x1(m)^2)^3*(1-y1(m)^2)^3-(6)*(35)^2/(32^2*n*h^
    f(i,j,1)=f(i,j,2);
    else f(i,j,1)=f(i,j,1);
    end
    end
    f1(i,j)=f(i,j,1);
    end
end
for i=1:n
    for j=1:n
        if f1(i,j)>=0
        f1(i,j)=f1(i,j)/(x(i)^2*y(j)^2);
        else f1(i,j)=0;
        end
    end
end
m=((4-0.000001)/n)^2;
f3(n,n)=f1(n,n)*m;
f3(n,n-1)=(f1(n,n)+f1(n,n-1))*m;
f3(n-1,n)=(f1(n,n)+f1(n-1,n))*m;
for i= n-1:-1:1
    for j= n-1:-1:1
        f3(i,j)=f1(i,j)*m + f3(i+1,j) + f3(i,j+1) - f3(i+1,j+1);
    end
end
mesh(x,y,f3)
%--------------------------------------------------------------------------
%Gaussian
n=100;
sample=zeros(n,2);
h=2;
rand( seed ,1);
randn( state ,1);
s1=exprnd(1,n,2);
s2=normrnd(0,5,n,2);
sample=s1+s2;
x1=zeros(n,1);
y1=zeros(n,1);
x=linspace(0.1,4,n);
y=linspace(0.1,4,n);
f=zeros(n,n,2);
f1=zeros(n,n);
f3=zeros(n,n);
for i=1:n
```

```
    for j=1:n
        for m=1:n
    x1(m)=(-log(x(i))-sample(m,1))/h;
    y1(m)=(-log(y(j))-sample(m,2))/h;
    f(i,j,2)=f(i,j,1)+1/(2*3.14*n*h^2)*exp(-1/2*(y1(m)^2))*exp(-1/2*(x1(m)^2))-1/(2*3.14*n*h^3)*ex
    f(i,j,1)=f(i,j,2);
    end
    f1(i,j)=f(i,j,1);
    end
end

for i=1:n
   for j=1:n
       if f1(i,j)>0
           f1(i,j)=f1(i,j)/(x(i)^2*y(j)^2);
       else f1(i,j)=0;
       end
   end
end
m=((4-0.000001)/n)^2;
f3(n,n)=f1(n,n)*m;
f3(n,n-1)=(f1(n,n)+f1(n,n-1))*m;
f3(n-1,n)=(f1(n,n)+f1(n-1,n))*m;
for i= n-1:-1:1
    for j= n-1:-1:1
        f3(i,j)=f1(i,j)*m + f3(i+1,j) + f3(i,j+1) - f3(i+1,j+1);
    end
end
mesh(x,y,f3)
 rotate3d on
%--------------------------------------------------------------------------
%Cosinus
n=100;
sample=zeros(n,2);
h=2.5;
rand( seed ,1);
randn( state ,1);
s1=exprnd(1,n,2);
sample=s1;
x1=zeros(n,1);
y1=zeros(n,1);
x=linspace(0.1,4,n);
y=linspace(0.1,4,n);
f=zeros(n,n,2);
f1=zeros(n,n);
for i=1:n
    for j=1:n
        for m=1:n
```

```
    x1(m)=(log(x(i))-sample(m,1))/h;
    y1(m)=(log(y(j))-sample(m,2))/h;
    if abs(x1(m))<=1 & abs(y1(m))<=1
    f(i,j,2)=f(i,j,1)+3.14^2/(16*n*h^2)*cos(3.14/2*y1(m))*cos(3.14/2*x1(m))-3.14^3/(32*n*h^
    f(i,j,1)=f(i,j,2);
    else f(i,j,1)=f(i,j,1);
    end
    end
    f1(i,j)=f(i,j,1);
    end
end
for i=1:n
   for j=1:n
       if f1(i,j)>0
           f1(i,j)=f1(i,j)/(x(i)^2*y(j)^2);
       else f1(i,j)=0;
       end
   end
end
m=((4-0.1)/n)^2;
f3(n,n)=f1(n,n)*m;
f3(n,n-1)=(f1(n,n)+f1(n,n-1))*m;
f3(n-1,n)=(f1(n,n)+f1(n-1,n))*m;
for i= n-1:-1:1
    for j= n-1:-1:1
        f3(i,j)=f1(i,j)*m + f3(i+1,j) + f3(i,j+1) - f3(i+1,j+1);
    end
end
mesh(x,y,f3)
```

Listing C.1: decreasing density estimation

Bibliography

[1] G.Biau and L.Devroye, On the risk of estimates for block decreasing densities, *Journal of multivariate analysis*, 86:143-165, (2003)

[2] K.Behnen, The Randle-Hogg test and an alternative proposal, *Communications in Statistics*, 4:203-238, (1975)

[3] J.E.Chacon, T.Duong,and M.P.Wand, Asymptotics for general multivariate kernel density derivative estimators, *Statistica Sinica*,21:807-840, (2011)

[4] A.J.van Es and A.R. Kok,Simple kernel estimators for certain nonparametric deconvolution problems, *Statistics and Probability Letters*, 39:151-160, (1998)

[5] J.Fan and J.S.Marron, Fast Implementations of Nonparametric Curve Estimators, *Journal of Computational and Graphical Statistics*, 3: 35-56, (1994).

[6] W.Feller , An Introduction to Probability Theory and Its Applications, *Wiley*, (1971)

[7] P.Groeneboom and H.P.Lopuhaa, Isotonic estimators of monotone densities and distribution functions:basic facts, *Statistica Neerlandica*, 47:175-183, (1993)

[8] P.Groeneboom and J.A.Wellner, Information bounds and nonparametric maximum likelihoodestimation, (1992)

[9] U.Grenander, On the theory of mortality measurement, *Skandinavisk Aktuarietidskrift*, 39:125-153, (1956)

[10] P.Hall and M.P.Wand, On the Accuracy of Binning Approximations to Kernel Density Estimators, *Working Paper Series 93-003, University of New South Wales*, Australian Graduate School of Management, (1993)

[11] T.J.Hastie and C.Loader, Local Regression: Automatic Kernel Carpentry, *Statistical Science*, 8: 120-143, (1993)

[12] G.Jongloed,Exponential deconvolution:Two asymptotically equivalent estimators, *Statistica Neerlandica*, 52:6-17, (1998)

[13] D.Krycha, Two-dimensional Nonparametric Deconvolution, *Master thesis of VU*, (2011)

[14] W. Polonik, The Silhouette concentration functions and ML-density estimation under order restrictions,*The annals of Statistics*, 26:1857-1877, (1998)

[15] W. Polonik, Desity Estimation under Qualitative Assumptions in Higher Dimensions, *Journal of multivariate analysis*, 55:61-81, (1995)

[16] B.W.Silverman, Density Estimation for Statistics and Data Analysis, *Chapman and Hall*, (1998)

[17] F.W.Scholz, Combining independent p-values,*P.J.Bickl et al.(eds.)*, (1983)

[18] Stein and G.Weiss, Introduction to Fourier Analysis on Euclidean Spaces, *Princeton University Press*, (1971)

[19] M.P.Wand, Fast Computation of Multivariate Kernel Estimator, *Journal of Computational and Graphical Statistics*, 3:433-445, (1994)

[20] M.P.Wand, M.C.Jones, Kernel Smoothing, *Chapman and Hall*,(1995)