



Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Bivariate density estimation in an Oriented Cylinder Model

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

**MASTER OF SCIENCE
in
APPLIED MATHEMATICS**

by

S. BOERSMA

**Delft, The Netherlands
December 2013**

Copyright © 2013 by S. Boersma. All rights reserved.



MSc thesis APPLIED MATHEMATICS

“Bivariate density estimation in an Oriented Cylinder Model”

S. Boersma

Delft University of Technology

Daily supervisors

Prof.dr.ir. G. Jongbloed

Dr. P.J.J. Kok

Responsible professor

Prof.dr.ir. G. Jongbloed

Other thesis committee members

Dr.ir. F.H. van der Meulen

Prof.dr.ir. J. Sietsma

December, 2013

Delft, The Netherlands

Abstract

The microstructure of an object of dual phase steel is modelled with an Oriented Cylinder Model, where the martensite grains in the steel object are modelled as oriented cylinders that are randomly distributed within a big block. By cutting this block parallel to the cylinder axes, rectangular visible profiles of the cut cylinders can be observed from the cut-plane. To translate the two-dimensional information about the rectangles to three-dimensional information about the cylinders, an inverse relationship between the bivariate probability density function of the height and squared half-width of a rectangle and the bivariate density of the height and squared radius of a cylinder is established. The former density is estimated by a bivariate modified Gamma kernel density estimator, and the relationship between the two densities is used to numerically transform this estimator into an estimator for the latter density. For the selection of the bandwidth parameters of the kernel density estimator, both a method based on the Integrated Squared Error of the kernel density estimator and a method based on the Integrated Squared Error of the transformed estimator are considered. The estimator for the height and squared radius of a cylinder is constructed for the underlying density of an example data set as well as for the underlying density of an experimental data set. These bivariate estimators are used to construct an estimator for the marginal density of the squared radius of a cylinder and the density of the volume of a cylinder.

Contents

1	Introduction	7
1.1	Introduction to the problem	7
1.2	Main idea of the Oriented Cylinder Model	8
1.3	Goal of this thesis	9
1.4	Thesis outline	9
2	Oriented Cylinder Model	10
2.1	General setting and notation	10
2.2	Important distributions and their relationship	12
3	Kernel Density Estimation	17
3.1	Univariate kernel density estimation	17
3.1.1	Univariate kernel density estimator	17
3.1.2	Bandwidth parameter selection	21
3.1.3	Example of a univariate kernel density estimator	23
3.2	Bivariate kernel density estimation	25
3.2.1	Bivariate kernel density estimator	25
3.2.2	Bandwidth matrix selection	28
3.2.3	Example of a bivariate kernel density estimator	29
3.3	Boundary problems	32
3.3.1	Gamma kernel density estimator	33
3.3.2	Modified Gamma kernel density estimator	40
3.3.3	Examples of a univariate modified Gamma kernel density estimator	44
3.3.4	Bivariate Gamma kernel density estimator	46
3.4	Estimator for the density of the height and squared half-width of a rectangle	49
3.4.1	Data simulation	49
3.4.2	Construction of the estimator	51
4	Estimator for the density of the height and squared radius of a cylinder	54
4.1	Transformation procedure	54
4.1.1	Numerical integration with the Trapezoidal Rule	54
4.1.2	Numerical differentiation with Finite Differences	56
4.2	Construction of the estimator	57
4.3	Bandwidth selection based on the density of interest	58
5	Application to experimental data	64
6	Important densities	67
6.1	Marginal density of the squared radius of a cylinder	67
6.2	Density of the volume of a cylinder	69
7	Conclusion and discussion	71
8	Recommendations	73
	References	74
A	Central moments of the $\text{Gamma}(a, b)$ distribution	75

1 Introduction

1.1 Introduction to the problem

An important material used in many products is steel. Think of various means of transportation like cars, trains, ships and planes, or packaging, like cans for example. It is not hard to imagine that these different types of products need different types of steel. For a can the steel should be thin but waterproof, while for a car the steel should have certain safety properties as well. It is therefore of great importance that the steel that is used for a particular product has the right properties. If this is not the case, possible errors may lead to the sinking of a ship or to the leakage of a can for example. To prevent this from happening, steel is studied extensively in order to assure that the steel that is used has the required properties.

Steel is an alloy, which means that it consists of multiple elements, which are mainly iron and carbon. The properties of a steel object depend very much on the concentrations of these elements, which are prescribed in the production process, and the microstructure that they form. One thus needs to know this microstructure in order to determine the properties of the steel object. However, it is impossible to see the inside of a steel object, because steel is an opaque medium. Therefore, the microstructure is not observable either. There are ways to obtain three-dimensional information about the interior of a steel object, for example by means of serial sectioning [1] or synchrotron radiation [2], but both methods have disadvantages. The former method is time-consuming and destructive to the steel, while the latter method is expensive for example. For this reason, it is regarded impossible to get the desired three-dimensional information.

When one of the two ways for obtaining three-dimensional information is used a little bit differently, the method will become useful. Although serial sectioning requires a long period of time, making only one section can be done relatively quick, which makes it an obvious way to obtain information. Furthermore, just one section will only destruct a small part of the steel, so the steel can still be used. From the obtained cut-plane it is possible to extract two-dimensional information that is naturally related to the corresponding three-dimensional information. This stereological problem is related to “The Corpuscle problem” of Wicksell [3], where an inverse relation between the diameters of corpuscles in an opaque body and the circular contours of these corpuscles on a cut-plane of the body is established.

The particular steel that is considered in this thesis is dual-phase steel, which is a high-strength steel that consists of two phases, with ferrite (a materials science term for pure iron) as the primary phase and martensite (a very hard form of steel crystalline structure) as the secondary phase. The microstructure of the steel is therefore constituted by the grains of ferrite and the grains of martensite. An example of such a cut-plane of dual-phase steel is given in Figure 1, where the white grains are ferrite and the black grains are martensite.

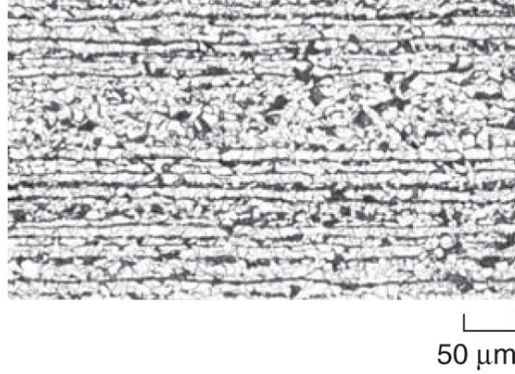


Figure 1: A real cut-plane (white: ferrite, black: martensite).

Note that this microstructure has horizontal bands, formed by the martensite grains. These bands appear in the microstructure of steel, because for the construction of steel, it is rolled. A consequence of the banded microstructure, is that the constructed steel is anisotropic, which means that the properties of the steel depend on the direction the forces work on. As a result, the steel is highly susceptible to cracking and corrosion.

Due to the necessary processing conditions it is currently not possible to prevent the formation of bands. Knowing the properties of these bands is therefore crucial to industry for dealing with the banded microstructures. Information about the microstructure can be observed on the cut-plane of a steel object, like the one in Figure 1. The two-dimensional information obtained from that cut-plane can be translated to three-dimensional information about the whole steel object, which thus leads to information about the microstructure of the steel object. To be able to use and translate this information, a model is needed. The model used for this is the Oriented Cylinder Model proposed by K.S. McGarrity, J. Sietsma and G. Jongbloed [4]. The main idea of the model will be given in the next section, while this model will be discussed in detail in Chapter 2.

1.2 Main idea of the Oriented Cylinder Model

As mentioned above, the information from the cut-plane of a steel object can be processed using the Oriented Cylinder Model. Although dual-phase steel consists of two phases, only one phase has to be considered, since then the properties of the other phase are known as well. In this model, the grains of martensite in the steel object are modelled as oriented cylinders in a big block, which represents the steel object. This block consists of cylinders with the same orientation representing martensite grains and ferrite in the remainder of the block.

As can be seen in Figure 1, the martensite grains have different heights and widths. So, although the orientation of all cylinders is the same, the heights and radii of these cylinders have to be different from each other. In the Oriented Cylinder Model, these heights and radii are randomly distributed according to an unknown bivariate distribution. The focus of this thesis lies in this distribution, because knowledge of this distribution leads to knowledge of the microstructure, and thus to knowledge of the properties of the steel object.

However, there is no three-dimensional information available, so neither the height nor the radius of each cylinder is directly known. To obtain two-dimensional information, the steel object is cut parallel to the cylinder axes, which results in visible profiles of the cut cylinders on the cut-plane. Because the cylinders are cut parallel to their axes, these visible profiles are rectangles. From this cut-plane the two-dimensional information, i.e. the heights and widths of these rectangles, can be obtained. Note that the height of the cylinder simply equals the height of the rectangle. For the radius of the cylinder, though, it is only known that this radius is greater than or equal to half the width of the corresponding rectangle. Since these two pairs of random variables are related, the bivariate distributions of these pairs are also related. The relationship between the densities of these two distributions is established in [4].

1.3 Goal of this thesis

In contrast to previous work, which focuses on univariate densities, the focus of this thesis lies in bivariate densities. More specifically, the goal of this thesis is to construct an estimator for the bivariate density of the height and squared radius of a cylinder. However, as mentioned above, there is no data available of heights and squared radii, so the density cannot be estimated directly. By making use of the discussed relationship between this density and the density of the height and squared half-width of a rectangle observed from the cut-plane, an estimator for the former density can be constructed by first estimating the latter density.

Since data of the rectangles on the cut-plane are observable, an estimator for the bivariate density of the height and squared half-width of a rectangle can be constructed. The density of this distribution is estimated using a bivariate kernel density estimator. Because the height and squared half-width are both always positive, the estimator for this density should be zero outside the first quadrant. However, if there are data points close enough to the boundary, a kernel density estimator with Normal kernels for example will give mass to points outside this quadrant. This problem, known as the boundary problem, is therefore solved first, in order to obtain an unbiased estimator of the bivariate density of the height and squared half-width of a rectangle. After that, the bivariate density of the height and squared radius of a cylinder can be estimated by transforming this estimator via the relationship between the two densities.

1.4 Thesis outline

In the next chapter the Oriented Cylinder Model will be discussed in detail. After that, in Chapter 3, the kernel density estimator will be explained in stages. First, an ordinary univariate kernel density estimator will be introduced. After that, this estimation method will be extended to the second dimension, resulting in a bivariate kernel density estimator. Then the boundary problem for our specific problem will be addressed and a solution to this problem will be given. In the last section, the bivariate Gamma kernel density estimator for the density of the heights and squared half-widths of the rectangles will be constructed, based on an example data set of these quantities. After that, in Chapter 4, the transformation procedure, including numerical integration and differentiation, for transforming the constructed estimator into an estimator for the density of interest will be discussed and applied to the estimator constructed for the underlying density of the example data set. In Chapter 5, the estimator discussed in the previous chapters will be applied to an experimental data set. In Chapter 6, the bivariate Gamma kernel density estimator will be used to estimate univariate densities. This thesis ends with a conclusion and discussion, and various recommendations will be made.

2 Oriented Cylinder Model

In this chapter, the Oriented Cylinder Model will be explained in detail. First the general setting and notation will be discussed, together with the relationships between the height and squared half-width of a rectangle and the height and squared radius of the corresponding cylinder. Based on the model and these relationships, the relationship between the bivariate densities of the two mentioned pairs of random variables will be established.

2.1 General setting and notation

Consider a big block of size $M \times M \times M$. This block represents an object of dual phase steel. As can be seen from Figure 1, the biggest part of dual phase steel consists of ferrite, and only a relatively small amount of martensite grains are present. These martensite grains are modelled as cylinders, with the centres of these cylinders distributed randomly within the block according to a low intensity homogeneous Poisson process. The remainder of the block is constituted by the ferrite grains. Furthermore, each cylinder is oriented in the same way, namely in the upward direction. Since each grain does not have the same height and width, the height H and squared radius X of a cylinder are random variables, distributed according to an unknown bivariate distribution with density function f . This bivariate density f is the function of interest of this thesis.

In Figure 2 an example of such a block is shown, with $M = 1$. Note that all heights as well as all radii differ from each other, but that all the cylinders have an upward orientation. Remember that the cylinders represent the grains of martensite, the remainder of the block represents ferrite.

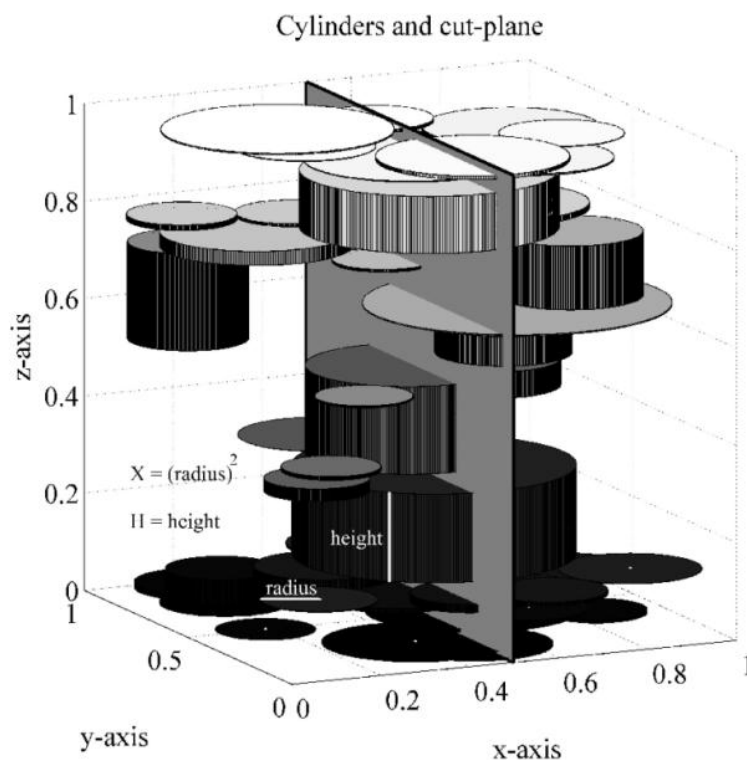


Figure 2: Cylinders in the block.

As mentioned above, the sizes of these martensite grains have a considerable influence on the properties of the particular object of dual phase steel. In this model, these sizes translate to the sizes of the oriented cylinders. Due to the opaqueness of the steel, the three-dimensional grains are not observable, so the cylinders shown in Figure 2 are not observable either. As a result, there is no information about the height and squared radius of a cylinder directly known. However, the information that is observable, is the information from the cut-plane obtained by cutting the block parallel to the axes of the cylinders. In this example the block is cut at $x = 0.5$, represented by the grey plane at this location in Figure 2. The visible profiles of the cut cylinders observed on the cut-plane is shown below in Figure 3.

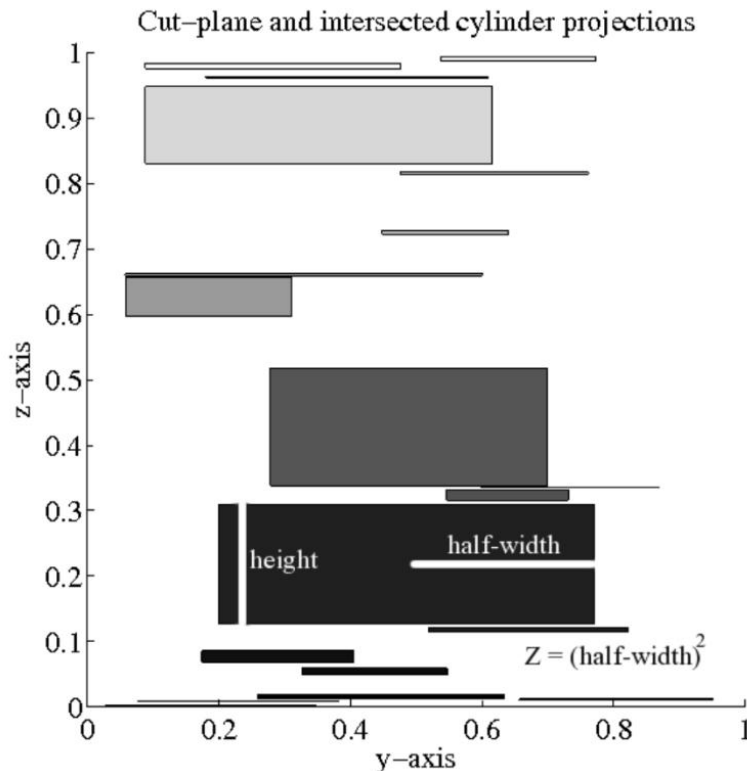


Figure 3: The visible profiles of the cut cylinders on the cut-plane.

Because the cutting direction is the same as the orientation of the cylinders, the profiles of the cut cylinders are observed as rectangles on the cut-plane. Since the rectangles are observable, the information about the heights and half-widths of these rectangles is known. These heights and half-widths of the rectangles relate naturally to the heights and radii of the corresponding cylinders. Since the steel object is cut parallel to the cylinder axes, the height of the cylinder is equal to the height of the rectangle, so both the height of the cylinder and the height of the rectangle will be denoted by H . The relationship between the squared radius X of the cylinder and the squared half-width Z of the rectangle is slightly more complicated. The width, and therefore the squared half-width, of a rectangle depends on precisely where the cylinder is cut. The width of the rectangle will be maximal when the cylinder is cut precisely in its centre. In that case, the squared half-width is equal to the squared radius. If the cylinder is cut at a different location, the squared half-width of the corresponding rectangle will be smaller than the

squared radius of that cylinder. However, the location where the cylinder is cut is not known, so in general it is only known that the squared half-width is less than or equal to the squared radius, or in notation: $Z \leq X$. The relationship between Z and X can be made more precise by considering the top view of a cut cylinder, shown in Figure 4.

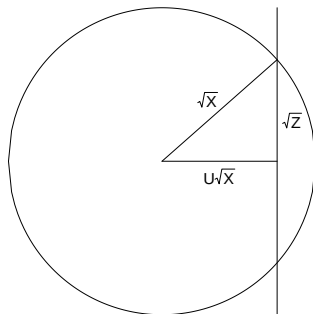


Figure 4: Top view of a cylinder that is cut by the cut-plane.

Due to symmetry, it can be assumed that the cylinder was cut in the right half of the circle in the figure above. Since the cylinders are randomly distributed within the block, the location of the cut is uniformly distributed on this right half. The distance of the centre of the circle to the plane can thus be expressed as $U\sqrt{X}$, with $\sqrt{X}$ the radius of the circle and $U \sim \text{Unif}(0, 1)$, i.e. U is a uniform random variable on the interval $(0, 1)$. Furthermore, the half-width $\sqrt{Z}$ of the corresponding rectangle completes the right-angled triangle shown in the figure above. Then the squared half-width of the observed rectangle can be expressed in terms of X and U by using Pythagoras' method:

$$Z = (\sqrt{Z})^2 = (\sqrt{X})^2 - (U\sqrt{X})^2 = (1 - U^2)X. \quad (1)$$

This relationship will be used in the next section, where the relationship between the density of the height and squared radius of a cylinder and the density of the height and squared half-width of a rectangle will be established.

2.2 Important distributions and their relationship

Since the width and height of a rectangle can be obtained from the cut-plane, full information about the rectangles is known. As a result, an estimator for the bivariate density function of the height and squared half-width of a rectangle, denoted by g , can be constructed. However, there is no full information available about the cylinders, since only a lower bound for the radius is known, so an estimator for the bivariate density function of the height and squared radius of a cylinder, denoted by f , cannot be directly constructed. The relationship between the two above-mentioned densities established in [4] makes it possible to construct an estimator for f based on an estimator for g . This relationship will be derived in the remaining of this section.

Whether a cylinder in the big cubic block of size M is cut by the plane, depends on the location of the centre of the cylinder within that block. Consider the block of size $M = 1$ of the previous section, where the block is cut at $x = 0.5$. Figure 5 below shows the cut-plane at that location in black:

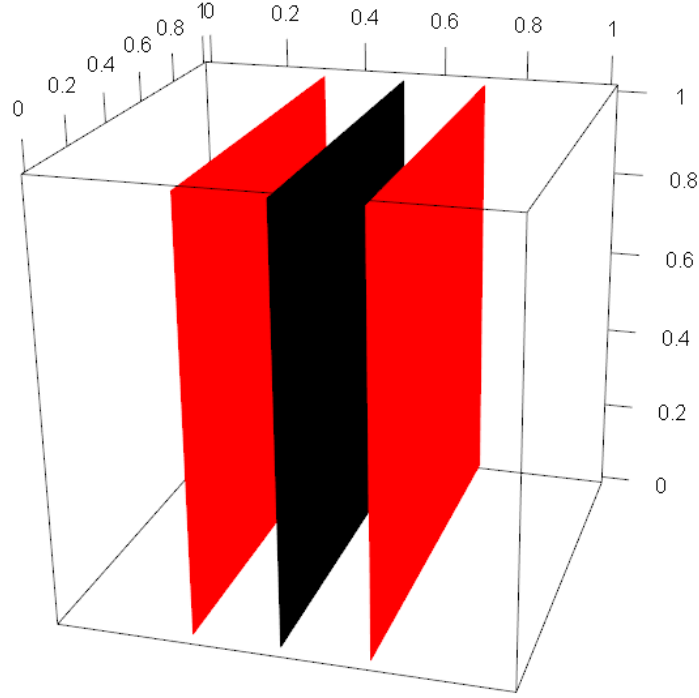


Figure 5: Cubic block of size $M = 1$, cut at $x = 0.5$.

In this block, a cylinder is then cut by this black plane, if the distance of the centre of this cylinder to the cut-plane is less than or equal to the radius. This means that a cylinder with radius equal to 0.2 is cut if its centre lies between the two red planes located at $x = 0.3$ and $x = 0.7$. In general, given that $X = x$ and $H = h$, a cylinder is cut if its centre lies between the two planes parallel to the cut-plane, with a distance of the radius $\sqrt{x}$ of the cylinder to that cut-plane. Given that $X = x$ and $H = h$, the probability that a cylinder is cut by the cut-plane is therefore equal to the probability that the centre of the cylinder lies between the two above-mentioned planes. Since the centre of a cylinder is distributed randomly within the block, this probability is equal to the ratio of the volume between these two planes to the total volume of the block:

$$P(\text{cylinder is cut} | X \in dx, H \in dh) = \frac{\text{Volume between the two planes}}{\text{Total volume of the block}} = \frac{2\sqrt{x}M^2}{M^3} = \frac{2\sqrt{x}}{M}.$$

Note that the condition $\{X \in dx, H \in dh\}$ is used instead of the condition $\{X = x, H = h\}$. Since X and H are continuous random variables, the probability of the latter event is equal to zero. The former event is an abbreviation for $\{X \in (x, x + dx), H \in (h, h + dh)\}$, where dx and dh are infinitesimally small numbers. Since this event is the continuous equivalent of $\{X = x, H = h\}$, the probability is conditioned on this event. Moreover, the probability does not depend on the value of H .

From the definition of conditional probabilities it follows that

$$\begin{aligned}
P(X \leq x, H \leq h | \text{cylinder is cut}) &= \frac{P(X \leq x, H \leq h, \text{cylinder is cut})}{P(\text{cylinder is cut})} \\
&= \frac{\int_{y=0}^x \int_{m=0}^h P(X \in dy, H \in dm, \text{cylinder is cut})}{\int_{y=0}^{\infty} \int_{m=0}^{\infty} P(X \in dy, H \in dm, \text{cylinder is cut})} \\
&= \frac{\int_{y=0}^x \int_{m=0}^h P(\text{cylinder is cut} | X \in dy, H \in dm) P(X \in dy, H \in dm)}{\int_{y=0}^{\infty} \int_{m=0}^{\infty} P(\text{cylinder is cut} | X \in dy, H \in dm) P(X \in dy, H \in dm)} \\
&= \frac{\int_{y=0}^x \int_{m=0}^h \frac{2\sqrt{y}}{M} f(y, m) dm dy}{\int_{y=0}^{\infty} \int_{m=0}^{\infty} \frac{2\sqrt{y}}{M} f(y, m) dm dy} \\
&= \frac{\int_{y=0}^x \int_{m=0}^h \sqrt{y} f(y, m) dm dy}{\int_{y=0}^{\infty} \int_{m=0}^{\infty} \sqrt{y} f(y, m) dm dy} \\
&= \frac{1}{E[\sqrt{X}]} \int_{y=0}^x \int_{m=0}^h \sqrt{y} f(y, m) dm dy.
\end{aligned}$$

In contrast to the previous probability, this probability does not depend on the size M of the block. Differentiation with respect to both x and h of the expression above yield

$$f^w(x, h) = \frac{\partial^2}{\partial x \partial h} P(X \leq x, H \leq h | \text{cylinder is cut}) = \frac{\sqrt{x}}{E[\sqrt{X}]} f(x, h),$$

where f^w is the density function f weighted by the ratio of the radius of the cylinder to the expected radius. Furthermore, from (1) it is known that $Z = (1 - U^2)X$, where $U \sim \text{Unif}(0, 1)$. The cumulative distribution function F_V of the random variable $V = 1 - U^2$ is then also known, since

$$F_V(v) = P(1 - U^2 \leq v) = P(U^2 \geq 1 - v) = P(U \geq \sqrt{1 - v}) = 1 - \sqrt{1 - v}, \quad \text{for } v \in [0, 1].$$

By differentiating this expression with respect to v , the probability density function f_V of V is obtained:

$$f_V(v) = \frac{d}{dv} F_V(v) = \frac{1}{2}(1 - v)^{-\frac{1}{2}}, \quad \text{for } v \in (0, 1),$$

which is the density of the $Beta(1, 1/2)$ distribution. The distribution of Z and H can then be derived via the distribution of X and H . Because cylinders with a larger radius have a higher probability of being cut, the observations of the rectangles on the cut-plane are size-biased. Due to this fact, the weighted density f^w is used in the following computations. The cumulative distribution function G of Z and H can be expressed as

$$\begin{aligned}
G(z, h) &= P(Z \leq z, H \leq h) = P(VX \leq z, H \leq h) \\
&= \int_0^{\infty} \int_0^h \int_0^{\frac{z}{x}} f^w(x, y) f_V(v) dv dy dx = \int_0^{\infty} \int_0^h \int_0^z \frac{1}{x} f^w(x, y) f_V\left(\frac{s}{x}\right) ds dy dx \\
&= \int_0^h \int_0^z \int_0^{\infty} \frac{1}{x} f^w(x, y) f_V\left(\frac{s}{x}\right) dx ds dy.
\end{aligned}$$

The density g is then obtained by differentiating the cumulative distribution function with respect to z and h :

$$\begin{aligned} g(z, h) &= \frac{\partial^2}{\partial z \partial h} G(z, h) = \int_0^\infty \frac{1}{x} f^w(x, h) f_V\left(\frac{z}{x}\right) dx = \int_z^\infty \frac{1}{x} \frac{\sqrt{x}}{E[\sqrt{X}]} f(x, y) \frac{1}{2} \left(1 - \frac{z}{x}\right)^{-\frac{1}{2}} dx \\ &= \frac{1}{2E[\sqrt{X}]} \int_z^\infty x^{-\frac{1}{2}} \left(1 - \frac{z}{x}\right)^{-\frac{1}{2}} f(x, h) dx = \frac{1}{2E[\sqrt{X}]} \int_z^\infty (x - z)^{-\frac{1}{2}} f(x, h) dx. \end{aligned}$$

Here it is also used that f_V is zero outside its support $(0, 1)$. The result is an expression for the density g in terms of the density f . However, since the random variable X is not observable, the density f can not be estimated directly. On the contrary, both Z and H are observable variables, which makes it possible to estimate the density g . The inverse relation of the relation above is therefore of more interest. To find the expression for f in terms of g , consider the following expression, depending on x and h :

$$\begin{aligned} \int_x^\infty (z - x)^{-\frac{1}{2}} g(z, h) dz &= \int_x^\infty (z - x)^{-\frac{1}{2}} \frac{1}{2E[\sqrt{X}]} \int_z^\infty (y - z)^{-\frac{1}{2}} f(y, h) dy dz \\ &= \frac{1}{2E[\sqrt{X}]} \int_x^\infty \int_z^\infty (z - x)^{-\frac{1}{2}} (y - z)^{-\frac{1}{2}} f(y, h) dy dz \\ &= \frac{1}{2E[\sqrt{X}]} \int_x^\infty f(y, h) \int_x^y (z - x)^{-\frac{1}{2}} (y - z)^{-\frac{1}{2}} dz dy \\ &= \frac{1}{2E[\sqrt{X}]} \int_x^\infty f(y, h) \int_0^1 ((y - x)u)^{-\frac{1}{2}} ((y - x)(1 - u))^{-\frac{1}{2}} (y - x) du dy \\ &= \frac{1}{2E[\sqrt{X}]} \int_x^\infty f(y, h) \int_0^1 u^{-\frac{1}{2}} (1 - u)^{-\frac{1}{2}} du dy \\ &= \frac{1}{2E[\sqrt{X}]} \int_x^\infty f(y, h) B\left(\frac{1}{2}, \frac{1}{2}\right) dy \\ &= \frac{1}{2E[\sqrt{X}]} \int_x^\infty f(y, h) \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(1)} dy \\ &= \frac{1}{2E[\sqrt{X}]} \int_x^\infty f(y, h) \frac{\sqrt{\pi}\sqrt{\pi}}{1!} dy \\ &= \frac{\pi}{2E[\sqrt{X}]} \int_x^\infty f(y, h) dy, \end{aligned} \tag{2}$$

where B is the Beta function and Γ the Gamma function, i.e. $B(a, b) = \int_0^1 t^{a-1} (1 - t)^{b-1} dt$ and

$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$. The partial derivative with respect to x of the above equation is equal to

$$\frac{\partial}{\partial x} \int_x^\infty (z-x)^{-\frac{1}{2}} g(z, h) dz = \frac{\pi}{2E[\sqrt{X}]} \frac{\partial}{\partial x} \int_x^\infty f(y, h) dy = -\frac{\pi}{2E[\sqrt{X}]} f(x, h).$$

Multiplication by $-E[Z^{-\frac{1}{2}}]^{-1}$ results in:

$$-\frac{1}{E[Z^{-\frac{1}{2}}]} \frac{\partial}{\partial x} \int_x^\infty (z-x)^{-\frac{1}{2}} g(z, h) dz = \frac{\pi}{2E[Z^{-\frac{1}{2}}]E[\sqrt{X}]} f(x, h). \quad (3)$$

The previous result (2), with $x = 0$, can be used to express $E[Z^{-\frac{1}{2}}]$ in terms of $E[\sqrt{X}]$:

$$E[Z^{-\frac{1}{2}}] = \int_0^\infty \int_0^\infty z^{\frac{1}{2}} g(z, h) dz dh = \frac{\pi}{2E[\sqrt{X}]} \int_0^\infty \int_0^\infty f(y, h) dy dh = \frac{\pi}{2E[\sqrt{X}]},$$

or written differently:

$$\frac{\pi}{2E[Z^{-\frac{1}{2}}]E[\sqrt{X}]} = 1.$$

By plugging this result into the expression (3), the final relationship between f and g is established:

$$f(x, h) = -\frac{1}{E[Z^{-\frac{1}{2}}]} \frac{\partial}{\partial x} \int_x^\infty (z-x)^{-\frac{1}{2}} g(z, h) dz. \quad (4)$$

An estimator for f can then be constructed by making use of this relationship. Based on data of heights H and squared half-widths Z of the rectangles observed on the cut-plane, an estimator for the bivariate probability density function g will be constructed. Similar to the solution of [5] to the Corpuscle problem, where univariate kernel density estimation is used to construct an estimator for the probability density function of the radii of spheres in a medium, given the radii of their visible profiles in a random cross-section, it is chosen to use a bivariate kernel density estimator to estimate the density g . The kernel density estimation method will be discussed in detail in the next chapter. Furthermore, the expected value in (4) can be estimated with the data of Z only. Then the constructed estimator for g can be transformed into an estimator for f with the relationship above, which will be done in Chapter 4.

3 Kernel Density Estimation

In this chapter, the method of kernel density estimation will be discussed. In the first section the general setting for univariate kernel density estimation will be introduced, while in Section 3.2 bivariate kernel density estimation will be discussed. In both sections the choice of the bandwidth parameter will be treated and an example will be given. In Section 3.3, the boundary problem that occurs in the specific problem of this thesis will be discussed and a solution will be given. In the last section of this chapter the kernel density estimator for the bivariate density of the height and squared half-width of a rectangle will be constructed for the underlying density of an example data set.

3.1 Univariate kernel density estimation

In this section, the univariate kernel density estimator, which is extensively discussed by Silverman [6] and Wand and Jones [7] for example, will be introduced. The kernel function and the bandwidth parameter, which will be introduced in the first section, form the basis of the kernel density estimator. Different kernel functions will be discussed and compared, together with the influence of the bandwidth parameter on the scaled kernel functions. After that, the univariate kernel density estimator will be constructed, and a simple example of such an estimator will be given. In Section 3.1.2, the bandwidth parameter selection method that is used to obtain an optimal value of the bandwidth parameter will be discussed. In Section 3.1.3, the — for the previously discussed bandwidth selection method — optimal univariate kernel density estimator will be constructed for a larger data set, and its performance will be graphically investigated.

3.1.1 Univariate kernel density estimator

Kernel density estimation is a non-parametric estimation method to estimate the probability density function of a random variable. This method is related to histograms, but kernel density estimation has the main advantage that, when using a smooth and continuous kernel function, the resulting estimators are smooth and continuous. A kernel function is a weighting function that can be used to spread mass over an area around a data point. The mass given to a point depends on the chosen kernel function and on the distance of that point to the data point. When the uniform kernel is chosen, the mass is equal for all points within a certain area, and the mass is zero for points outside that area. When a different kernel is used, the mass given to a point within the area is bigger for points closer to the data point. Furthermore, the mass given to a point also depends on the bandwidth parameter: the bigger the bandwidth parameter is, the wider the mass is spread around the data point. Often, since the kernel function is a weighting function, the function is chosen to be a density, which ensures that the total mass that is given by the kernel function is equal to 1. A result of this choice is that the corresponding kernel density estimator will then be a density function as well.

First different kernel functions will be compared. Consider the following six kernel functions K , which are the most commonly used kernel functions:

Uniform:	$K(x) = \frac{1}{2} \mathbb{1}_{\{ x \leq 1\}}$
Triangular:	$K(x) = (1 - x) \mathbb{1}_{\{ x \leq 1\}}$
Epanechnikov:	$K(x) = \frac{3}{4} (1 - x^2) \mathbb{1}_{\{ x \leq 1\}}$
Biweight:	$K(x) = \frac{15}{16} (1 - x^2)^2 \mathbb{1}_{\{ x \leq 1\}}$
Triweight:	$K(x) = \frac{35}{32} (1 - x^2)^3 \mathbb{1}_{\{ x \leq 1\}}$
Normal:	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

Note that for all these kernels $\int_{-\infty}^{\infty} K(x) dx = 1$, that all kernels except for the first are continuous on $\mathbb{R}$, and that only the last three kernels are smooth on $\mathbb{R}$. A plot of these different kernels can be found in Figure 6.

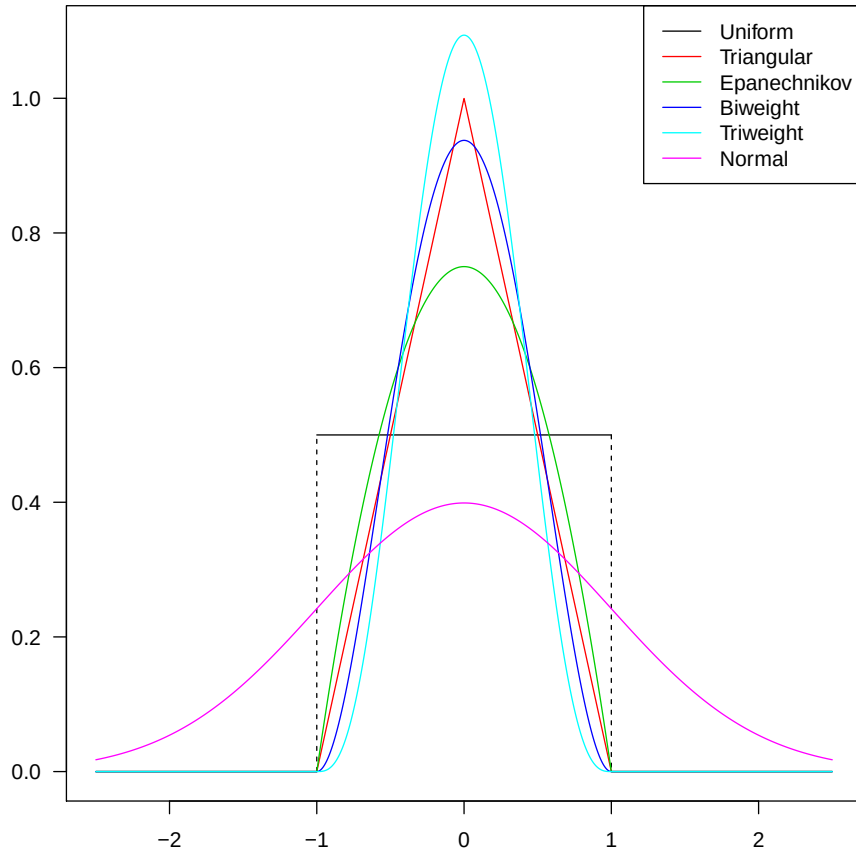


Figure 6: Different kernel functions.

In contrast to the Uniform kernel, the Triangular, Epanechnikov, Biweight and Triweight kernel give more weight to points closer to the data point. This also holds for the Normal kernel, but this kernel gives nonzero weight to all points, while the other kernels are restricted to the finite support $[-1, 1]$. Although the Biweight and Triweight kernel functions are also smooth kernels, the Normal kernel is the most used kernel function.

Besides that the kernel density estimator depends on the choice of the kernel function, it also depends on the choice of the bandwidth parameter $b > 0$. Moreover, it is known from literature that the choice of the bandwidth parameter is even more important than the choice of kernel function, see [8] for example. In the remainder of this section, the Normal kernel is chosen as kernel function. For the construction of the kernel density estimator, the kernel function is scaled on the basis of the bandwidth parameter, where the scaled kernel function K_b is defined as:

$$K_b(x) = \frac{1}{b} K\left(\frac{x}{b}\right).$$

Note that for $b = 1$, the scaled kernel function is equal to the kernel function itself. The influence of the bandwidth parameter on the scaled kernel function can be observed in Figure 7, which shows the scaled Normal kernel for different values of the bandwidth parameter.

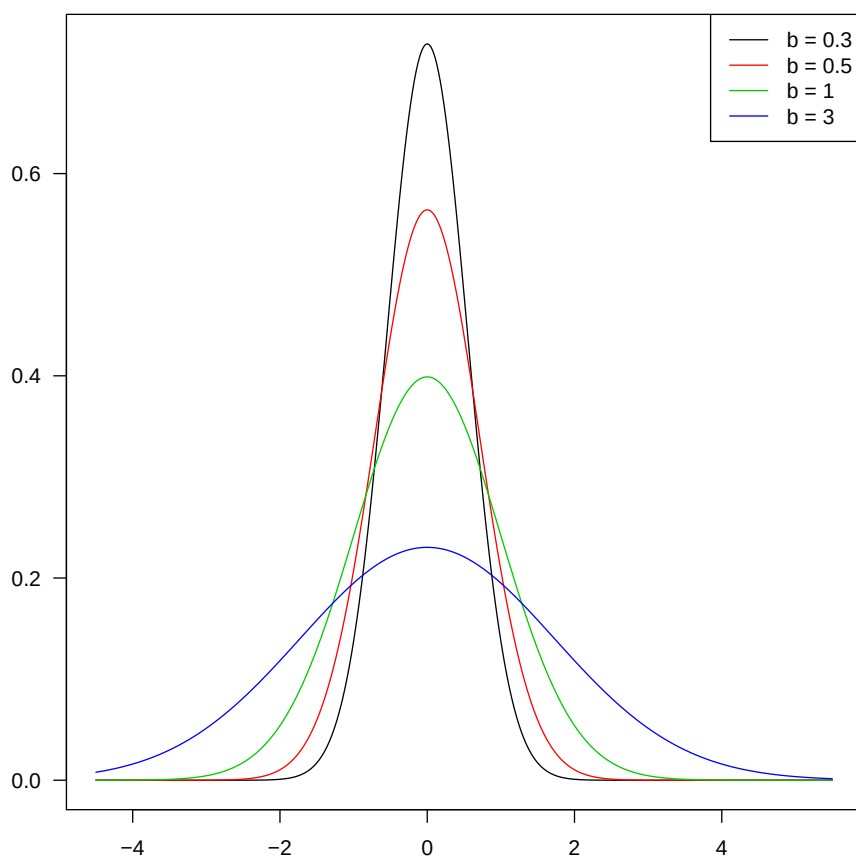


Figure 7: Scaled Normal kernel for different values of the bandwidth parameter.

It is clear that a bigger bandwidth parameter results in a more levelled kernel function, since the mass is then more equally spread over the whole real line. The effect on the other above-mentioned kernels is similar, with the main difference for kernels that give nonzero mass to a finite interval, that this interval changes. When the support of the kernel function is $[-1, 1]$, the

support of the corresponding scaled kernel function with bandwidth parameter b is $[-b, b]$.

For constructing the kernel density estimator, these scaled kernel functions are centered around the data points. If x_i is an observed data point, then the corresponding centered scaled kernel function will be $K_b(x - x_i)$. The final kernel density estimator is then obtained by taking the sum of the centered scaled kernel functions of all data points divided by the number of data points. Then the definition of a univariate kernel density estimator can be stated:

Let X be a random variable with unknown density function f , and let $x_1, \dots, x_n$ be n observed realisations of X . Let K be a kernel function and $b > 0$ the bandwidth parameter. The univariate kernel density estimator $\hat{f}$ for the density f is then defined by:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_b(x - x_i) = \frac{1}{nb} \sum_{i=1}^n K\left(\frac{x - x_i}{b}\right).$$

Although $\hat{f}$ depends on the choice of K and b , it is chosen to omit this dependence in the notation of the estimator. Since each kernel function is smoother for a bigger bandwidth parameter, also the kernel density estimator will be smoother when the value of the bandwidth parameter is bigger. This statement will be strengthened with an example:

Let X again be a random variable with unknown density f , and let $x_1, \dots, x_8$ be realisations of X , with:

$$(x_1, \dots, x_8) = (0.1, 0.2, 0.3, 0.6, 0.6, 0.7, 1.0, 1.2). \quad (5)$$

Based on this data, the density function of X can be estimated with histogram estimation and kernel density estimation. Other methods to estimate this density are available as well, but will not be discussed in this thesis. For both methods choices have to be made to construct the estimator. For histogram estimation the breakpoints have to be chosen. Even for equidistant breakpoints, which means that each bin has the same width, but a different starting point the resulting histogram estimator can be very different. To construct a kernel density estimator, the kernel function and a bandwidth parameter have to be chosen. The figure below shows two examples of a histogram estimator and two examples of a kernel density estimator for different choices of the parameters.

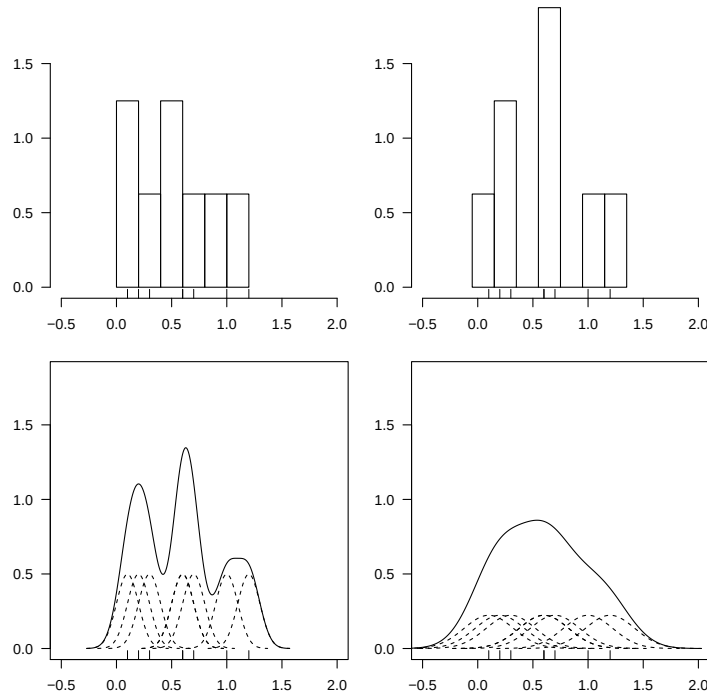


Figure 8: Top: Histogram estimators. Bottom: Kernel density estimators.

The top two plots of Figure 8 give examples of histogram estimators based on the data (5), where the data is shown as small vertical lines. Both histograms have bins of width 0.2, but the left histogram starts at 0, while the right histogram starts at -0.05. Although this difference is relatively small, the resulting histograms are very different. The bottom two plots of Figure 8 give examples of kernel density estimators for the same data set. Both estimators are constructed with Normal kernels, where the bandwidth parameter is equal to 0.01 for the left estimator and equal to 0.05 for the right estimator. In both plots, the individual kernel functions are shown with dotted lines. Also for this estimator, the choice of bandwidth parameter has a considerable influence on the resulting estimator.

From the figure it is clear that, with the same data set, very different kernel density estimators can be constructed by changing the bandwidth parameter. It is therefore important for kernel density estimation that the bandwidth parameter is chosen carefully. In Section 3.1.2 the method to find an “optimal bandwidth” will be discussed.

3.1.2 Bandwidth parameter selection

First note that there is no general “optimal bandwidth”, as the optimality of the bandwidth parameter depends on the measure on which the optimality criterion is based. In this thesis, the bandwidth parameter is chosen based on the Least Squares Cross-Validation (*LSCV*) method as described in [9], where it is also shown that the obtained bandwidth is asymptotically optimal for minimizing the Mean Integrated Squared Error (*MISE*), and that the optimal rate of b is equal to $O(n^{-\frac{1}{5}})$. This method is based on the minimization of the Integrated Squared Error (*ISE*) of the kernel density estimator $\hat{f}$. This quantity is defined as a function of the bandwidth

parameter b in the following way:

$$\begin{aligned} ISE(b) &= \int_{-\infty}^{\infty} \left(\hat{f}(x) - f(x) \right)^2 dx \\ &= \int_{-\infty}^{\infty} \hat{f}(x)^2 dx - 2 \int_{-\infty}^{\infty} \hat{f}(x) f(x) dx + \int_{-\infty}^{\infty} f(x)^2 dx. \end{aligned}$$

Although omitted in the notation, remember that $\hat{f}$ depends on the bandwidth parameter b , and the true density f does not. The last term of the right hand side does therefore not depend on the bandwidth parameter, so the minimization amounts to minimizing the sum of the first two terms. Instead of ISE , the following expression is minimized:

$$ISE(b) - \int_{-\infty}^{\infty} f(x)^2 dx = \int_{-\infty}^{\infty} \hat{f}(x)^2 dx - 2 \int_{-\infty}^{\infty} \hat{f}(x) f(x) dx. \quad (6)$$

Note that the second term of the right hand side depends on the unknown density function f , so this term cannot be computed. In the *LSCV* method this term is therefore estimated, which is done as follows: Let $X, X_1, \dots, X_n \sim f$ be independent, and let $\hat{f}$ be the kernel density estimator based on $X_1, \dots, X_n$. Since then $\hat{f}$ does not depend on X , it is known that $E_f[\hat{f}(X)|X_1, \dots, X_n] = \int_{-\infty}^{\infty} \hat{f}(x) f(x) dx$. The last term of (6) can thus be estimated by estimating this expected value. Let $\hat{f}_{-i}$ denote the leave-one-out kernel density estimator, which is the estimator based on the whole data set $X_1, \dots, X_n$ except for the i th observation, so that the estimator is independent of X_i . Let $x_1, \dots, x_n$ be observed realisations of $X_1, \dots, X_n$, then the expected value can be estimated as follows:

$$\begin{aligned} E_f[\widehat{\hat{f}(X)}|x_1, \dots, x_n] &= \frac{1}{n} \sum_{i=1}^n \hat{f}_{-i}(x_i) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n-1} \sum_{j=1, j \neq i}^n K_b(x_i - x_j) \right) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_b(x_i - x_j). \end{aligned}$$

By plugging this estimator into (6), an estimator, denoted by CV , for the function that has to be minimized is obtained. This estimate is thus defined as

$$\begin{aligned} CV(b) &= \int_{-\infty}^{\infty} \hat{f}(x)^2 dx - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_b(x_i - x_j) \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{n} \sum_{i=1}^n K_b(x - x_i) \right)^2 dx - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_b(x_i - x_j). \end{aligned}$$

The bandwidth that minimizes this function is the optimal bandwidth $\hat{b}$ with respect to the *LSCV* method, so:

$$\hat{b} = \underset{b>0}{\operatorname{argmin}} \{ CV(b) \}.$$

In the next section, this bandwidth selection method is used to find the *LSCV*-optimal kernel density estimator for a particular data set. This data set is drawn from a known distribution, so the estimator can be compared to the true density.

3.1.3 Example of a univariate kernel density estimator

Let $x_1, \dots, x_{100}$ be 100 realisations of the $Norm(1, 2)$ distribution, i.e. the Normal distribution with mean 1 and variance 2. The quantity $CV(b)$ can then be computed for every b , where the integral of the first term of CV is approximated with numerical integration using the Trapezoidal Rule (see Section 4.1.1). The error made in the approximation of this integral is $O(b_T^2)$, where b_T is the step size in the numerical integration. By choosing this step size small enough, the numerical approximation of the integral is accurate enough for approximating the bandwidth parameter up to two decimals. For $b = 0.01, 0.02, \dots, 2$ the value of $CV(b)$ is computed, and the result is shown in Figure 9.

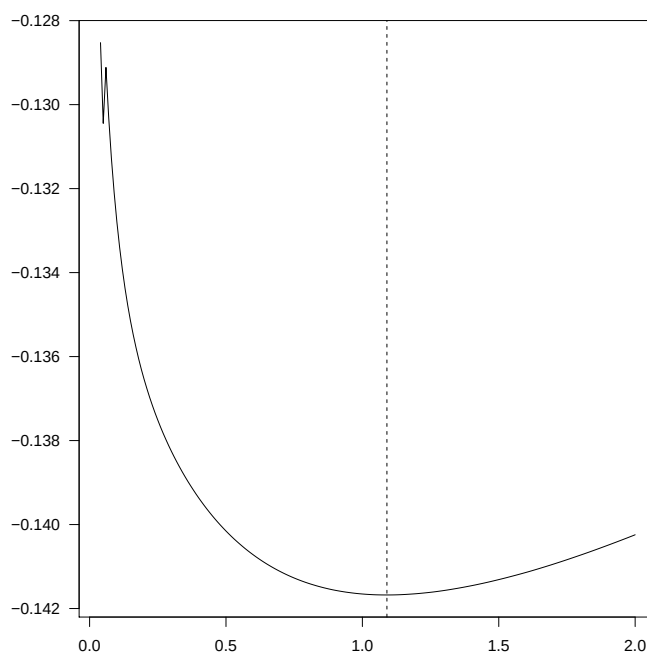


Figure 9: The function values of CV for a sample of size 100 of the $Norm(1, 2)$ distribution.

From the computations it follows that the function CV for this data set has its minimum in $\hat{b} = 1.09$. As can be seen from the figure, the function has a local minimum as well, but the interest lies in the global minimum. Furthermore it can be seen that, in this example, a small change in the bandwidth parameter near its minimum does not change the value of CV very much. As a result, the estimator will not become much better by approximating the minimum even more accurate. It is possible, though, that for other data sets this small change results in a bigger change in the value of CV and therefore in a bigger change in the accuracy of the estimator, or, the other way around, that even a big change in the value of the bandwidth parameter has only a small effect on the value of the function CV and thus only a small effect on the accuracy of the estimator.

The *LSCV*-optimal kernel density estimator, with Normal kernels as kernel function, can then be constructed with this optimal bandwidth parameter. This estimator is plotted together with the true density $\phi_{1,2}$ of the *Norm*(1, 2) distribution, defined as

$$\phi_{1,2}(x) = \frac{1}{2\sqrt{\pi}} e^{-\frac{(x-1)^2}{4}}, \quad x \in \mathbb{R},$$

in Figure 10.

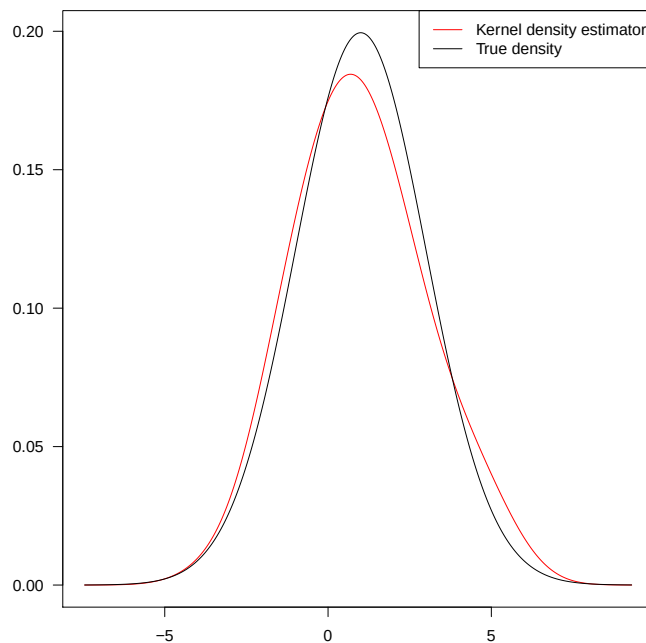


Figure 10: The *LSCV*-optimal kernel density estimator together with the true density.

The kernel density estimator is close to the true density, but slightly positively skewed. This is a consequence of the fact that the average of the data is smaller than the true mean. For a bigger data set, this effect will become smaller, which will result in a better estimator for the density.

Using the above-described method to obtain an optimal value of the bandwidth parameter, one can find the kernel density estimator, for a univariate data set, that is optimal in the sense that it minimizes the estimated Integrated Squared Error. However, the data set that is considered in this thesis is a bivariate data set, consisting of the heights and squared radii of observed rectangles. This means that each observation is a data pair (x_i, y_i) instead of a single observation x_i . As a consequence, the method described above has to be adjusted to be able to estimate a bivariate density. In the next section, the bivariate kernel density estimator will be introduced. This estimator is able to estimate a bivariate density based on bivariate observations. Similar to Section 3.1.2, an optimal bandwidth can be found, to optimize the bivariate kernel density estimator.

3.2 Bivariate kernel density estimation

The structure of this section is similar to that of the previous section. First, the bivariate kernel density estimator will be introduced, and the properties of the kernel functions and bandwidth parameter will be discussed. In Section 3.2.2, a similar method as before will be used to obtain an optimal bandwidth. In the last section, an example of a bivariate kernel density estimator will be given.

3.2.1 Bivariate kernel density estimator

Compared to univariate kernel density estimation, two differences occur in bivariate kernel density estimation. First, the kernel function is a bivariate function, instead of a univariate one. Second, the bandwidth parameter translates in the second dimension to a two by two bandwidth matrix B . The definition of a bivariate kernel density estimator can then be stated:

Let $\mathbf{X}$ be a bivariate random variable with unknown bivariate density f , and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be n observed realisations of $\mathbf{X}$. Let B be a symmetric and positive definite two by two bandwidth matrix, K a bivariate kernel function, and K_B the corresponding scaled kernel function. The bivariate kernel density estimator $\hat{f}$ for the density f , based on the observed realisation $\mathbf{x}_1, \dots, \mathbf{x}_n$, is then defined by:

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n K_B(\mathbf{x} - \mathbf{x}_i) = \frac{1}{n} |B|^{-1} \sum_{i=1}^n K(B^{-1}(\mathbf{x} - \mathbf{x}_i)),$$

where $|B|$ is the determinant of B . Note that, since B is a symmetrical and positive definite matrix, $|B| \neq 0$.

Similar to univariate kernel density estimation, different kernel functions can be used. Although the density function of every bivariate distribution can be used as a kernel function, the bivariate Normal density is again the most commonly used, and will thus be considered in the remainder of this section. Regarding the bandwidth, there are more options than in the univariate situation. The simplest choice for the bandwidth matrix is the identity matrix multiplied by a constant. Other possible choices are a diagonal matrix with different diagonal elements or a full matrix. The contour plots of Normal kernel functions with these different kinds of bandwidth matrices are shown in Figure 11. The interpretation of the different options are discussed below this figure.

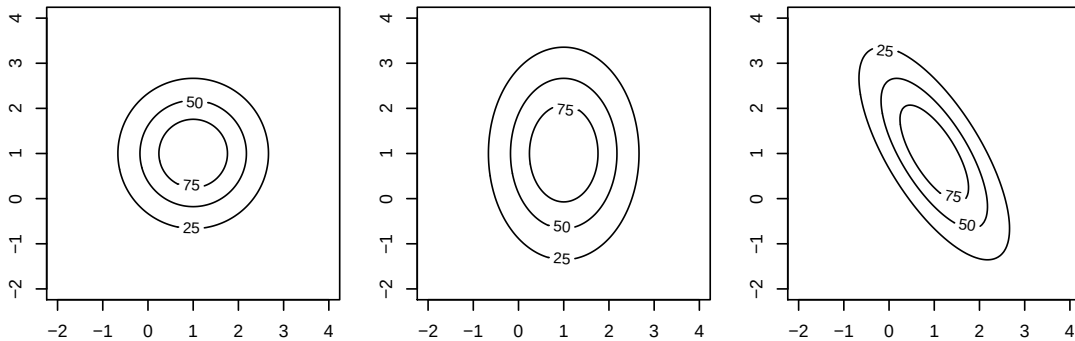


Figure 11: Contour plots of the scaled bivariate Normal kernel with different bandwidth matrices.

The different bandwidth matrices that were used are

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix},$$

where B_1 is used in the left plot, B_2 in the middle plot and B_3 in the right plot. The diagonal elements of the bandwidth matrix can be seen as the individual bandwidth parameters for the different variables, which are X and Y if $\mathbf{X} = (X, Y)^T$. When the bandwidth matrix is chosen to be (a constant times) the identity matrix, like B_1 , the bandwidths are the same in both the x - and y -direction. This results in circular contour lines, as can be seen in the left plot of Figure 11. This means that the amount of smoothing is the same in each direction. However, one variable is often spread more than the other, which indicates that a different amount of smoothing for different variables is often a better choice. The contour plot of the Normal kernel after changing the diagonal element of the bandwidth matrix corresponding to Y to 2, is shown in the middle plot. In this plot, the contour lines are ellipsoidal, with its orthogonal axes parallel to the X and Y axes. Taking a full matrix as bandwidth matrix gives even more freedom, since the ellipse can then be turned. Its axes are no longer parallel to the X and Y axes, so the mass given by the kernel function is oriented in a specific direction. The contour plot in this case is shown in the right plot. This last result can also be achieved by considering a joint kernel for the two variables in combination with a diagonal bandwidth matrix.

In this thesis, the diagonal matrices with different diagonal elements as bandwidth matrix will be used for the construction of the bivariate kernel density estimator. Moreover, a product kernel is used as kernel function, which means that the bivariate kernel function K can be written as a product of two univariate kernels K_1 and K_2 : $K(x, y) = K_1(x)K_2(y)$. Note that this does not mean that the two random variables X and Y are assumed to be independent. With both these restrictions, there is a loss of generality in the sense that the orientation of the kernel function cannot be changed. The considered bivariate kernel density estimator with bandwidth matrix

$$B = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}$$

and $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ can then be expressed as:

$$\begin{aligned} \hat{f}(\mathbf{x}) &= \frac{1}{n} |B|^{-1} \sum_{i=1}^n K(B^{-1}(\mathbf{x} - \mathbf{x}_i)) = \frac{1}{nb_1b_2} \sum_{i=1}^n K\left(\begin{pmatrix} b_1^{-1} & 0 \\ 0 & b_2^{-1} \end{pmatrix} \begin{pmatrix} x - x_i \\ y - y_i \end{pmatrix}\right) \\ &= \frac{1}{nb_1b_2} \sum_{i=1}^n K\left(\begin{pmatrix} b_1^{-1}(x - x_i) \\ b_2^{-1}(y - y_i) \end{pmatrix}\right) = \frac{1}{nb_1b_2} \sum_{i=1}^n K_1\left(\frac{x - x_i}{b_1}\right) K_2\left(\frac{y - y_i}{b_2}\right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{b_1} K_1\left(\frac{x - x_i}{b_1}\right) \frac{1}{b_2} K_2\left(\frac{y - y_i}{b_2}\right) = \frac{1}{n} \sum_{i=1}^n K_{1,b_1}(x - x_i) K_{2,b_2}(y - y_i), \end{aligned}$$

where K_{1,b_1} and K_{2,b_2} are the scaled versions of the kernel functions K_1 and K_2 respectively. Note that these functions are ordinary univariate kernel functions. Moreover, from this expression, it is clear that the diagonal elements can be seen as the individual bandwidth parameters of the corresponding variable. Furthermore, the marginal densities of X and Y can be obtained from this expression. Since a product kernel is used, these marginals are equal to the univariate kernel density estimators.

Figure 12 shows both a perspective plot and a contour plot of the bivariate kernel density estimator, with a bivariate Normal kernel and the identity matrix of dimension 2 as bandwidth matrix, based on the data pairs $(x_1, y_1), \dots, (x_8, y_8)$, where $x_1, \dots, x_8$ are the same as in (5), and

$$(y_1, \dots, y_8) = (0.3, 0.3, 0.4, 0.6, 1.4, 2.5, 2.7, 2.8).$$

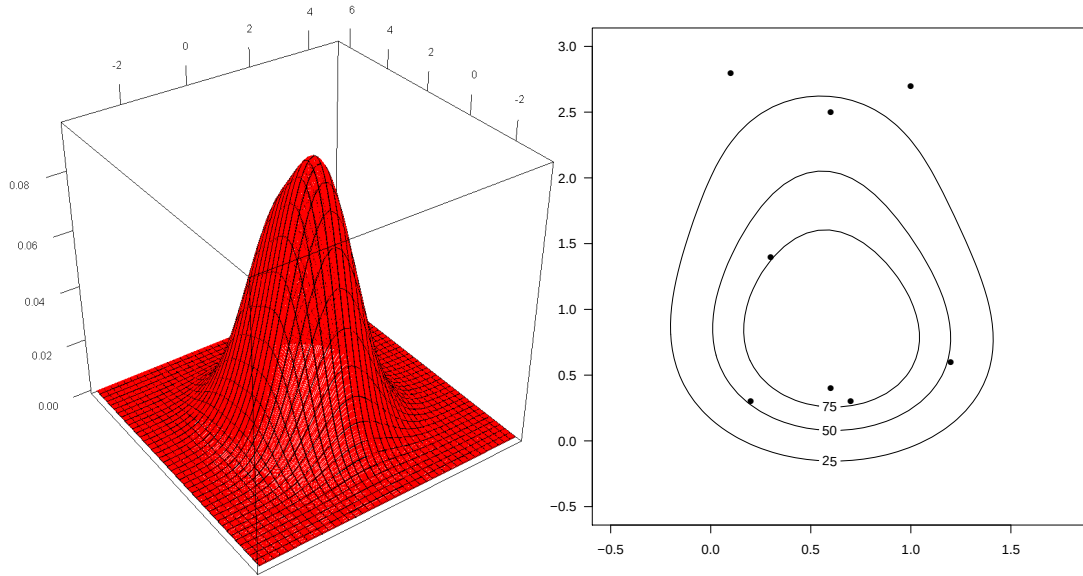


Figure 12: Perspective (left) and contour (right) plot of the bivariate kernel density estimator.

Also in the second dimension, the bandwidth is an important parameter. Compare the perspective plots of Figure 13 to the perspective plot of the figure above, where it should be noted that the axes in the different plots have a different scale. Instead of the identity matrix, the identity matrix multiplied by 0.05 and 5 are used to construct the bivariate kernel density estimator in the plots of the figure below. Although these estimators are based on the same data set, the resulting estimators are very different. The left estimator of Figure 13 is too rough, while the right estimator is too smooth.

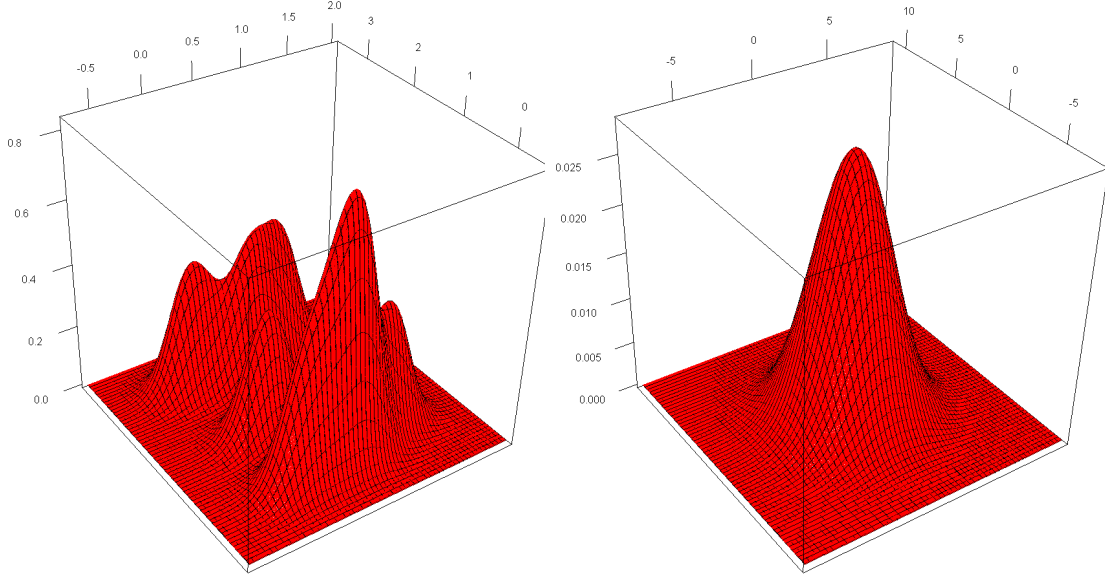


Figure 13: Perspective plot of kernel density estimator with identity matrix multiplied by 0.05 (left) and 5 (right) as bandwidth matrix.

The bandwidth matrix selection method for the bivariate kernel density estimation will be discussed in the next section. This method is similar to the bandwidth parameter selection method of Section 3.1.2, where for the bivariate estimator a bandwidth matrix has to be found. Since a diagonal bandwidth matrix is used, only the two diagonal elements of the bandwidth matrix have to be considered. The bandwidth matrix is therefore written as the vector (b_X, b_Y) , where b_X is the bandwidth parameter corresponding to the random variable X , and b_Y the bandwidth parameter corresponding to Y .

3.2.2 Bandwidth matrix selection

To compute the kernel density estimator in a concrete situation, the two diagonal elements of the bandwidth matrix have to be chosen. Analogous to the univariate bandwidth selection, an optimal choice for the bandwidth can be based on the Integrated Squared Error of the kernel density estimator. In this case, the *ISE* is a two dimensional function of b_X and b_Y :

$$\begin{aligned} ISE(b_X, b_Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\hat{f}(x, y) - f(x, y) \right)^2 dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x, y)^2 dx dy - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x, y) f(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)^2 dx dy. \end{aligned}$$

Similar as before, the last term does not depend on the bandwidth parameters, so the minimization amounts to minimizing the sum of the first two terms:

$$ISE(b_X, b_Y) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)^2 dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x, y)^2 dx dy - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x, y) f(x, y) dx dy. \quad (7)$$

Again, the second term of the right hand side depends on the unknown density function f , and therefore has to be estimated. In the *LSCV* method this term is estimated with the same method as in the univariate case: Let $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n) \sim f$ be independent, and $\hat{f}$ the bivariate kernel density estimator based on $(X_1, Y_1), \dots, (X_n, Y_n)$. Then the second term is equal to $E[\hat{f}(X, Y)|(X_1, Y_1), \dots, (X_n, Y_n)]$, since $\hat{f}$ is then independent of X . Let $(x_1, y_1), \dots, (x_n, y_n)$ be observed realisations of $(X_1, Y_1), \dots, (X_n, Y_n)$, and $\hat{f}_{-i}$ the leave-on-out bivariate kernel density estimator, then this term can be estimated by

$$\begin{aligned}\widehat{E}[\hat{f}(X, Y)|(x_1, y_1), \dots, (x_n, y_n)] &= \frac{1}{n} \sum_{i=1}^n \hat{f}_{-i}(x_i, y_i) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n-1} \sum_{j=1, j \neq i}^n \{K_{b_1}(x_i - x_j)K_{b_2}(y_i - y_j)\} \right) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n (K_{b_1}(x_i - x_j)K_{b_2}(y_i - y_j)).\end{aligned}$$

By plugging in this estimator into the previous equation, the function CV on $(0, \infty)^2$ is obtained:

$$CV(b_X, b_Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x, y)^2 dx dy - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n (K_{b_1}(x_i - x_j)K_{b_2}(y_i - y_j)).$$

The bandwidths $\hat{b}_X$ and $\hat{b}_Y$ that minimize this expression are the optimal bandwidths with respect to the *LSCV* method, so

$$(\hat{b}_X, \hat{b}_Y) = \underset{(b_X, b_Y) > 0}{\operatorname{argmin}} \{CV(b_X, b_Y)\}.$$

It is shown in [10] that this choice for the bandwidth parameter is asymptotically optimal in terms of the Mean Integrated Squared Error. In this case, the optimal rate of the bandwidths is $O(n^{-\frac{1}{6}})$, see [11] for example.

In the next section, the above-described method will be used to construct the *LSCV*-optimal bivariate kernel density estimator for the underlying density of a data set of which the true distribution is known. As a result, the constructed estimator can be compared to the true density.

3.2.3 Example of a bivariate kernel density estimator

Consider the data $(x_1, y_1), \dots, (x_n, y_n)$, for $n = 100$, where the x'_i 's are realisations of the random variable X , and the y'_i 's realisations of the random variable Y , where X and Y are both standard Normal and independent random variables. Analogous to the example for univariate kernel density estimation, the function values of CV are computed on a discrete grid. However, this grid is two-dimensional in this case, which results in much more work for the same accuracy. Therefore, it is chosen to approximate the bandwidths first on a coarse grid. After that, the bandwidths are approximated more accurately in a neighbourhood around the coarse approximated optimal bandwidths.

The coarse grid is chosen to be $\{0.1, 0.2, \dots, 2\}^2$, on which the CV is evaluated. The function values on this grid, linearly connected to each other, are shown in the left plot of Figure 14. The minimum of the function CV on this grid is found at $(0.5, 0.6)$, indicated by the black dot in the figure. After that, a better approximation to the $LSCV$ -optimal bandwidths can be found by minimizing the function CV over the grid $\{0.41, \dots, 0.59\} \times \{0.51, \dots, 0.69\}$. The function values on this grid, together with the minimum value of CV on this grid, are shown in the right plot of the figure below. Moreover, the contour plot of the function on this grid is shown in the plot below the two three-dimensional figures. The minimization over this finer grid results in the approximate optimal bandwidth parameters $\hat{b}_X = 0.50$ and $\hat{b}_Y = 0.56$.

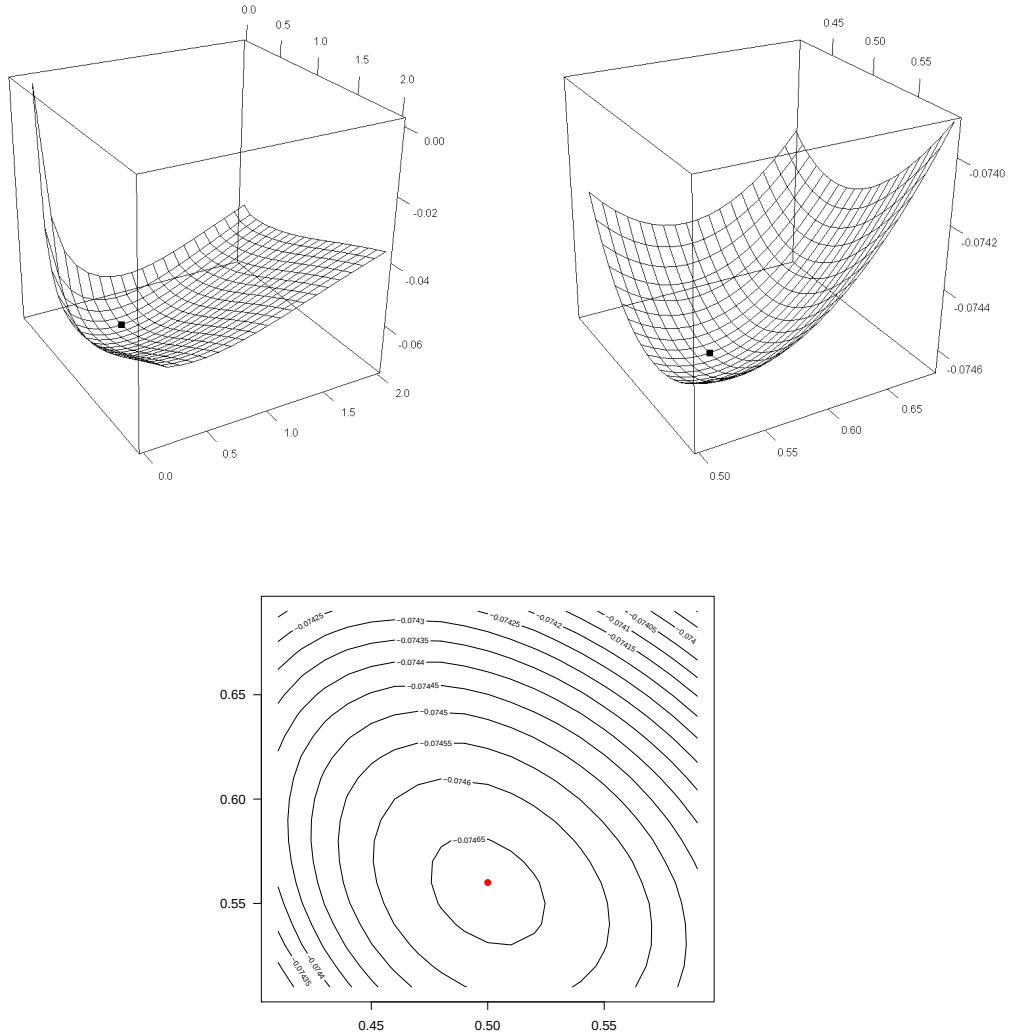


Figure 14: The function CV on a coarse grid (left) and fine grid (right) and a contour plot of CV on the fine grid (lower).

Note that the obtained approximate optimal bandwidth parameters are smaller than the optimal bandwidth parameter found in the univariate example. The distribution chosen in the univariate example is the Normal distribution with a variance of 2, while in this example both variables are standard Normal, so these variables have a variance of 1. This smaller variance results in the fact that a smaller bandwidth parameter suffices to accurately estimate the underlying density.

The obtained approximate optimal bandwidths $\hat{b}_X$ and $\hat{b}_Y$ can be used to construct the *LSCV*-optimal bivariate Normal kernel density estimator for this example. A plot of this estimator, together with the true density of the bivariate standard Normal distribution, can be found in Figure 15 below.

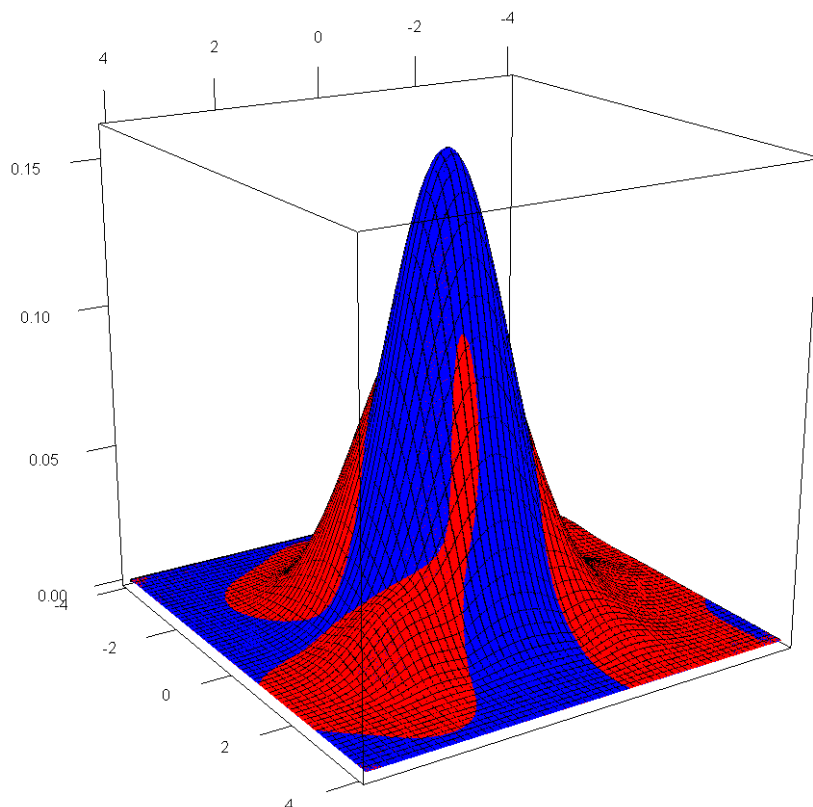


Figure 15: The true density (blue) and the *LSCV*-optimal Normal kernel density estimator (red).

Note that the estimator is close to the true density, but especially near the origin $(0,0)$ the estimator is lower than the true density. Due to the relatively small data set, the bandwidth parameters are relatively large, so the mass is spread more equally over the area. This leads to an estimator that cannot estimate the high peak of the Normal density accurately. If a larger data set is considered, the bandwidth parameter can be chosen smaller, which will result in a more accurate estimator.

Although an estimator for a bivariate density is then constructed, there is a problem: In the contour plot of Figure 12, it can be seen that the kernel density estimator gives mass outside the first quadrant. This is not necessarily a problem, but if the data represents quantities that cannot be negative, e.g. lengths or ages, the probability density function should be zero for negative input. However, this is not the case, even though all data is positive. This problem is called the boundary problem. In the Oriented Cylinder Model, the variables represent lengths, so the boundary problem occurs in estimating the density of these quantities. Moreover, since the obtained density function will be used as input for a numerical procedure to obtain an estimator for the density of interest, this problem is of real significance. A solution to this boundary problem will be discussed in the next section.

3.3 Boundary problems

Consider the contour plot of the bivariate kernel density estimator of the previous section, with two dotted lines added at $x = 0$ and $y = 0$, in Figure 16:

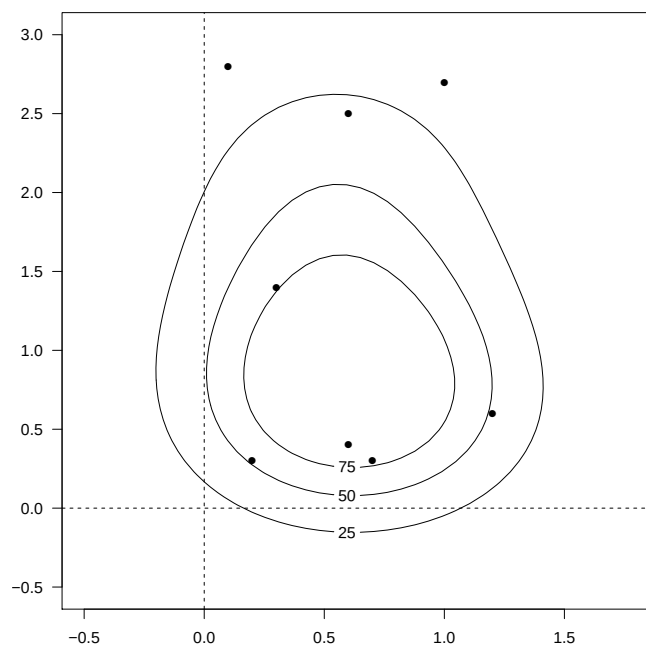


Figure 16: A bivariate kernel density estimator with bivariate Normal kernels.

Since the contour line of level 25 intersects the dotted lines, it is clear that the kernel density estimator is nonzero for certain negative values of x or y . This means that, even though the example data set contains only positive data, the probability that X or Y is negative, is nonzero according to the estimator. In our case, all data is also positive, since the data concerns lengths (heights and squared half-widths or squared radii), so the probability that one or more of these variables are negative should therefore be zero. The general kernel density estimator with Normal kernels does therefore not meet this requirement.

For univariate kernel density estimation this boundary problem is studied extensively, and multiple solutions are proposed, like data reflection [12], boundary kernels [13], data transformation

[14] and the use of Gamma kernels [15] for example. This last solution, which is extended to the second dimension in [10], will be used to solve the boundary problem. First, a specific Gamma kernel will be introduced, after which the bias of the univariate kernel density estimator with this Gamma kernel will be studied asymptotically. Then, in Section 3.3.2, this Gamma kernel will be modified to improve the properties of the estimator based on this kernel. After that, some examples of univariate estimators with this modified Gamma kernel will be given. In the last section, Section 3.3.4, this solution to the boundary problem will be extended to solve the boundary problem in bivariate kernel density estimation.

3.3.1 Gamma kernel density estimator

Consider the following Gamma kernel, which is proposed in [15], in the univariate situation:

$$K_{\frac{x}{b}+1,b}(t) = \frac{t^{\frac{x}{b}} e^{-\frac{t}{b}}}{b^{\frac{x}{b}+1} \Gamma\left(\frac{x}{b} + 1\right)}, \quad t > 0.$$

In words, $K_{\frac{x}{b}+1,b}$ is the probability density function of a *Gamma* $(\frac{x}{b} + 1, b)$ distribution. The parameters of a Gamma distribution should be positive, so $\frac{x}{b} + 1 > 0$ and $b > 0$, which leads for x to the inequality $x > -b$. Moreover, since $K_{\frac{x}{b}+1,b}$ is a probability density,

$$\int_0^{\infty} K_{\frac{x}{b}+1,b}(t) dt = 1.$$

The Gamma kernel density estimator $\hat{f}_1$ for the observations $x_1, \dots, x_n$ based on this kernel, is defined as

$$\hat{f}_1(x) = \frac{1}{n} \sum_{i=1}^n K_{\frac{x}{b}+1,b}(x_i).$$

The kernel density estimator based on the discussed Gamma kernel is thus the average of these kernels evaluated in the different observations. Since the support of each Gamma kernel is $[0, \infty)$, the support of the corresponding estimator is also equal to $[0, \infty)$. As a consequence, the Gamma kernel density estimator will not give weight to the negative real half-line. This property is the reason for the choice of the Gamma kernel.

To ensure that the kernel density estimator is a density itself, the integral over x of $\hat{f}_1$ should be equal to 1. This implies that $\hat{f}_1$ is a density if $\int_0^{\infty} K_{\frac{x}{b}+1,b}(t) dx = 1$ for every t , which is not as trivial as the integral over t . In fact, in contrast to the rescaled and translated kernels considered in Section 3.1.1, it is not even the case that this holds for every b and t . However, Table 1 provides empirical evidence that the integral over x of the Gamma kernel equals 1 for t fixed and b tending to 0. The absolute difference $|\int_0^{\infty} K_{\frac{x}{b}+1,b}(t) dx - 1|$ is computed numerically for different values of b and t :

		t			
		0.5	1	2	4
b	0.01	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
	0.1	$2.161 \cdot 10^{-3}$	$1.255 \cdot 10^{-5}$	$< 10^{-10}$	$< 10^{-10}$
	0.5	$1.662 \cdot 10^{-1}$	$5.298 \cdot 10^{-2}$	$6.169 \cdot 10^{-3}$	$9.720 \cdot 10^{-5}$
	1	$3.119 \cdot 10^{-1}$	$1.662 \cdot 10^{-1}$	$5.298 \cdot 10^{-2}$	$6.169 \cdot 10^{-3}$

Table 1: Absolute differences $|\int_0^{\infty} K_{\frac{x}{b}+1,b}(t) dx - 1|$.

It can be seen in the table above that for the given values of t , the absolute difference seems to tend to 0 as the value of b becomes smaller. Next, it will be proved that for $b \downarrow 0$ the integral over x of the Gamma kernel is indeed 1 for every t .

For this proof, the integral of interest is rewritten as

$$\int_0^\infty K_{\frac{x}{b}+1,b}(t)dx = \int_0^\infty \frac{t^{\frac{x}{b}} e^{-\frac{t}{b}}}{b^{\frac{x}{b}+1} \Gamma\left(\frac{x}{b}+1\right)} dx = \frac{1}{b} \int_0^\infty \frac{\left(\frac{t}{b}\right)^{\frac{x}{b}} e^{-\frac{t}{b}}}{\Gamma\left(\frac{x}{b}+1\right)} dx = \int_0^\infty \frac{\left(\frac{t}{b}\right)^v e^{-\frac{t}{b}}}{\Gamma(v+1)} dv = \int_0^\infty \varphi_{t,b}(v) dv,$$

where

$$\varphi_{t,b}(v) = \frac{\left(\frac{t}{b}\right)^v e^{-\frac{t}{b}}}{\Gamma(v+1)}.$$

Then the derivative of the logarithm of this function is given by

$$\begin{aligned} \frac{\partial}{\partial v} \log(\varphi_{t,b}(v)) &= \frac{\partial}{\partial v} \left[v \log\left(\frac{t}{b}\right) - \frac{t}{b} - \log(\Gamma(v+1)) \right] \\ &= \log\left(\frac{t}{b}\right) - \frac{\Gamma'(v+1)}{\Gamma(v+1)} = \log\left(\frac{t}{b}\right) - \psi(v+1), \end{aligned} \quad (8)$$

where ψ is the Digamma function, which is defined as the derivative of the logarithm of the Gamma function. In [16] it is shown that ψ can be written as

$$\psi(s) = -\gamma + \sum_{n=0}^{\infty} \left(\frac{1}{n+1} - \frac{1}{n+s} \right),$$

for $s > 0$, and γ the Euler-Mascheroni constant. Since the Digamma function is continuous and $\lim_{s \downarrow 0} \psi(s) = -\infty$ and $\lim_{s \rightarrow \infty} \psi(s) = \infty$, there is at least one point in which the derivative of the logarithm of $\varphi_{t,b}$ is equal to zero. Furthermore, the Digamma function is a strictly increasing function, since each term of the sum increases if s increases, so there is at most one point in which (8) equals zero. Because the derivative of $\varphi_{t,b}$ itself is also equal to zero when the derivative of its logarithm equals zero, there is exactly one point, denoted by $v_{t,b}^*$, such that

$$\frac{\partial}{\partial v} \varphi_{t,b}(v_{t,b}^*) = 0.$$

Together with the fact that the derivative of the logarithm of $\varphi_{t,b}$ is decreasing, since ψ is increasing, it can be concluded that $\varphi_{t,b}$ is a unimodal function with its maximum in $v = v_{t,b}^*$. As a consequence, there is a $k_{t,b}^* \in \mathbb{N}$, namely $k_{t,b}^* = \max\{k | k \leq v_{t,b}^*, k \in \mathbb{N}\}$, such that $\varphi_{t,b}$ is increasing on the interval $[0, k_{t,b}^*]$ and decreasing on $[k_{t,b}^* + 1, \infty)$. With this result, bounds can be constructed for the integral of $\varphi_{t,b}$ over $[0, \infty)$. The integral of $\varphi_{t,b}$ over the first interval is bounded by

$$\sum_{k=0}^{k_{t,b}^*-1} \varphi_{t,b}(k) = \sum_{k=0}^{k_{t,b}^*-1} \int_k^{k+1} \varphi_{t,b}(k) dv \leq \int_0^{k_{t,b}^*} \varphi_{t,b}(v) dv \leq \sum_{k=0}^{k_{t,b}^*-1} \int_k^{k+1} \varphi_{t,b}(k+1) dv = \sum_{k=0}^{k_{t,b}^*-1} \varphi_{t,b}(k+1),$$

where the lower and upper bound are a left and right Riemann sum respectively. Since the function is decreasing in the second interval, the integral can be bounded from below by the right Riemann sum and bounded from above by the left Riemann sum. The interval in between,

i.e. $[k_{t,b}^*, k_{t,b}^* + 1]$, is simply bounded by the minimum and maximum on that interval of length 1. This results in the inequalities:

$$\begin{aligned} \sum_{k=k_{t,b}^*+1}^{\infty} \varphi_{t,b}(k+1) &\leq \int_{k_{t,b}^*+1}^{\infty} \varphi_{t,b}(v)dv \leq \sum_{k=k_{t,b}^*+1}^{\infty} \varphi_{t,b}(k) \\ \min\{\varphi_{t,b}(k_{t,b}^*), \varphi_{t,b}(k_{t,b}^* + 1)\} &\leq \int_{k_{t,b}^*}^{k_{t,b}^*+1} \varphi_{t,b}(v)dv \leq \varphi_{t,b}(v_{t,b}^*). \end{aligned}$$

The whole integral $\int_0^{\infty} \varphi_{t,b}(v)dv$ is then bounded from below by the three lower bounds added together and bounded from above by the three upper bounds added together. This leads to the following lower bound $L_t(b)$ and upper bound $U_t(b)$:

$$\begin{aligned} L_t(b) &= \sum_{k=0}^{k_{t,b}^*-1} \varphi_{t,b}(k) + \sum_{k=k_{t,b}^*+1}^{\infty} \varphi_{t,b}(k+1) + \min\{\varphi_{t,b}(k_{t,b}^*), \varphi_{t,b}(k_{t,b}^* + 1)\} \\ &= \sum_{k=0}^{k_{t,b}^*-1} \varphi_{t,b}(k) + \sum_{k=k_{t,b}^*+2}^{\infty} \varphi_{t,b}(k) + \min\{\varphi_{t,b}(k_{t,b}^*), \varphi_{t,b}(k_{t,b}^* + 1)\} \\ &= \sum_{k=0}^{\infty} \varphi_{t,b}(k) - \max\{\varphi_{t,b}(k_{t,b}^*), \varphi_{t,b}(k_{t,b}^* + 1)\} \\ U_t(b) &= \sum_{k=0}^{k_{t,b}^*-1} \varphi_{t,b}(k+1) + \sum_{k=k_{t,b}^*+1}^{\infty} \varphi_{t,b}(k) + \varphi_{t,b}(v_{t,b}^*) \\ &= \sum_{k=1}^{k_{t,b}^*} \varphi_{t,b}(k) + \sum_{k=k_{t,b}^*+1}^{\infty} \varphi_{t,b}(k) + \varphi_{t,b}(v_{t,b}^*) \\ &= \sum_{k=0}^{\infty} \varphi_{t,b}(k) + (\varphi_{t,b}(v_{t,b}^*) - \varphi_{t,b}(0)). \end{aligned}$$

Since the sum that is present in both bounds equals

$$\sum_{k=0}^{\infty} \varphi_{t,b}(k) = \sum_{k=0}^{\infty} \frac{\left(\frac{t}{b}\right)^k e^{-\frac{t}{b}}}{\Gamma(k+1)} = e^{-\frac{t}{b}} \sum_{k=0}^{\infty} \frac{\left(\frac{t}{b}\right)^k}{k!} = e^{-\frac{t}{b}} e^{\frac{t}{b}} = 1,$$

and since $\varphi_{t,b}(v_{t,b}^*) = \sup_{v>0} \{\varphi_{t,b}(v)\}$, the following inequalities hold:

$$1 - \varphi_{t,b}(v_{t,b}^*) \leq L_t(b) \leq \int_0^{\infty} \varphi_{t,b}(v)dv \leq U_t(b) \leq 1 + \varphi_{t,b}(v_{t,b}^*)$$

Then it only has to be proved that $\lim_{b \downarrow 0} \varphi_{t,b}(v_{t,b}^*) = 0$. By writing $\varphi_{t,b}(v) = e^{-\frac{t}{b} + v \log(\frac{t}{b}) - \log(\Gamma(v+1))}$ and deducing the lower bound for the logarithm of the Gamma function, with $\alpha = 0.5$,

$$\begin{aligned} \log(\Gamma(v+1)) &\geq \log(\alpha^{-\alpha} e^{-v}(v+\alpha)^{v+\alpha}) \\ &\geq -v + (v+\alpha) \log(v+\alpha) \end{aligned}$$

from [17], the following upper bound for the function $\varphi_{t,b}$ is established:

$$\varphi_{t,b}(v) \leq e^{-\frac{t}{b} + v \log\left(\frac{t}{b}\right) - (-v + (v + \alpha) \log(v + \alpha))}.$$

To determine the maximum value of the right hand side, which is maximal where its logarithm is maximal, the derivative of the logarithm is computed:

$$\frac{\partial}{\partial v} \left[-\frac{t}{b} + v \log\left(\frac{t}{b}\right) - (-v + (v + \alpha) \log(v + \alpha)) \right] = \log\left(\frac{t}{b}\right) + 1 - 1 - \log(v + \alpha),$$

which is zero for $v_0 = \frac{t}{b} - \alpha$. Since the second derivative is equal to $-\frac{1}{v + \alpha}$, the upper bound for $\varphi_{t,b}$ has its maximum in v_0 , and this maximum is equal to

$$e^{-\frac{t}{b} + \frac{t}{b} \log\left(\frac{t}{b}\right) - \left(-\frac{t}{b} + \left(\frac{t}{b} + \alpha\right) \log(v + \alpha)\right)} = e^{-\alpha \log\left(\frac{t}{b}\right) + \alpha \log(\alpha) - \alpha}.$$

For all $b > 0$, the maximum of $\varphi_{t,b}$ is also bounded by this maximum, so

$$0 \leq \varphi_{t,b}(v_{t,b}^*) \leq e^{-\alpha \log\left(\frac{t}{b}\right) + \alpha \log(\alpha) - \alpha}.$$

Because the limit for $b \downarrow 0$ of the right hand side equals 0, also $\lim_{b \downarrow 0} \varphi_{t,b}(v_{t,b}^*) = 0$. Moreover, then

$$\begin{aligned} \lim_{b \downarrow 0} (1 - \varphi_{t,b}(v_{t,b}^*)) &= 1 \\ \lim_{b \downarrow 0} (1 + \varphi_{t,b}(v_{t,b}^*)) &= 1. \end{aligned}$$

From the Squeeze theorem, and the result that

$$1 - \varphi_{t,b}(v_{t,b}^*) \leq \int_0^\infty \varphi_{t,b}(v) dv \leq 1 + \varphi_{t,b}(v_{t,b}^*),$$

with the limit of each bound equal to 1, it then follows that

$$\lim_{b \downarrow 0} \int_0^\infty \varphi_{t,b}(v) dv = 1.$$

Remember that this integral is equal to the integral of interest, so the final result is then also achieved:

$$\lim_{b \downarrow 0} \int_0^\infty K_{\frac{x}{b} + 1, b}(t) dx = 1.$$

This means that the Gamma kernel is for $b \downarrow 0$ a density with respect to the variable x . This means that if b is small enough, the Gamma kernel will integrate to a value close to 1, so the kernel function will approximately be a density. As a result, the corresponding kernel density estimator will then also approximately be a density. Note that, in contrast, the kernels discussed in 3.1.1, like the Normal kernel or Epanechnikov kernel for example, are a density for every b .

Another difference between the Gamma kernel and the other kernels mentioned above is the influence of the bandwidth parameter b on the shape of the Gamma kernel function, and thus on the shape of the Gamma kernel density estimator. For one observation, denoted by x_1 , the Gamma kernel is shown in Figure 17 for different values of b and $x_1 = 1$ fixed.

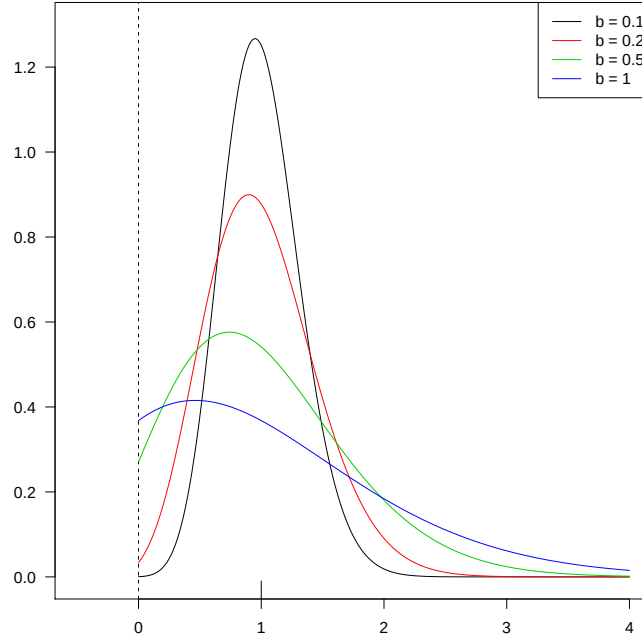


Figure 17: The Gamma kernel density estimator for different bandwidth parameters.

Although all kernels of Section 3.1.1 are symmetric around the data point, the Gamma kernels are clearly asymmetric. A bigger bandwidth parameter does not only increase the amount of smoothing of the Gamma kernel, but also shifts the mode of the kernel to the left of the data point. The Gamma kernel therefore becomes more symmetrical and more similar to a Normal density for smaller values of the bandwidth parameter. For the kernels of Section 3.1.1 the bandwidth parameter only influences the amount of smoothing, where the kernel is more levelled for bigger values of the bandwidth parameter.

The shape of the Gamma kernel is not only influenced by the bandwidth parameter, but also by the value of the observed data point. This is in contrast to other kernels, where a data point only determines the location of the centered scaled kernel function. In Figure 18, the Gamma kernel is shown for different values of the observation x_1 , with $b = 0.2$ fixed.

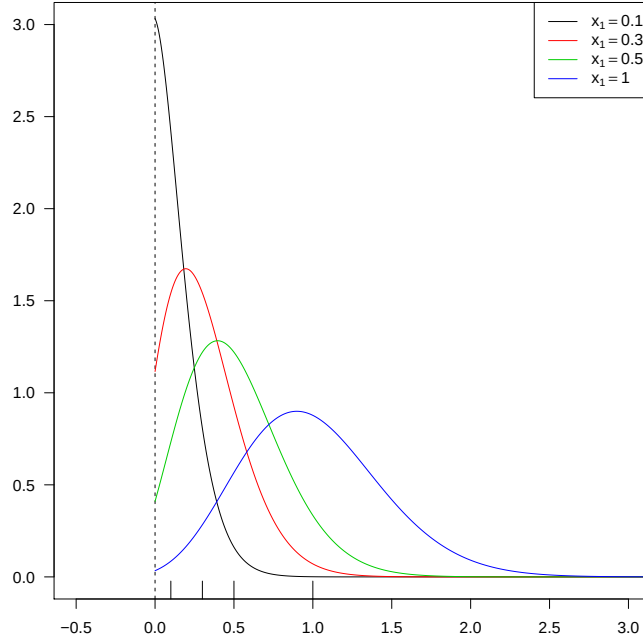


Figure 18: The Gamma kernel density estimator for different observations.

Regarding an observation x_1 , it can be concluded that the mode of the Gamma kernel function is close to the observation, but in each case to the left of it. Furthermore, a bigger value of x_1 results in a kernel function that is more symmetric and smoother. This last result also occurs in a different approach of solving the boundary problem, namely data transformation. When transforming positive data to its logarithm, large data values will be compressed, so the constant bandwidth covers a wider range of original observations for larger observations. Performing kernel density estimation with a kernel of Section 3.1.1 on the transformed data set on the whole real line will therefore result in a smoother estimator for larger observed values.

To summarize, both the bandwidth parameter and the value of the observation have an influence on the shape of the Gamma kernel. Regarding the smoothness of the kernel, a bigger value of either the bandwidth parameter or the observation will lead to a smoother kernel. Their influence on the symmetry of the Gamma kernel, though, is opposite to each other: A smaller bandwidth parameter leads to a more symmetric kernel, while a smaller observation leads to a less symmetric kernel.

To investigate the performance of the Gamma kernel density estimator, its bias is studied asymptotically. For this bias, defined as the difference between the expected value of the estimator and the true value, the expected value of the estimator in a point $x > 0$ is considered. Let $Y_1 \sim \text{Gamma}(\frac{x}{b} + 1, b)$, E_f the expectation under f , and E_K and Var_K the expectation and variance under $K_{\frac{x}{b}+1, b}$ respectively. Then it is known that $E_K[Y_1] = (\frac{x}{b} + 1)b = x + b$ and $\text{Var}_K(Y_1) = (\frac{x}{b} + 1)b^2 = b(x + b)$. Furthermore, the density f is assumed to be three times continuously differentiable, with a Lipschitz continuous third derivative. From this last assumption it follows that there is a $C \geq 0$ such that for all $r, s > 0$ it holds that

$$|f'''(r) - f'''(s)| \leq C|r - s|.$$

Then the expected value of the estimator under f is equal to:

$$\begin{aligned}
E_f[\hat{f}_1(x)] &= E_f \left[\frac{1}{n} \sum_{i=1}^n K_{\frac{x}{b}+1,b}(X_i) \right] = \frac{1}{n} \sum_{i=1}^n E_f[K_{\frac{x}{b}+1,b}(X_i)] = E_f[K_{\frac{x}{b}+1,b}(X_1)] \\
&= \int_0^\infty K_{\frac{x}{b}+1,b}(y) f(y) dy = E_K[f(Y_1)] = E_K[f(E_K[Y_1] + (Y_1 - E_K[Y_1]))] \\
&= E_K \left[f(E_K[Y_1]) + f'(E_K[Y_1])(Y_1 - E_K[Y_1]) + \frac{1}{2} f''(E_K[Y_1])(Y_1 - E_K[Y_1])^2 \right. \\
&\quad \left. + \frac{1}{6} f'''(\xi_1)(Y_1 - E_K[Y_1])^3 \right] \\
&= f(E_K[Y_1]) + \frac{1}{2} f''(E_K[Y_1]) E_K[(Y_1 - E_K[Y_1])^2] + \frac{1}{6} E_K[f'''(\xi_1)(Y_1 - E_K[Y_1])^3] \\
&= f(E_K[Y_1]) + \frac{1}{2} f''(E_K[Y_1]) E_K[(Y_1 - E_K[Y_1])^2] + \frac{1}{6} f'''(E_K[Y_1]) E_K[(Y_1 - E_K[Y_1])^3] \\
&\quad + \frac{1}{6} E_K[f'''(\xi_1) - f'''(E_K[Y_1])](Y_1 - E_K[Y_1])^3, \tag{9}
\end{aligned}$$

where ξ_1 is between Y_1 and $E_K[Y_1]$. For the first three terms, note that:

$$\begin{aligned}
f(E_K[Y_1]) &= f(x+b) = f(x) + f'(x)b + o(b) \\
\frac{1}{2} f''(E_K[Y_1]) E_K[(Y_1 - E_K[Y_1])^2] &= \frac{1}{2} f''(x+b) \text{Var}_K(Y_1) \\
&= \frac{1}{2} (f''(x) + f'''(x)b + o(b)) (b(x+b)) \\
&= \frac{1}{2} b x f''(x) + o(b) \\
\frac{1}{6} f'''(E_K[Y_1]) E_K[(Y_1 - E_K[Y_1])^3] &= \frac{1}{6} f'''(x+b) E_K[(Y_1 - E_K[Y_1])^3] \\
&= \frac{1}{6} (f'''(x) + o(1)) \left(2 \left(\frac{x}{b} + 1 \right) b^3 \right) \\
&= \frac{1}{6} (f'''(x) + o(1)) (2b^2(x+b)) \\
&= o(b),
\end{aligned}$$

where for the third term the known central moments of a Gamma distribution in Appendix A are used. For the fourth term of (9), the fact that the third derivative of f is Lipschitz continuous is used. From this property it follows that

$$\begin{aligned}
\frac{1}{6} E_K[f'''(\xi_1) - f'''(E_K[Y_1])](Y_1 - E_K[Y_1])^3 &\leq \frac{C}{6} E_K[|\xi_1 - E_K[Y_1]| (Y_1 - E_K[Y_1])^3] \\
&\leq \frac{C}{6} E_K[|Y_1 - E_K[Y_1]| (Y_1 - E_K[Y_1])^3] \\
&= \frac{C}{6} E_K[(Y_1 - E_K[Y_1])^4] \\
&= \frac{C}{6} \left(6 \left(\frac{x}{b} + 1 \right) b^4 + 3 \left(\frac{x}{b} + 1 \right)^2 b^4 \right) \\
&= \frac{C}{6} (6b^3(x+b) + 3b^2(x+b)) = o(b),
\end{aligned}$$

where again the central moments in Appendix A are used. By adding the four terms, the expected value of $\hat{f}_1(x)$ is obtained:

$$\begin{aligned} E_f[\hat{f}_1(x)] &= f(x) + f'(x)b + \frac{1}{2}bx f''(x) + o(b) \\ &= f(x) + b \left(f'(x) + \frac{1}{2}x f''(x) \right) + o(b). \end{aligned}$$

The bias of the estimator is therefore equal to:

$$\text{Bias}(\hat{f}_1(x)) = E_f[\hat{f}_1(x) - f(x)] = b \left(f'(x) + \frac{1}{2}x f''(x) \right) + o(b).$$

As can be seen from this equation, the bias is $O(b)$ for every value of x . As a result, the Gamma kernel density estimator is consistent for every x , so also near the boundary. In contrast, it is shown, in [7] for example, that for the kernels discussed in Section 3.1.1, if $x = 0$ is a boundary, it holds that

$$E[\hat{f}(0)] = \frac{1}{2}f(0) + O(b),$$

which means that these kernels are inconsistent in the boundary. In other words, these kernels are biased near the boundary.

Although the Gamma kernel leads to an unbiased estimator, it is chosen analogous to [15] to modify this Gamma kernel. Since the mean of the kernel $K_{\frac{x}{b}+1,b}$ is $x + b$, the Gamma kernel $K_{\frac{x}{b},b}$ is considered, which has x as its mean. However, this kernel is unbounded near $x = 0$, so a combination of the two Gamma kernels is used to construct a modified Gamma kernel. In the next section, this modified Gamma kernel will be introduced. Furthermore, the kernel density estimator based on this modified Gamma kernel will be constructed and its bias and variance will be studied asymptotically.

3.3.2 Modified Gamma kernel density estimator

As mentioned at the end of the previous section, the previously described Gamma kernel will be modified in this section. The two mentioned Gamma kernels are combined as done in [15]. The modified Gamma kernel is defined as

$$K_{\rho_b(x),b}(t) = \frac{t^{\rho_b(x)-1} e^{-\frac{t}{b}}}{b^{\rho_b(x)} \Gamma(\rho_b(x))}, \quad t > 0,$$

with

$$\rho_b(x) = \begin{cases} \frac{x}{b}, & \text{if } x \geq 2b, \\ \frac{1}{4} \left(\frac{x}{b} \right)^2 + 1, & \text{if } x \in [0, 2b). \end{cases}$$

Note that for values of x away from the boundary, i.e. $x \geq 2b$, the modified Gamma kernel is equal to the Gamma kernel $K_{\frac{x}{b},b}$. Moreover, ρ_b is a continuous and smooth function on the whole positive real half-line. The modified Gamma kernel density estimator, based on the observations $x_1, \dots, x_n$, then becomes:

$$\hat{f}_2(x) = \frac{1}{n} \sum_{i=1}^n K_{\rho_b(x),b}(x_i).$$

The bias of this estimator can be computed similarly as before. Let $Y_2 \sim \text{Gamma}(\rho_b(x), b)$, then $E_K[Y_2] = b\rho_b(x)$ and $\text{Var}_K(Y_2) = b^2\rho_b(x)$, where E_K and Var_K are the expected value and variance with respect to $K_{\rho_b(x), b}$ respectively. Under the same assumptions as before, the result (9) of the previous derivation can be used:

$$\begin{aligned} E_f[\hat{f}_2(x)] &= f(E_K[Y_2]) + \frac{1}{2}f''(E_K[Y_2])E_K[(Y_2 - E_K[Y_2])^2] + \frac{1}{6}f'''(E_K[Y_2])E_K[(Y_2 - E_K[Y_2])^3] \\ &\quad + \frac{1}{6}E_K[(f'''(\xi_2) - f'''(E_K[Y_2]))(Y_2 - E_K[Y_2])^3] \\ &= f(b\rho_b(x)) + \frac{1}{2}f''(b\rho_b(x))b^2\rho_b(x) + \frac{1}{6}f'''(b\rho_b(x))(2\rho_b(x)b^3) \\ &\quad + \frac{1}{6}E_K[(f'''(\xi_2) - f'''(E_K[Y_2]))(Y_2 - E_K[Y_2])^3], \end{aligned}$$

with ξ_2 between Y_2 and $E_K[Y_2]$. For the expression of the third central moment, Appendix A is used. In the same way as before, again under the assumption that f''' is Lipschitz continuous, the last term can be estimated with:

$$\begin{aligned} \frac{1}{6}E_K[(f'''(\xi_2) - f'''(E_K[Y_2]))(Y_2 - E_K[Y_2])^3] &\leq \frac{C}{6}E[(Y_2 - E_K[Y_2])^4] \\ &= \frac{C}{6}(6\rho_b(x)b^4 + 3\rho_b(x)^2b^4) \\ &= o(b), \end{aligned}$$

where again Appendix A is used. Moreover, the last equation holds, since $\rho_b(x) = \frac{x}{b} = O(b^{-1})$ for $x \geq 2b$ and $\rho_b(x) = \frac{1}{4}(\frac{x}{b})^2 + 1 = O(1)$ for $x \leq 2b$, since then $x = O(b)$.

The first three terms are considered separately for $x \geq 2b$ and $x \in [0, 2b)$. In the first case $\rho_b(x) = \frac{x}{b}$, so the third term equals

$$\frac{1}{6}f'''(b\rho_b(x))(2\rho_b(x)b^3) = \frac{1}{3}f'''(x)b^2x = o(b).$$

Since $b\rho_b(x) = x$, the expression of the first two terms follow directly. The expected value of $\hat{f}_2(x)$ for $x \geq 2b$ is then known:

$$E_f[\hat{f}_2(x)] = f(x) + \frac{1}{2}bx f''(x) + o(b), \quad \text{for } x \geq 2b.$$

For $x \in [0, 2b)$, x is written as $x = 2bu$, with $u \in [0, 1)$. Then $\rho_b(x) = \rho_b(2bu) = u^2 + 1 = O(1)$, and so $b\rho_b(x) - x = b(u^2 + 1) - 2bu = O(b)$. Furthermore, by using Taylor expansion, the first two terms can be expressed as

$$\begin{aligned} f(b\rho_b(x)) &= f(x) + f'(x)(b\rho_b(x) - x) + o(b), \quad \text{for } x \in [0, 2b) \\ \frac{1}{2}f''(b\rho_b(x))b^2\rho_b(x) &= \frac{1}{2}(f''(x) + O(b))b^2\rho_b(x) = o(b), \quad \text{for } x \in [0, 2b) \end{aligned}$$

Since the third term is then equal to $\frac{1}{6}(f'''(x) - O(b))2b^3\rho_b(x) = o(b)$, the expected value for $\hat{f}_2(x)$ for $x \in [0, 2b)$ is

$$E_f[\hat{f}_2(x)] = f(x) + (b\rho_b(x) - x)f'(x) + o(b), \quad \text{for } x \in [0, 2b).$$

The expected value of $\hat{f}_2(x)$ can thus be expressed in general as:

$$E_f[\hat{f}_2(x)] = \begin{cases} f(x) + \frac{1}{2}bx f''(x) + o(b), & \text{if } x \geq 2b, \\ f(x) + (b\rho_b(x) - x)f'(x) + o(b), & \text{if } x \in [0, 2b]. \end{cases}$$

Therefore, the bias of the estimator $\hat{f}$ is:

$$\text{Bias}(\hat{f}_2(x)) = \begin{cases} \frac{1}{2}bx f''(x) + o(b), & \text{if } x \geq 2b, \\ (b\rho_b(x) - x)f'(x) + o(b), & \text{if } x \in [0, 2b]. \end{cases}$$

Note that in [15] the first derivative is multiplied by $b\xi_b(x)$, with $\xi_b(x) = \frac{(1-x)(\rho_b(x)-\frac{x}{b})}{1+b\rho_b(x)-x}$, but since

$$\begin{aligned} b\xi_b(x) &= \frac{(1-x)(b\rho_b(x)-x)}{1+b\rho_b(x)-x} = b\rho_b(x) - x - \frac{b\rho_b(x)(b\rho_b(x)-x)}{1+b\rho_b(x)-x} \\ &= b\rho_b(x) - x - \frac{b(u^2+1)(b(u^2+1)-2bu)}{1+b(u^2+1)-2bu} = b\rho_b(x) - x - \frac{b^2(u^2+1)(u-1)^2}{1+b(u-1)^2} \\ &= b\rho_b(x) - x + o(b), \end{aligned}$$

the result is the same.

It can be seen that f' is only present in the bias in a small area near the origin, while the second derivative is only present in the bias in the interior. To further investigate the performance of the estimator based on this modified Gamma kernel, the variance of the estimator $\hat{f}_2(x)$ is computed:

$$\begin{aligned} \text{Var}_f(\hat{f}_2(x)) &= \text{Var}_f\left(\frac{1}{n} \sum_{i=1}^n K_{\rho_b(x),b}(X_i)\right) = \frac{1}{n^2} n \text{Var}_f(K_{\rho_b(x),b}(X_1)) \\ &= \frac{1}{n} (E_f[K_{\rho_b(x),b}(X_1)^2] - E_f[K_{\rho_b(x),b}(X_1)]^2) \\ &= \frac{1}{n} E_f[K_{\rho_b(x),b}(X_1)^2] + O(n^{-1}). \end{aligned} \tag{10}$$

Let $Z \sim \text{Gamma}(2\rho_b(x) - 1, \frac{b}{2})$, then:

$$\begin{aligned} E_f[K_{\rho_b(x),b}(X_1)^2] &= \int_0^\infty K_{\rho_b(x),b}(z)^2 f(z) dz \\ &= \int_0^\infty \frac{z^{2\rho_b(x)-2} e^{-\frac{2z}{b}}}{b^{2\rho_b(x)} \Gamma(\rho_b(x))^2} f(z) dz \\ &= \int_0^\infty \frac{b^{-1} \Gamma(2\rho_b(x) - 1)}{2^{2\rho_b(x)-1} \Gamma(\rho_b(x))^2} \frac{z^{2\rho_b(x)-2} e^{-\frac{2z}{b}}}{(\frac{b}{2})^{2\rho_b(x)-1} \Gamma(2\rho_b(x) - 1)} f(z) dz \\ &= B_b(x) E[f(Z)]. \end{aligned} \tag{11}$$

Let $R(z) = \frac{\sqrt{2\pi} e^{-z} z^{z+\frac{1}{2}}}{\Gamma(z+1)}$ for $z \geq 0$. Then B_b can be expressed in terms of R :

$$B_b(x) = \frac{(\rho_b(x) - 1)^{-\frac{1}{2}}}{2\sqrt{\pi}} \frac{R(\rho_b(x) - 1)^2}{R(2(\rho_b(x) - 1))} b^{-1}.$$

It is shown in [18] that R is a monotonic increasing function that converges to 1 as $z \rightarrow \infty$ and that $R(z) < 1$ for $z > 0$. Therefore $\frac{R(\rho_b(x)-1)^2}{R(2(\rho_b(x)-1))} < 1$, and so:

$$B_b(x) \leq \frac{(\rho_b(x) - 1)^{-\frac{1}{2}}}{2\sqrt{\pi}} b^{-1}, \quad (12)$$

so B_b is bounded from above by the term on the right hand side. Moreover, for b small enough, and C a nonnegative constant:

$$B_b(x) \sim \begin{cases} \frac{(\rho_b(x)-1)^{-\frac{1}{2}}}{2\sqrt{\pi}} b^{-1} & \text{if } \frac{x}{b} \rightarrow \infty, \\ \frac{\Gamma(2\rho_b(x)-1)}{2^{2\rho_b(x)-1}\Gamma(\rho_b(x))^2} b^{-1}, & \text{if } \frac{x}{b} \rightarrow C. \end{cases} \quad (13)$$

It can therefore be concluded for the variance of $\hat{f}_2(x)$, that

$$\text{Var}_f(\hat{f}_2(x)) \sim \begin{cases} \frac{(\rho_b(x)-1)^{-\frac{1}{2}}}{2\sqrt{\pi}} b^{-1} n^{-1} f(x) & \text{if } \frac{x}{b} \rightarrow \infty, \\ \frac{\Gamma(2\rho_b(x)-1)}{2^{2\rho_b(x)-1}\Gamma(\rho_b(x))^2} b^{-1} n^{-1} f(x), & \text{if } \frac{x}{b} \rightarrow C. \end{cases}$$

The latter term is

$$\begin{aligned} & \frac{\Gamma(2C-1)}{2^{2C-1}\Gamma(C)^2} b^{-1} n^{-1} f(x) \quad \text{for } C \geq 2 \\ & \frac{\Gamma(\frac{1}{2}C^2+1)}{2^{\frac{1}{2}C^2+1}\Gamma(\frac{1}{4}C^2+1)^2} b^{-1} n^{-1} f(x) \quad \text{for } C < 2. \end{aligned}$$

Then both the bias and the variance of the estimator is known. These two quantities are combined in the Mean Squared Error (MSE) of the estimator, which is defined as the expected value of the square of the difference between the estimator and the true value. The MSE can be written in terms of the bias and the variance of the estimator:

$$\begin{aligned} MSE(\hat{f}_2(x)) &= E \left[\left(\hat{f}_2(x) - f(x) \right)^2 \right] \\ &= E \left[\hat{f}_2(x)^2 \right] - 2f(x)E[\hat{f}_2(x)] + f(x)^2 \\ &= E \left[\hat{f}_2(x)^2 \right] - E[\hat{f}_2(x)]^2 + E[\hat{f}_2(x)]^2 - 2f(x)E[\hat{f}_2(x)] + f(x)^2 \\ &= \left(E \left[\hat{f}_2(x)^2 \right] - E[\hat{f}_2(x)]^2 \right) + E[\hat{f}_2(x) - f(x)]^2 \\ &= \text{Var}(\hat{f}_2(x)) + \text{Bias}(\hat{f}_2(x))^2. \end{aligned}$$

Note that the variance of the estimator increases if b decreases, but that the bias of the estimator decreases if b decreases. The optimal rate of b for the MSE is then obtained through the Bias-Variance trade-off:

$$\begin{aligned} \text{Bias}(\hat{f}_2(x))^2 &\sim \text{Var}(\hat{f}_2(x)) \\ b^2 &\sim b^{\frac{1}{2}} b^{-1} n^{-1} \\ b^{\frac{5}{2}} &\sim n^{-1} \\ b &\sim n^{-\frac{2}{5}}. \end{aligned}$$

So the optimal rate for b is $b = O(n^{-\frac{2}{5}})$. With this rate for the bandwidth parameter, it can be concluded that $MSE(\hat{f}_2(x)) = O(n^{-\frac{4}{5}})$. For the kernels discussed in Section 3.1.1, the optimal rate for the MSE is the same, since in that case $b = O(n^{-\frac{2}{5}})$ and the bias is $O(b^2)$, see [7] for example. The rate of convergence of the Gamma kernel density estimator is therefore the same as the rate of those kernels, with the advantage that there is no boundary bias.

In the next section an example of a univariate modified Gamma kernel density estimator will be given. First this estimator will be constructed for the data set (5), with an arbitrary bandwidth parameter. After that, the estimator will be constructed for a bigger data set, drawn from a known distribution, with the bandwidth parameter chosen based on the *LSCV* method discussed in Section 3.1.2.

3.3.3 Examples of a univariate modified Gamma kernel density estimator

The modified Gamma kernel density estimator discussed in the previous section will be applied to the example data set (5). By visual comparison, the bandwidth parameter is chosen to be 0.03. The resulting estimator is shown in Figure 19.

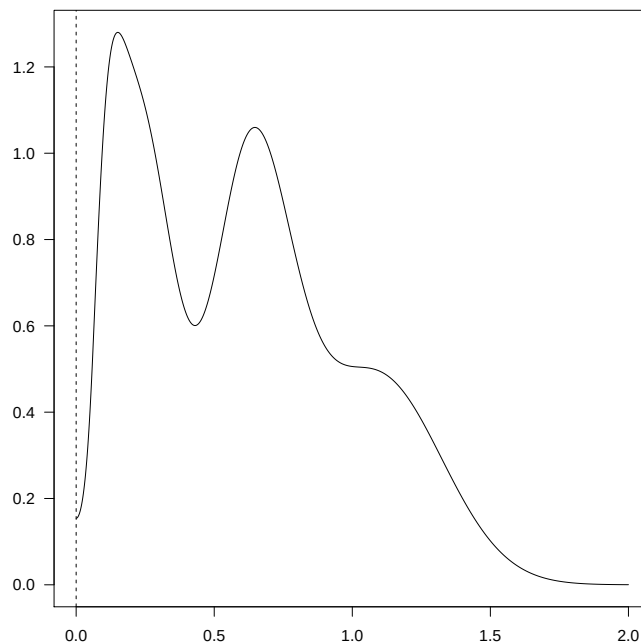


Figure 19: Modified Gamma kernel density estimator for the data set (5).

From the figure above it can be concluded that this estimator does not give mass to the left of $x = 0$, which is in contrast to both kernel density estimators with Normal kernels of Figure 8. Therefore, if it is known that the data are realisations of a positive random variable, the estimator with the modified Gamma kernels will be the preferred estimator.

However, in this example the true distribution of the underlying random variable is not known, so the estimator cannot be compared to the true density. To be able to compare the estimator

to the true density, the data for the next example will be drawn from a known distribution. To obtain only positive data, the exponential distribution is chosen. Let $X \sim \text{Exp}(1)$, i.e. X is an exponentially distributed with parameter 1, and let $x_1, \dots, x_{100}$ be 100 realisations of X . Based on this data, the *LSCV*-optimal bandwidth parameter can be chosen. For this, the method of Section 3.1.2 will be used. To use this method, the modified Gamma kernel density estimator is plugged in as the estimator. Moreover, since the Gamma kernel is only nonzero for positive arguments, the lower bound of the integrals are 0 instead of $-\infty$.

The *LSCV*-optimal value of the bandwidth parameter obtained with this method, is equal to 0.17. The modified Gamma kernel density estimator with this bandwidth parameter based on the data $x_1, \dots, x_{100}$ is shown in Figure 20, together with the true density and the kernel density estimator with Normal kernels and the same bandwidth parameter.

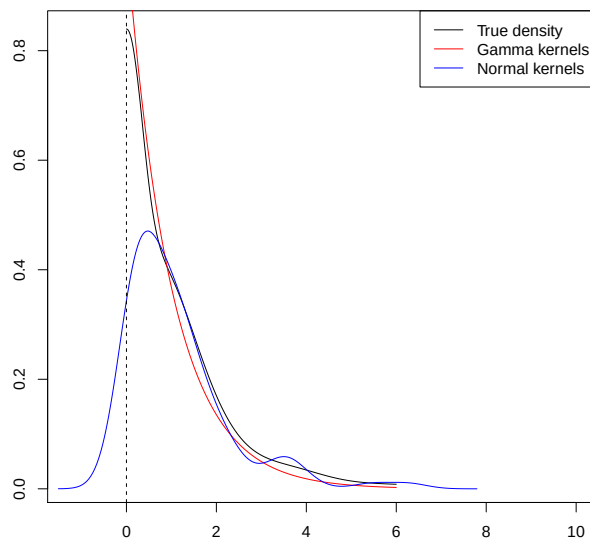


Figure 20: Kernel density estimator with Gamma kernels, Normal kernels, and the true density.

As can be seen in the figure above, both estimators are close to the true density for relatively large values, while the estimators differ from each other in the area close to the boundary. The estimator with Gamma kernels is also close to the true density near the boundary, but the estimator with Normal kernels is not accurate at all in that area. Due to the fixed shaped kernels of the latter estimator, this estimator does not only give weight to the negative axis, but also fails to accurately estimate the high values of the true density near the boundary. In this example it is clear that the Gamma kernel density estimator is a better estimator for the density of the considered positive random variable than the estimator with Normal kernels. Since the data in the Oriented Cylinder Model are also positive data, the Gamma kernel density estimator is used to estimate the densities in this model. However, these densities are bivariate densities, so a bivariate kernel density estimator has to be constructed based on the modified Gamma kernel. This will be done in the next section.

3.3.4 Bivariate Gamma kernel density estimator

In this section, the solution to the univariate boundary problem will be used to solve the boundary problem for the bivariate kernel density estimator. As explained in Section 3.2.1, the bivariate product kernel density estimator will be used as bivariate estimator, since then for each variable the modified Gamma kernel can be used as kernel function. Moreover, the bandwidth matrix is a diagonal matrix with the bandwidth parameters b_X and b_Y , corresponding to the variables X and Y respectively, on its diagonal. For the observed realisations $(x_1, y_1), \dots, (x_n, y_n)$ of (X, Y) the bivariate Gamma kernel density estimator is defined as:

$$\hat{f}(x, y) = \frac{1}{n} \sum_{i=1}^n (K_X(x_i) K_Y(y_i)),$$

where K_X and K_Y are short for $K_{\rho_{b_X}(x), b_X}$ and $K_{\rho_{b_Y}(y), b_Y}$. First, analogous to the univariate estimator, the bias of this estimator is investigated. Let $V \sim \text{Gamma}(\rho_{b_X}(x), b_X)$ and $W \sim \text{Gamma}(\rho_{b_Y}(y), b_Y)$ be independent, and let $f^{(i,j)}(x, y) = \frac{\partial^{i+j}}{\partial x^i \partial y^j} f(x, y)$. Furthermore, assume that $f^{(l, 5-l)}$ is Lipschitz continuous for $l = 0, 1, \dots, 5$. For b_X and b_Y smaller and smaller, it will hold for fixed x and y that $x \geq 2b_X$ and $y \geq 2b_Y$, so in the derivation of the bias it is therefore assumed that these relations are true, which implies that $b_X \rho_{b_X}(x) = x$ and $b_Y \rho_{b_Y}(y) = y$. Then $E_{K_X}[V] = x$, $E_{K_Y}[W] = y$, $\text{Var}_{K_X}(V) = b_X x$ and $\text{Var}_{K_Y}(W) = b_Y y$, where E_{K_X} (E_{K_Y}) and Var_{K_X} (Var_{K_Y}) are the expected value and variance with respect to K_X (K_Y) respectively. Furthermore, let E_K denote the expected value with respect to K , with $K(x, y) = K_X(x) K_Y(y)$. Then the expected value of the bivariate Gamma kernel density estimator equals:

$$\begin{aligned} E_f[\hat{f}(x, y)] &= E_f \left[\frac{1}{n} \sum_{i=1}^n (K_X(X_i) K_Y(Y_i)) \right] = E_f [K_X(X_1) K_Y(Y_1)] \\ &= \int_0^\infty \int_0^\infty K_X(v) K_Y(w) f(v, w) dv dw = E_K[f(V, W)] \\ &= E_K \left[f(x, y) + \sum_{k=1}^4 \frac{1}{k!} \left(\sum_{l=0}^k \binom{k}{l} f^{(l, k-l)}(x, y) (V-x)^l (W-y)^{k-l} \right) \right. \\ &\quad \left. + \frac{1}{5!} \sum_{l=0}^5 \binom{5}{l} f^{(l, 5-l)}(\xi_V, \xi_W) (V-x)^l (W-y)^{5-l} \right] \\ &= f(x, y) + \sum_{k=1}^5 \frac{1}{k!} \sum_{l=0}^k \binom{k}{l} f^{(l, k-l)}(x, y) E_{K_X} [(V-x)^l] E_{K_Y} [(W-y)^{k-l}] \\ &\quad + \frac{1}{5!} \sum_{l=0}^5 \binom{5}{l} E_K \left[\left(f^{(l, 5-l)}(\xi_V, \xi_W) - f^{(l, 5-l)}(x, y) \right) (V-x)^l (W-y)^{5-l} \right] \end{aligned} \quad (14)$$

with ξ_V between V and x and ξ_W between W and y . For the derivations below, Appendix A is used for the central moments. Note that for $k = 1$ the two terms of the inner sum are equal to 0, since then either one of the two expected values equals 0, while the other is equal to 1. For $k = 2$, only the first and last term are nonzero (both expectations are equal to 0 in the second term), which results in the terms:

$$\begin{aligned} \frac{1}{2} f^{(0,2)}(x, y) E_K [(W-y)^2] &= \frac{1}{2} f^{(0,2)}(x, y) b_Y y \\ \frac{1}{2} f^{(2,0)}(x, y) E_K [(V-x)^2] &= \frac{1}{2} f^{(2,0)}(x, y) b_X x. \end{aligned}$$

For $k = 3$, the second and third term of the inner sum are again equal to 0, while the first and last term are:

$$\begin{aligned}\frac{1}{6}f^{(0,3)}(x,y)E_K[(W-y)^3] &= \frac{1}{6}f^{(0,3)}(x,y)2\rho_{b_Y}(y)b_Y^3 = \frac{2}{6}f^{(0,3)}(x,y)b_Y^2y = o(b_Y) \\ \frac{1}{6}f^{(3,0)}(x,y)E_K[(V-x)^3] &= \frac{1}{6}f^{(3,0)}(x,y)2\rho_{b_X}(x)b_X^3 = \frac{2}{6}f^{(3,0)}(x,y)b_X^2x = o(b_X).\end{aligned}$$

Then, $k = 4$ leads to the following three nonzero terms:

$$\begin{aligned}\frac{1}{24}f^{(0,4)}(x,y)E_K[(W-y)^4] &= \frac{1}{24}f^{(0,4)}(x,y)(6\rho_{b_Y}(y)b_Y^4 + 3\rho_{b_Y}(y)^2b_Y^4) = o(b_Y) \\ \frac{1}{4}f^{(2,2)}(x,y)E_K[(V-x)^2]E_K[(W-y)^2] &= \frac{1}{4}f^{(2,2)}(x,y)b_Xxb_Yy \\ \frac{1}{24}f^{(4,0)}(x,y)E_K[(V-x)^4] &= \frac{1}{24}f^{(4,0)}(x,y)(6\rho_{b_X}(x)b_X^4 + 3\rho_{b_X}(x)^2b_X^4) = o(b_X),\end{aligned}$$

and $k = 5$ to:

$$\begin{aligned}\frac{1}{120}f^{(0,5)}(x,y)E_K[(W-y)^5] &= \frac{1}{120}f^{(0,5)}(x,y)(24\rho_{b_Y}(y)b_Y^5 + 20\rho_{b_Y}(y)^2b_Y^5) = o(b_Y) \\ \frac{1}{12}f^{(2,3)}(x,y)E_K[(V-x)^2]E_K[(W-y)^3] &= \frac{1}{12}f^{(2,3)}(x,y)2b_Xxb_Y^2y = o(b_Y) \\ \frac{1}{12}f^{(3,2)}(x,y)E_K[(V-x)^3]E_K[(W-y)^2] &= \frac{1}{12}f^{(3,2)}(x,y)2b_X^2xb_Yy = o(b_X) \\ \frac{1}{120}f^{(5,0)}(x,y)E_K[(V-x)^5] &= \frac{1}{120}f^{(5,0)}(x,y)(24\rho_{b_X}(x)b_X^5 + 20\rho_{b_X}(x)^2b_X^5) = o(b_X).\end{aligned}$$

To obtain the final expression for the expected value of $\hat{f}(x,y)$, only the last term has to be computed. As mentioned above, for this term it is assumed that, for $l = 0, \dots, 5$, $f^{(l,5-l)}$ is Lipschitz continuous. This means that there is a $C_l > 0$ such that

$$|f^{(l,5-l)}(r,s) - f^{(l,5-l)}(t,u)| \leq C_l\sqrt{(r-t)^2 + (s-u)^2} \leq C_l(|r-t| + |s-u|),$$

where the last inequality follows from the triangle inequality. Furthermore, note that $|\xi_V - x| \leq |V - x|$ and $|\xi_W - y| \leq |W - y|$. Then the expected value in the last term of (14), denoted by T_l , equals

$$\begin{aligned}T_l &= E_K\left[\left(f^{(l,5-l)}(\xi_V, \xi_W) - f^{(l,5-l)}(x,y)\right)(V-x)^l(W-y)^{5-l}\right] \\ &\leq E_K\left[C_l(|\xi_V - x| + |\xi_W - y|)(V-x)^l(W-y)^{5-l}\right] \\ &\leq E_K\left[C_l(|V-x| + |W-y|)(V-x)^l(W-y)^{5-l}\right] \\ &= C_l\left(E_K[|V-x|(V-x)^l]E_K[(W-y)^{5-l}] + E_K[(V-x)^l]E_K[|W-y|(W-y)^{5-l}]\right).\end{aligned}$$

From the Cauchy-Schwarz inequality it is known that

$$|E_K[|V-x|(V-x)^l]| \leq \sqrt{E_K[(V-x)^2]}\sqrt{E_K[(V-x)^{2l}]}.\quad (15)$$

In Appendix A, the first ten central moments of a $\text{Gamma}(a,b)$ distributed random variable can be found. From these moments it follows that (15), and thus the first term of T_l , is $o(h_X)$ for $l \geq 2$. Analogous, it holds that the second term of T_l is $o(h_Y)$ for $l \leq 3$. Furthermore, since

$$\begin{aligned}E_K[(W-y)^{5-l}] &= o(h_Y) \quad \text{for } l = 1, 2, \\ E_K[(V-x)^l] &= o(h_X) \quad \text{for } l = 4, 5,\end{aligned}$$

it can be concluded that both terms of T_l are either $o(h_X)$ or $o(h_Y)$, or even both, for each l . As a consequence,

$$T_l = o(h_X) \text{ or } T_l = o(h_Y) \text{ for all } l.$$

The expected value of $\hat{f}(x, y)$ can thus, by adding all the terms, be expressed as:

$$E_f \left[\hat{f}(x, y) \right] = f(x, y) + \frac{1}{2} f^{(2,0)}(x, y) b_X x + \frac{1}{2} f^{(0,2)}(x, y) b_Y y + \frac{1}{4} f^{(2,2)}(x, y) b_X b_Y xy + o(b_X + b_Y).$$

The bias of the estimator is therefore equal to

$$Bias \left(\hat{f}(x, y) \right) = \frac{1}{2} f^{(2,0)}(x, y) b_X x + \frac{1}{2} f^{(0,2)}(x, y) b_Y y + \frac{1}{4} f^{(2,2)}(x, y) b_X b_Y xy + o(b_X + b_Y),$$

which is $O(b_X + b_Y)$. If it is assumed that $b_X = O(n^{-\alpha})$ and $b_Y = O(n^{-\alpha})$, then the bias of the bivariate Gamma kernel density estimator is $O(n^{-\alpha})$.

Also the variance of the estimator can be derived similarly to the variance of the univariate estimator. Let for this derivation $V \sim \text{Gamma}(2\rho_{b_X}(x) - 1, \frac{b_X}{2})$ and $W \sim \text{Gamma}(2\rho_{b_Y}(y) - 1, \frac{b_Y}{2})$ be independent, with their densities denoted by K_V and K_W respectively. From the previous result (10) for the univariate estimator, it is known that

$$\text{Var}_f \left(\hat{f}(x, y) \right) = \frac{1}{n} E_f \left[(K_X(X_1) K_Y(Y_1))^2 \right] + O(n^{-1}).$$

Analogous to the univariate situation (11), the expected value can be derived as follows:

$$\begin{aligned} E_f \left[(K_X(X_1) K_Y(Y_1))^2 \right] &= \int_0^\infty \int_0^\infty K_X(v)^2 K_Y(w)^2 f(v, w) dv dw \\ &= \int_0^\infty \int_0^\infty \frac{v^{2\rho_{b_X}(x)-2} e^{-\frac{2v}{b_X}}}{b_X^{2\rho_{b_X}(x)} \Gamma(\rho_{b_X}(x))^2} \frac{w^{2\rho_{b_Y}(y)-2} e^{-\frac{2w}{b_Y}}}{b_Y^{2\rho_{b_Y}(y)} \Gamma(\rho_{b_Y}(y))^2} f(v, w) dv dw \\ &= \int_0^\infty \int_0^\infty \frac{b_X^{-1} \Gamma(2\rho_{b_X}(x) - 1)}{2^{2\rho_{b_X}(x)-1} \Gamma(\rho_{b_X}(x))^2} \frac{b_Y^{-1} \Gamma(2\rho_{b_Y}(y) - 1)}{2^{2\rho_{b_Y}(y)-1} \Gamma(\rho_{b_Y}(y))^2} K_V(v) K_W(w) f(v, w) dv dw \\ &= B_{b_X, b_Y}(x, y) E_K[f(V, W)], \end{aligned}$$

with $B_{b_X, b_Y}(x, y) = \frac{b_X^{-1} \Gamma(2\rho_{b_X}(x) - 1)}{2^{2\rho_{b_X}(x)-1} \Gamma(\rho_{b_X}(x))^2} \frac{b_Y^{-1} \Gamma(2\rho_{b_Y}(y) - 1)}{2^{2\rho_{b_Y}(y)-1} \Gamma(\rho_{b_Y}(y))^2}$, and E_K the expectation under K_{VW} , with $K_{VW}(v, w) = K_V(v) K_W(w)$, the joint density of V and W . For $B_{b_X, b_Y}(x, y)$ a bound can be established as done in (12) for the variance of the univariate estimator:

$$B_{b_X, b_Y}(x, y) \leq \frac{(\rho_{b_X}(x) - 1)^{-\frac{1}{2}} (\rho_{b_Y}(y) - 1)^{-\frac{1}{2}}}{4\pi} b_X^{-1} b_Y^{-1}.$$

It is still assumed that $x \geq 2b_X$ and $y \geq 2b_Y$, so it can then be concluded similar to (13) that for $\frac{x}{b_X} \rightarrow \infty$ and $\frac{y}{b_Y} \rightarrow \infty$

$$B_{b_X, b_Y}(x, y) \sim \frac{(\rho_{b_X}(x) - 1)^{-\frac{1}{2}} (\rho_{b_Y}(y) - 1)^{-\frac{1}{2}}}{4\pi} b_X^{-1} b_Y^{-1}.$$

As a result,

$$\text{Var}_f \left(\hat{f}(x, y) \right) \sim \frac{(\rho_{b_X}(x) - 1)^{-\frac{1}{2}} (\rho_{b_Y}(y) - 1)^{-\frac{1}{2}}}{4\pi} b_X^{-1} b_Y^{-1} n^{-1} f(x, y).$$

Since $b_X = O(n^{-\alpha})$ and $b_Y = O(n^{-\alpha})$, it follows that $\rho_{b_X}(x) - 1 = O(n^\alpha)$ and $\rho_{b_Y}(y) - 1 = O(n^\alpha)$ and thus that $\text{Var}_f \left(\hat{f}(x, y) \right) = O(n^{\alpha-1})$. Then the optimal rate for the *MSE* of the bivariate Gamma kernel density estimator can be obtained. Since

$$MSE(\hat{f}(x, y)) = \text{Bias}(\hat{f}(x, y))^2 + \text{Var}_K(\hat{f}(x, y)) = O(n^{-2\alpha}) + O(n^{\alpha-1}),$$

the optimal rate is determined by equating the orders. These orders are equal if $-2\alpha = \alpha - 1$, so for $\alpha = \frac{1}{3}$, which results in an optimal rate of $O(n^{-\frac{2}{3}})$. Compared to the optimal rate of the *MSE* of the univariate kernel density estimator with modified Gamma kernels, the rate is less. However, this is a well-known phenomenon often referred to as the Curse of Dimensionality, which states that the best rate for the *MSE* decreases as the dimension increases. As shown in [7], this best rate is equal to $O(n^{-\frac{4}{4+d}})$, where d is the number of dimensions. The rate of the *MSE* of the bivariate ($d = 2$) kernel density estimator with modified Gamma kernels is therefore equal to the best possible rate.

In the next section, a data set will be given as an example of observed data from the Oriented Cylinder Model. Based on this example data set of observed heights H and squared half-widths Z of rectangles on the cut-plane, the bivariate modified Gamma kernel density estimator for the bivariate distribution of Z and H will be constructed. Moreover, this example will be continued in the subsequent chapters.

3.4 Estimator for the density of the height and squared half-width of a rectangle

The methods discussed in the previous sections will be applied to an example data set in this section. This data set is drawn from a known distribution that is more realistic than the previous explanatory examples. Details about this distribution, and about the way the data is drawn, can be found in Section 3.4.1. In the subsequent section this data set will be used to construct the *LSCV*-optimal bivariate modified Gamma kernel density estimator $\hat{g}$ for this data. Since the true density g is also known, the estimator $\hat{g}$ can be compared to this density, which is also done in Section 3.4.2.

3.4.1 Data simulation

First the distributions of X and of H given $X = x$ are chosen. For X the *Gamma*(3, 1) distribution is chosen and for H given $X = x$ the triangular distribution on $[0, x]$. This leads to the following densities:

$$\begin{aligned} f_X(x) &= \frac{1}{2}x^2e^{-x}, & x \geq 0 \\ f_{H|X}(h|x) &= \frac{2}{x^2}(x-h), & 0 \leq h \leq x. \end{aligned} \tag{16}$$

The joint density can then be derived from these densities by multiplying the two former densities:

$$f(x, h) = f_{X,H}(x, h) = f_{H|X}(h|x)f_X(x) = e^{-x}(x-h), \quad \text{for } x \geq 0, 0 \leq h \leq x.$$

The bivariate density of Z and H can then be calculated from the bivariate density of X and H , by making use of the known relationship (4). This leads to the following densities, for $z, h \geq 0$:

$$\begin{aligned}
g_Z(z) &= \frac{4}{15} \left(z^2 + z + \frac{3}{4} \right) e^{-z} \\
g_{H|Z}(h|z) &= \frac{2 \left(\frac{1}{2} + z - h \right)}{z^2 + z + \frac{3}{4}} 1_{\{0 < h < z\}} + \frac{2 \left[\left(\frac{1}{2} + z - h \right) I_G \left(\frac{1}{2}, h - z \right) + e^{-(h-z)} \sqrt{h-z} \right]}{\sqrt{\pi} \left(z^2 + z + \frac{3}{4} \right)} 1_{\{h > z\}} \\
&= g_1(h|z) 1_{\{0 < h < z\}} + g_2(h|z) 1_{\{h > z\}} \\
g(z, h) &= g_{Z,H}(z, h) = g_{H|Z}(h|z) g_Z(z) = \frac{8}{15} \left(\frac{1}{2} + z - h \right) e^{-z} 1_{\{0 < h < z\}} \\
&\quad + \frac{8e^{-z}}{15\sqrt{\pi}} \left[\left(\frac{1}{2} + z - h \right) I_G \left(\frac{1}{2}, h - z \right) + e^{-(h-z)} \sqrt{h-z} \right] 1_{\{h > z\}}.
\end{aligned}$$

In this case the true densities f and g are known, so an estimator for one of these densities can be compared with the true density. Because g_Z and $g_{Z|H}$ are known, it is possible to draw a sample from these distributions. The distribution of H is conditional on Z , so the realisations of Z are drawn first. Observe that the density of Z is a mixture of three Gamma distributions, namely $\text{Gamma}(3, 1)$, $\text{Gamma}(2, 1)$ and $\text{Gamma}(1, 1)$, where the density γ_α of a $\text{Gamma}(\alpha, 1)$ distribution equals

$$\gamma_\alpha(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \quad x \geq 0.$$

Then the density of Z can be expressed as a finite mixture of these Gamma densities:

$$g_Z(z) = \frac{8}{15} \gamma_3(z) + \frac{4}{15} \gamma_2(z) + \frac{3}{15} \gamma_1(z).$$

As a result, drawing from g_Z is the same as drawing from γ_3 , γ_2 and γ_1 with a probability of $\frac{8}{15}$, $\frac{4}{15}$ and $\frac{3}{15}$ respectively. Because these distributions are standard distributions, for which functions are available in the statistical software package *R*, the sample of Z can then be drawn.

It is more difficult to draw a sample of H . As mentioned above, the distribution of H is conditional on Z , and is therefore different for all realisations. Moreover, each density is divided into two parts, depending on whether H is less than or bigger than Z . The probability that H is less than Z , given that $Z = z$, is equal to:

$$\begin{aligned}
P_z &= P(H < Z | Z = z) = P(H < z | Z = z) \\
&= \int_0^z g_{H|Z}(h|z) dh = \int_0^z \frac{2 \left(\frac{1}{2} + z - h \right)}{z^2 + z + \frac{3}{4}} dh \\
&= \left[\frac{2 \left(\frac{1}{2} h + zh - \frac{h^2}{2} \right)}{z^2 + z + \frac{3}{4}} \right]_{h=0}^{h=z} = \frac{z^2 + z}{z^2 + z + \frac{3}{4}}.
\end{aligned}$$

Each realisation of H is then drawn from a distribution proportional to either g_1 or g_2 with probability P_z and $1 - P_z$ respectively. In the first case, the Inverse Sampling method can be used. This method uses the relationship $G_1(h|z) = u \Leftrightarrow h = G_1^{-1}(u|z)$, where G_1 is the cumulative distribution function related to the density proportional to g_1 , and $0 < u < 1$. First u is drawn from a uniform distribution on $[0, 1]$, then the second relationship is used to obtain

a realisation h . It is not always possible to get an expression for the inverse of the cumulative distribution function, but for $H < Z$, it is possible:

$$g_1(h|z) \propto \frac{1}{2} + z - h \text{ leads to } h = z + \frac{1}{2} - \frac{1}{2}\sqrt{(2z+1)^2 - 4uz(1+z)} = G_1^{-1}(u|z).$$

A realisation h of H corresponding to a realisation z of Z can then be constructed by drawing u from a uniform distribution on $[0, 1]$ and by using these z and u in the equation above to compute h .

In the second case, $H > Z$, Rejection Sampling is used to draw a sample of H . For this method, a u_1 is drawn from a density ϕ that lies completely above a constant c times the function g_2 (note: not necessarily a density) from which the realisation originally should be drawn. After that a u_2 is drawn from a uniform distribution on $[0, \phi(u_1)]$. If u_2 is bigger than $cg_2(u_1)$ the draw u_1 will be rejected, otherwise u_1 is kept as a realisation of H . Then only a density ϕ has to be found with the above-mentioned property. Consider $Y \sim \text{Cauchy}(0, 0.5)$, then the density of the random variable $|Y|$ satisfies these properties. This density is equal to twice the density of the $\text{Cauchy}(0, 0.5)$ distribution restricted to $[0, \infty)$. Since this is a standard distribution, there are functions available in R to draw from this distribution.

In summary, for both cases a method is available to draw the data from the distributions mentioned above. Inverse Sampling is used in the case that $H < Z$, otherwise Rejection Sampling is used. Furthermore, the probability that $H < Z$ is known, so a sample can be drawn from the density g .

3.4.2 Construction of the estimator

With the above-mentioned procedure a data set of size $n = 100$ of heights and squared half-widths of rectangles is simulated, which is exactly the observable data. With this data set a bivariate modified Gamma kernel density estimator for the bivariate density of the height and squared half-widths of a rectangle can be constructed. Because also the true density is known, the estimator can be compared with this true density.

The *LSCV*-optimal bandwidths for Z and H can be determined as described in Section 3.2.2, with the same adjustments as for the univariate Gamma kernel density estimator: The estimator is replaced with the bivariate modified Gamma kernel density estimator, and the lower bounds of the integrals are equal to 0. In Figure 21, a contour plot of the corresponding function CV is shown.

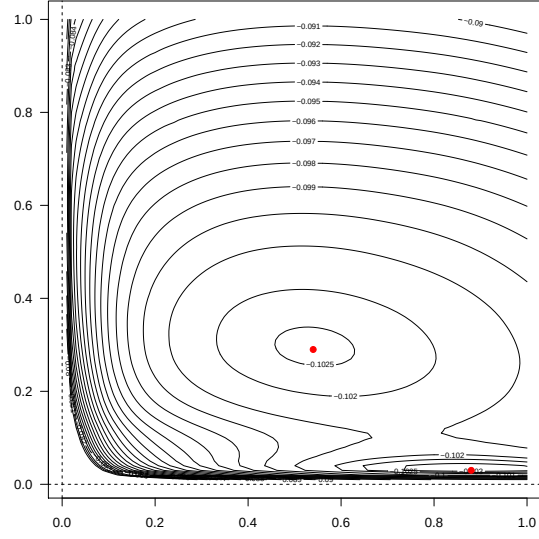


Figure 21: Contour plot of the values of CV .

As can be seen from the figure above, the function CV has two minima, one local minimum in $(b_Z, b_H) = (0.54, 0.29)$ and a global minimum in $(b_Z, b_H) = (0.88, 0.03)$. The bivariate Gamma kernel density estimator, as described in the previous sections, is constructed for both these pairs of bandwidth parameters, and shown in Figure 22.

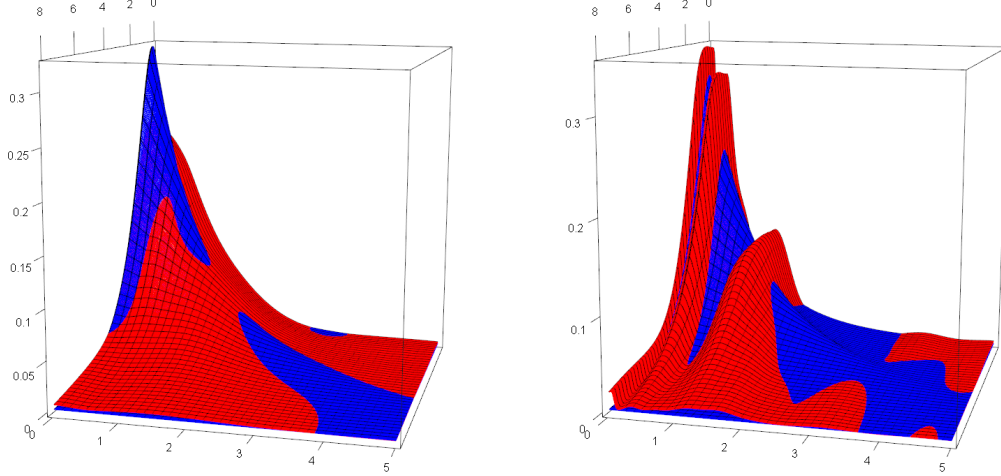


Figure 22: The true density (blue) and the estimator for g (red) with $(b_Z, b_H) = (0.54, 0.29)$ in the left plot and $(b_Z, b_H) = (0.88, 0.03)$ in the right plot.

It can be clearly seen that the estimator in the right plot is much rougher, which is explained by the much smaller bandwidth corresponding to the variable H . Although the bandwidth parameter corresponding to Z is bigger in this case, resulting in a smoother estimator in that direction,

the estimator is overall too rough. Since for the estimation of f this estimator has to be transformed, it is chosen to use the smoother estimator. This estimator is close to the true density for larger values, while the estimator is less accurate near the origin. Note that the rougher estimator is, due to the relatively small bandwidth b_H , much closer to the true density near the origin.

The chosen *LSCV*-optimal bandwidths are thus equal to $\hat{b}_Z = 0.54$ and $\hat{b}_H = 0.29$. The bivariate Gamma kernel density estimator with these bandwidth parameters, denoted by $\hat{g}$, is again plotted in Figure 23, where the estimator is shown from different angles to graphically investigate the estimator further.

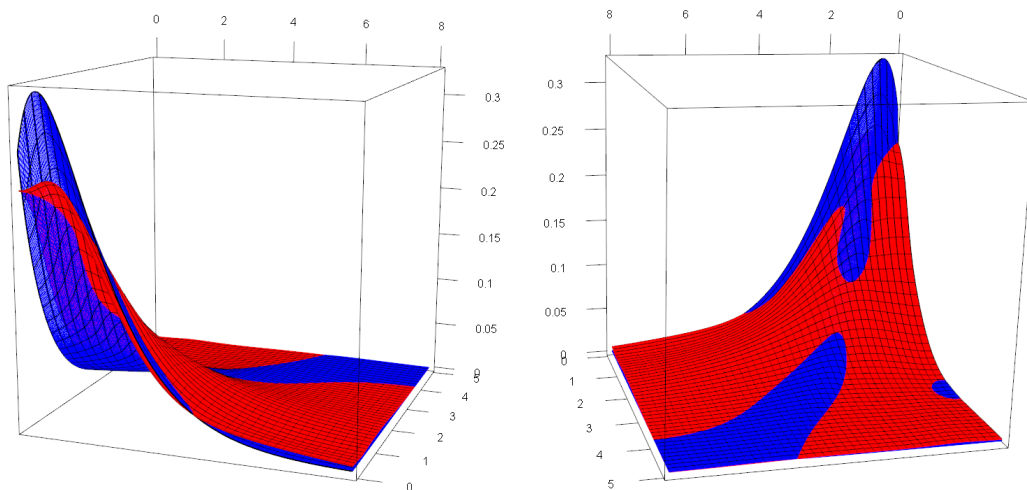


Figure 23: The estimator for g (red) with the true density (blue) from different angles.

As mentioned above, the estimator is very similar to the true density, while especially around the origin the differences are relatively big. Near the boundary, only the high peak of the true density is not estimated accurately. However, even for a relatively small data set of size $n = 100$, the Gamma kernel density estimator performs well. The first step in constructing an estimator for f is established: An unbiased estimator for the density g of the height and squared half-width of a rectangle is constructed.

In the next chapter this estimator will be used to obtain an estimator for the bivariate density of the height H and squared radius X of a cylinder. This last density is obtained by transforming the estimator $\hat{g}$ into the estimator $\hat{f}$ with a numerical procedure. It is therefore important to find the best estimator, since each error made by this estimator might be strengthened by the numerical procedure. The estimator that is constructed in this section minimizes the Mean Squared Error, and is therefore a good starting point for the construction of the estimator $\hat{f}$.

4 Estimator for the density of the height and squared radius of a cylinder

In the previous chapter an estimator for the bivariate density of the height and squared half-width of a rectangle was constructed. In this chapter, the procedure to transform this density into the density of the height and squared radius of a cylinder will be described. For this transformation the relationship (4) between f and g , will be used. To construct the estimator $\hat{f}$, the true density g is replaced by the estimator $\hat{g}$:

$$\hat{f}(x, h) = -\frac{1}{E[Z^{-\frac{1}{2}}]} \frac{\partial}{\partial x} \int_{z=x}^{\infty} (z-x)^{-\frac{1}{2}} \hat{g}(z, h) dz.$$

From this equation it can be seen that the transformation of $\hat{g}$ into $\hat{f}$ involves integration and differentiation. Because the expression of the kernel density estimator $\hat{g}$ that was found in the previous chapter is complicated, the integration and differentiation cannot be done analytically, but have to be done numerically. After that, only the expectation $E[Z^{-\frac{1}{2}}]$ has to be estimated. For the observed realisations $z_1, \dots, z_n$ of Z , this expected value is estimated by

$$\hat{E}[Z^{-\frac{1}{2}}] = \frac{1}{n} \sum_{i=1}^n z_i^{-\frac{1}{2}}. \quad (17)$$

First, in Section 4.1, the numerical procedure behind this transformation will be discussed. The methods that were used for the numerical integration and differentiation will be explained. After that, in Section 4.2, the transformation will be carried out to construct the estimator for f based on the estimator $\hat{g}$ constructed with the example data set of the previous chapter. In the last section, a different approach to choose the optimal bandwidth parameters will be discussed.

4.1 Transformation procedure

First the estimator $\hat{g}$, multiplied by a function of z and x , is numerically integrated with the Trapezoidal Rule. An example will be considered in Section 4.1.1 to explain this method. The constructed integral will then be differentiated with Finite Differences, which will be discussed in Section 4.1.2.

4.1.1 Numerical integration with the Trapezoidal Rule

For the numerical integration the Trapezoidal Rule is used. Consider the following figure for an explanation of this method.

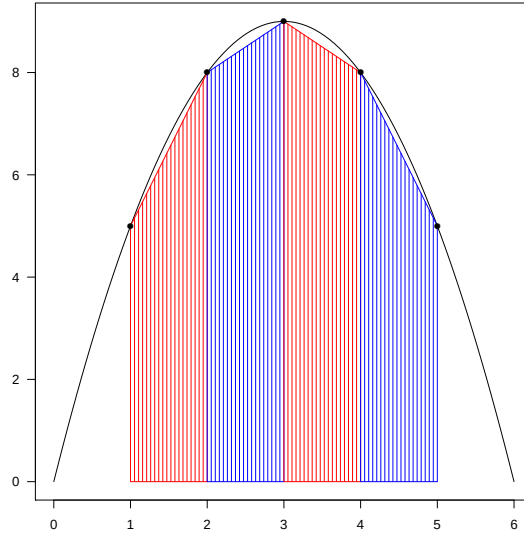


Figure 24: Trapezoidal Rule example.

Suppose the integral from 1 to 5 of the function (black line) in Figure 24 has to be computed. The function values are only known on a certain grid (thick points), which in this case consists of the points $x = 1, 2, 3, 4$ and 5 . Then the integral of interest can be estimated by taking the sum of the areas of the drawn red and blue shaded trapezoids. The area of each trapezoid can be calculated by taking the mean of the left and right function value and multiplying it by the width Δx . This results in the following estimate of the integral:

$$\sum_{i=1}^{n-1} \frac{f(x_i) + f(x_{i+1})}{2} \Delta x.$$

For the example the function $f(x) = -x^2 + 6x$ was chosen, so the integral can also be computed analytically. The true value of the integral is therefore known:

$$\int_1^5 f(x) dx = 30 \frac{2}{3}.$$

In the example $\Delta x = 1$, $n = 5$ and $x_i = i$ for $i = 1, \dots, 5$. The estimated value of the integral with the Trapezoidal Rule is therefore:

$$\sum_{i=1}^4 \frac{f(i) + f(i+1)}{2} = 30.$$

Note that this estimate is based on a coarse grid ($\Delta x = 1$), but the relative error is nevertheless only $2.22 \cdot 10^{-2}$. When considering a finer grid, for instance a grid with $\Delta x = 0.1$, the relative error will become even smaller, namely $2.22 \cdot 10^{-4}$. A decrease by a factor 10 in the step size leads to a decrease by a factor 100 in relative error.

This idea of estimating an integral can be applied to our case, where a three-dimensional function of the variables h , x and z has to be integrated over the variable z . For fixed variables h and x ,

the function amounts to a one-dimensional function of z . So for each fixed h and x this integral can be evaluated, by means of the above-mentioned method, which results in a two-dimensional function of the variables h and x .

4.1.2 Numerical differentiation with Finite Differences

Numerical differentiation at interior points of the grid is done by using Central Differences. This means that the derivative of f at an interior point x is approximated with the slope of the straight line through the values of the function at the grid points to the left and right of x . On a grid that consists of the ordered points $x_1, \dots, x_k$, the derivative at interior points can be estimated by

$$\hat{f}'_c(x_j) = \frac{f(x_{j+1}) - f(x_{j-1}))}{2\Delta x}, \text{ with } \Delta x = x_{j+1} - x_j,$$

for $j = 2, \dots, k-1$.

However, at the points on the edge of the grid, there is no point either to the left or to the right of that point. For this reason, for these points Forward and Backward Differences are used respectively. A forward difference can be calculated as the slope of the straight line through the values of the function at the grid point itself and the grid point to the right. Analogous to this, for backward difference, the grid point to the left is used instead of the one to the right. So at these points the derivative is approximated by:

$$\begin{aligned} \hat{f}'_f(x_1) &= \frac{f(x_2) - f(x_1)}{\Delta x} \\ \hat{f}'_b(x_k) &= \frac{f(x_k) - f(x_{k-1}))}{\Delta x}. \end{aligned}$$

These methods are graphically described in Figure 25, where the same function as in the example of the previous section is used.

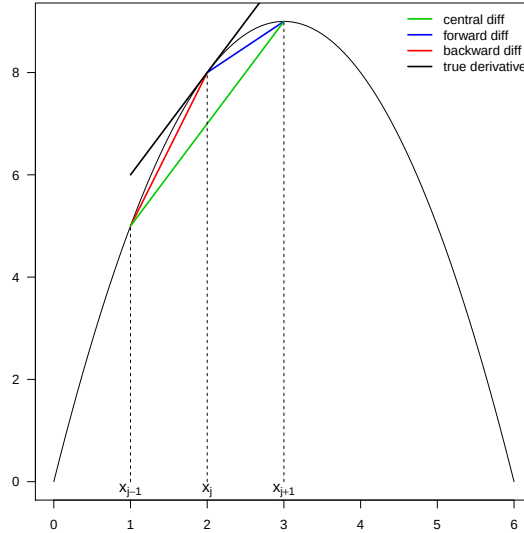


Figure 25: Forward, Central and Backward Differences example.

As can be seen from the figure above, the forward and backward differences result in an estimate that is less accurate than the central difference, which in general is the case. As a consequence, the estimates of the derivative at the boundaries will be worse. However, the accuracy will increase as the step size decreases.

In the next section, the above-described methods for numerical integration and differentiation will be applied to the relationship mentioned above to construct the estimator $\hat{f}$.

4.2 Construction of the estimator

The transformation of the density is divided into three parts: estimating the expectation, integrating the function and differentiating the obtained integral. After these three computations have been done, the estimator for the bivariate distribution of the height and squared radius of a cylinder will be constructed.

First, the expectation is estimated as described above. The data set of Section 3.4, consisting of $n = 100$ pairs (z_i, h_i) of heights and squared half-widths, results in an estimated expected value equal to

$$\widehat{E}[Z^{-\frac{1}{2}}] = \frac{1}{n} \sum_{i=1}^n z_i^{-\frac{1}{2}} = 0.9254308.$$

The next step is to compute the integral over z of the three-dimensional function $(z-x)^{-\frac{1}{2}}\hat{g}(z, h)$. Since $\hat{g}$ has already been constructed, the whole integrand is known and therefore the integral can be computed numerically. By fixing h and x , the three-dimensional function becomes a one-dimensional function, which can be integrated using the Trapezoidal Rule. For every h and x the integral value can be approximated, which will lead to a function of h and x .

After this has been done, the obtained function of x and h can be differentiated with respect to x . Similar to the integration, the variable h is then fixed, in order to obtain a one-dimensional function of x . Then the numerical differentiation method discussed in the previous section can be applied to this function, which will lead to the approximate function $\hat{k}$ of h and x . Then the final step of constructing the estimator for f can be taken. According to the relationship, $\hat{k}$ has to be divided by minus the expectation, or in notation:

$$\hat{f}(x, h) = -\frac{\hat{k}(x, h)}{\widehat{E}[Z^{-\frac{1}{2}}]}.$$

When the transformation procedure is performed, the estimator for f is found on a discrete grid of x and h . Since also the true density f is known, the constructed estimator $\hat{f}$ can be compared to the true density f . Both the estimator and the true density are plotted in Figure 26 from different angles.

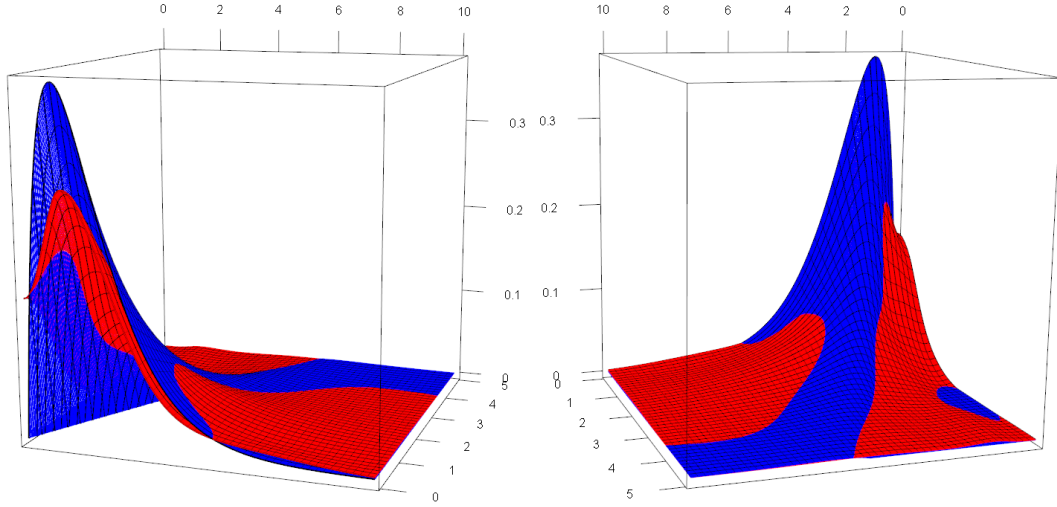


Figure 26: The estimator for f (red) with the true density (blue) from different angles.

As can be seen in Figure 26, the estimator $\hat{f}$ has the same properties as the estimator for the density g . For relative large values of x and h the estimator is close to the true density, while close to the boundary the estimator is less accurate. Furthermore, both estimators are about equally smooth, which is not a strange result, since the estimators can, though not easily, be transformed into each other. However, a big difference between the two estimators is that the error made in constructing $\hat{f}$ consists of two parts instead of only one. First g is estimated by $\hat{g}$, which involves an estimation error. After that, this estimator is used for numerical integration and differentiation, in which an approximation error is made. Due to these errors, it is possible that $\hat{f}$ becomes negative for some values. In the figure above, this is indeed the case, for example in the right plot, where on the right the estimator is below the true density, which has a value of 0 there. A possible solution to reduce the error made in estimating f , is to optimize the bandwidth parameters in a different way. This different method to find optimal bandwidths for the bivariate Gamma kernel density estimator of g will be derived in the last section of this chapter.

4.3 Bandwidth selection based on the density of interest

In the previous sections the optimal bandwidths were found by minimizing an estimate of the Integrated Squared Error of $\hat{g}$. However, the interest lies in the density f , so an alternative approach to obtain optimal bandwidths is to minimize the *ISE* of $\hat{f}$ instead of $\hat{g}$. Note that $\hat{f}$, and thus also its Integrated Squared Error, depends on the bandwidth parameters b_Z and b_H , since the estimator is constructed by transforming $\hat{g}$, which depends on these parameters. The *ISE* of $\hat{f}$ can be derived similarly as (7), which leads to the following expression that has to be minimized:

$$ISE(b_Z, b_H) = \int_0^\infty \int_0^\infty f(x, h)^2 dx dh = \int_0^\infty \int_0^\infty \hat{f}(x, h)^2 dx dh - 2 \int_0^\infty \int_0^\infty \hat{f}(x, h) f(x, h) dx dh.$$

The first term of the right hand side can be computed by constructing $\hat{f}$ as described before, and integrate both integrals numerically with the Trapezoidal Rule. The second term involves

the unknown density again, but estimating this term is harder than before. The method used before, involving the leave-one-out estimator, cannot be used in this case, since there is no data available of X . This term is therefore expressed differently, so that the term can be estimated. First the expression of f in terms of g is plugged into the equation, which yields

$$\int_0^\infty \int_0^\infty \hat{f}(x, h) f(x, h) dx dh = - \int_0^\infty \int_0^\infty \hat{f}(x, h) \frac{1}{E[Z^{-\frac{1}{2}}]} \frac{\partial}{\partial x} \int_x^\infty (z - x)^{-\frac{1}{2}} g(z, h) dz dx dh. \quad (18)$$

The partial derivative with respect to x that occurs in the expression can also be expressed differently. Under the assumption that g is Lipschitz continuous, the partial derivative can be rewritten as:

$$\begin{aligned} \frac{\partial}{\partial x} \int_x^\infty (z - x)^{-\frac{1}{2}} g(z, h) dz &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \left[\int_{x+\varepsilon}^\infty (z - x - \varepsilon)^{-\frac{1}{2}} g(z, h) dz - \int_x^\infty (z - x)^{-\frac{1}{2}} g(z, h) dz \right] \\ &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \left[\int_x^\infty (z - x)^{-\frac{1}{2}} g(z + \varepsilon, h) dz - \int_x^\infty (z - x)^{-\frac{1}{2}} g(z, h) dz \right] \\ &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \left[\int_x^\infty (z - x)^{-\frac{1}{2}} \{g(z + \varepsilon, h) - g(z, h)\} dz \right] \\ &= \int_x^\infty (z - x)^{-\frac{1}{2}} g^{(1,0)}(z, h) dz, \end{aligned}$$

where the limit and integration can be interchanged due to the fact that g is assumed to be Lipschitz continuous. Plugging the obtained result into (18) leads to

$$\begin{aligned} \int_0^\infty \int_0^\infty \hat{f}(x, h) f(x, h) dx dh &= \frac{-1}{E[Z^{-\frac{1}{2}}]} \int_0^\infty \int_0^\infty \hat{f}(x, h) \int_x^\infty (z - x)^{-\frac{1}{2}} g^{(1,0)}(z, h) dz dx dh \\ &= \frac{-1}{E[Z^{-\frac{1}{2}}]} \int_0^\infty \int_0^\infty \int_x^\infty \hat{f}(x, h) (z - x)^{-\frac{1}{2}} g^{(1,0)}(z, h) dz dx dh \\ &= \frac{-1}{E[Z^{-\frac{1}{2}}]} \int_0^\infty \int_0^\infty \int_0^z \hat{f}(x, h) (z - x)^{-\frac{1}{2}} g^{(1,0)}(z, h) dx dz dh \\ &= \frac{-1}{E[Z^{-\frac{1}{2}}]} \int_0^\infty \int_0^\infty \int_0^z \hat{f}(x, h) (z - x)^{-\frac{1}{2}} dx g^{(1,0)}(z, h) dz dh. \quad (19) \end{aligned}$$

The integral over z and x can be computed with integration by parts, which results in:

$$\begin{aligned}
\int_0^\infty \int_0^z \hat{f}(x, h)(z-x)^{-\frac{1}{2}} dx g^{(1,0)}(z, h) dz &= \int_0^z \hat{f}(x, h)(z-x)^{-\frac{1}{2}} dx g(z, h) \Big|_{z=0}^{z=\infty} \\
&\quad - \int_0^\infty \frac{\partial}{\partial z} \left[\int_0^z \hat{f}(x, h)(z-x)^{-\frac{1}{2}} dx \right] g(z, h) dz \\
&= - \int_0^\infty \frac{\partial}{\partial z} \left[\int_0^z \hat{f}(x, h)(z-x)^{-\frac{1}{2}} dx \right] g(z, h) dz. \quad (20)
\end{aligned}$$

The partial derivative with respect to z can be rewritten as before, where it is assumed that also $\hat{f}$ is Lipschitz continuous:

$$\frac{\partial}{\partial z} \int_0^z \hat{f}(x, h)(z-x)^{-\frac{1}{2}} dx = \int_0^z \hat{f}^{(1,0)}(x, h)(z-x)^{-\frac{1}{2}} dx.$$

By plugging in (20) with this last result into (19), the final expression can be derived:

$$\begin{aligned}
\int_0^\infty \int_0^\infty \hat{f}(x, h) f(x, h) dx dh &= \frac{1}{E[Z^{-\frac{1}{2}}]} \int_0^\infty \int_0^\infty \int_0^z \hat{f}^{(1,0)}(x, h)(z-x)^{-\frac{1}{2}} dz g(z, h) dz dh \\
&= \frac{1}{E[Z^{-\frac{1}{2}}]} E_g \left[\int_0^Z \hat{f}^{(1,0)}(x, H)(Z-x)^{-\frac{1}{2}} dx \right].
\end{aligned}$$

The *ISE* of f is therefore minimized, when

$$\int_0^\infty \int_0^\infty \hat{f}(x, h)^2 dx dh - \frac{2}{E[Z^{-\frac{1}{2}}]} E_g \left[\int_0^Z \hat{f}^{(1,0)}(x, H)(Z-x)^{-\frac{1}{2}} dx \right]$$

is minimized. For observed realisations $(z_1, h_1), \dots, (z_n, h_n)$, the estimator $\hat{f}_{-i}$ is defined as the estimator constructed by transforming $\hat{g}_{-i}$, the estimator of the density g of (Z, H) based on all data except for (z_i, h_i) . Then the second expectation in the expression above, denoted by μ , can be estimated by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \left(\int_0^{z_i} \hat{f}_{-i}^{(1,0)}(x, h_i)(z_i - x)^{-\frac{1}{2}} dx \right).$$

Note that $\hat{\mu}$ depends on the bandwidth parameters b_Z and b_H , since $\hat{f}_{-i}^{(1,0)}$ depends on these parameters. Furthermore, the expected value of $Z^{-\frac{1}{2}}$ is estimated with (17). The optimal bandwidths based on the *ISE* of $\hat{f}$ can then be found by minimizing the function CV_f of b_Z and b_H , defined as

$$CV_f(b_Z, b_H) = \int_0^\infty \int_0^\infty \hat{f}(x, h)^2 dx dh - \frac{2\hat{\mu}}{E[Z^{-\frac{1}{2}}]}.$$

Each of the quantities in this function can be determined, so CV_f can be computed on a discrete grid. For this method, the computations are more time-consuming than before, since the computation of $\hat{\mu}$ involves n transformations of $\hat{g}_{-i}$ into $\hat{f}_{-i}$, and each of the latter estimates has to be transformed again, involving numerical differentiation and numerical integration. The function CV_f will therefore not be computed on a fine discrete grid, but first on a coarse grid, and after that in a neighbourhood around the minimum found on the coarse grid. In this way, the number of computations that have to be done will be relatively low. The contour plot of the function near its minimum can be found in the figure below.

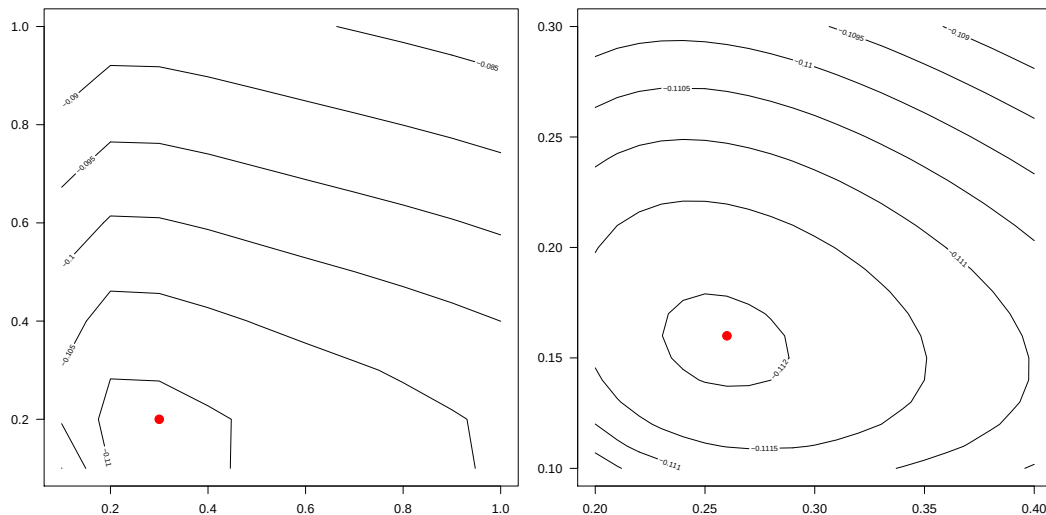


Figure 27: Contour plot of the function CV_f on a coarse grid (left) and fine grid (right).

As can be seen, on the coarse grid the function CV_f is minimal for the bandwidth parameters equal to $\hat{b}_Z = 0.3$ and $\hat{b}_H = 0.2$. After computing CV_f on the finer grid near this minimum, shown in the right plot, the optimal bandwidth parameters are approximated more accurately. The resulting bandwidth parameters are equal to $\hat{b}_Z = 0.26$ and $\hat{b}_H = 0.16$. Compared to the bandwidths found in Section 3.4.2, these bandwidth parameters are about half the value. These bandwidth parameters are therefore much smaller, which will lead to a rougher estimator. First the estimator $\hat{g}$ is constructed with these bandwidths, and shown in Figure 28.

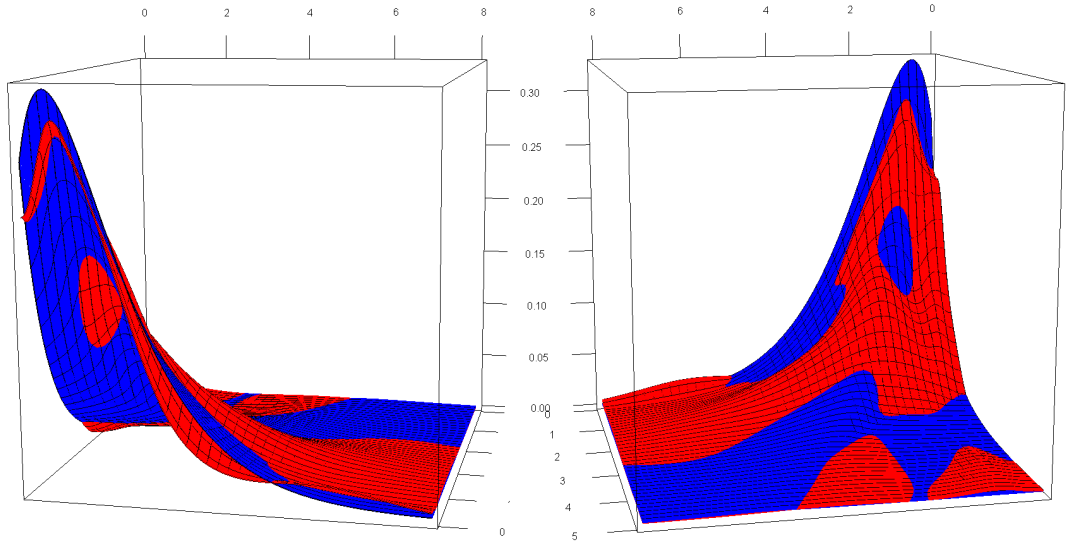


Figure 28: The estimator for g (red) with the true density (blue) from different angles, where the bandwidth is chosen based on the ISE of $\hat{f}$.

As can be seen from the figure below, the estimator is rougher than the estimator shown in Figure 23. However, the estimator is more accurate near the high peak of the true density. The estimator $\hat{f}$ constructed with these bandwidths, i.e. the transformed estimator $\hat{g}$ of Figure 28, is shown in Figure 29.

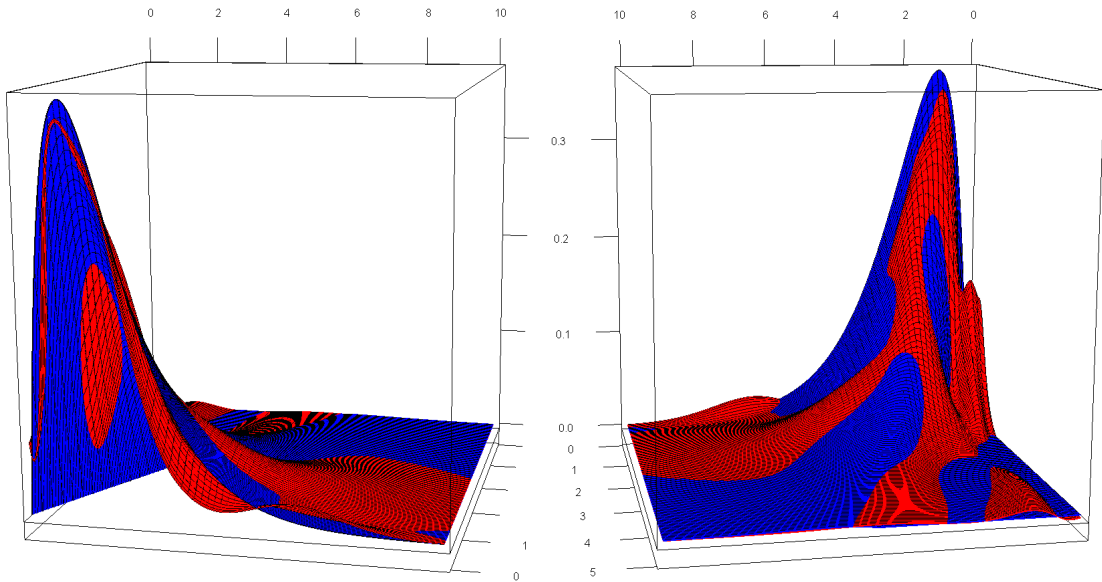


Figure 29: The estimator for f (red) with the true density (blue) from different angles, where the bandwidth is chosen based on the ISE of $\hat{f}$.

Compared to the estimator of Figure 26, constructed with bandwidth parameters chosen as minimizers of the *ISE* of $\hat{g}$, the shapes of the estimators are very similar. However, the estimator of the figure above is rougher, due to the smaller bandwidths. As a result, the estimator for f is more accurate near the high peak of this underlying density. Since the true density is known, for both estimators the true Integrated Squared Error can be computed. For the estimator constructed with bandwidth parameters chosen based on the *ISE* of $\hat{g}$ this error is equal to 0.0216, while for the estimator constructed with bandwidths chosen based on the *ISE* of $\hat{f}$ this error is equal to 0.0129. The latter error is significantly smaller, so it can be concluded that it is beneficial to consider this method to obtain the optimal bandwidths.

However, computations to obtain the optimal bandwidths with this method are very time-consuming, so this method is not yet applicable in practical situations. This method will thus not be used in the next chapter, where an estimator for the underlying density of an experimental data set will be constructed. The bandwidth parameters for this estimator are therefore chosen as described in Section 3.2.2. However, the estimators that are already constructed in this section with the alternative method for choosing the bandwidth parameters will still be used in Chapter 6 for the estimation of densities that can be derived from the bivariate density f .

5 Application to experimental data

In this chapter, the bivariate modified Gamma kernel density estimator for the density function of the height and squared radius of a cylinder will be constructed, based on experimental data gathered by K.S. McGarrity. This data was gathered by means of serial sectioning a steel plate, and examining the obtained cut-planes. Since the observed information on these cut-planes will not contain true rectangles of martensite grains, bounding boxes are used to translate the information on the cut-planes to information about rectangles, see [19] for more details. The experimental data consists of different quantities, of which only the width and height of the rectangles will be used. Moreover, only the data from the first cut-plane of the serial sectioning will be used. The data, which contain 89 widths and heights of rectangles observed from the first cut-plane, are graphically shown in the scatterplot below, where the squared half-widths are plotted against the heights.

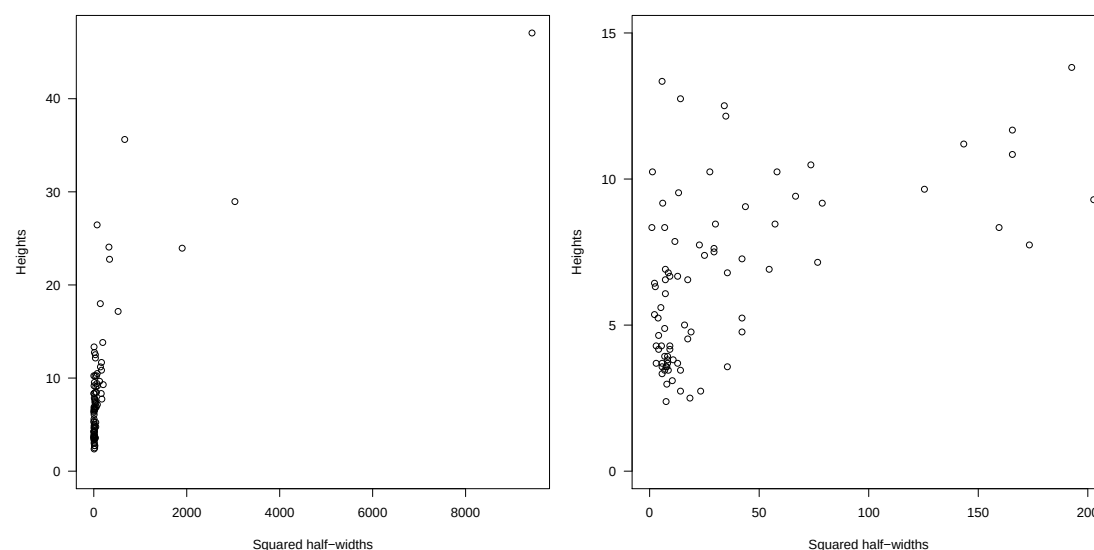


Figure 30: The experimental data of heights and squared radii of rectangles.

As can be seen from the right plot, much of the data is located in the area $[0, 200] \times [0, 15]$, while there are a couple of data points with a relatively large value. Large values for the squared half-width correspond to a large value for the height as well. This means that when a large width of the rectangle is observed, its height will also be relatively large. In other words, low but wide rectangles are not observed. Furthermore, the scale of both variables are different: The height takes on a maximum value of about 50, while the maximum value of the squared half-width is about 10000. However, the heights are spread more widely over the interval than the squared half-widths.

The underlying density of this experimental data set can be estimated with the previously described bivariate modified Gamma kernel density estimator. As mentioned above, the optimal bandwidths are chosen based on the *LSCV* method discussed in Section 3.2.2. A contour plot of *CV* is shown in Figure 31, where the minimum of *CV* is indicated by the red dot.

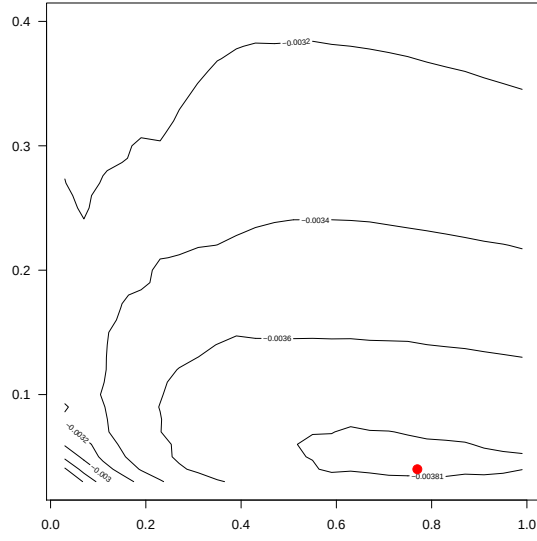


Figure 31: Contour plot of CV for the experimental data.

As can be seen from this figure, the experimental data leads to the $LSCV$ -optimal bandwidths $\hat{b}_Z = 0.77$ and $\hat{b}_H = 0.04$. Note that the two values differ much from each other. This is a result from the fact that the observed values of Z are much wider spread than the values of H . The obtained bandwidths are used to construct the estimator for the density of the height and squared half-width of a rectangle, after which this estimator is transformed into the corresponding estimator for the density of the height and squared radius of a cylinder. Both estimators are plotted in the figure below, the former in the left plot, the latter in the right plot.

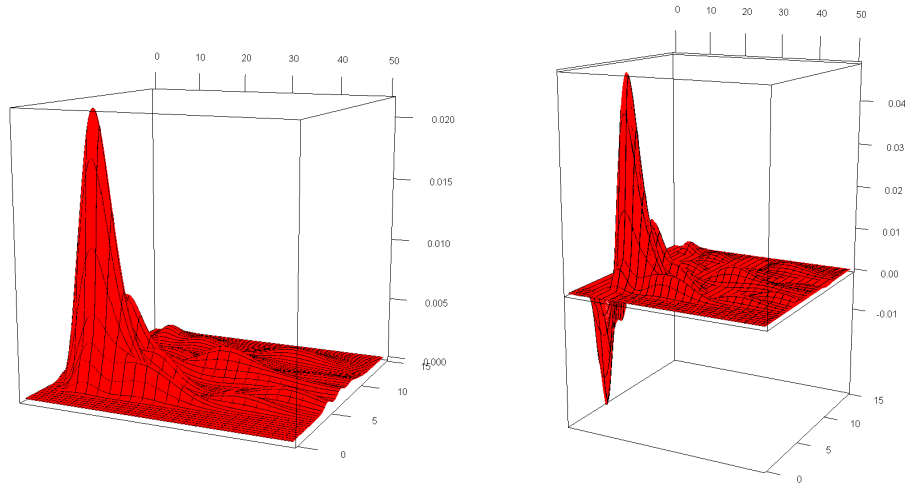


Figure 32: The estimator $\hat{g}$ (left) and $\hat{f}$ (right) based on the experimental data.

As can be seen in Figure 32, much of the mass is centered around the same area for both estimators. Moreover, both estimators are quite rough, with one high peak on the location where

much of the data is clustered (see Figure 30). However, the estimator for f becomes negative on a part of the domain. Since the estimator for g is not negative, because it is a bivariate kernel density estimator with positive kernels, the fact that f becomes negative is a result of the numerical transformation procedure. In this procedure, both errors in the estimation of g and errors in the approximations of the integration and differentiation can lead to negative values of the estimator for f .

Since the data of Z contains extreme outliers, it is also possible to remove these outliers, and construct an estimator based on the data of the right plot of Figure 30. The optimal bandwidth parameters for the data set without the outliers are equal to $\hat{b}_Z = 0.54$ and $\hat{b}_H = 0.05$. Since the data of Z are much closer to each other in this case, the *LSCV*-optimal bandwidth for Z is smaller than when the whole data set is considered. On the other hand, the *LSCV*-optimal bandwidth for H is more or less the same. The resulting estimators are shown in the figure below.

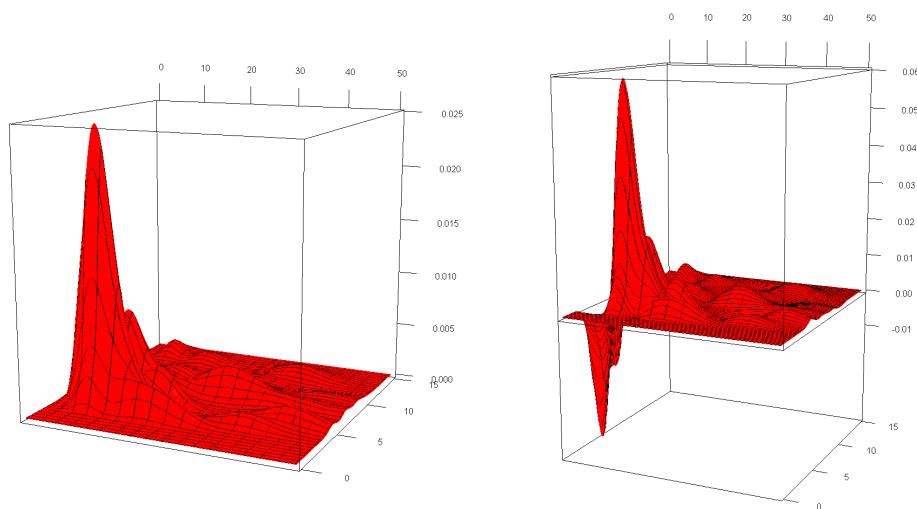


Figure 33: The estimator $\hat{g}$ (left) and $\hat{f}$ (right) based on the experimental data without the outliers.

The estimators in Figure 33 are almost equal to the estimators in Figure 32, where the latter estimators have larger tails due to the large observations. Furthermore, both the estimators for the density f are very similar, where both estimators show a large peak downwards, resulting in negative values of the estimates. The similarities between the estimators is not strange, since the two obtained *LSCV*-optimal bandwidth do not differ much from each other.

From the bivariate density of the height and squared radius of a cylinder, important marginal densities can be deduced. For example, the marginal density of the squared radius, or the marginal density of the volume of the cylinder. In the next section, these marginals will be constructed for the example data set of Section 3.4 as well as for the experimental data set of this section.

6 Important densities

From the bivariate density of the heights and squared radius of a cylinder, multiple densities can be deduced. In this chapter, two of these densities will be estimated based on the estimator $\hat{f}$. The first density that will be considered is the marginal density f_X of the squared radius X . This marginal can be computed by integrating the variable h out of the bivariate density, so

$$f_X(x) = \int_0^{\infty} f(x, h) dh.$$

Again, the integration will be done numerically, using the Trapezoidal Rule. More details about the estimation of the marginal density of X via the bivariate density f can be found in Section 6.1. The second density that will be discussed is the density of the volume V of a cylinder. The estimator for this density can be determined similar to the marginal of X , but the bivariate density f has to be integrated differently. Details about the estimation of this density will be discussed in Section 6.2. Although the density of each random variable that is a function of X and H can be deduced from the bivariate density f , this thesis is restricted to the two above-mentioned densities. The estimator for both these densities will be constructed for the underlying density of the experimental data set of Chapter 5 as well as for the underlying density of the example data set of Section 3.4.

6.1 Marginal density of the squared radius of a cylinder

First, the underlying marginal density of the squared radius X of the experimental data set will be estimated in this section. As described above, integrating the bivariate estimator $\hat{f}$ of the right plot of Figure 32 with respect to h results in the estimator for the marginal density of X . This marginal is also estimated based on the same data set in [19]. In the figure below, the estimator constructed via the estimator of the bivariate density f is shown in the left plot, while the estimators constructed in [19] are shown in the right plot.

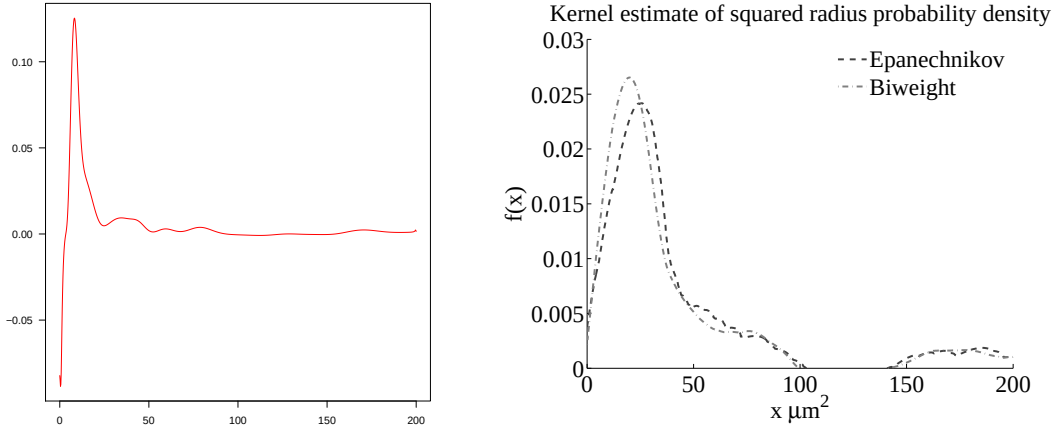


Figure 34: Estimators for the marginal density of the squared radius of a cylinder.

Note that the estimators in Figure 34 are very similar, with a high peak near the boundary for all three estimators. However, the peak of the estimator constructed in this thesis is higher and

closer to the boundary than the other two estimators. This is a result of the relatively small *LSCV*-optimal bandwidth parameters, which leads to a rough estimator. Furthermore, the estimator in the left figure is negative close to the boundary, which is a consequence of the fact that the corresponding bivariate estimator is negative near the boundary.

The marginal density of the squared radius can also be estimated based on the bivariate estimator $\hat{f}$ constructed with the example data set described in Section 3.4. Both the estimator constructed in Section 4.2 and the estimator constructed in Section 4.3 are used to estimate the marginal density of X . Since the true underlying density of the squared radius is also known, namely (16), the constructed estimators can be compared to the true density. Both the estimators and the true density are shown in the figure below.

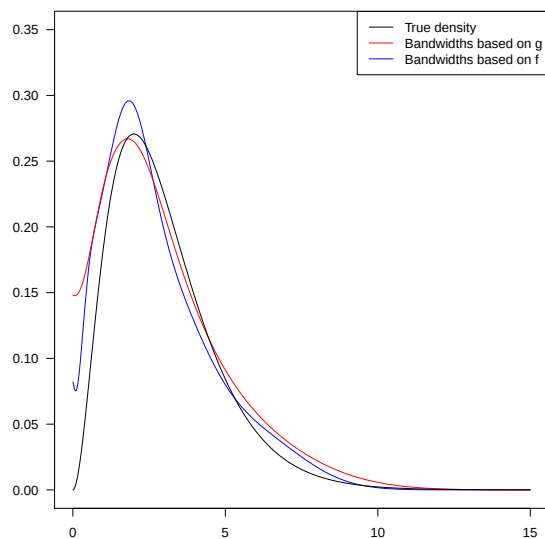


Figure 35: Estimators for the marginal density of the squared radius of a cylinder, together with the true density.

As can be seen in Figure 35, the estimators are very close to the true density. The estimator constructed with bandwidths chosen based on the *ISE* of $\hat{g}$ is smoother, while the estimator constructed with bandwidths chosen based on the *ISE* of $\hat{f}$ is rougher and, as a result, closer to the true density near the boundary. Remember that the estimators are constructed by estimating the density g , transforming this estimator into $\hat{f}$, and finally numerically integrating this estimator to obtain the marginal density. Despite the approximations that occur in the transformations, and the estimation of g based on a relatively small data set for bivariate estimation, the estimators are accurate. Note that especially close to the origin both estimators are too high. This may be a result of the approximations near the boundary, which are less accurate than the approximations in the interior.

6.2 Density of the volume of a cylinder

The other density function that will be estimated, is the density of the volume of a cylinder. The volume V of a cylinder is equal to

$$V = \pi X H.$$

The cumulative distribution function F_V of the volume V can thus be computed by:

$$F_V(v) = P(V \leq v) = P(\pi X H \leq v) = \int_0^\infty \int_0^{\frac{v}{\pi h}} f(x, h) dx dh.$$

The corresponding density f_V of V can then be obtained by differentiating the expression above with respect to v :

$$f_V(v) = \frac{d}{dv} \int_0^\infty \int_0^{\frac{v}{\pi h}} f(x, h) dx dh = \int_0^\infty f\left(\frac{v}{\pi h}, h\right) \frac{1}{\pi h} dh.$$

By plugging in the estimator $\hat{f}$, the estimator for f_V is obtained. First the estimator is constructed for the underlying density of the experimental data set of Chapter 5. Both the estimator constructed via $\hat{f}$ and the estimators constructed in [19] are shown in Figure 36 below.

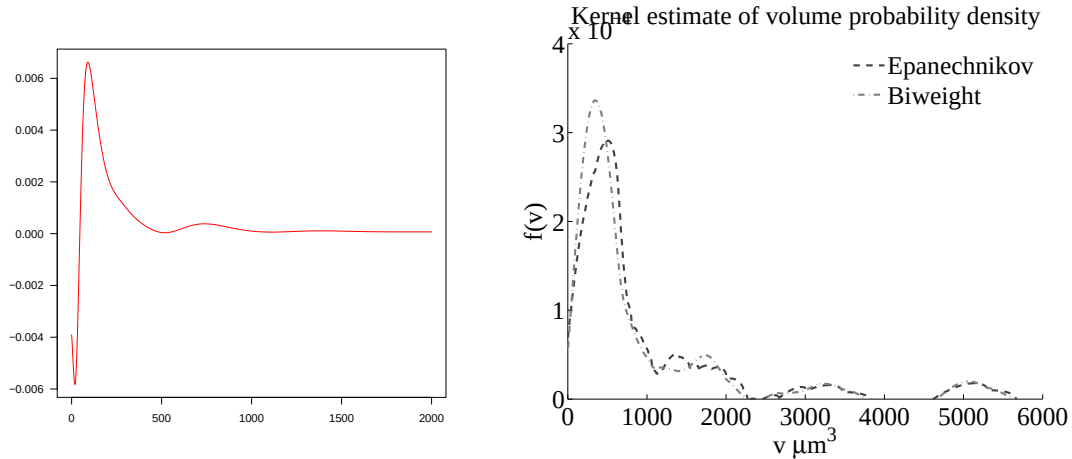


Figure 36: Estimators for the density of the volume of a cylinder.

Similar to the estimators for the marginal density of the squared radius, the peak of the estimator constructed via the bivariate estimator $\hat{f}$ is much narrower than the peak of the estimator constructed in [19]. This is again a result of the relatively small *LSCV*-optimal bandwidth parameters. Furthermore, the former estimator is again negative close to the boundary, while the latter estimator becomes negative for some larger values.

The estimator for the density of the volume of a cylinder can also be constructed for the underlying density of the example data set of Section 3.4, which is again known. Both the estimators constructed with the different methods for choosing the optimal bandwidth parameters and the true density are shown in Figure 37 below.

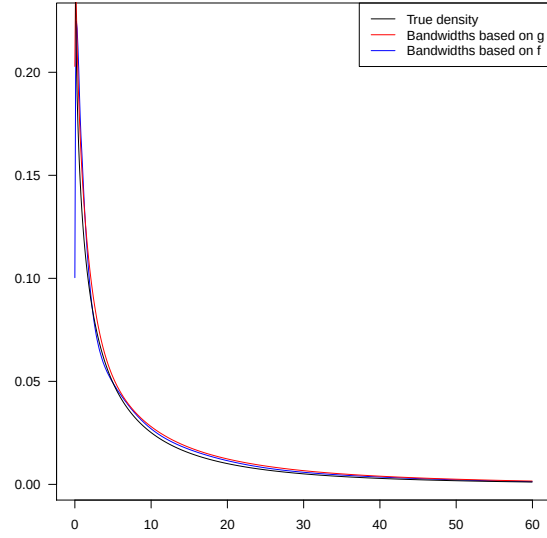


Figure 37: Estimators for the density of the volume of a cylinder, together with the true density.

As can be seen from the figure, the estimators are close to the true density. Also for the estimation of this density, which involves a slightly more complicated numerical integration, the resulting estimators are accurate. Note that in this case, the estimator constructed with bandwidths chosen based on the *ISE* of $\hat{g}$ is near the boundary closer to the true estimator than the estimator constructed with the bandwidths selected based on the *ISE* of $\hat{f}$, while away from the boundary, the latter estimator is closer. Overall, both estimators are accurate estimators for the true underlying density.

7 Conclusion and discussion

The microstructure of an object of dual phase steel is modelled with the Oriented Cylinder Model, where the oriented cylinders in a big block represent the martensite grains in the steel object. To obtain information about the microstructure, the steel object is cut. The two-dimensional information from the cut-plane of the steel object is translated to two-dimensional information about the rectangular visible profiles of the oriented cylinders on the cut-plane within the model. This information contains the heights and squared half-widths of the observed rectangles. For a data set of these random variables, an estimator is constructed for the underlying bivariate density function of the height and squared half-width of a rectangle. This density is estimated with a bivariate kernel density estimator, for which the bandwidth parameters are chosen with the Least Squares Cross-Validation method. This bandwidth selection method is based on the Integrated Squared Error of the estimator, where the optimal bandwidths are found by minimizing an estimate of the Integrated Squared Error. Since the data consists of lengths, all data is positive. The domain of the underlying bivariate density is therefore restricted to the first quadrant, which results in a boundary problem when using kernel density estimation. This boundary problem is solved by using a modified Gamma kernel as kernel function for the bivariate kernel density estimator. The resulting bivariate modified Gamma kernel density estimator for the height and squared half-width of a rectangle has no boundary bias and the Mean Integrated Squared Error converges to zero at optimal rate.

Via the Oriented Cylinder Model and the relationship between the height and squared half-width of a rectangle and the height and squared radius of a cylinder, a relationship between the bivariate densities of the two pairs of random variables is constructed. Through this relationship, the bivariate modified Gamma kernel density estimator of the former density is transformed into an estimator for the latter density. Since the microstructure of the steel object is modelled with the cylinders, the density of the height and squared radius of a cylinder gives information about the microstructure. Due to the complicated expression of the Gamma kernel density estimator and the complexity of the transformation, a numerical procedure is constructed to perform the transformation. Although both an estimation error (in the kernel density estimator) and an approximation error (in the numerical transformation) is made, the resulting estimator turned out to be an accurate estimator, even for small data sets. However, the mentioned errors can lead to an estimator that becomes negative for some values, which is a property the underlying density function cannot have.

Analogous to the bandwidth selection method based on the Integrated Squared Error of the bivariate modified Gamma kernel density estimator, a method is described to select the bandwidth parameters based on the Integrated Squared Error of the density of interest, i.e. the bivariate density of the height and squared radius of a cylinder. When the bandwidth parameters are optimized with this method, the resulting estimator for this density is more accurate, even though the estimator for the height and squared half-width of a rectangle is in that case less accurate. Since the interest lies in the former density, this alternative method for finding the optimal bandwidth parameters is an improvement in the estimation of this density. However, it should be noted that this method is time-consuming, due to the several numerical integrations and differentiations that have to be performed.

The bivariate modified Gamma kernel density estimator is constructed for an experimental data set, after which this estimator is transformed into an estimator for the height and squared radius of a cylinder. The shape of the estimators of both bivariate densities were very similar, with the main difference that the estimator for the latter density is negative for certain values close to the boundary.

For the estimation of further derivations of the bivariate modified Gamma kernel density estimator, like the considered marginal density of the squared radius of a cylinder and the density of the volume of a cylinder, the bandwidth parameters were optimized with respect to both bandwidth selection methods. These densities were estimated based on the example data set as well as on the experimental data set. The estimators constructed for the underlying densities of the experimental data set were compared to the estimators constructed in [19]. In this comparison, it can be concluded that all estimators agree on the shape, while the estimators constructed in this thesis were rougher, due to the smaller bandwidth. Since the underlying density of the example data set is known, it was possible to compare the estimator to the true density of that data set. For both bandwidth selection methods the resulting estimators were close to the true densities, while the estimators constructed with the bandwidth parameters chosen with the method based on the Integrated Squared Error of the estimator of the density of the height and squared radius of a cylinder slightly outperforms the other.

8 Recommendations

In this section some recommendations will be made for the improvement of the methods described in this thesis and for future research.

- The numerical procedure for transforming an estimator $\hat{g}$ into an estimator $\hat{f}$ is currently based on the Trapezoidal Rule for the integration and Finite Differences for the differentiation. Other numerical methods can be investigated to improve the numerical transformation of the estimator.
- With the bandwidth parameter selection method based on the Integrated Squared Error of $\hat{f}$, an increase in the accuracy of the estimator of the bivariate density f can be realised. Although this method is still time-consuming, this method can be improved by making the numerical procedure more efficient or by finding an alternative analytical expression for the function that has to be evaluated for example.
- When the interest lies in a univariate density that is derived from the bivariate density f , another method for selecting the bandwidth parameters can be considered. Similar to the alternative method of this thesis, a method to obtain optimal bandwidth parameters for the estimation of such a density derived from the bivariate density estimator $\hat{f}$ can be investigated.
- For the alternative bandwidth selection method based on the Integrated Squared Error of the estimator $\hat{f}$, the inverse relation between f and g was used. This method can possibly be generalized to other inverse relations to obtain improved bandwidth selection methods for estimation of the corresponding densities.

References

- [1] Fujisaki, K., Yamashita, N. and Yokota, H. (2011), *An automated three-dimensional internal structure observation system based on high-speed serial sectioning of steel materials*, Precision Engineering, Vol. 36, p. 315-321
- [2] Ginzburg, V.L. and Syrovatskii, S.I. (1965), *Cosmic Magnetobremssstrahlung (Synchrotron Radiation)*, Annual Review of Astronomy and Astrophysics, Vol. 3, p. 297-350
- [3] Wicksell, S.D. (1925), *The Corpuscle Problem: A mathematical study of a biometric problem*, Biometrika, Vol. 17, p. 84-99
- [4] McGarrrity, K.S., Sietsma, J. and Jongbloed, G. (2012), *Nonparametric Inference in a Stereological Model with Oriented Cylinders Applied to Dual Phase Steel*, submitted.
- [5] Van Es, B. and Hoogendoorn, A. (1990), *Kernel estimation in Wicksell's corpuscle problem*, Biometrika, Vol. 77, p. 139-145
- [6] Silverman, B.W. (1986), *Density estimation for statistics and data analysis*, Chapman-Hall
- [7] Wand, M.P. and Jones, M.C. (1995), *Kernel smoothing*, Chapman-Hall
- [8] Izenman, A.J. (1991), *Review Papers: Recent developments in nonparametric density estimation*, Journal of the American Statistical Association, Vol. 86, p. 205-224
- [9] Rudemo, M. (1982) *Empirical choice of histograms and kernel density estimates*, Scandinavian Journal of Statistics, Vol. 9, p. 65-78
- [10] Bouezmarni, T. and Rombouts, J. (2010), *Nonparametric density estimation for multivariate bounded data*, Journal of Statistical Planning and Inference, Vol. 140, p. 139-152
- [11] Wand, M.P. and Jones, M.C. (1993), *Comparison of smoothing parameterizations in bivariate kernel density estimation*, Journal of the American Statistical Association, Vol. 88, p. 520-528
- [12] Schuster, E.F. (1985), *Incorporating support constraints into nonparametric estimators of densities*, Communications in Statistics - Theory and Methods, p. 1123-1136
- [13] Müller, H. (1991), *Smooth optimum kernel estimators near endpoints*, Biometrika, Vol. 78, p. 521-530
- [14] Wand, M.P., Marron, J.S. and Ruppert D. (1991), *Transformations in density estimation*, Journal of the American Statistical Association, Vol. 86, p. 343-353
- [15] Chen, S.X. (2000), *Probability density function estimation using gamma kernels*, Annals of the Institute of Statistical Mathematics, Vol. 52, p. 471-480
- [16] Spouge, J.L. (1994), *Computation of the Gamma, Digamma, and Trigamma functions*, SIAM Journal of Numerical Analysis, Vol. 31, p. 931-944
- [17] Batir, N. (2008), *Inequalities for the Gamma Function*, Archiv der Mathematik, Vol. 91, p. 554-563
- [18] Brown, B.M. and Chen, S.X. (1999), *Beta-Bernstein smoothing for regression curves with compact supports*, Scandinavian Journal of Statistics, Vol. 26, p. 47-59
- [19] McGarrrity, K.S. (2013), *Stereological Estimation of Anisotropic Microstructural Features: Applying an Oriented Cylinder Model to Dual Phase Steel*, Doctoral thesis

A Central moments of the $Gamma(a, b)$ distribution

For the derivation of the bias of the bivariate modified Gamma kernel density estimator, the first ten central moments of the Gamma distribution are needed. The central moments μ_n of a random variable X are defined as

$$\mu_n = E[(X - E[X])^n], \quad \text{for } n = 1, 2, \dots$$

Let $X \sim Gamma(a, b)$, with density function $\gamma_{a,b}$, and known expected value $E[X] = ab$. The first ten central moments of X are determined by analytically computing the corresponding integral with Maple:

$$\int_0^{\infty} (x - ab)^n \gamma_{a,b}(x) dx.$$

The results can be found in Table 2:

n	μ_n	n	μ_n
1	0	6	$120ab^6 + 130a^2b^6 + 15a^3b^6$
2	ab^2	7	$720ab^7 + 924a^2b^7 + 210a^3b^7$
3	$2ab^3$	8	$5040ab^8 + 7308a^2b^8 + 2380a^3b^8 + 105a^4b^8$
4	$6ab^4 + 3a^2b^4$	9	$40320ab^9 + 64224a^2b^9 + 26432a^3b^9 + 2520a^4b^9$
5	$24ab^5 + 20a^2b^5$	10	$362880ab^{10} + 623376a^2b^{10} + 303660a^3b^{10} + 44100a^4b^{10} + 945a^5b^{10}$

Table 2: The first ten moments of the $Gamma(a, b)$ distribution.