A General Approach to the Study of L^1 Asymptotic Unbiasedness of Kernel Density Estimators in $\mathbf{R}^d$

by

Mark Stinner

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg

Abstract

A technique for establishing L^1 asymptotic unbiasedness of a kernel density estimator in $\mathbf{R}^d$ that does not depend on the form of the kernel function will be demonstrated. We will introduce the concept of a region sequence of a sequence of kernel functions and show how this can be used to give necessary and sufficient conditions for L^1 asymptotic unbiasedness. These results are then applied to kernel density estimators whose form is given and a number of known and novel results are obtained.

Acknowledgments

I thank the Department of Statistics at the University of Manitoba and the Provincial Government of Manitoba for their financial aid during the writing of this thesis. I also thank Dr. Alex Leblanc for his patience and valuable comments.

Contents

Chapter 1

Introduction

Let g_n be a density that depends on a random sample of size n drawn from a population in $\mathbf{R}^d$ with density f. This random function g_n is a density estimator. We would like to know how well g_n estimates the density f. One property that we would like g_n to have is the property of L^1 asymptotic unbiasedness. This means that $\int |\mathbf{E}g_n - f|$ tends to 0 as n tends to ∞ where $\mathbf{E}g_n$ is the expectation of g_n.

The L^1 norm provides a useful measure of error in the context of density estimation in $\mathbf{R}^d$. This is due to Scheffé's Theorem (Theorem A.5) which states that for densities f and g we have $\int |f - g| = 2 \sup_A \left| \int_A f - \int_A g \right|$. This shows that bounding the L^1 difference between f and g will also bound the difference between any probabilities calculated with these densities. Other measures of distance, including the L^2 norm, do not have this property.

There is a large variety of density estimators in use. All of the common ones may be written in the form $g_n(y) = \frac{1}{n} \sum_{i=1}^{n} g_{nX_i}(y)$ where X_i is the ith observation and $g_{nx}(y)$ is the kernel function. These estimators are known as kernel density estimators. In order to study the various kinds of L^1 convergence of such a large class of estimators many different techniques have been developed. These techniques tend to require that the form of the kernel function is known and so apply only to a special subclass of these estimators.

Here we will demonstrate a general method for establishing L^1 asymptotic unbiasedness that can be applied to many different kernel density estimators. This will be done using the concept of a region sequence of the sequence of kernel functions which will allow us to ignore, to some extent, the form of the kernel function used. Region sequences describe, roughly speaking, regions that are almost the support of the kernel functions. It will be shown that the property of asymptotic unbiasedness depends on the properties of these regions. Once this is shown we will then determine particular region sequences for different classes of kernel density estimators and deduce necessary and sufficient conditions for asymptotic unbiasedness that are described in terms of the parameters of the kernel functions and not the region sequences. This will show that the concept of a region sequence may be used as an intermediary to deriving results of interest. In addition a number of technical results will need to be shown and these are included in the Appendix.

Chapter 2

Preliminaries

Before we establish the general results regarding asymptotic unbiasedness, we will need to do some preliminary work. This includes describing some of the notation used in this document and introducing some concepts that will prove useful in subsequent chapters.

Definition 2.1 (Some Notation)**.** The following notational conventions and definitions are used throughout this document. Let $\mathbf{N}$, $\mathbf{Z}$, and $\mathbf{R}$ denote the set of natural numbers, the set of integers, and the set of real numbers respectively. Let $\mathbf{P}(E)$ denote the probability of event E. For a random variable X let $\mathbf{E}(X)$ and $\mathbf{V}(X)$ denote the expectation and variance of X respectively. Let d be in $\mathbf{N}$. Let A be a subset of $\mathbf{R}^d$. We define the following:

1. Let $\lambda(A)$ be the Lebesgue measure of A.

2. Let $\delta(A)$ be the diameter of A, that is, $\delta(A) = \sup_{x_1,x_2 \in A} \|x_1 - x_2\|$ where $\|x\|$ denotes the Euclidean norm of x.

3. Let $\mathring{A}$ be the interior of A.

4. Let $\overline{A}$ be the closure of A.

5. Let A' be the complement of A.

6. Let I_A be the characteristic function of A.

7. For any y in $\mathbf{R}^d$ and positive real number ϵ, let $\mathrm{B}(y, \epsilon)$ be the open ball of radius ϵ around y.

8. If f is a real valued function with domain D then the *support of f* is defined as the closure of the set $\{x \in D : f(x) \neq 0\}$.

9. For a real number t let $[t]$ denote the greatest integer less than or equal to t.

All subsets of $\mathbf{R^d}$ and functions are assumed to be Lebesgue measurable. Also we will often use the same notation to indicate a sequence and the value of a sequence. It should be clear from the context which is being denoted.

Definition 2.2 (Kernel Function Sequence)**.** For $G, H \subseteq \mathbf{R}^d$, we define a *sequence of kernel functions on $G \times H$* to be a sequence of functions

$$s_n : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$$

where n in $\mathbf{N}$ (with s_n being the nth *kernel function*), if the following hold:

1. For all x and y in $\mathbf{R}^d$ and n in $\mathbf{N}$, $s_n(x, y) \geq 0$.

2. For all n in $\mathbf{N}$, we have that

$$\int_H g_{nx}(y)dy = \begin{cases} 1 & x \in G \\ 0 & x \notin G \end{cases}$$

where g_{nx} is defined by the equation

$$g_{nx}(y) = s_n(x, y).$$

3. For all n in $\mathbf{N}$, we have that

$$\int_G h_{ny}(x)dx = \begin{cases} 1 & y \in H \\ 0 & y \notin H \end{cases}$$

where h_{ny} is defined by the equation

$$h_{ny}(x) = s_n(x, y).$$

Remark. In other words a kernel function sequence is a sequence of functions whose cross-sections are densities. This makes it a sequence of doubly-stochastic functions. In addition since these cross-sections are integrable, double integrals of products involving these functions may be evaluated in any order provided that the other factors make the product integrable. This will be made use of in subsequent calculations.

Definition 2.3 (Kernel Density Estimator)**.** For each i in $\mathbf{N}$ let X_i be a continuous random variable with density f. Let s_n be a sequence of kernel functions. We define the *kernel density estimator* of the density f from the sequence of random variables X_i, for each n in $\mathbf{N}$ by the equation

$$g_n(y) = \frac{1}{n} \sum_{i=1}^{n} g_{nX_i}(y).$$

Remark. The above definition of a kernel density estimator will be used in this document. This definition is broader than the definition used in many other works.

In particular this definition includes rectangular histograms. For other examples see Chapter 4. Also note that G and H will typically be the same set in any application. The notational distinction will be kept even in this case to aid in the understanding of the proofs.

Definition 2.4 (Region Sequence)**.** Let s_n be a sequence of kernel functions on $G \times H$, let ϵ be a positive real number and let V be a subset of both G and H. Let S_n be a sequence of subsets of $\mathbf{R}^d \times \mathbf{R}^d$ and denote the cross-sections of S_n by the following:

1. $G_{nx} = \{y \in H : (x, y) \in S_n\}$ for x in G.

2. $H_{ny} = \{x \in G : (x, y) \in S_n\}$ for y in H.

We say that S_n is a *region sequence of significance ϵ on V for s_n (ϵ, V-RS for s_n)* if the following hold:

1. If $x \in G$, $y \in H$, and $x = y$ then $(x, y) \in S_n$.

2. $\delta(G_{nx}) < \infty$ for all x in G and n in $\mathbf{N}$.

3. $\delta(H_{ny}) < \infty$ for all y in H and n in $\mathbf{N}$.

4. There is an N in $\mathbf{N}$ such that for all n in $\mathbf{N}$, x in V, y in V, if $n > N$ we have

$$\int_{G'_{nx}} g_{nx}(y)dy < \epsilon \text{ and } \int_{H'_{ny}} h_{ny}(x)dx < \epsilon.$$

In addition the number ϵ is referred to as the *level of significance* of the region sequence and V is referred to as the *region of significance* of the region sequence.

Remark. Intuitively a region sequence for a sequence of kernel functions is a sequence of regions that are almost the supports of the cross-sections of the corresponding kernel functions. The level and region of significance determine what counts as "almost". The region of significance determines which cross-sections of the region sequence are close to being the supports and the level of significance determines how close they are to being the supports. For an example of such a region sequence see Section 4.1. Also note that condition 1 states that the diagonal belongs to the region sequence.

Definition 2.5 (Narrowing Region Sequence)**.** An ϵ, V-RS for s_n is *narrowing* (ϵ, V-NRS for s_n) if

$$\overline{\lim_n} \sup_{y \in H} \delta(H_{ny}) = 0.$$

Definition 2.6 (Regular Region Sequence)**.** For a region sequence S_n that is an ϵ, V-RS for s_n we define the following regularity conditions:

1. The region sequence S_n is *regular in size* if

$$\underline{\lim_n} \frac{\inf_{y \in H} \lambda(H_{ny})}{\sup_{y \in H} \delta(H_{ny})^d} > 0.$$

2. The region sequence S_n is *regular in height* if

$$\overline{\lim_n} \sup_{y \in V} \lambda(H_{ny}) \sup_{x \in H_{ny}} h_{ny}(x) < \infty.$$

3. The region sequence S_n is *regular in shape* if for each y in $\mathring{H}$ there is a positive real number R and N in $\mathbf{N}$ such that for all n in $\mathbf{N}$ if $n > N$

$$B\left(y, R \sup_{y \in H} \delta(H_{ny})\right) \subseteq H_{ny}.$$

Finally an ϵ, V-RS for s_n is *regular* if it is regular in size, height, and shape. This will be abbreviated by ϵ, V-RRS. If the region sequence is also narrowing, we will write ϵ, V-NRRS.

Remark. Note that in the limits in the above definition, the level of significance ϵ and the region of significance V for the region sequence are fixed as n tends to infinity. Also while the statements of these conditions may seem complicated, they all have reasonable geometric interpretations and are usually easy to apply to particular examples. In particular condition 1, regularity in size, requires that the cross-sectional volumes of the region sequence ultimately do not decrease faster than the corresponding diameters. Condition 2, regularity in height, requires that the height of the cross-sectional densities ultimately do not increase faster than the rate at which the corresponding volumes of their supports decrease. Note that this only needs to work for y in V and not for all y in H. And condition 3, regularity in shape, requires a ball around the diagonal in each cross-section whose diameter ultimately changes uniformly with the diameter of the cross-section. Note that the radius of this ball may depend on y through R but that the rate at which the radius changes must be uniform.

Definition 2.7 (Relevant Region)**.** For a subset V of $\mathbf{R}^d$, we call V *relevant for* s_n if $V \subseteq \mathring{H}$, $V \subseteq \mathring{G}$, V is compact, and $\lambda(V) > 0$.

Definition 2.8 (Relevant Density)**.** A density $f : \mathbf{R}^d \to \mathbf{R}$ is *relevant for* s_n if the support of f lies in both H and G.

Remark. Note that in the above two definitions, relevance is dependent on the sequence of kernel functions under discussion. This is different from significance which relates to a sequence of regions for the sequence of kernel functions. Roughly speaking relevant regions will be the only regions that need to be considered when discussing regions of significance for a region sequence. All others will be degenerate in some way. Similarly relevant densities will be the only ones worth considering as others will be impossible to estimate in an asymptotically unbiased manner by the kernel density estimator under consideration.

Chapter 3

Asymptotic Unbiasedness

In this chapter we will prove three theorems that we will later use to derive necessary and sufficient conditions for asymptotic unbiasedness for some common types of density estimators.

The first of these theorems, Theorem 3.1, provides a sufficient condition in terms of the existence of certain region sequences for asymptotic L^1 unbiasedness of the kernel density estimator. Applications of this theorem will simply require the construction of such region sequences and in Chapter 4 we will show how this can be done.

The second theorem, Theorem 3.2, provides something close to the converse of Theorem 3.1. However, since the conclusion of this theorem simply states the existence of a region sequence with certain properties, it is difficult to use in application as we cannot conclude directly that it applies to our constructed region sequence.

The third theorem, Theorem 3.3 will show us how to deal with this issue. It states that, under the appropriate conditions, if there are region sequences with a certain property, then all region sequences have this property.

Theorem 3.1 (A Sufficient Condition for Asymptotic Unbiasedness)**.** *Let s_n be a sequence of kernel functions. If for each positive real number ϵ and each relevant region V, there is a sequence of regions S_n such that S_n is an ϵ, V-NRS for s_n , then for any relevant density f, we have that*

$$\int_{\mathbf{R}^d} |\mathbf{E}g_n(y) - f(y)|dy \to 0 \text{ as } n \to \infty$$

where $\mathbf{E}g_n(y) = \int g_{nx}(y)f(x)dx$.

Remark. This theorem will provide us with our basic tool for establishing a sufficient condition for asymptotic unbiasedness. In rough outline it follows the proof of Theorem 2.3 in [Dev85] but the details are somewhat different due to the more general setting. It involves using Theorem B.1, the uniform approximation of a density, whose proof is given in the Appendix, to choose an appropriate approximation of the given density and then show that the resulting differences can all be made sufficiently small. In particular we will show that $\int_{\mathbf{R}^d} |\mathbf{E}g_n^*(y) - f^*(y)|dy$ is small for n sufficiently large where $\mathbf{E}g_n^*(y) = \int g_{nx}(y)f^*(x)dx$ and f^* is the approximating density. To do this we will take an appropriately large region sequence and find an upper bound for the part of this integral over the region sequence and another for the part not over the region sequence. Note that no regularity conditions are used to establish this result.

Proof of Theorem 3.1. Let ϵ be a positive real number. Using Theorem B.1, choose a density f^* with compact support V such that

$$\int_{\mathbf{R}^d} |f(x) - f^*(x)|dx < \frac{\epsilon}{3}, \tag{3.1}$$

$V \subseteq \mathring{S} \subseteq G, H$ and f^* is uniformly continuous, where S is the support of f. Note that V is a relevant region for s_n.

Let $W = \bigcup_{x \in V} \{y \in H : \|x - y\| < 1\}$ and let $M = \lambda(W)$. Since V is compact, V is bounded which makes W bounded and so $0 < M = \lambda(W) < \infty$. Using the uniform continuity of f^* we may choose a positive real number δ such that for any x_1, x_2 in $\mathbf{R}^d$ if $\|x_1 - x_2\| < \delta$ then

$$|f^*(x_1) - f^*(x_2)| < \frac{\epsilon}{9M}. \tag{3.2}$$

Now choose a region sequence S_n such that S_n is an $\frac{\epsilon}{9}, V$-NRS for s_n. Since S_n is a narrowing sequence, we may choose an N_0 in $\mathbf{N}$ such that for any n in $\mathbf{N}$ if $n > N_0$,

$$\sup_{y \in H} \delta(H_{ny}) < \min(\delta, 1).$$

This implies that for y in H and x in H_{ny}, if $n > N_0$, we know that $\|x - y\| \leq \delta(H_{ny}) < \min(\delta, 1) \leq \delta$, since $y \in H_{ny}$. Using inequality (3.2) this implies that

$$|f^*(x) - f^*(y)| < \frac{\epsilon}{9M}. \tag{3.3}$$

Now let V_n be a sequence of sets defined by

$$V_n = \{y \in H : H_{ny} \cap V \neq \emptyset\}.$$

Again $y \in H_{ny}$ so we see that if $n > N_0$,

$$\begin{aligned} y \in V_n &\Rightarrow H_{ny} \cap V \neq \emptyset \\ &\Rightarrow \text{there is an } x \text{ in } V \text{ such that } x \in H_{ny} \\ &\Rightarrow \text{there is an } x \text{ in } V \text{ such that } \|x - y\| \leq \delta(H_{ny}) < \min(\delta, 1) \leq 1 \\ &\Rightarrow y \in W \end{aligned}$$

This shows that $V_n \subseteq W$ and hence

$$\lambda(V_n) \leq \lambda(W) = M. \tag{3.4}$$

Furthermore if $y \in H$ but $y \notin V_n$ then

$$y \notin V \text{ and } H_{ny} \cap V = \emptyset. \tag{3.5}$$

In addition since S_n is a region sequence, we may choose N_1 in $\mathbf{N}$ such that, for any n in $\mathbf{N}$ and x, y in V, if $n > N_1$,

$$\int_{H'_{ny}} h_{ny}(x)dx < \frac{\epsilon}{9} \text{ and } \int_{G'_{nx}} g_{nx}(y)dy < \frac{\epsilon}{9}. \tag{3.6}$$

We will now show that the above inequalities and propositions lead to the desired conclusion. Let $N = \max(N_0, N_1)$ and suppose that $n \in \mathbf{N}$ and that $n > N$. Consider the following:

$$\begin{aligned}
\int_{\mathbf{R}^d} |\mathbf{E}g_n^*(y) - f^*(y)|dy &= \int_H \left| \int_G g_{nx}(y)f^*(x)dx - f^*(y) \right| dy \\
&= \int_H \left| \int_G h_{ny}(x)f^*(x)dx - \int_G h_{ny}(x)f^*(y)dx \right| dy \\
&\leq \int_H \int_G h_{ny}(x)|f^*(x) - f^*(y)|dxdy \\
&= \int_H \int_{H_{ny}} h_{ny}(x)|f^*(x) - f^*(y)|dxdy \\
&\quad + \int_H \int_{H'_{ny}} h_{ny}(x)|f^*(x) - f^*(y)|dxdy \\
&= I + II
\end{aligned}$$

Using inequalities (3.3) and (3.4), and statement (3.5), we have that

$$
\begin{aligned}
I &= \int_{V_n} \int_{H_{ny}} h_{ny}(x)|f^*(x) - f^*(y)|dxdy \\
&\leq \int_{V_n} \int_{H_{ny}} h_{ny}(x)\frac{\epsilon}{9M}dxdy \\
&\leq \frac{\epsilon}{9M} \int_{V_n} dy \\
&= \frac{\epsilon}{9M}\lambda(V_n) \\
&\leq \frac{\epsilon}{9M}M = \frac{\epsilon}{9}.
\end{aligned}
$$

Furthermore using statement (3.6) we have that

$$
\begin{aligned}
II &= \int_{H} \int_{H'_{ny}} h_{ny}(x)|f^*(x) - f^*(y)|dxdy \\
&\leq \int_{H} \int_{H'_{ny}} h_{ny}(x)f^*(x)dxdy + \int_{H} \int_{H'_{ny}} h_{ny}(x)f^*(y)dxdy \\
&= \int_{G} \int_{G'_{nx}} g_{nx}(y)f^*(x)dydx + \int_{H} \int_{H'_{ny}} h_{ny}(x)f^*(y)dxdy \\
&\leq \frac{\epsilon}{9} \int_{G} f^*(x)dx + \frac{\epsilon}{9} \int_{H} f^*(y)dy \\
&= \frac{\epsilon}{9} + \frac{\epsilon}{9} = \frac{2\epsilon}{9}.
\end{aligned}
$$

It then follows that

$$
\int_{\mathbf{R}^d} |\mathbf{E}g_n^*(y) - f^*(y)|dy \leq I + II < \frac{\epsilon}{9} + \frac{2\epsilon}{9} = \frac{\epsilon}{3}. \tag{3.7}
$$

Finally note that

$$\int_{\mathbf{R}^d} |\mathbf{E}g_n(y) - \mathbf{E}g_n^*(y)|dy = \int_H \left| \int_G g_{nx}(y)f(x)dx - \int_G g_{nx}(y)f^*(x)dx \right| dy$$
$$\leq \int_G \int_H g_{nx}(y)|f(x) - f^*(x)|dydx$$
$$= \int_{\mathbf{R}^d} |f(x) - f^*(x)|dx$$

and thus using inequalities (3.1) and (3.7) we have

$$\int_{\mathbf{R}^d} |\mathbf{E}g_n(y) - f(y)|dy$$
$$\leq \int_{\mathbf{R}^d} |\mathbf{E}g_n(y) - \mathbf{E}g_n^*(y)|dy + \int_{\mathbf{R}^d} |\mathbf{E}g_n^*(y) - f^*(y)|dy + \int_{\mathbf{R}^d} |f^*(y) - f(y)|dy$$
$$\leq \int_{\mathbf{R}^d} |\mathbf{E}g_n^*(y) - f^*(y)|dy + 2\int_{\mathbf{R}^d} |f^*(y) - f(y)|dy$$
$$< \frac{\epsilon}{3} + 2\frac{\epsilon}{3} = \epsilon.$$

The conclusion follows. □

Theorem 3.2 (A Necessary Condition for Asymptotic Unbiasedness)**.** *Let s_n be a sequence of kernel functions. Suppose that for each positive real number ϵ and each relevant region V there is a region sequence S_n such that S_n is an ϵ, V-RRS for s_n. If for all relevant densities f,*

$$\int_{\mathbf{R}^d} |\mathbf{E}g_n(y) - f(y)|dy \to 0$$

as $n \to \infty$, then for each positive real number ϵ and relevant region V there is a region sequence T_n such that T_n is an ϵ, V-NRRS.

Remark. This theorem will give us partial converse to Theorem 3.1. Note that the regularity conditions play a role in this theorem but did not in the previous theorem. Also note that while the theorem is stated in terms of regular region sequences, no use of regularity in shape is actually used. The proof is by contradiction. The main idea of this proof is to use the non-narrowing of a regular region sequence to find a density that cannot be estimated in an asymptotically unbiased manner.

Proof of Theorem 3.2. Suppose the antecedent is true and the consequent is false. Choose a positive real number ϵ_0 and relevant region V_0 such that, for any region sequence T_n, T_n is not an ϵ_0, V_0-NRRS. Let the number ϵ_1 be given by the equation $\epsilon_1 = \min(\epsilon_0, \frac{1}{6})$. Note that the number $\frac{1}{6}$ is somewhat arbitrary. It is chosen to make ϵ_1 sufficiently small so that the contradiction is easily obtained. Now choose a region sequence S_n such that S_n is an ϵ_1, V_0-RRS. Note that S_n cannot be narrowing and so

$$\overline{\lim_n} \sup_{y \in H} \delta(H_{ny}) > 0.$$

Thus we may choose a positive real number A such that for each N in $\mathbf{N}$ there is a number m in $\mathbf{N}$ such that $m > N$ and

$$\sup_{y \in H} \delta(H_{my}) > A. \tag{3.8}$$

Since S_n is regular in height we may choose a positive real number B and N_0 in $\mathbf{N}$ such that for each n in $\mathbf{N}$ if $n > N_0$,

$$\sup_{y \in V_0} \lambda(H_{ny}) \sup_{x \in H_{ny}} h_{ny}(x) < B. \tag{3.9}$$

Since S_n is regular in size we may choose a positive real number C and N_1 in $\mathbf{N}$ such that for each n in $\mathbf{N}$ if $n > N_1$ then

$$\inf_{y\in H} \lambda(H_{ny}) > C \sup_{y\in H} \delta(H_{ny})^d. \tag{3.10}$$

Furthermore, since S_n is a region sequence for s_n we may choose N_2 in $\mathbf{N}$ such that for each n in $\mathbf{N}$ if $n > N_2$ then

$$\int_{G'_{nx}} g_{nx}(y)dy < \epsilon_1 \text{ and } \int_{H'_{ny}} h_{ny}(x)dx < \epsilon_1 \tag{3.11}$$

for any x, y in V_0.

Now let $D = \frac{B}{CA^d}$. Choose a subset V_1 of V_0 such that V_1 is a relevant region with a volume small enough so that

$$0 < \lambda(V_1) < \frac{1 - 3\epsilon_1}{D}. \tag{3.12}$$

Note that number on the right hand side is positive since $\epsilon_1 \leq \frac{1}{6}$. Define a function $f : \mathbf{R}^d \to \mathbf{R}$ by

$$f(x) = \begin{cases} \frac{1}{\lambda(V_1)} & \text{if } x \in V_1 \\ 0 & \text{otherwise.} \end{cases}$$

Clearly f is a relevant density. By assumption we can choose N_3 in $\mathbf{N}$ such that for any n in $\mathbf{N}$ if $n > N_3$, we have

$$\int_{\mathbf{R}^d} |\mathbf{E}g_n(y) - f(y)|dy < \epsilon_1. \tag{3.13}$$

Now $N = \max(N_0, N_1, N_2, N_3)$. Using inequality (3.8) we can choose a number m in $\mathbf{N}$ such that $m > N$ and

$$\sup_{y\in H} \delta(H_{my}) > A. \tag{3.14}$$

Now we will show that the above inequalities lead to a contradiction. Using inequalities (3.9), (3.10), and (3.14), we see that for any y in V_1 and x in H_{my} we have

$$h_{my}(x) \leq \sup_{x \in H_{my}} h_{my}(x) < \frac{B}{\lambda(H_{my})} \leq \frac{B}{\inf_{y \in H} \lambda(H_{my})}$$

$$< \frac{B}{C(\sup_{y \in H} \delta(H_{my}))^d} < \frac{B}{CA^d} = D$$

and so for y in V_1 we have

$$\int_{H_{my}} h_{my}(x) f(x) dx = \frac{1}{\lambda(V_1)} \int_{H_{my} \cap V_1} h_{my}(x) dx$$

$$\leq \frac{D\lambda(H_{my} \cap V_1)}{\lambda(V_1)} \leq D. \tag{3.15}$$

Note that $D < \frac{1}{\lambda(V_1)}$ and so using inequalities (3.12) and (3.15) we have

$$\int_H \left| \int_{H_{my}} h_{my}(x) f(x) dx - f(y) \right| dy$$

$$\geq \int_{V_1} \left| \int_{H_{my}} h_{my}(x) f(x) dx - f(y) \right| dy \geq \int_{V_1} (\frac{1}{\lambda(V_1)} - D) dy$$

$$= (\frac{1}{\lambda(V_1)} - D)\lambda(V_1) = 1 - D\lambda(V_1) > 1 - D\frac{1 - 3\epsilon_1}{D} = 3\epsilon_1.$$

But using inequalities (3.11) and (3.13) we have

$$\int_H \left| \int_{H_{my}} h_{my}(x) f(x) dx - f(y) \right| dy$$

$$\leq \int_H \left| \int_{H_{my}} h_{my}(x) f(x) dx - \int_G h_{my}(x) f(x) dx \right| dy$$

$$+ \int_H \left| \int_G h_{my}(x) f(x) dx - f(y) \right| dy$$

$$= \int_H \int_{H'_{my}} h_{my}(x) f(x) dx dy + \int_{\mathbf{R}^d} |\mathbf{E} g_m(y) - f(y)| dy$$

$$= \int_G \left(\int_{G'_{mx}} g_{mx}(y) dy \right) f(x) dx + \int_{\mathbf{R}^d} |\mathbf{E} g_m(y) - f(y)| dy$$

$$\leq \epsilon_1 \int_G f(x) dx + \int_{\mathbf{R}^d} |\mathbf{E} g_m(y) - f(y)| dy$$

$$< \epsilon_1 + \epsilon_1 = 2\epsilon_1$$

and so

$$3\epsilon_1 < \int_H \left| \int_{H_{my}} h_{my}(x) f(x) dx - f(y) \right| dy < 2\epsilon_1,$$

which gives us the wanted contradiction. □

Theorem 3.3 (Narrowing Region Sequences)**.** *Let s_n be a sequence of kernel functions. Suppose that for each positive real number ϵ and relevant region V, there is a region sequence S_n for s_n such that S_n is an ϵ, V-NRRS. Then for each positive real number ϵ and relevant region V, if T_n is an ϵ, V-RRS then T_n is an ϵ, V-NRRS.*

Remark. Roughly speaking this theorem states that if there is an narrowing regular region sequence then every regular region sequence is narrowing, or alternatively that narrowing is really a property of the sequence of kernel functions and not any particular region sequence. Like Theorem 3.2, the proof is by contradiction. Note that regularity in shape plays a crucial role in the proof of this theorem.

Proof of Theorem 3.3. Suppose the antecedent is true and the consequent is false. Then choose a positive real number ϵ_0 and relevant region V_0 and region sequence T_n such that T_n is an ϵ_0, V_0-RRS but not an NRRS. Let $\epsilon_1 = \min(\epsilon_0, 1/2)$ so that $0 < \epsilon_1 < 1$. Choose a region sequence S_n such that S_n is an ϵ_1, V_0-NRRS. Hence there is an N_0 such that, for any n in $\mathbf{N}$, and y in V_0, if $n > N_0$ we have

$$\int_{H'_{ny}} h_{ny}(x)dx < \epsilon_1. \tag{3.16}$$

Let $Q_{ny} = \{x \in G : (x, y) \in T_n\}$. Since T_n is not narrowing we can choose a positive real number A such that for any N in $\mathbf{N}$ there is a number m in $\mathbf{N}$ such that $m > N$ and

$$\sup_{y \in H} \delta(Q_{my}) > A. \tag{3.17}$$

We know that T_n is regular in height and so we may choose a positive real number B and a number N_1 in $\mathbf{N}$ such that for any n in $\mathbf{N}$ and y in V_0 if $n > N_1$, we have

$$\lambda(Q_{ny}) \sup_{x \in Q_{ny}} h_{ny}(x) < B. \tag{3.18}$$

Since T_n is regular in size, there is a positive real number C and a number N_2 in $\mathbf{N}$ such that for any n in $\mathbf{N}$ if $n > N_2$, we have

$$\inf_{y \in H} \lambda(Q_{ny}) > C(\sup_{y \in H} \delta(Q_{ny}))^d. \tag{3.19}$$

Now choose a point y_0 in V_0. Since T_n is regular in shape we know that there is a positive real number R_0 and a number N_3 in $\mathbf{N}$ such that for any n in $\mathbf{N}$ if $n > N_3$ we have

$$\mathrm{B}\left(y_0, R_0 \sup_{y\in H} \delta(Q_{ny})\right) \subseteq Q_{ny_0}. \tag{3.20}$$

Now define a positive real number ϵ_2 by the equation

$$\epsilon_2 = \min\left(\left(\frac{1-\epsilon_1}{2B}C\right)^{\frac{1}{d}} A, \frac{R_0 A}{2}\right). \tag{3.21}$$

Since S_n is narrowing, we may choose N_4 in $\mathbf{N}$ such that for any n in $\mathbf{N}$, if $n > N_4$, we have

$$\sup_{y\in H} \delta(H_{ny}) < \epsilon_2. \tag{3.22}$$

Let $N = \max(N_1, N_2, N_3, N_4)$. Using statement (3.17) we can choose m in $\mathbf{N}$ such that $m > N$ and

$$\sup_{y\in H} \delta(Q_{my}) > A. \tag{3.23}$$

Now we will show that the above inequalities lead to a contradiction. Using equality (3.21) and inequality (3.22) we have the inequalities

$$\delta(H_{my_0}) \leq \sup_{y\in H} \delta(H_{my}) < \epsilon_2 \leq \frac{R_0 A}{2}. \tag{3.24}$$

Then using the fact that $y_0 \in H_{my_0}$, statement (3.20), and inequalities (3.23) and (3.24), we have the following containments:

$$H_{my_0} \subseteq \mathrm{B}(y_0, R_0 A) \subseteq \mathrm{B}(y_0, R_0 \sup_{y\in H} \delta(Q_{my})) \subseteq Q_{my_0}. \tag{3.25}$$

This shows that

$$\sup_{x\in H_{my_0}} h_{my_0}(x) \leq \sup_{x\in Q_{my_0}} h_{my_0}(x). \tag{3.26}$$

In addition using inequality (3.16) we have

$$\lambda(H_{my_0}) \sup_{x\in H_{my_0}} h_{my_0}(x) \geq \int_{H_{my_0}} h_{my_0}(x)dx = 1 - \int_{H'_{my_0}} h_{my_0}(x)dx > 1 - \epsilon_1$$

which can be rearranged so that

$$\sup_{x\in H_{my_0}} h_{my_0}(x) > \frac{1-\epsilon_1}{\lambda(H_{my_0})}. \tag{3.27}$$

Now using inequalities (3.17) and (3.19) we have

$$CA^d < C \sup_{y\in H} \delta(Q_{my})^d < \inf_{y\in H} \lambda(Q_{my}). \tag{3.28}$$

Using inequalities (3.18), (3.26), and (3.27) we have

$$\lambda(Q_{my_0}) < \frac{B}{\sup_{x\in Q_{my_0}} h_{my_0}(x)} < \frac{B}{\sup_{x\in H_{my_0}} h_{my_0}(x)} < \frac{B\lambda(H_{my_0})}{1-\epsilon_1}. \tag{3.29}$$

Using inequality (3.22), definition (3.21), and the fact that in general, for any region E of $\mathbf{R}^d$, $\lambda(E) < \delta(E)^d$, we have the inequalites

$$\frac{B\lambda(H_{my_0})}{1-\epsilon_1} < \frac{B\delta(H_{my_0})^d}{1-\epsilon_1} < \frac{B\epsilon_2^d}{1-\epsilon_1} \leq \frac{B}{1-\epsilon_1}\left(\frac{1-\epsilon_1}{2B}C\right)A^d = \frac{1}{2}CA^d. \tag{3.30}$$

Finally combining the inequalities (3.28), (3.29), and (3.30), we have

$$CA^d < \frac{1}{2}CA^d,$$

the wanted contradiction. □

Chapter 4

Kernel Density Estimators

In this chapter we will use the theory developed in the previous chapters to establish useful conditions for the asymptotic unbiasedness of some commonly used density estimators.

In each case we will define a sequence of kernel functions s_n and then define a region sequence S_n for this sequence of functions for any level and region of significance. These region sequences will be defined carefully using the parameters of the given kernel functions so that the conditions of narrowing and regularity are equivalent to easily verified propositions. In fact the final statements of this chapter will not make reference to region sequences at all. This will show that the notion of a region sequence may be used as an intermediate step to get some well known as well as novel results concerning the L^1 asymptotic unbiasedness of commonly used density estimators.

The proof of each result is similar in outline. Each sequence of kernel functions is related to one or more known probability density functions and this, together with Boole's and Chebyshev's inequalities (Theorems A.4 and A.3 respectively) in some cases, is used to establish that the given definition of S_n is an ϵ, V-RS. The conditions of narrowing and regularity are stated in terms of diameters and volumes of the cross-sections H_{ny} so expressions for these quantities (or bounds for these quantities) will be established. Finally using these expressions we will establish propositions in terms of the parameters of s_n that are equivalent to the conditions of narrowing and regularity. The arguments used to do this sometimes make use of some technical results that can be found in the Appendix. In particular dealing with regularity in size will make use of Lemma E.1 and regularity in height will sometimes require one of the uniform approximation theorems (Theorems C.1 and D.1). Finally we use the theorems of Chapter 3 to derive a result concerning the asymptotic unbiasedness of the particular kernel density estimator which does not reference any region sequences and is expressed solely in terms of the parameters of the sequence of kernel functions.

We will now introduce some notation. For a given sequence of kernel functions s_n, for each y in H, let X_{yn} be a random variable with density given by $f_{X_{yn}}(x) = s_n(x, y)$ and, for each x in G, let Y_{xn} be a random variable with density given by $f_{Y_{xn}}(y) = s_n(x, y)$. The kth components of the random variables of X_{yn} and Y_{xn} are denoted by X_{ynk} and Y_{xnk} respectively.

4.1 Standard Kernel Density Estimator

For each $k = 1, \ldots, d$ let h_k be a sequence of positive real numbers, depending on n, where n is in $\mathbf{N}$. Let H be a $d \times d$ diagonal matrix whose kk entry is h_k, that is, $H = \operatorname{diag}(h_1, \ldots, h_d)$. Let K be a bounded density on $\mathbf{R}^d$ with finite marginal expectations and variances. Then define the sequence of functions s_n by

$$s_n(x, y) = \det(H^{-1})K(H^{-1}(y - x))$$

where $x, y \in \mathbf{R}^d$. It is clear that s_n is a sequence of kernel functions on $\mathbf{R}^d \times \mathbf{R}^d$. Using this sequence of functions, the density estimator g_n from Definition 2.3 is given by

$$g_n(y) = \frac{1}{n}\sum_{i=1}^{n} \det(H^{-1})K(H^{-1}(y - X_i))$$

where X_i is the ith sample observation. We will call this estimator the *standard kernel density estimator.*

To use the theory from the previous chapters we need to define a region sequence for the given sequence of kernel functions s_n. To this end let A_k be a positive constant for $k = 1, \ldots, d$ and let S_n be a sequence of sets defined by

$$S_n = \{(x, y) \in \mathbf{R}^d \times \mathbf{R}^d : |y_k - x_k| \leq A_k h_k, \text{ for all } k = 1, \ldots, d\}. \tag{4.1}$$

Furthermore, let Z_k be a random variable with density given by the kth margin of K so that

$$X_{ynk} \sim y_k - h_k Z_k \text{ and } Y_{xnk} \sim x_k + h_k Z_k$$

for any x in $\mathbf{R}^d$ and y in $\mathbf{R}^d$.

Theorem 4.1 (Region Sequences for the Standard Kernel Density Estimator)**.** *For each positive real number ϵ, and relevant region V, there are positive real numbers A_k, for $k = 1, \ldots, d$ such that S_n is an ϵ, V-RS for s_n. Furthermore, the following hold:*

1. $\lim_n h_k = 0$ *for* $k = 1, \ldots, d$ *if and only if* S_n *is narrowing.*

2. $\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0$ *if and only if* S_n *is regular.*

Remark. Note that this case is the easiest of those considered to work with. This is mainly due to the fact that the cross-sectional densities are just shifted and scaled versions of the same densities, the marginal densities of K. This leads to the construction of a region sequence that is very easy to work with. In particular we will be able to easily calculate cross-sectional diameters and volumes that do not depend on y. In addition the region of significance V plays essentially no role in the proof. This is to be expected as the theorems of Chapter 3 are based on proofs found in [Dev85] that deal with standard kernel density estimators.

Proof of Theorem 4.1. Let ϵ be a positive real number and V be a relevant region. For each $k = 1, \ldots, d$ choose positive real numbers A_k such that

$$A_k > \left(\frac{d}{\epsilon}(\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2)\right)^{\frac{1}{2}}.$$

Then, using Boole's and Chebyschev's inequalities (Theorems A.4 and A.3), for y in V we have

$$
\begin{aligned}
\int_{H'_{ny}} h_{ny}(x)dx &= \mathbf{P}(X_{yn} \in H'_{ny}) \\
&= \mathbf{P}(|y_k - X_{ynk}| > A_k h_k \text{ for some } k = 1, \ldots, d) \\
&\leq \sum_{k=1}^{d} \mathbf{P}(|y_k - X_{ynk}| > A_k h_k) \\
&\leq \sum_{k=1}^{d} \frac{\mathbf{V}(X_{ynk}) + (\mathbf{E}(X_{ynk}) - y_k)^2}{A_k^2 h_k^2} \\
&= \sum_{k=1}^{d} \frac{h_k^2(\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2)}{A_k^2 h_k^2} \\
&= \sum_{k=1}^{d} \frac{\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2}{A_k^2} \\
&< \sum_{k=1}^{d} \frac{\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2}{\frac{d}{\epsilon}(\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2)} \\
&= \sum_{k=1}^{d} \frac{\epsilon}{d} = \epsilon.
\end{aligned}
$$

Similarly for x in V we have

$$\begin{aligned}
\int_{G'_{nx}} g_{nx}(y)dy &= \mathbf{P}(Y_{xn} \in G'_{nx}) \\
&= \mathbf{P}(|Y_{xnk} - x_k| > A_k h_k \text{ for some } k = 1, \ldots, d) \\
&\leq \sum_{k=1}^{d} \mathbf{P}(|Y_{xnk} - x_k| > A_k h_k) \\
&\leq \sum_{k=1}^{d} \frac{\mathbf{V}(Y_{xnk}) + (\mathbf{E}(Y_{xnk}) - x_k)^2}{A_k^2 h_k^2} \\
&= \sum_{k=1}^{d} \frac{h_k^2(\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2)}{A_k^2 h_k^2} \\
&= \sum_{k=1}^{d} \frac{\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2}{A_k^2} \\
&\leq \sum_{k=1}^{d} \frac{\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2}{\frac{d}{\epsilon}(\mathbf{V}(Z_k) + \mathbf{E}(Z_k)^2)} \\
&= \sum_{k=1}^{d} \frac{\epsilon}{d} = \epsilon.
\end{aligned}$$

Thus for each positive real number ϵ and relevant region V, we can find positive real numbers A_k, for $k = 1, \ldots, d$, such that S_n is an ϵ, V-RS for s_n.

We will now find expressions for the volumes and diameters of the cross-sections. Note that H_{ny} is a d-dimensional rectangle whose kth side has length $2A_k h_k$ and so

$$\delta(H_{ny}) = 2\left(\sum_{k=1}^{d} A_k^2 h_k^2\right)^{\frac{1}{2}} \tag{4.2}$$

$$\lambda(H_{ny}) = 2^d \prod_{k=1}^{d} A_k h_k. \tag{4.3}$$

To see that S_n is regular in height, note that

$$\begin{aligned}
\sup_{y\in V} \lambda(H_{ny}) \sup_{x\in H_{ny}} h_{ny}(x) &= \left(2^d \prod_{k=1}^{d} A_k h_k\right) \sup_{y\in V} \sup_{x\in H_{ny}} \det(H^{-1})K(H^{-1}(y-x)) \\
&\leq \left(2^d \prod_{k=1}^{d} A_k\right)\left(\prod_{k=1}^{d} h_k\right) \frac{1}{\prod_{k=1}^{d} h_k} \sup_z K(z) \\
&= \left(2^d \prod_{k=1}^{d} A_k\right) \sup_z K(z) < \infty.
\end{aligned}$$

In addition

$$\begin{aligned}
S_n \text{ is narrowing} &\Leftrightarrow \sup_y \delta(H_{ny}) \to 0 \text{ as } n \to \infty \\
&\Leftrightarrow 2\left(\sum_{k=1}^{d} A_k^2 h_k^2\right)^{\frac{1}{2}} \to 0 \text{ as } n \to \infty \\
&\Leftrightarrow h_k \to 0 \text{ as } n \to \infty \text{ for all } k = 1, \ldots, d.
\end{aligned}$$

Furthermore we have, using Lemma E.1 and equalities (4.2) and (4.3),

$$
\begin{aligned}
S_n \text{ is regular in size } &\Leftrightarrow \underline{\lim}_n \frac{\inf_{y\in H} \lambda(H_{ny})}{\sup_{y\in H} \delta(H_{ny})^d} > 0 \\
&\Leftrightarrow \underline{\lim}_n \frac{\prod_{k=1}^d A_k^2 h_k^2}{\left(\sum_{k=1}^d A_k^2 h_k^2\right)^d} > 0 \\
&\Leftrightarrow \underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0.
\end{aligned}
$$

Finally we need to show that S_n is regular in shape if $\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0$. Choose a positive real number B and a number N in $\mathbf{N}$ such that, for any n in $\mathbf{N}$, if $n > N$,

$$
\min_k h_k > B \max_k h_k.
$$

Now let

$$
R = \frac{B \min_k A_k}{2\left(\sum_{k=1}^d A_k^2\right)^{\frac{1}{2}}}.
$$

Then for all n in $\mathbf{N}$, if $n > N$, we have, again using equality (4.2),

$$
\begin{aligned}
R \sup_{y\in H} \delta(H_{ny}) &= 2R \left(\sum_{k=1}^d A_k^2 h_k^2\right)^{\frac{1}{2}} \\
&= \frac{B \min_k A_k}{\left(\sum_{k=1}^d A_k^2\right)^{\frac{1}{2}}} \left(\sum_{k=1}^d A_k^2 h_k^2\right)^{\frac{1}{2}} \\
&\leq \frac{B \min_k A_k}{\left(\sum_{k=1}^d A_k^2\right)^{\frac{1}{2}}} \left(\sum_{k=1}^d A_k^2\right)^{\frac{1}{2}} \max_k h_k \\
&= B(\min_k A_k) \max_k h_k \\
&< (\min_k A_k) \min_k h_k
\end{aligned}
$$

implying that $R \sup_{y \in H} \delta(H_{ny}) \leq A_k h_k$ for all $k = 1, \ldots, d$, and so

$$\mathrm{B}\left(y, R \sup_{y \in H} \delta(H_{ny})\right) \subseteq H_{ny},$$

that is, S_n is regular in shape. □

Theorem 4.2 (Asymptotic Unbiasedness of the Standard Kernel Density Estimator)**.** *Given that* $\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0$*, the standard kernel density estimator is asymptotically unbiased if and only if* $\lim h_k = 0$ *for* $k = 1, \ldots, d$*.*

Proof of Theorem 4.2. Suppose that $\lim_n h_k = 0$ for $k = 1, \ldots, d$. Then by Theorem 4.1 we know that for each positive real number ϵ and relevant region V, there is an ϵ, V-NRS, and so by Theorem 3.1 the standard kernel density estimator is asymptotically unbiased.

Now suppose that $\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0$ and that we have asymptotic unbiasedness. By Theorem 4.1 we know that for each positive real number ϵ and relevant region V there is an ϵ, V-RRS. Hence, using the assumption of asymptotic unbiasedness, by Theorem 3.2, for each positive real number ϵ and relevant region V there is an ϵ, V-NRRS. Using Theorem 3.3, we then know that for each positive real number ϵ and relevant region V, S_n (given by definition 4.1) is an ϵ, V-NRRS, since it is ϵ, V-RRS, and hence we have that $\lim_n h_k = 0$ for $k = 1, \ldots, d$. □

Remark. This result is similar to Theorem 2.3 in [Dev85]. That result provides only a sufficient condition and considers only the case where $h_k = h$ for $k = 1, \ldots, d$. In this case the sufficient condition for asymptotic unbiasedness of Theorem 4.2 reduces to that of Theorem 2.3 in [Dev85]. There are also similarities in the proofs used to establish these theorems as the proof of the main theorem used, Theorem 3.1, is based on the the proof of Devroye's theorem.

4.2 Rectangular Histogram Density Estimator

Let h_k be a sequence of positive real numbers, depending on n, for each $k = 1, \ldots, d$, where n is in $\mathbf{N}$. For each j in $\mathbf{Z}^d$, y in $\mathbf{R}^d$, $k = 1, \ldots, d$, and n in $\mathbf{N}$, let $A_{njk} = [j_k h_k, (j_k + 1)h_k)$, $A_{nj} = A_{nj1} \times \cdots \times A_{njd}$, $p_{nj}(y) = I_{A_{nj}}(y)/\lambda(A_{nj})$, and finally let

$$s_n(x, y) = \sum_j p_{nj}(y) I_{A_{nj}}(x)$$

where $x \in \mathbf{R}^d$. The sequence of functions s_n is a sequence of kernel functions on $\mathbf{R}^d \times \mathbf{R}^d$. Using this sequence of functions, the density estimator g_n from Definition 2.3 is given by

$$g_n(y) = \frac{1}{n} \sum_{i=1}^{n} \sum_j \frac{I_{A_{nj}}(y)}{\lambda(A_{nj})} I_{A_{nj}}(X_i)$$

where X_i is the ith sample observation. We will call this estimator the *rectangular histogram density estimator*.

As before to use the theory from the previous chapters we need to define a region sequence for this sequence of kernel functions. For n in $\mathbf{N}$, x in $\mathbf{R}^d$ and $k = 1, \ldots, d$, let $j_{xk} = [\frac{x_k}{h_k}]$, $j_x = (j_{xk})_{k=1,\ldots,d}$, and

$$S_n = \{(x, y) \in \mathbf{R}^d \times \mathbf{R}^d : |j_{xk} - j_{yk}| \leq 1 \text{ for } k = 1, \ldots, d\}.$$

Furthermore note that, for any x in $\mathbf{R}^d$ and y in $\mathbf{R}^d$, we have

$$X_{ynk} \sim \text{Uniform}(j_{yk} h_k, (j_{yk} + 1)h_k) \text{ and } Y_{xnk} \sim \text{Uniform}(j_{xk} h_k, (j_{xk} + 1)h_k)$$

and in addition we have $j_{X_{yn}k} = j_{yk}$ and $j_{Y_{xn}k} = j_{xk}$ with probability one.

Theorem 4.3 (Region Sequences for the Rectangular Histogram Density Estimator)**.** *For each positive real number ϵ, and relevant region V, S_n is an ϵ, V-RS for s_n. Also the following hold:*

1. $\lim_n h_k = 0$ *for* $k = 1, \ldots, d$ *if and only if* S_n *is narrowing.*

2. $\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0$ *if and only if* S_n *is regular.*

Remark. This case is also easy to work with. As in Theorem 4.1, the cross-sectional diameters and volumes are easy to calculate and the the region of significance V plays essentially no role. The region sequence S_n defined above may seem unnecessarily large at first (its cross-sections cover the supports of the functions h_{ny} by a large margin), but its size is actually required to allow for S_n to be regular in shape.

Proof of Theorem 4.3. Let V be a relevant region. Then, using Boole's inequality (Theorem A.4), for y in V we have

$$\int_{H'_{ny}} h_{ny}(x)dx = \mathbf{P}(X_{yn} \in H'_{ny})$$

$$= \mathbf{P}(|j_{X_{yn}k} - j_{yk}| > 1 \text{ for some } k = 1, \ldots, d)$$

$$\leq \sum_{k=1}^{d} \mathbf{P}(|j_{X_{yn}k} - j_{yk}| > 1) = 0.$$

Similarly for x in V we have

$$\int_{G'_{nx}} g_{nx}(y)dy = \mathbf{P}(Y_{xn} \in G'_{nx})$$

$$= \mathbf{P}(|j_{Y_{xn}k} - j_{xk}| > 1 \text{ for some } k = 1, \dots, d)$$

$$\leq \sum_{k=1}^{d} \mathbf{P}(|j_{Y_{xn}k} - j_{xk}| > 1) = 0.$$

So trivially, for each positive real number ϵ and relevant region V, S_n is an ϵ, V-RS for s_n.

We will now find expressions for the diameters and volumes of the cross-sections. Each cross-section H_{ny} is a d-dimensional rectangle whose kth side has length $3h_k$. To see this note that for a fixed y_k, if $|j_{xk} - j_{yk}| \leq 1$ then either $j_{xk} = j_{yk} - 1$, j_{yk}, or $j_{yk} + 1$. Each case corresponds to an interval of length h_k. Thus

$$\delta(H_{ny}) = 3 \left(\sum_{k=1}^{d} h_k^2 \right)^{\frac{1}{2}} \tag{4.4}$$

$$\lambda(H_{ny}) = 3^d \prod_{k=1}^{d} h_k. \tag{4.5}$$

Thus, using equality (4.4), we see that

$$S_n \text{ is narrowing} \Leftrightarrow \sup_y \delta(H_{ny}) \to 0$$

$$\Leftrightarrow 3 \left(\sum_{k=1}^{d} h_k^2 \right)^{\frac{1}{2}} \to 0 \text{ as } n \to \infty$$

$$\Leftrightarrow h_k \to 0 \text{ for all } k = 1, \dots, d.$$

In addition S_n is regular in height since, using equality (4.5),

$$\sup_{y\in V}\lambda(H_{ny})\sup_{x\in H_{ny}} h_{ny}(x) = 3^d\prod_{k=1}^d h_k\frac{1}{\prod_{k=1}^d h_k} = 3^d < \infty.$$

Furthermore, using Lemma E.1 and equalities (4.4) and (4.5), we see that

$$\begin{aligned} S_n \text{ is regular in size} &\Leftrightarrow \varliminf_n \frac{\inf_{y\in H}\lambda(H_{ny})}{\sup_{y\in H}\delta(H_{ny})^d} > 0 \\ &\Leftrightarrow \varliminf_n \frac{\prod_{k=1}^d h_k^2}{\left(\sum_{k=1}^d h^2\right)^d} > 0 \\ &\Leftrightarrow \varliminf_n \frac{\min_k h_k}{\max_k h_k} > 0. \end{aligned}$$

Finally we will show that S_n is regular in shape if $\varliminf_n \frac{\min_k h_k}{\max_k h_k} > 0$. To see this choose a positive real number B and N in $\mathbf{N}$ such that for any n in $\mathbf{N}$, if $n > N$,

$$\min_k h_k > B\max_k h_k.$$

Let $R = \frac{B}{3d^{\frac{1}{2}}}$ so that using equality (4.4), if $n > N$,

$$R\sup_{y\in H}\delta(H_{ny}) = 3R\left(\sum_{k=1}^d h_k^2\right)^{\frac{1}{2}} = \frac{B}{d^{\frac{1}{2}}}\left(\sum_{k=1}^d h_k^2\right)^{\frac{1}{2}} = \frac{B}{d^{\frac{1}{2}}}d^{\frac{1}{2}}\max_k h_k < \min_k h_k$$

implying that $R\sup_{y\in H}\delta(H_{ny}) < h_k$ for all $k = 1,\ldots,d$. So if $x \in \mathrm{B}\left(y, R\sup_{y\in H}\delta(H_{ny})\right)$ then for each $k = 1,\ldots,d$, $|y_k - x_k| \le \|y - x\| < h_k$ which implies that

$$\left|\frac{y_k}{h_k} - \frac{x_k}{h_k}\right| < 1 \Rightarrow |j_{yj} - j_{xk}| \le 1 \Rightarrow x \in H_{ny}.$$

Thus S_n is regular in shape. □

Theorem 4.4 (Asymptotic Unbiasedness and the Rectangular Histogram Density Estimator)**.** *Given that* $\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0$*, the rectangular histogram density estimator is asymptotically unbiased if and only if* $\lim_n h_k = 0$ *for* $k = 1, \ldots, d$*.*

Proof. The argument here is essentially the same as the above argument in Theorem 4.2 for the standard kernel density estimator with Theorem 4.3 replacing Theorem 4.1. □

Remark. The asymptotic unbiasedness of histograms has been studied before. In particular Abou-Jaoude in [AJ76] provides necessary and sufficient conditions for the asymptotic unbiasedness of histograms. The method of proof used by Abou-Jaoude is completely different from that used above. The histograms he considers are built on an arbitrary partition of a space and not necessarily a rectangular partition as in Theorem 4.4. In the case of a rectangular partition and equality of rate parameters, that is, $h_k = h$ for $k = 1, \ldots, d$, there is agreement in these two results. In addition the results of Abou-Jaoude do not seem to require anything analogous to the regularity conditions used in Theorem 4.4. This may be due to the general regularity conditions being too strong in the case of a rectangular histogram. However Theorem 4.4 may be extended in various ways to kernel functions whose cross-sectional supports are contained in rectangles and are not necessarily uniform. It is not readily clear how one might do this with the results in [AJ76].

4.3 Gamma Smoothed Histogram Density Estimator

Let m_k be a sequence of positive integers, depending on n, for each $k = 1, \ldots, d$, where n is in $\mathbf{N}$. For each j in $\mathbf{Z}^d$ with $j_k \geq 0$, y in $(0, \infty)^d$, $k = 1, \ldots, d$, let $A_{njk} = [\frac{j_k}{m_k}, \frac{j_k+1}{m_k})$, $A_{nj} = A_{nj1} \times \cdots \times A_{njd}$, $p_{njk}(y_k) = m_k \frac{(m_k y_k)^{j_k}}{j_k!} e^{-m_k y_k}$, $p_{nj}(y) = \prod_{k=1}^d p_{njk}(y_k)$, and finally let

$$s_n(x, y) = \sum_j p_{nj}(y) I_{Anj}(x)$$

where $x \in (0, \infty)^d$. The sequence of functions s_n is a sequence of kernel functions on $(0, \infty)^d \times (0, \infty)^d$. Using this sequence of functions, the density estimator g_n from Definition 2.3 is given by

$$g_n(y) = \frac{1}{n} \sum_{i=1}^n \sum_j \prod_{k=1}^d m_k \frac{(m_k y_k)^{j_k}}{j_k!} e^{-m_k y_k} I_{A_{nj}}(X_i) \tag{4.6}$$

where X_i is the ith observation. We will call this estimator the *Gamma smoothed histogram density estimator.*

Next we define a region sequence for this sequence of kernel functions. Let $j_{xk} = [x_k m_k]$ and $j_x = (j_{xk})_{k=1,\ldots,d}$. Let A_k be a positive constant for $k = 1, \ldots, d$ and let

$$S_n = \left\{ (x, y) \in (0, \infty)^d \times (0, \infty)^d : \left| \frac{j_{xk}}{m_k} - y_k \right| < A_k m_k^{-\frac{1}{2}}, \text{ for all } k = 1, \ldots, d \right\}.$$

It is easy to see that

$$Y_{nxk} \sim \text{Gamma}\left(j_{xk} + 1, \frac{1}{m_k} \right) \text{ and } j_{X_{ny}k} \sim \text{Poisson}(m_k y_k).$$

Theorem 4.5 (Region Sequences for the Gamma Smoothed Histogram Density Estimator)**.** *For each positive real number ϵ, and relevant region V, there are constants A_k for $k = 1, \ldots, d$ such that S_n is an ϵ, V-RS for s_n. Also the following hold:*

1. $\lim_n m_k = \infty$ *for* $k = 1, \ldots, d$ *if and only if* S_n *is narrowing.*

2. $\overline{\lim}_n \frac{\max_k m_k}{\min_k m_k} < \infty$ *if and only if* S_n *is regular.*

Remark. There are number of challenges in proving this result that did not appear in Theorems 4.1 and 4.3. The cross-sectional diameters, volumes, and probabilities outside S_n are not uniform and depend on what cross-section is being considered. In addition establishing regularity in height is technically difficult and requires a theorem found in the Appendix. The compactness of the region of significance V will be important in overcoming these challenges.

Proof of Theorem 4.5. Let ϵ be a positive real number and V be a relevant region. Since V is compact, for each $k = 1, \ldots, d$ there is a real number v_k that is an upper bound for each y_k, where $y \in V$. For each $k = 1, \ldots, d$ choose a positive real number A_k such that

$$A_k > \left(\frac{v_k + 2}{\epsilon} \right)^{\frac{1}{2}}.$$

Then, using Boole's and Chebyshev's inequalities (Theorems A.4 and A.3), for y in V, we have

$$
\begin{aligned}
\int_{H'_{ny}} h_{ny}(x)dx &= \mathbf{P}(X_{ny} \in H'_{ny}) \\
&= \mathbf{P}\left(\left|\frac{j_{X_{ny}k}}{m_k} - y_k\right| > A_k m_k^{-\frac{1}{2}} \text{ for some } k = 1, \ldots, d\right) \\
&\leq \sum_{k=1}^{d} \mathbf{P}\left(\left|\frac{j_{X_{ny}k}}{m_k} - y_k\right| > A_k m_k^{-\frac{1}{2}}\right) \\
&= \sum_{k=1}^{d} \mathbf{P}(|j_{X_{ny}k} - m_k y_k| > A_k m_k^{\frac{1}{2}}) \\
&= \sum_{k=1}^{d} \frac{\mathbf{V}(j_{X_{ny}k})}{A_k^2 m_k} = \sum_{k=1}^{d} \frac{m_k y_k}{A_k^2 m_k} = \sum_{k=1}^{d} \frac{y_k}{A_k^2} \\
&\leq \sum_{k=1}^{d} \frac{v_k + 2}{A_k^2} < \sum_{k=1}^{d} \frac{v_k + 2}{\frac{v_k+2}{\epsilon} d} = \epsilon,
\end{aligned}
$$

and for x in V we have

$$\begin{aligned}
\int_{G'_{nx}} g_{nx}(y)dy &= \mathbf{P}(Y_{nx} \in G'_{nx}) \\
&= \mathbf{P}\left(\left|\frac{j_{xk}}{m_k} - Y_{nx}\right| > A_k m_k^{-\frac{1}{2}} \text{ for some } k = 1, \ldots, d\right) \\
&\leq \sum_{k=1}^{d} \mathbf{P}\left(\left|\frac{j_{xk}}{m_k} - Y_{nx}\right| > A_k m_k^{-\frac{1}{2}}\right) \\
&\leq \sum_{k=1}^{d} \frac{\mathbf{V}(Y_{xnk}) + \left(\mathbf{E}(Y_{xnk}) - \frac{j_{xk}}{m_k}\right)^2}{A_k^2 m_k^{-1}} \\
&= \sum_{k=1}^{d} \frac{\frac{j_{xk}+1}{m_k^2} + \left(\frac{j_{xk}+1}{m_k} - \frac{j_{xk}}{m_k}\right)^2}{A_k^2 m_k^{-1}} \\
&= \sum_{k=1}^{d} \frac{\frac{j_{xk}+1}{m_k} + \frac{1}{m_k}}{A_k^2} = \sum_{k=1}^{d} \frac{\frac{j_{xk}}{m_k} + \frac{2}{m_k}}{A_k^2} \leq \sum_{k=1}^{d} \frac{\frac{m_k x_k}{m_k} + \frac{2}{m_k}}{A_k^2} \\
&\leq \sum_{k=1}^{d} \frac{x_k + 2}{A_k^2} \leq \sum_{k=1}^{d} \frac{v_k + 2}{A_k^2} < \sum_{k=1}^{d} \frac{v_k + 2}{\frac{v+2}{\epsilon} d} = \epsilon.
\end{aligned}$$

Thus for each positive real number ϵ and relevant region V, there are positive real numbers A_k for $k = 1, \ldots, d$ such that S_n is an ϵ, V-RS for s_n.

In order to deal with the diameter and volume of the cross-sections H_{ny}, we will find related regions that are simpler to deal with. Note that

$$j_{xk} = [m_k x_k] \leq m_k x_k < [m_k x_k] + 1 = j_{xk} + 1$$

so that

$$\left|\frac{j_{xk}}{m_k} - x_k\right| < \frac{1}{m_k}. \tag{4.7}$$

Now for each $k = 1, \ldots, d$ choose $\underline{A}_k$ such that $0 < \underline{A}_k$ and $\underline{A}_k + 1 \leq A_k$ and choose $\overline{A}_k$ such that $\overline{A}_k \geq 1 + A_k$ and define the following sets:

$$\underline{H}_{ny} = \{x \in (0, \infty)^d : |x_k - y_k| \leq \underline{A}_k m_k^{-\frac{1}{2}} \text{ for all } k = 1, \ldots, d\}$$

$$\overline{H}_{ny} = \{x \in (0, \infty)^d : |x_k - y_k| \leq \overline{A}_k m_k^{-\frac{1}{2}} \text{ for all } k = 1, \ldots, d\}$$

We have $\underline{H}_{ny} \subseteq H_{ny}$ since, using inequality (4.7), for $k = 1, \ldots, d$

$$\begin{aligned}
\left| \frac{j_{xk}}{m_k} - y_k \right| &\leq \left| \frac{j_{xk}}{m_k} - x_k \right| + |x_k - y_k| \\
&< \frac{1}{m_k} + \underline{A}_k m_k^{-\frac{1}{2}} \\
&\leq (1 + \underline{A}_k) m_k^{-\frac{1}{2}} \\
&< A_k m_k^{-\frac{1}{2}}.
\end{aligned}$$

Also we have $H_{ny} \subseteq \overline{H}_{ny}$ since, for $k = 1, \ldots, d$

$$\begin{aligned}
|x_k - y_k| &\leq \left| \frac{j_{xk}}{m_k} - x_k \right| + \left| \frac{j_{xk}}{m_k} - y_k \right| \\
&\leq \frac{1}{m_k} + A_k m_k^{-\frac{1}{2}} \\
&\leq (1 + A_k) m_k^{-\frac{1}{2}} \\
&\leq \overline{A}_k m_k^{-\frac{1}{2}}.
\end{aligned}$$

It is clear that $\{x_k \in \mathbf{R} : y_k \leq x_k \leq y_k + \underline{A}_k m_k^{-\frac{1}{2}}\}$ is contained in the kth side of $\underline{H}_{ny}$. Thus $\underline{H}_{ny}$ always contains a d-dimensional rectangle whose kth side has length $\underline{A}_k m_k^{-\frac{1}{2}}$. In addition $\{x_k \in \mathbf{R} : y_k - \overline{A}_k m_k^{-\frac{1}{2}} \leq x_k \leq y_k + \overline{A}_k m_k^{-\frac{1}{2}}\}$ contains the kth side of $\overline{H}_{ny}$ and so $\overline{H}_{ny}$ is always contained in a d-dimensional rectangle whose kth side has length $2\overline{A}_k m_k^{-\frac{1}{2}}$. Thus the following hold:

$$\prod_{k=1}^{d} \underline{A}_k m_k^{-\frac{1}{2}} \leq \lambda(\underline{H}_{ny}) \leq \lambda(H_{ny}) \leq \lambda(\overline{H}_{ny}) \leq 2^d \prod_{k=1}^{d} \overline{A}_k m_k^{-\frac{1}{2}}. \tag{4.8}$$

$$\left(\sum_{k=1}^{d} \underline{A}_k^2 m_k^{-1}\right)^{\frac{1}{2}} \leq \delta(\underline{H}_{ny}) \leq \delta(H_{ny}) \leq \delta(\overline{H}_{ny}) \leq 2\left(\sum_{k=1}^{d} \overline{A}_k^2 m_k^{-1}\right)^{\frac{1}{2}}. \tag{4.9}$$

Using statement (4.9) it is clear that $\sup_{y\in H} \delta(H_{ny}) \to 0$ if and only if $m_k \to \infty$ for $k = 1, \ldots, d$ and so S_n is narrowing if and only if $m_k \to \infty$ for $k = 1, \ldots, d$.

Furthermore using statements (4.8) and (4.9), we have that

$$\frac{\prod_{k=1}^{d} \underline{A}_k m_k^{-\frac{1}{2}}}{2^d \left(\sum_{k=1}^{d} \overline{A}_k^2 m_k^{-1}\right)^{\frac{d}{2}}} \leq \frac{\inf_{y\in H} \lambda(H_{ny})}{\sup_{y\in H} \delta(H_{ny})^d} \leq \frac{2^d \prod_{k=1}^{d} \overline{A}_k m_k^{-\frac{1}{2}}}{\left(\sum_{k=1}^{d} \underline{A}_k^2 m_k^{-1}\right)^{\frac{d}{2}}} \tag{4.10}$$

and we see that S_n is regular in size if and only if

$$\underline{\lim}_n \frac{\min_k m_k^{-\frac{1}{2}}}{\max_k m_k^{-\frac{1}{2}}} > 0$$

which is equivalent to

$$\overline{\lim}_n \frac{\max_k m_k}{\min_k m_k} < \infty.$$

We will now show that S_n is regular in shape if $\overline{\lim}_n \frac{\max_k m_k}{\min_k m_k} < \infty$. For each y in $(0, \infty)^d$ and $k = 1, \ldots, d$ choose a positive real number B_k such that $y_k > B_k$ and

$\underline{A}_k > B_k$. Also choose a positive real number B and a number N in $\mathbf{N}$ such that for all n in $\mathbf{N}$, if $n > N$ we have

$$\min_k m_k^{-\frac{1}{2}} > B \max_k m_k^{-\frac{1}{2}}.$$

Now let

$$R = \frac{B \min_k B_k}{2\left(\sum_{k=1}^d \overline{A}^2\right)^{\frac{1}{2}}}.$$

It follows that, using statement (4.9), for all n in $\mathbf{N}$, if $n > N$,

$$\begin{aligned}
R \sup_{y\in H} \delta(H_{ny}) &\le 2R \left(\sum_{k=1}^d \overline{A}_k^2 m_k^{-1}\right)^{\frac{1}{2}} \\
&= \frac{B \min_k B_k}{2\left(\sum_{k=1}^d \overline{A}_k^2\right)^{\frac{1}{2}}} 2 \left(\sum_{k=1}^d \overline{A}_k^2 m_k^{-1}\right)^{\frac{1}{2}} \\
&\le B(\min_k B_k)(\max_k m_k^{-\frac{1}{2}}) \\
&\le (\min_k B_k)(\min_k m_k^{-\frac{1}{2}}) \\
&\le B_k m_k^{-\frac{1}{2}}
\end{aligned}$$

Thus if $x \in \mathrm{B}\left(y, R \sup_{y\in H} \delta(H_{ny})\right)$ we have, for $k = 1, \ldots, d$,

$$|x_k - y_k| \le \|x - y\| \le B_k m_k^{-\frac{1}{2}} < \underline{A}_k m_k^{-\frac{1}{2}}$$

and

$$|x_k - y_k| \le B_k m_k^{-\frac{1}{2}} \le B_k$$

so that

$$0 < y_k - B_k \leq x_k.$$

This shows that $\mathrm{B}\left(y, R\sup_{y\in H} \delta(H_{ny})\right) \subseteq \underline{H}_{ny} \subseteq H_{ny}$ and so S_n is regular in shape.

Finally we will show that S_n is regular in height. For each $k = 1, \ldots, d$ choose a compact set W_k contained in $(0, \infty)$ such that $V \subseteq W_1 \times \cdots \times W_d$ and let a_k be the smallest value in W_k. Now note that

$$x \in H_{ny} \Leftrightarrow x \in (0, \infty)^d \text{ and } \left|\frac{j_{xk}}{m_k} - y_k\right| \leq A_k m_k^{-\frac{1}{2}}, k = 1, \ldots, d.$$

Thus using Theorem C.1, there is an N in $\mathbf{N}$ such that for all y in V and x in H_{ny}, if $n > N$, then

$$P_{nj_x}(y) < 2F_{nj_x}(y), \tag{4.11}$$

where the functions P and F are defined as in the referenced theorem. Now define a positive real number C by

$$C = (2\pi)^{\frac{d}{2}} \prod_{k=1}^{d} a_k^{\frac{1}{2}}$$

so that, using the definition of F we have,

$$\left(\prod_{k=1}^{d} m_k\right)^{\frac{1}{2}} F_{nj_x}(y)$$

$$= \frac{\prod_{k=1}^{d} m_k^{\frac{1}{2}}}{(2\pi)^{\frac{d}{2}} \left(\prod_{k=1}^{d} m_k y_k\right)^{\frac{1}{2}}} \exp\left(-\sum_{k=1}^{d} \frac{m_k}{2y_k}\left(\frac{j_{xk}}{m_k} - y_k\right)^2\right)$$

$$\leq \frac{1}{(2\pi)^{\frac{d}{2}} \left(\prod_{k=1}^{d} y_k\right)^{\frac{1}{2}}}$$

$$\leq \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{k=1}^{d} a_k^{\frac{1}{2}}} = \frac{1}{C} \tag{4.12}$$

for y in V, x in H_{ny}, and n in $\mathbf{N}$. In addition we have that

$$h_{ny}(x) = p_{nj_x}(y)$$

$$= \prod_{k=1}^{d} m_k \frac{(m_k y_k)^{j_{xk}}}{j_{xk}!} e^{-m_k y_k}$$

$$= \left(\prod_{k=1}^{d} m_k\right)\left(\prod_{k=1}^{d} \frac{(m_k y_k)^{j_{xk}}}{j_{xk}!} e^{-m_k y_k}\right)$$

$$= \left(\prod_{k=1}^{d} m_k\right) P_{nj_x}(y). \tag{4.13}$$

Then, using inequalities (4.8), (4.11), and (4.12), and equality (4.13), it follows that, if $n > N$,

$$\sup_{y\in V}\lambda(H_{ny})\sup_{x\in H_{ny}} h_{ny}(x)$$

$$\leq 2^d\left(\prod_{k=1}^d \overline{A}_k m_k^{-\frac{1}{2}}\right)\sup_{y\in V}\sup_{x\in H_{ny}}\left(\prod_{k=1}^d m_k\right)P_{nj_x}(y)$$

$$\leq 2^{d+1}\left(\prod_{k=1}^d \overline{A}_k m_k^{-\frac{1}{2}}\right)\left(\prod_{k=1}^d m_k\right)^{\frac{1}{2}}\sup_{y\in V}\sup_{x\in H_{ny}}\left(\prod_{k=1}^d m_k\right)^{\frac{1}{2}}F_{nj_x}(y)$$

$$\leq \frac{2^{d+1}}{C}\prod_{k=1}^d \overline{A}_k$$

Thus S_n is regular in height. □

Theorem 4.6 (Asymptotic Unbiasedness of the Gamma Smoothed Histogram Density Estimator). *Given that $\overline{\lim}_n \frac{\max_k m_k}{\min_k m_k} < \infty$, the gamma smoothed histogram density estimator is asymptotically unbiased if and only if $\lim_n m_k = \infty$ for $k = 1, \ldots, d$.*

Proof. Again the argument here is essentially the same as the above argument in Theorem 4.2 for the standard kernel density estimator with Theorem 4.5 replacing Theorem 4.1. □

Remark. Gamma smoothed histograms have not been widely studied. They are mentioned only in a small number of articles and few seem to deal with L^1 asymptotic unbiasedness. In [Sta83], Stadtmüller considers weak uniform consistency of a variety of smoothed histograms, including the Gamma smoothed histogram. In [BS05], Bouezmarni and Scaillet consider the one-dimensional case of the Gamma smoothed histogram density estimator and study L^1 consistency of the estimator and asymptotic unbiasedness but provide only a sufficient condition. Their method of proof relies on specific characteristics of the Gamma distribution.

4.4 Beta Smoothed Histogram Density Estimator

Let m_k be a sequence of positive integers, depending on n, for each $k = 1, \ldots, d$, where n is in $\mathbf{N}$. For each y in $(0,1)^d$, $k = 1, \ldots, d$, j in $\mathbf{Z}^d$ with $0 \le j_k \le m_k$, and n in $\mathbf{N}$, let $A_{njk} = [\frac{j_k}{m_k+1}, \frac{j_k+1}{m_k+1})$, $A_{nj} = A_{nj1} \times \cdots \times A_{njd}$, $p_{njk}(y_k) = (m_k + 1)\binom{m_k}{j_k} y_k^{j_k}(1-y_k)^{m_k-j_k}$, $p_{nj}(y) = \prod_{k=1}^d p_{njk}(y_k)$, and finally let

$$s_n(x,y) = \sum_j p_{nj}(y) I_{Anj}(x)$$

where $x \in (0,1)^d$. The sequence of functions s_n is a sequence of kernel functions on $(0,1)^d \times (0,1)^d$. Using this sequence of functions, the density estimator g_n from Definition 2.3 is given by

$$g_n(y) = \frac{1}{n} \sum_{i=1}^n \sum_j \prod_{k=1}^d (m_k+1) \binom{m_k}{j_k} y_k^{j_k} (1-y_k)^{m_k-j_k} I_{A_{nj}}(X_i)$$

where X_i is the ith observation. We will call this estimator the the *Beta smoothed histogram density estimator.*

Next we define a region sequence for this sequence of kernel functions. Let $j_{xk} = [x_k(m_k+1)]$ and $j_x = (j_{xk})_{k=1,\ldots,d}$. Let A_k be a positive constant for $k = 1, \ldots, d$ and let

$$S_n = \left\{ (x,y) \in (0,1)^d \times (0,1)^d : \left| \frac{j_{xk}}{m_k} - y_k \right| < A_k m_k^{-\frac{1}{2}}, \text{ for all } k = 1, \ldots, d \right\}.$$

It is easy to see that

$$Y_{xnk} \sim \text{Beta}(j_{xk}+1, m_k - j_{xk}+1) \text{ and } j_{X_{yn}k} \sim \text{Binomial}(m_k, y_k).$$

Theorem 4.7 (Region Sequences for the Beta Smoothed Histogram Density Estimator)**.** *For each positive real number ϵ, and relevant region V, there are positive real numbers A_k for $k = 1, \ldots, d$ such that S_n is an ϵ, V-RS for s_n. Also the following hold:*

1. $\lim_n m_k = \infty$ *for* $k = 1, \ldots, d$ *if and only if* S_n *is narrowing.*

2. $\overline{\lim}_n \frac{\max_k m_k}{\min_k m_k} < \infty$ *if and only if* S_n *is regular.*

Proof of Theorem 4.7. Let ϵ be a positive real number and V be a relevant region. For each $k = 1, \ldots, d$ choose a real number A_k such that $A_k > \max((\frac{5d}{\epsilon})^{\frac{1}{2}}, 3)$. Then, using Boole's and Chebyshev's inequalities (Theorems A.4 and A.3) for y in V we have

$$
\begin{aligned}
\int_{H'_{ny}} h_{ny}(x)dx &= \mathbf{P}(X_{ny} \in H'_{ny}) \\
&= \mathbf{P}\left(\left| \frac{j_{X_{ny}k}}{m_k} - y_k \right| > A_k m_k^{-\frac{1}{2}} \text{ for some } k = 1, \ldots, d \right) \\
&\leq \sum_{k=1}^{d} \mathbf{P}\left(\left| \frac{j_{X_{ny}k}}{m_k} - y_k \right| > A_k m_k^{-\frac{1}{2}} \right) \\
&= \sum_{k=1}^{d} \mathbf{P}\left(\left| j_{X_{ny}k} - m_k y_k \right| > A_k m_k^{\frac{1}{2}} \right) \\
&\leq \sum_{k=1}^{d} \frac{\mathbf{V}(j_{X_{ny}k})}{A_k^2 m_k} = \sum_{k=1}^{d} \frac{m_k y_k (1-y_k)}{A_k^2 m_k} = \sum_{k=1}^{d} \frac{y_k(1-y_k)}{A_k^2} \\
&\leq \frac{1}{4} \sum_{k=1}^{d} \frac{1}{A_k^2} < \frac{1}{4} \sum_{k=1}^{d} \frac{1}{\frac{5d}{\epsilon}} < \frac{1}{d} \sum_{k=1}^{d} \epsilon = \epsilon.
\end{aligned}
$$

Similarly for x in V, we have

$$
\begin{aligned}
\int_{G'_{nx}} g_{nx}(y)dy &= \mathbf{P}(Y_{nx} \in G'_{nx}) \\
&= \mathbf{P}\left(\left|\frac{j_{xk}}{m_k} - Y_{nx}\right| > A_k m_k^{-\frac{1}{2}} \text{ for some } k = 1, \ldots, d\right) \\
&\leq \sum_{k=1}^{d} \mathbf{P}\left(\left|\frac{j_{xk}}{m_k} - Y_{nxk}\right| > A_k m_k^{-\frac{1}{2}}\right) \\
&\leq \sum_{k=1}^{d} \frac{\mathbf{V}(Y_{nxk}) + \left(\mathbf{E}(Y_{nxk}) - \frac{j_{xk}}{m_k}\right)^2}{A_k^2 m_k^{-1}} \\
&\leq \sum_{k=1}^{d} \frac{\frac{(j_{xk}+1)(m_k - j_{xk}+1)}{(m_k+2)^2(m_k+3)} + \left(\frac{j_{xk}+1}{m_k+2} - \frac{j_{xk}}{m_k}\right)^2}{A_k^2 m_k^{-1}} \\
&= \sum_{k=1}^{d} \frac{\frac{\left(\frac{j_{xk}+1}{m_k}\right)\left(1 - \frac{j_{xk}-1}{m_k}\right)}{\left(1+\frac{2}{m_k}\right)^2\left(1+\frac{3}{m_k}\right)} + m_k\left(\frac{1 - \frac{2j_{xk}}{m_k}}{m_k+2}\right)^2}{A_k^2} \\
&\leq \sum_{k=1}^{d} \frac{2^2+1}{A_k^2} = \sum_{k=1}^{d} \frac{5}{A_k^2} < \sum_{k=1}^{d} \frac{5}{\frac{5d}{\epsilon}} = \epsilon.
\end{aligned}
$$

So for each positive real number ϵ and relevant region V, there are positive real numbers A_k for $k = 1, \ldots, d$ such that S_n is an ϵ, V-RS for s_n.

Now in order to deal with the diameter and volume of the cross-sections H_{ny}, we will find related regions that are simpler to deal with. To this end note that

$$
j_{xk} = [x_k(m_k+1)] \leq x_k(m_k+1) < [x_k(m_k+1)] + 1 = j_{xk} + 1,
$$

and so

$$
\frac{j_{xk}}{m_k} \leq \frac{m_k+1}{m_k} x_k < \frac{j_{xk}+1}{m_k},
$$

which implies that

$$\left|\frac{m_k+1}{m_k}x_k - \frac{j_{xk}}{m_k}\right| \leq \frac{1}{m_k}.$$

Hence

$$\left|\frac{j_{xk}}{m_k} - x_k\right| \leq \left|\frac{m_k+1}{m_k}x_k - \frac{j_{xk}}{m_k}\right| + \left|\frac{m_k+1}{m_k}x_k - x_k\right| < \frac{1}{m_k} + \frac{1}{m_k} = \frac{2}{m_k}. \tag{4.14}$$

Choose positive real numbers $\overline{A}_k$ and $\underline{A}_k$ such that $0 < \underline{A}_k < \frac{1}{2}$ and $\overline{A}_k > A_k + 2$. We define the following sets:

$$\underline{H}_{ny} = \{x \in (0,1)^d : |x_k - y_k| \leq \underline{A}_k m_k^{-\frac{1}{2}} \text{ for all } k = 1, \ldots, d\}$$

$$\overline{H}_{ny} = \{x \in (0,1)^d : |x_k - y_k| \leq \overline{A}_k m_k^{-\frac{1}{2}} \text{ for all } k = 1, \ldots, d\}$$

Note that $\underline{H}_{ny} \subseteq H_{ny}$ since, if $x \in \underline{H}_{ny}$, for any $k = 1, \ldots, d$, we have, using inequality (4.14), that

$$\begin{aligned}
\left|\frac{j_{xk}}{m_k} - y_k\right| &\leq \left|\frac{j_{xk}}{m_k} - x_k\right| + |x_k - y_k| \\
&< \frac{2}{m_k} + \underline{A}_k m_k^{-\frac{1}{2}} \\
&\leq (2 + \underline{A}_k) m_k^{-\frac{1}{2}} \\
&< 3 m_k^{-\frac{1}{2}} \\
&\leq A_k m_k^{-\frac{1}{2}}.
\end{aligned}$$

Similarly, we have $H_{ny} \subseteq \overline{H}_{ny}$, since, if $x \in H_{ny}$, for any $k = 1, \ldots, d$, we have that

$$
\begin{aligned}
|x_k - y_k| &\le \left|\frac{j_{xk}}{m_k} - x_k\right| + \left|\frac{j_{xk}}{m_k} - y_k\right| \\
&< \frac{2}{m_k} + A_k m_k^{-\frac{1}{2}} \\
&\le \overline{A}_k m_k^{-\frac{1}{2}}.
\end{aligned}
$$

Since, for each $k = 1, \ldots, d$, $\underline{A}_k < \frac{1}{2}$, we see that either $\{x_k \in \mathbf{R} : y_k \le x_k \le y_k + \underline{A}_k m_k^{-\frac{1}{2}}\}$ or $\{x_k \in \mathbf{R} : y_k - \underline{A}_k m_k^{-\frac{1}{2}} \le x_k \le y_k\}$ is contained in the kth side of $\underline{H}_{ny}$. Thus $\underline{H}_{ny}$ always contains a d-dimensional rectangle whose kth side has length $\underline{A}_k m_k^{-\frac{1}{2}}$. In addition $\{x_k \in \mathbf{R} : y_k - \overline{A}_k m_k^{-\frac{1}{2}} \le x_k \le y_k + \overline{A}_k m_k^{-\frac{1}{2}}\}$ contains the kth side of $\overline{H}_{ny}$ and so $\overline{H}_{ny}$ is always contained in a d-dimensional rectangle whose kth side has length $2\overline{A}_k m_k^{-\frac{1}{2}}$. Thus

$$
\prod_{k=1}^{d} \underline{A}_k m_k^{-\frac{1}{2}} \le \lambda(\underline{H}_{ny}) \le \lambda(H_{ny}) \le \lambda(\overline{H}_{ny}) \le 2^d \prod_{k=1}^{d} \overline{A}_k m_k^{-\frac{1}{2}} \tag{4.15}
$$

$$
\left(\sum_{k=1}^{d} \underline{A}_k^2 m_k^{-1}\right)^{\frac{1}{2}} \le \delta(\underline{H}_{ny}) \le \delta(H_{ny}) \le \delta(\overline{H}_{ny}) \le 2\left(\sum_{k=1}^{d} \overline{A}_k^2 m_k^{-1}\right)^{\frac{1}{2}}. \tag{4.16}
$$

Using statement (4.16) it is easy to see that $\sup_{y \in H} \delta(H_{ny}) \to 0$ if and only if $m_k \to \infty$ for $k = 1, \ldots, d$ so S_n is narrowing if and only if $m_k \to \infty$ for $k = 1, \ldots, d$.

Also using statements (4.15) and (4.16) we see that

$$
\frac{\prod_{k=1}^{d} \underline{A}_k m_k^{-\frac{1}{2}}}{2^d \left(\sum_{k=1}^{d} \overline{A}_k^2 m_k^{-1}\right)^{\frac{d}{2}}} \le \frac{\inf_{y \in H} \lambda(H_{ny})}{\sup_{y \in H} \delta(H_{ny})^d} \le \frac{2^d \prod_{k=1}^{d} \overline{A}_k m_k^{-\frac{1}{2}}}{\left(\sum_{k=1}^{d} \underline{A}_k^2 m_k^{-1}\right)^{\frac{d}{2}}}
$$

and so using Lemma E.1 we see that S_n is regular in size if and only if

$$\underline{\lim}_n \frac{\min_k m_k^{-\frac{1}{2}}}{\max_k m_k^{-\frac{1}{2}}} > 0$$

and this is equivalent to

$$\overline{\lim}_n \frac{\max_k m_k}{\min_k m_k} < \infty.$$

We will now show that S_n is regular in shape if $\overline{\lim}_n \frac{\max_k m_k}{\min_k m_k} < \infty$. For y in $(0,1)^d$ and each $k = 1, \ldots, d$, choose $B_k > 0$ such that $y_k - B_k > 0$, $y_k + B_k < 1$ and $B_k < \underline{A_k}$ for each $k = 1, \ldots, d$. Now choose a positive real number B and a number N in $\mathbf{N}$ such that for each n in $\mathbf{N}$, if $n > N$,

$$\min_k m_k^{-\frac{1}{2}} > B \max_k m_k^{-\frac{1}{2}}.$$

Now let

$$R = \frac{B \min_k B_k}{2 \left(\sum_{k=1}^{d} \overline{A}_k^2 \right)^{\frac{1}{2}}}$$

so that, using (4.16), for $n > N$,

$$R \sup_{y\in H} \delta(H_{ny}) \le 2R \left(\sum_{k=1}^{d} \overline{A}_k^2 m_k^{-1} \right)^{\frac{1}{2}}$$

$$= \frac{B \min_k B_k}{2\left(\sum_{k=1}^{d} \overline{A}_k^2\right)^{\frac{1}{2}}} 2 \left(\sum_{k=1}^{d} \overline{A}_k^2 m_k^{-1} \right)^{\frac{1}{2}}$$

$$\le B(\min_k B_k)(\max_k m_k^{-\frac{1}{2}})$$

$$\le (\min_k B_k)(\min_k m_k^{-\frac{1}{2}})$$

$$\le B_k m_k^{-\frac{1}{2}}.$$

Hence if $x \in \mathrm{B}\left(y, R \sup_{y\in H} \delta(H_{ny})\right)$ then

$$|x_k - y_k| \le B_k m_k^{-\frac{1}{2}} < \underline{A}_k m_k^{-\frac{1}{2}}$$

and

$$|x_k - y_k| \le B_k m_k^{-\frac{1}{2}} \le B_k.$$

This implies that $0 < y_k - B_k \le x_k \le y_k + B_k < 1$ and so

$$\mathrm{B}\left(y, R \sup_{y\in H} \delta(H_{ny})\right) \subseteq \underline{H}_{ny} \subseteq H_{ny}.$$

Thus S_n is regular in shape.

Finally we will show that S_n is regular in height. For each $k = 1, \ldots, d$ choose a compact set W_k contained in $(0, 1)$ such that $V \subseteq W_1 \times \cdots \times W_d$ and let a_k and

b_k be the smallest and largest values respectively in W_k. Now note that

$$x \in H_{ny} \Leftrightarrow x \in (0,1)^d \text{ and } \left| \frac{j_{xk}}{m_k} - y_k \right| \leq A_k m_k^{-\frac{1}{2}}, k = 1, \ldots, d.$$

Thus using Theorem D.1, there is an N in $\mathbf{N}$ such that for all y in V and x in H_{ny}, if $n > N$, then

$$P_{nj_x}(y) < 2F_{nj_x}(y), \tag{4.17}$$

where the functions P and F are defined as in the referenced theorem. Now define a positive real number C by

$$C = \pi^{\frac{d}{2}} \prod_{k=1}^{d} (a_k(1-b_k))^{\frac{1}{2}}$$

so that, using the definition of F we have,

$$\begin{aligned}
&\left(\prod_{k=1}^{d} (m_k + 1) \right)^{\frac{1}{2}} F_{nj_x}(y) \\
&\quad = \frac{\prod_{k=1}^{d} (m_k+1)^{\frac{1}{2}}}{(2\pi)^{\frac{d}{2}} \left(\prod_{k=1}^{d} m_k y_k (1-y_k) \right)^{\frac{1}{2}}} \exp\left(- \sum_{k=1}^{d} \frac{m_k}{2y_k(1-y_k)} \left(\frac{j_{xk}}{m_k} - y_k \right)^2 \right) \\
&\quad \leq \frac{\prod_{k=1}^{d} \left(1 + \frac{1}{m_k}\right)^{\frac{1}{2}}}{(2\pi)^{\frac{d}{2}} \left(\prod_{k=1}^{d} y_k(1-y_k) \right)^{\frac{1}{2}}} \\
&\quad \leq \frac{1}{\pi^{\frac{d}{2}} \left(\prod_{k=1}^{d} a_k (1-b_k) \right)^{\frac{1}{2}}} = \frac{1}{C}
\end{aligned} \tag{4.18}$$

for y in V, x in H_{ny}, and n in $\mathbf{N}$. In addition we have that

$$
\begin{aligned}
h_{ny}(x) &= p_{nj}(y) \\
&= \prod_{k=1}^{d} (m_k+1) \binom{m_k}{j_{xk}} y_k^{j_{xk}} (1-y_k)^{m_k - j_{xk}} \\
&= \left(\prod_{k=1}^{d} (m_k+1) \right) \left(\prod_{k=1}^{d} \binom{m_k}{j_{xk}} y_k^{j_{xk}} (1-y_k)^{m_k - j_{xk}} \right) \\
&= \left(\prod_{k=1}^{d} (m_k+1) \right) P_{nj_x}(y). \qquad (4.19)
\end{aligned}
$$

Then, using inequalities (4.15), (4.17), and (4.18), and equality (4.19), it follows that, if $n > N$,

$$
\begin{aligned}
&\sup_{y \in V} \lambda(H_{ny}) \sup_{x \in H_{ny}} h_{ny}(x) \\
&\leq 2^d \left(\prod_{k=1}^{d} \overline{A}_k m_k^{-\frac{1}{2}} \right) \sup_{y \in V} \sup_{x \in H_{ny}} \left(\prod_{k=1}^{d} (m_k+1) \right) P_{nj_x}(y) \\
&\leq 2^{d+1} \left(\prod_{k=1}^{d} \overline{A}_k m_k^{-\frac{1}{2}} \right) \left(\prod_{k=1}^{d} (m_k+1) \right)^{\frac{1}{2}} \sup_{y \in V} \sup_{x \in H_{ny}} \left(\prod_{k=1}^{d} (m_k+1) \right)^{\frac{1}{2}} F_{nj_x}(y) \\
&\leq \frac{2^{d+1}}{C} \prod_{k=1}^{d} \overline{A}_k \left(1 + \frac{1}{m_k} \right)^{\frac{1}{2}} \\
&\leq \frac{2^{\frac{3d}{2}+1}}{C} \prod_{k=1}^{d} \overline{A}_k.
\end{aligned}
$$

Thus S_n is regular in height. □

Theorem 4.8 (Asymptotic Unbiasedness of the Beta Smoothed Histogram Density Estimator). *Given that* $\overline{\lim}_n \frac{\max_k m_k}{\min_k m_k} < \infty$, *the beta smoothed histogram density estimator is asymptotically unbiased if and only if* $\lim_n m_k = \infty$ *for* $k = 1, \ldots, d$.

Proof of Theorem 4.8. Again the argument here is essentially the same as the above argument in Theorem 4.2 for the standard kernel density estimator with Theorem 4.7 replacing Theorem 4.1. □

Remark. Beta smoothed histogram density estimators have been extensively studied. These estimators are often referred to by other names such as Bernstein or Kantorovich polynomial density estimators. Vitale in [Vit75] proposed using this estimator and established some of the basic convergence properties of this estimator. This work was extended by Babu et al. in [GJB02] where results on uniform strong consistency and other types of convergence for the one-dimensional case are obtained. In [Ten95], Tenbusch considers the multidimensional case and obtains the same result as Theorem 4.8 for the case of uniform rate parameters, that is, $m_k = m$ for $k = 1, \ldots, d$. The method of proof is very different from that used in the above theorem and is very reliant on specific properties of the Bernstein polynomials.

Appendix A

Useful Results

Listed here are some standard results used in mathematical analysis and statistics that are made use of in this document. Proofs of these results can be found in many books on analysis and topology and mathematical statistics. In particular Theorem A.1 and Lemma A.2 can be found in [Rud87] and Theorems A.3 an A.4 can be found in [BE00]. Theorem A.5 can be found in [Dev85].

Theorem A.1 (Monotone Convergence Theorem)**.** *Let f_n be a sequence of measurable functions on measure space X with measure μ and suppose that the following hold:*

1. $0 \leq f_n(x) \leq f_{n+1}(x)$ *for all x in X and n in* $\mathbf{N}$*.*
2. $\lim_n f_n(x) = f(x)$ *for all x in X.*

Then the function f is measurable and

$$\lim_n \int_X f_n d\mu = \int_X f d\mu.$$

Lemma A.2 (Urysohn's Lemma)**.** *Suppose that X is a locally compact Hausdorff space, V is open in X, $K \subseteq V$, and K is compact. Then there exists a function f such that the following hold:*

1. *f is continuous.*

2. *$I_K \leq f \leq I_V$.*

3. *The support of f is compact.*

where I_S is the characteristic function of S.

Theorem A.3 (Chebyshev's Inequality)**.** *Let X be a real valued random variable, x be a real number and A be a positive real number. Then*

$$\mathbf{P}(|X - x| > A) \leq \frac{\mathbf{V}(X) + (\mathbf{E}(X) - x)^2}{A^2}.$$

Theorem A.4 (Boole's Inequality)**.** *For each $k = 1, \ldots, d$ let E_k be an event. Then*

$$\mathbf{P}\left(\bigcup_{k=1}^{d} E_k\right) \leq \sum_{k=1}^{d} \mathbf{P}(E_k).$$

Theorem A.5 (Scheffé's Theorem)**.** *Let f and g be densities. Then*

$$\int |f - g| = 2 \sup_A \left| \int_A f - \int_A g \right|.$$

Appendix B

Uniform Approximation of a Density

The following theorem is essentially a version of the theorem that compactly supported continuous functions are dense in L^1. It plays an important role in establishing the basic result on asymptotic unbiasedness, Theorem 3.1.

Theorem B.1 (Approximation of a Density by a Compactly Supported Uniformly Continuous Density)**.** *Let $f : \mathbf{R}^d \to \mathbf{R}$ be a density with support S. Then for any positive real number ϵ, there is a density g with compact support T such that $T \subset \mathring{S}$, g is uniformly continuous, and*

$$\int_{\mathbf{R}^d} |f - g| < \epsilon.$$

Remark. In real analysis a standard theorem states that functions in L^p may be approximated by compactly supported continuous functions. Usually this is stated by saying that compactly supported continuous functions are dense in L^p. The above theorem differs slightly from this in that it states that the approximating function may be chosen so that its support lies inside the interior of the support of the original function and that if the original function is a density, the approximating function may be chosen to be a density as well. The proof is largely the same and roughly follows that of theorem 3.14 given in [Rud87]. Only a few details need to be changed to show that the approximating function may be chosen to have the desired properties.

Proof of Theorem B.1. Let f be a density with support S and let ϵ be a positive real number. For each i in $\mathbf{N}$ define a subset F_i of $\mathbf{R}^d$ by

$$F_i = \left(\bigcup_{x \in S'} \mathrm{B}(x, 1/i) \right)' \cap \overline{\mathrm{B}(0, i)}.$$

It is clear that F_i is compact and that $F_i \subseteq \mathring{S}$ for each such i. In addition $F_i \subseteq F_{i+1}$ for each i in $\mathbf{N}$, which makes F an increasing sequence of sets with limit $\mathring{S}$.

Define the function $f_i : \mathbf{R}^d \to \mathbf{R}$ for each i in $\mathbf{N}$ by

$$f_i = \min(i, I_{F_i} \cdot f).$$

The functions f_i define an increasing sequence of bounded integrable functions that converge pointwise to f. So by the Monotone Convergence Theorem A.1 we can choose a k in $\mathbf{N}$ such that

$$\int_{\mathbf{R}^d} |f - f_k| < \epsilon/3. \tag{B.1}$$

Now define the function $\phi_i : [0, \infty] \to \mathbf{R}$ for each i in $\mathbf{N}$ by

$$\phi_i(t) = a_i \left[\frac{t}{a_i} \right]$$

where $a_i = k/2^{i-1}$. We then have an increasing sequence of integrable functions $s_i : \mathbf{R}^d \to \mathbf{R}$ where $s_i = \phi_i \circ f_k$ which converge pointwise to f_k. Again by the monotone convergence theorem we can choose a p in $\mathbf{N}$ such that

$$\int_{\mathbf{R}^d} |s_p - f_k| < \epsilon/3. \tag{B.2}$$

Define the function $t_i : \mathbf{R}^d \to \mathbf{R}$ for each i in $\mathbf{N}$ by the equation $t_i = s_i - s_{i-1}$, where $s_0 = 0$. It can be shown that $t_i/a_i = I_{T_i}$ for some set T_i contained in the support of f_k. To see this note that

$$\begin{aligned}
\frac{t_i}{a_i} &= \frac{s_i - s_{i-1}}{a_i} \\
&= \frac{\phi_i(f_k) - \phi_{i-1}(f_k)}{a_i} \\
&= \left[\frac{f_k}{a_i}\right] - \frac{a_{i-1}}{a_i}\left[\frac{f_k}{a_{i-1}}\right] \\
&= \left[\frac{f_k}{a_i}\right] - 2\left[\frac{f_k}{2a_i}\right] \\
&= \begin{cases} 0 & \frac{f_k}{a_i} \in \bigcup_{n\in\mathbf{N}}[2n-2, 2n-1) \\ 1 & \frac{f_k}{a_i} \in \bigcup_{n\in\mathbf{N}}[2n-1, 2n). \end{cases}
\end{aligned}$$

Note that the number k in the above definition of a_i is needed to ensure that the above statement is true for $i = 1$. Furthermore $s_p = \sum_{i=1}^{p} t_i$.

Choose an open set W that contains the support of f_k, and hence the support of s_i for i in $\mathbf{N}$, such that $\bar{W} \subseteq \mathring{S}$. This can be done since the support of f_k is contained in F_k and F_k is compact and contained in $\mathring{S}$.

For each i in $\mathbf{N}$, select a compact set K_i and an open set V_i such that $K_i \subseteq T_i \subseteq V_i \subseteq W$ and

$$\lambda(V_i \backslash K_i) < \frac{\epsilon}{9 \cdot 2^i k}. \tag{B.3}$$

By Urysohn's Lemma A.2 there are continuous functions with compact support $h_i : \mathbf{R}^d \to \mathbf{R}$ such that $0 \leq h_i \leq 1$, $h_i = 1$ on K_i, and $h_i = 0$ outside of V_i. Define the function $g : \mathbf{R}^d \to \mathbf{R}$ by the equation $g = \sum_{i=1}^{p} a_i h_i$. By construction g is equal to s_p except on $\bigcup_{i=1}^{p} V_i \backslash K_i$, g is continuous with compact support contained in $\mathring{S}$, and $g \leq 2k$.

Thus using inequality (B.3) we have that

$$\lambda(\bigcup_{i=1}^{p} V_i \backslash K_i) \leq \sum_{i=1}^{p} \lambda(V_i \backslash K_i) < \sum_{i=1}^{p} \frac{\epsilon}{9 \cdot 2^i k} < \frac{\epsilon}{9k}, \tag{B.4}$$

and so, using inequalities (B.1), (B.2), and (B.4), it follows that

$$\int_{\mathbf{R}^d} |f - g| \leq \int_{\mathbf{R}^d} |f - f_k| + \int_{\mathbf{R}^d} |f_k - s_p| + \int_{\mathbf{R}^d} |s_p - g| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + 3k\frac{\epsilon}{9k} = \epsilon.$$

This shows that we can always approximate f by a non-negative continuous function g with compact support (and hence uniformly continuous) contained in the interior of S.

Finally, since f is a density and

$$\begin{aligned}
\int_{\mathbf{R}^d} |f - \frac{g}{\int_{\mathbf{R}^d} g}| &\leq \int_{\mathbf{R}^d} |f-g| + \int_{\mathbf{R}^d} \left| g - \frac{g}{\int_{\mathbf{R}^d} g} \right| \\
&= \int_{\mathbf{R}^d} |f-g| + \left| 1 - \frac{1}{\int_{\mathbf{R}^d} g} \right| \int_{\mathbf{R}^d} g \\
&= \int_{\mathbf{R}^d} |f-g| + \left| \int_{\mathbf{R}^d} g - 1 \right| \\
&= \int_{\mathbf{R}^d} |f-g| + \left| \int_{\mathbf{R}^d} g - \int_{\mathbf{R}^d} f \right| \\
&\leq 2 \int_{\mathbf{R}^d} |f-g|,
\end{aligned}$$

we see that g may be chosen to be sufficiently close to f so that normalizing g gives us a density with the desired properties. □

Appendix C

Uniform Approximation of a Poisson Distribution

Here we prove a theorem dealing with a uniform approximation of the Poisson distribution. This result does not rest on any the work done in the rest of this document but is useful in establishing the results for Gamma smoothed histogram density estimators (see Section 4.3), in particular, it is useful in establishing regularity in height.

Theorem C.1 (Uniform Approximation for the Poisson Distribution)**.** *Let m be in $\mathbf{N}^d$ and j be in $\mathbf{Z}^d$ with $j_k \geq 0$ for $k = 1, \ldots, d$ and suppose that m depends on n where n belongs to $\mathbf{N}$. Let the functions P_{nj} and F_{nj} be given by*

$$P_{nj}(y) = \prod_{k=1}^{d} \frac{(m_k y_k)^{j_k} \exp(-m_k y_k)}{j_k!},$$

$$F_{nj}(y) = \frac{1}{(2\pi)^{\frac{d}{2}} \left(\prod_{k=1}^{d} m_k y_k\right)^{\frac{1}{2}}} \exp\left(-\sum_{k=1}^{d} \frac{m_k}{2y_k} \left(\frac{j_k}{m_k} - y_k\right)^2\right)$$

where $y_k > 0$ for each $k = 1, \dots, d$. Let α be a positive real number such that $\alpha > \frac{1}{3}$. For each $k = 1, \dots, d$, let W_k be a compact subset of the open interval $(0, \infty)$, and A_k be a positive real number and suppose that $m_k \to \infty$ as $n \to \infty$. Then for any positive real number ϵ, there is an integer N, such that for all n in $\mathbf{N}$, if $n > N$,

$$\left| \frac{P_{nj}(y)}{F_{nj}(y)} - 1 \right| < \epsilon$$

for all y_k in W_k and $j_k = 0, \dots$ such that $\left| \frac{j_k}{m_k} - y_k \right| \leq A_k m_k^{-\alpha}$ for each $k = 1, \dots, d$.

Remark. Like Theorem D.1 the proof of this result follows that of Theorem 1.5.2 (Laplace's formula of the theory of probability) in [Lor86]. This theorem of Lorentz is stated for the binomial distribution (not the Poisson distribution), handles only the one-dimensional case, and deals with point-wise convergence (not uniform convergence). The generality stated here requires a number of changes to the proof.

Proof of Theorem C.1. Note that it suffices to prove that each factor of a finite product satisfies the above type of uniform convergence to prove that the product also satisfies it. We have

$$\frac{P_{nj}(y)}{F_{nj}} = \frac{\prod_{k=1}^{d} \frac{(m_k y_k)^{j_k} \exp(-m_k y_k)}{j_k!}}{\frac{1}{(2\pi)^{\frac{d}{2}} \left(\prod_{k=1}^{d} m_k y_k\right)^{\frac{1}{2}}} \exp\left(- \sum_{k=1}^{d} \frac{m_k}{2y_k} \left(\frac{j_k}{m_k} - y_k \right)^2 \right)}$$

$$= \prod_{k=1}^{d} \frac{(m_k y_k)^{j_k}}{j_k!} \exp(-m_k y_k) (2\pi m_k y_k)^{\frac{1}{2}} \exp\left(\frac{m_k}{2y_k} \left(\frac{j_k}{m_k} - y_k \right)^2 \right).$$

We will consider one factor of the above product. For ease of notation we will drop the subscript in what follows and let $u = \frac{j}{m} - y$. Let Q be the factor, so that

$$Q = \frac{(my)^j}{j!} \exp(-my)(2\pi my)^{\frac{1}{2}} \exp\left(\frac{m}{2y}u^2\right).$$

From Sterling's formula, for any n in $\mathbf{N}$, we have that

$$n! = (2\pi n)^{\frac{1}{2}} n^n e^n H_n \tag{C.1}$$

where H_n is some sequence of real numbers such that $H_n \to 1$ as $n \to \infty$. Rewriting Q using equation (C.1) we have

$$\begin{aligned} Q &= \frac{(my)^j \exp(-my)(2\pi my)^{\frac{1}{2}}}{(2\pi j)^{\frac{1}{2}} j^j \exp(-j) H_j} \exp\left(\frac{m}{2y}u^2\right) \\ &= \frac{1}{H_j}\left(\frac{my}{j}\right)^{\frac{1}{2}}\left(\frac{my}{j}\right)^j \exp\left(mu + \frac{m}{2y}u^2\right). \end{aligned}$$

We need to show that Q is a product of factors, each of which satisfies the type of uniform convergence from the statement of the theorem.

First we will show that the factor $\frac{1}{H_j}$ converges uniformly to 1. Let $y_* = \inf_{y \in W} y$. Since W is compact and contained in the interval $(0, \infty)$ we see that $y_* > 0$ and so for $\left|\frac{j}{m} - y\right| \leq Am^{-\alpha}$, we have

$$j \geq my - Am^{1-\alpha} \geq my_* - Am^{1-\alpha} = m(y_* - Am^{-\alpha}). \tag{C.2}$$

Let ϵ be a positive real number. Choose N_0 in $\mathbf{N}$ such that for all n in $\mathbf{N}$ if $n > N_0$ then $|H_m - 1| < \epsilon$. This can be done since $m \to \infty$ as $n \to \infty$. Since $m(y_* - Am^{-\alpha}) \to \infty$ as $n \to \infty$ we can choose N_1 in $\mathbf{N}$ such that if $n > N_1$ then $m(y_* - Am^{-\alpha}) > N_0$. Using inequality (C.2) we see that for $n > N_1$, and any y and j such that $\left|\frac{j}{m} - y\right| \leq Am^{-\alpha}$ we have $j \geq m(y_* - Am^{-\alpha}) > N_0$ and so $|H_j - 1| < \epsilon$. This shows that H_j converges uniformly to 1.

Now we will show that the factor $(\frac{my}{j})^{\frac{1}{2}}$ converges uniformly to 1. Using inequality (C.2) again we have, for y and j such that $|\frac{j}{m} - y| \leq Am^{-\alpha}$,

$$\left|\frac{y}{j/m} - 1\right| \leq \frac{Am^{1-\alpha}}{j} \leq \frac{Am^{1-\alpha}}{my_* - Am^{1-\alpha}} = \frac{A}{m^\alpha y_* - A}. \tag{C.3}$$

Since this bound tends to 0 as $n \to \infty$ we see that $\frac{my}{j}$ converges uniformly to 1.

Finally we will show that the remaining factor of Q,

$$R = \left(\frac{my}{j}\right)^j \exp\left(mu + \frac{m}{2y}u^2\right),$$

converges uniformly to 1. Using a Taylor expansion about zero, we have

$$\log(1+x) = x - \frac{1}{2}x^2\rho_1 \tag{C.4}$$

where $\rho_1 = 1 + \epsilon_1 x$, $\epsilon_1 = -\frac{2}{3}(1+\theta_1 x)^{-3}$. Here $0 < \theta_1 < 1$, θ_1 depends on x. Note that ϵ_1 is bounded for x near 0. Using equation (C.4)

$$
\begin{aligned}
-\log(R) &= j\log\left(\frac{j}{my}\right) - mu - \frac{m}{2y}u^2 \\
&= j\log\left(1+\frac{u}{y}\right) - mu - \frac{m}{2y}u^2 \\
&= j\left(\frac{u}{y} - \frac{u^2}{2y^2}\rho_1\right) - mu - \frac{m}{2y}u^2 \\
&= \left(\frac{j}{y} - m\right)u - \left(\frac{j\rho_1}{2y^2} + \frac{m}{2y}\right)u^2 \\
&= \frac{m}{y}u^2 - \left(\frac{m}{2y}\left(1+\frac{u}{y}\right)\left(1+\epsilon_1\frac{u}{y}\right) + \frac{m}{2y}\right)u^2 \\
&= \left(\frac{m}{y} - \frac{m}{2y} - \frac{m\epsilon_1}{2y^2}u - \frac{m}{2y^2}u - \frac{m\epsilon_1}{2y^3}u^2 - \frac{m}{2y}\right)u^2 \\
&= -\left(\frac{\epsilon_1}{2y} + \frac{1}{2y^2} + \frac{\epsilon_1}{2y^3}u\right)mu^3.
\end{aligned}
$$

The first factor of the above product is bounded for large n, and y and j such that $\left|\frac{j}{m} - y\right| \leq Am^{-\alpha}$ and $y \in W$. Let B be the bound so that

$$
|\log(R)| \leq Bm|u|^3 \leq BA^3m^{1-3\alpha},
$$

which converges to zero since $\alpha > \frac{1}{3}$. Thus R converges uniformly to 1 and this shows that Q converges uniformly to 1. □

Appendix D

Uniform Approximation of a Binomial Distribution

Here we prove a theorem dealing with a uniform approximation of the binomial distribution. This result is useful in establishing the results for Beta smoothed histogram density estimators (see Section 4.4), in particular, it is useful in establishing regularity in height. It also does not rest on any the work done in the rest of this document.

Theorem D.1 (Uniform Approximation for the Binomial Distribution)**.** *Let m be in $\mathbf{N}^d$ and j be in $\mathbf{Z}^d$ with $m_k \geq j_k \geq 0$ for $k = 1, \ldots, d$ and suppose that m depends on n where n belongs to $\mathbf{N}$. Let the functions P_{nj} and F_{nj} be given by*

$$P_{nj}(y) = \prod_{k=1}^{d} \binom{m_k}{j_k} y_k^{j_k} (1 - y_k)^{m_k - j_k},$$

$$F_{nj}(y) = \frac{1}{(2\pi)^{\frac{d}{2}} \left(\prod_{k=1}^{d} m_k y_k (1 - y_k) \right)^{\frac{1}{2}}} \exp\left(- \sum_{k=1}^{d} \frac{m_k}{2y_k(1 - y_k)} \left(\frac{j_k}{m_k} - y_k \right)^2 \right)$$

where $0 < y_k < 1$ *for each* $k = 1, \ldots, d$. *Let* α *be a positive real number such that* $\alpha > \frac{1}{3}$. *For each* $k = 1, \ldots, d$, *let* W_k *be a compact subset of the open interval* $(0, 1)$, *and* A_k *be a positive real number and suppose that* $m_k \to \infty$ *as* $n \to \infty$. *Then for each positive number* ϵ, *there is an integer* N *for any* n *in* $\mathbf{N}$ *such that if* $n > N$,

$$\left| \frac{P_{nj}(y)}{F_{nj}(y)} - 1 \right| < \epsilon$$

for all y_k *in* W_k *and* $j_k = 0, \ldots, m_k$ *such that* $\left| \frac{j_k}{m_k} - y_k \right| \leq A_k m_k^{-\alpha}$ *for each* $k = 1, \ldots, d$.

Remark. Like Theorem C.1 the proof of this result follows that of Theorem 1.5.2 (Laplace's formula of the theory of probability) in [Lor86]. This theorem of Lorentz handles only the one-dimensional case, and deals with point-wise convergence (not uniform convergence). The generality stated here requires a number of changes to the proof.

Proof of Theorem D.1. As in Theorem C.1 it suffices to prove that each factor of a finite product satisfies the above type of uniform convergence to prove that the product also satisfies it. To this end note that $\frac{P_{nj}(y)}{F_{nj}}$ is equal to

$$\frac{\prod_{k=1}^d \binom{m_k}{j_k} y_k^{j_k}(1-y_k)^{m_k-j_k}}{\frac{1}{(2\pi)^{\frac{d}{2}}\left(\prod_{k=1}^d m_k y_k(1-y_k)\right)^{\frac{1}{2}}} \exp\left(-\sum_{k=1}^d \frac{m_k}{2y_k(1-y_k)} \left(\frac{j_k}{m_k} - y_k\right)^2 \right)}$$

$$= \prod_{k=1}^d \binom{m_k}{j_k} y_k^{j_k}(1-y_k)^{m_k-j_k}(2\pi m_k y_k(1-y_k))^{\frac{1}{2}} \exp\left(\frac{m_k}{2y_k(1-y_k)} \left(\frac{j_k}{m_k} - y_k\right)^2 \right).$$

We will consider one factor of the above product. For ease of notation we will drop the subscript and let $k = m - j$, $z = 1 - y$, and $u = \frac{j}{m} - y$. Let Q be the factor, so that

$$Q = \binom{m}{j} y^j z^k (2\pi m y z)^{\frac{1}{2}} \exp\left(\frac{m}{2yz}u^2\right). \tag{D.1}$$

From Sterling's formula, for any n in $\mathbf{N}$,we have that

$$n! = (2\pi n)^{\frac{1}{2}} n^n e^n H_n \tag{D.2}$$

where H_n is some sequence of real numbers such that $H_n \to 1$ as $n \to \infty$. We can use equation (D.2) to rewrite the expression for Q so that

$$Q = \frac{(2\pi m)^{\frac{1}{2}} m^m e^{-m} H_m}{(2\pi j)^{\frac{1}{2}} j^j e^{-j} H_j (2\pi k)^{\frac{1}{2}} k^k e^{-k} H_k} y^j z^k (2\pi m y z)^{\frac{1}{2}} \exp\left(\frac{m}{2yz}u^2\right)$$

$$= \frac{H_m}{H_j H_k} \left(\frac{m^2 y z}{jk}\right)^{\frac{1}{2}} \frac{m^m}{j^j k^k} y^j z^k \exp\left(\frac{m}{2yz}u^2\right)$$

We now need to show that Q is a product of factors each of which satisfies the type of uniform convergence from the statement of the theorem.

First we will show that the factor $\frac{H_m}{H_j H_k}$ converges uniformly to 1. Note that trivially H_m converges uniformly to 1. Let $y_* = \inf_{y \in W} y$. Since W is compact and contained in the interval $(0, 1)$ we see that $y_* > 0$ and so for $\left|\frac{j}{m} - y\right| \le Am^{-\alpha}$, we have

$$j \ge my - Am^{1-\alpha} \ge my_* - Am^{1-\alpha} = m(y_* - Am^{-\alpha}) \tag{D.3}$$

Let ϵ be a positive real number. Choose N_0 in $\mathbf{N}$ such that for all n in $\mathbf{N}$ if $n > N_0$ then $|H_m - 1| < \epsilon$. Since $m(y_* - Am^{-\alpha}) \to \infty$ as $n \to \infty$ we can choose

N_1 in $\mathbf{N}$ such that if $n > N_1$ then $m(y_* - Am^{-\alpha}) > N_0$. Using inequality (D.3) we see that for $n > N_1$, and any y and j such that $\left|\frac{j}{m} - y\right| \leq Am^{-\alpha}$ we have $j \geq m(y_* - Am^{-\alpha}) > N_0$ and so $|H_j - 1| < \epsilon$. This shows that H_j converges uniformly to 1.

Similarly let $z_* = 1 - y^*$ where $y^* = \sup_{y \in W} y$ so that $y^* < 1$ and $z_* > 0$. For $\left|\frac{j}{m} - y\right| \leq Am^{-\alpha}$, we have

$$
\begin{aligned}
k = m - j &= m - (my + Am^{1-\alpha}) = mz - Am^{1-\alpha} \\
&\geq mz_* - Am^{1-\alpha} = m(z_* - Am^{-\alpha})
\end{aligned}
$$

and also $m(z_* - Am^{-\alpha}) \to \infty$ as $n \to \infty$. Arguing as before, we see that H_k converges uniformly to 1.

Now we will show that the factor $\left(\frac{m^2 yz}{jk}\right)^{\frac{1}{2}}$ converges uniformly to 1. For y and j such that $\left|\frac{j}{m} - y\right| \leq Am^{-\alpha}$ we have that

$$
\left|\frac{y}{j/m} - 1\right| \leq \frac{Am^{1-\alpha}}{j} \leq \frac{Am^{1-\alpha}}{my_* - Am^{1-\alpha}} = \frac{A}{m^\alpha y_* - A}
$$

and that

$$
\begin{aligned}
\left|\frac{1-y}{1-j/m} - 1\right| &= \frac{|j/m - y|}{1 - j/m} \leq \frac{Am^{-\alpha}}{1 - j/m} \\
&= \frac{Am^{1-\alpha}}{k} \leq \frac{Am^{1-\alpha}}{mz_* - Am^{1-\alpha}} = \frac{A}{m^\alpha z_* - A}
\end{aligned}
$$

Both bounds tend to 0 as $n \to \infty$ so both $\frac{y}{j/m}$ and $\frac{1-y}{1-j/m}$ converge uniformly to 1 and since

$$\left(\frac{m^2yz}{jk}\right)^{\frac{1}{2}} = \left(\frac{y}{j/m}\frac{z}{k/m}\right)^{\frac{1}{2}} = \left(\frac{y}{j/m}\frac{1-y}{1-j/m}\right)^{\frac{1}{2}}$$

the factor converges uniformly to 1 as well.

Finally we will show that the remaining factor of Q,

$$R = \frac{m^m}{j^j k^k} y^j z^k \exp\left(\frac{m}{2yz}u^2\right) = \left(\frac{my}{j}\right)^j \left(\frac{mz}{k}\right)^k \exp\left(\frac{m}{2yz}u^2\right),$$

converges uniformly to 1. Note that, using Taylor's expansion about zero, we have

$$\log(1+x) = x - \frac{1}{2}x^2\rho_1 \text{ and } \log(1-y) = -y - \frac{1}{2}y^2\rho_2 \tag{D.4}$$

where $\rho_1 = 1 + \epsilon_1 x$, $\epsilon_1 = -\frac{2}{3}(1+\theta_1 x)^{-3}$, $\rho_2 = 1 + \epsilon_2 y$, and $\epsilon_2 = \frac{2}{3}(1-\theta_2 y)^{-3}$. Here $0 < \theta_i < 1$ for $i = 1, 2$, θ_1 depends on x and θ_2 depends on y. Note that ϵ_1 is bounded for x near 0, and ϵ_2 is bounded for y near 0. Using the equations in statement (D.4), we have

$$
\begin{aligned}
-\log(R) &= j\log\left(\frac{j}{my}\right) + k\log\left(\frac{k}{mz}\right) - \frac{m}{2yz}u^2 \\
&= j\log\left(1+\frac{u}{y}\right) + k\log\left(1-\frac{u}{z}\right) - \frac{m}{2yz}u^2 \\
&= j\left(\frac{u}{y} - \frac{u^2}{2y^2}\rho_1\right) + k\left(-\frac{u}{z} - \frac{u^2}{2z^2}\rho_2\right) - \frac{m}{2yz}u^2 \\
&= \left(\frac{j}{y} - \frac{k}{z}\right)u - \left(\frac{j\rho_1}{2y^2} + \frac{k\rho_2}{2z^2}\right)u^2 - \frac{m}{2yz}u^2 \\
&= \left(m\left(1+\frac{u}{y}\right) - m\left(1-\frac{u}{z}\right)\right)u \\
&\quad - \left(\frac{m}{2y}\left(1+\frac{u}{y}\right)\left(1+\epsilon_1\frac{u}{y}\right) + \frac{m}{2z}\left(1-\frac{u}{z}\right)\left(1+\epsilon_2\frac{u}{z}\right)\right)u^2 - \frac{m}{2yz}u^2 \\
&= \frac{m}{yz}u^2 - \left(\frac{m}{2y} + \frac{m\epsilon_1}{2y^2}u + \frac{m}{2y^2}u + \frac{m\epsilon_1}{2y^3}u^2\right. \\
&\quad \left. + \frac{m}{2z} + \frac{m\epsilon_2}{2z^2}u - \frac{m}{2z^2}u - \frac{m\epsilon_2}{2z^3}u^2\right)u^2 - \frac{m}{2yz}u^2 \\
&= -\left(\frac{\epsilon_1}{2y^2} + \frac{1}{2y^2} + \frac{\epsilon_1}{2y^3}u + \frac{\epsilon_2}{2z^2} - \frac{1}{2z^2} - \frac{\epsilon_2}{2z^3}u\right)mu^3.
\end{aligned}
$$

The first factor of the above product is bounded for large n and y and j such that $\left|\frac{j}{m} - y\right| \le Am^{-\alpha}$ and $y \in W$. Let B be the bound so that

$$
|\log(R)| \le Bm|u|^3 \le BA^3m^{1-3\alpha},
$$

which converges to zero since $\alpha > \frac{1}{3}$. Thus R converges uniformly to 1 and this shows that Q converges uniformly to 1. □

Appendix E

A Characterization of a Limit of a Ratio of Volumes

The following technical lemma allows us to restate a limit that involves a ratio of volumes in terms of a limit that involves only minimums and maximums of certain parameters. It is used to simplify regularity conditions that show up when dealing with particular density estimators.

Lemma E.1 (A Characterization of a Limit of a Ratio of Volumes)**.** *For each $k = 1, \ldots, d$ let h_k be a sequence of positive numbers that depends on n, which belongs to $\mathbf{N}$, and A_k a positive constant. Then*

$$\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0 \textit{ if and only if } \underline{\lim}_n \frac{\prod_{k=1}^d A_k^2 h_k^2}{(\sum_{k=1}^d A_k^2 h_k^2)^d} > 0$$

Proof of Lemma E.1. First, suppose that

$$\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0.$$

Then

$$\lim_n \frac{\prod_{k=1}^d A_k^2 h_k^2}{(\sum_{k=1}^d A_k^2 h_k^2)^d} \geq \frac{\prod_{k=1}^d A_k^2}{(\sum_{k=1}^d A_k^2)^d} \left(\lim_n \frac{\min_k h_k}{\max_k h_k} \right)^{2d} > 0.$$

Conversely, suppose that

$$\lim_n \frac{\prod_{k=1}^d A_k^2 h_k^2}{(\sum_{k=1}^d A_k^2 h_k^2)^d} > 0$$

so that, for each $k = 1, \ldots, d$, we have

$$\lim_n \frac{A_k^2 h_k^2}{(\sum_{k=1}^d A_k^2 h_k^2)^d} > 0.$$

Thus for each $k = 1, \ldots, d$ there is a positive real number B_k and an N_k in $\mathbf{N}$ such that for any n in $\mathbf{N}$ if $n > N_k$,

$$A_k^2 h_k^2 > B_k \sum_{k=1}^d A_k^2 h_k^2. \tag{E.1}$$

Now let the number N be given by the equation $N = \max_k N_k$. Suppose that $n \in \mathbf{N}$ and $n > N$. Using inequality (E.1) this implies that, for each $k = 1, \ldots, d$, we have

$$(\max_k A_k)^2 h_k^2 \geq A_k^2 h_k^2 > B_k \sum_{k=1}^d A_k^2 h_k^2 \geq (\min_k B_k) \sum_{k=1}^d A_k^2 h_k^2$$

so that

$$(\max_k A_k)^2 (\min_k h_k)^2 > (\min_k B_k) \sum_{k=1}^d A_k^2 h_k^2 > (\min_k B_k) A_k^2 h_k^2 \geq (\min_k B_k)(\min_k A_k)^2 h_k^2$$

and so

$$(\max_k A_k)^2(\min_k h_k)^2 > (\min_k B_k)(\min_k A_k)^2(\max_k h_k)^2.$$

It follows that

$$\frac{\min_k h_k}{\max_k h_k} > \frac{(\min_k B_k)^{\frac{1}{2}} \min_k A_k}{\max_k A_k} > 0.$$

This shows that

$$\underline{\lim}_n \frac{\min_k h_k}{\max_k h_k} > 0.$$

and the conclusion follows. □

Bibliography

[AJ76] Saab Abou-Jaoude, *Conditions nécessaires et suffisantes de convergence l1 en probabilité de l'histogramme pour une densité*, Annales de l'I.H.P. **12** (1976), 213–231. (Cited on page 36.)

[BE00] L. J. Bain and M. Engelhardt, *Introduction to probability and mathematical statistics*, Duxbury Press, Pacific Grove, CA, 2000. (Cited on page 58.)

[BS05] Taoufik Bouezmarni and Olivier Scaillet, *Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data*, Econometric Theory **21** (2005), 390–412. (Cited on page 46.)

[Dev85] Luc Devroye, *Nonparametric density estimation: The L_1 view*, Wiley, New York, NY, 1985. (Cited on pages 11, 26, 31 and 58.)

[GJB02] Yogendra P. Chaubey G. Jogesh Babu, Angelo J. Canty, *Application of bernstein polynomials for smooth estimation of a distribution and density function*, Journal of Statistical Planning and Inference **105** (2002), 377–392. (Cited on page 57.)

[Lor86] G. G. Lorentz, *Bernstein polynomials*, Chelsea Publishing Company, New York, NY, 1986. (Cited on pages 66 and 71.)

[Rud87] Walter Rudin, *Real and complex analysis*, WCB/McGraw-Hill, Boston, MA, 1987. (Cited on pages 58 and 61.)

[Sta83] U. Stadtmüller, *Asymptotic distributions of smoothed histograms*, Metrika **30** (1983), 145–158. (Cited on page 46.)

[Ten95] A. Tenbusch, *Nonparametric curve estimation with bernstein estimates*, Ph.D. thesis, University of Osnabrück, 1995. (Cited on page 57.)

[Vit75] Richard A. Vitale, *A bernstein polynomial approach to density estimation*, Statistical Processes and Related Topics **2** (1975), 87–100. (Cited on page 57.)