I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Professor Peter Hall
(Principal Adviser)

Dr Andrew Wood

Dr Alan Welsh

Approved for the University Committee on Graduate Studies:

Dean of Graduate Studies & Research
Declaration

I hereby declare that this thesis describes my own original work, supervised by Professor P.G. Hall and published jointly with him.

R.D. Murison

R.D. Murison
I wish to express appreciation and thanks to my supervisor, Professor Peter Hall, for guiding me through the research for this thesis and expanding my statistical horizons.

My research was made possible by the generous financial assistance from Wool Research and Development Corporation and from New South Wales Agriculture. I am very appreciative that both organisations see merit in statistics research and have confidence in me. It will be a pleasure to reciprocate their support in the future.

Dr Andrew Wood was always available and patient to discuss my research and assist me over the hurdles and to him I owe thanks. I also wish a special thanks to my friend and fellow student, Song Xi Chen, for his companionship and moral support as we pursued our research interests.

The final and deepest appreciation is reserved for my family, Mary, Greg and Julie, who relinquished a lot so that I could come to Canberra and were steadfast in their love, support and understanding.
Associated Paper

Abstract

The aim of this thesis is to provide two extensions to the theory of nonparametric kernel density estimation that increase the scope of the technique.

The basic ideas of kernel density estimation are not new, having been proposed by Rosenblatt [20] and extended by Parzen [17]. The objective is that for a given set of data, estimates of functions of the distribution of the data such as probability densities are derived without recourse to rigid parametric assumptions and allow the data themselves to be more expressive in the statistical outcome. Thus kernel estimation has captured the imagination of statisticians searching for more flexibility and eager to utilise the computing revolution. The abundance of data and computing power have revealed distributional shapes that are difficult to model by traditional parametric approaches and in this era, the computer intensive technique of kernel estimation can be performed routinely. Also we are aware that computing power can be harnessed to give improved statistical analyses. Thus a lot of modern statistical research involves kernel estimation from complex data sets and our research is concordant with that momentum.

The thesis contains three chapters. In Chapter 1 we provide an introduction to
kernel density estimation and we give an outline to our two research topics.

Our first extension to the theory is given in Chapter 2 where we investigate density estimation from independent data, using high order kernel functions. These kernel functions are designed for bias reduction but they have the penalty of yielding negative density estimates where data are sparse. In common practice, the negative estimates would arise in the tails of the density and we provide four ways of correcting this negativity to give bona fide estimates of the probability density. Our theory shows that the effects of these corrections are asymptotically negligible and thus opens the way for the regular use of bias reducing, high order kernel functions.

We also consider density estimation of continuous stationary stochastic processes and this is the content of Chapter 3. With this problem, the dependent nature of the data influences the accuracy of the kernel density estimator and we provide theory regarding the convergence of the kernel estimators of the density and its derivatives to the true functions. An important result from this study is that nonparametric density estimators from dependent processes can have the same rates of convergence as their parametric counterparts yet retain the flexibility of being independent of parametric assumptions. Our other results indicate that the convergence rate of the density estimator can be quite slow if there are large lag dependencies amongst the data and suggests that large samples would be required for reliable inference about such data.

The flexibility of kernel density estimation for continuous and discrete data, independent and dependent observations, means that it is a useful statistical tool. The
techniques given in this thesis are not restricted to the analysis of simple sets of data but may be employed in the construction of statistical models for complex data with a high degree of structure.
Contents

Declaration iv

Acknowledgements v

Abstract vi

1 Introduction 1

1.1 The Setting for Kernel Density Estimation 1

1.2 Outline of Thesis 2

1.3 Basic Concepts of Kernel Density Estimation 2

2 Correcting the Negativity of High-Order Kernel Density Estimators 6

2.1 Introduction 6

2.2 Positive Estimators Derived from \( \hat{f} \) 8

2.3 Main Results 12

2.3.1 Integrated Squared Error Properties of the Main Estimators 13

2.3.2 Densities with Regularly Varying Tails 17

2.3.3 Compactly Supported Densities 22
2.4 Simulation Study ...................................................... 25
  2.4.1 Simulation Details .............................................. 26
  2.4.2 Results ............................................................ 28
2.5 Proofs of Results and Theorems .............................. 41
  2.5.1 Proof of $\int \hat{f}^2 = \int f^2 + o_p(1)$ .......... 41
  2.5.2 Proof of $\int \hat{f}^2 I(\hat{f} > 0) = \int f^2 + o_p(1)$ 41
  2.5.3 Proof of $\int (\hat{f} - f)fI(\hat{f} < 0) \leq o_p\{(nh)^{-1} + h^2r\}$ 42
  2.5.4 Proof of $\int (\hat{f} - E\hat{f})f = o_p\{(nh)^{-1/2}\}$ 43
  2.5.5 Proof of $\int (E\hat{f} - f) f = -\frac{1}{2} A_k h^r + o(h^r)$ 44
  2.5.6 Proof of Theorem 2.1 ......................................... 45
  2.5.7 Proof of Theorem 2.2 ......................................... 56
  2.5.8 Proof of Theorem 2.3 ......................................... 61
  2.5.9 Proof of Theorem 2.4 ......................................... 62
3 Estimation of the Density and Mode of a Continuous-Time Stationary Stochastic Process 66
  3.1 Introduction .......................................................... 66
  3.2 Previous Developments in Density and Mode Estimation for Stationary Processes .................................. 70
    3.2.1 Density Estimation ......................................... 70
    3.2.2 Mode Estimation .............................................. 77
3.3 Estimation of the Marginal Density and its Derivatives of a Continuous-
Time Process ................................................................. 78

3.3.1 The Covariance Function of a Gaussian Process .......... 79

3.3.2 Kernel Estimators for Continuous-Time Data .............. 81

3.3.3 The Bias of $\hat{f}^{(j)}$ .................................................. 82

3.3.4 The Variance of $\hat{f}^{(j)}$ .............................................. 83

3.3.5 Asymptotic Variance of $\hat{f}^{(j)}$ for Gaussian Processes .... 84

3.4 Simulation Study of Kernel Density Estimators of a Gaussian Process 95

3.4.1 Generating the Data for a Gaussian Process ................. 95

3.4.2 Choosing the Bandwidth .............................................. 98

3.4.3 Numerical Study of $L^2$ Convergence of Kernel Estimators . 106

3.5 The Surface Roughness Data ........................................... 113

3.6 Proofs ............................................................................. 118

3.6.1 Proof of Theorem 3.1. .................................................. 118

3.6.2 Proof of Theorem 3.2. .................................................. 134

References ........................................................................ 136
List of Tables

2.1 Asymptotically optimal kernel functions for \( r = 2 \) and \( r = 4 \) .... 7

2.2 Values of relative difference \( (R_i) \) of \( \hat{f}_i \) for simulated Normal data. . . . 33

2.3 Values of relative difference \( (R_i) \) of \( \hat{f}_i \) for simulated Cauchy data. . . 34

2.4 Values of relative difference \( (R_i) \) of \( \hat{f}_i \) for simulated Students’t (2 df.)
data. ......................................................... 35

2.5 Values of relative difference \( (R_i) \) of \( \hat{f}_i \) for simulated Students’t (3 df.)
data. ......................................................... 36

2.6 Values of relative difference \( (R_i) \) of \( \hat{f}_i \) for simulated Students’t (5 df.)
data. ......................................................... 37

2.7 Values of relative difference \( (R_i) \) of \( \hat{f}_i \) for simulated Students’t (8 df.)
data. ......................................................... 38

2.8 Values of relative difference \( (R_i) \) of \( \hat{f}_i \) for simulated Gamma data. . . 39

2.9 Values of relative difference \( (R_i) \) of \( \hat{f}_i \) for simulated Beta data. . . . 40

3.1 Asymptotic variance components of \( \hat{f}^{(j)} \) due to short term dependence. 87

3.2 Asymptotic variance components of \( \hat{f}^{(j)} \) due to long term dependence. 87

3.3 Frequencies of \( h \) values that are optimal for simulated data. ....... 100
3.4 Least squares fits of the linear relationship between log(T) and log(\(\hat{h}_{\text{min}}\)) for different proportions of short and long term dependence. .......... 103

3.5 Least squares fits of the linear relationship between log(T) and log(\(\hat{h}_{\text{min}}\)) for different proportions of short and long term dependence. .......... 104

3.6 Comparisons of bandwidths that minimise the ISE and the bandwidths given by the rules for T = 400 and \(\sigma = 1\) ......... 105

3.7 Squared errors for the sample mode of Gaussian processes. ........ 108

3.8 Asymptotic and sample integrated variances (\(\times 10^4\)) of kernel estimators for Gaussian processes of length T=400. ............... 111

3.9 Asymptotic and sample integrated squared errors (\(\times 10^4\)) of kernel estimators for Gaussian processes of length T=400. ............... 112
List of Figures

2.1 Asymptotically optimal kernel functions for $r = 2$ and $r = 4$. 8
2.2 The high-order kernel density estimators $\hat{f}$ and $\hat{f}_2$. 9
2.3 The estimators $\hat{f}_3$ (with truncated tails) and $\hat{f}_4$ (with tapered tails). 11
2.4 Close up views of the spline taper applied to $\hat{f}_3$ to get $\hat{f}_4$. 11
2.5 Values of relative difference $(R_t)$ of $\hat{f}_i$ for sample sizes 50,100,200,400,1000 of Normal $(0,1)$, Cauchy $(0,1)$ and Student’s $t$ $(2 \text{ df})$ data. 30
2.6 Values of relative difference $(R_t)$ of $\hat{f}_i$ for sample sizes 50,100,200,400,1000 of Student’s $t$ $(3 \text{ df})$, Student’s $t$ $(5 \text{ df})$ and Student’s $t$ $(8 \text{ df})$ data. 31
2.7 Values of relative difference $(R_t)$ of $\hat{f}_i$ for sample sizes 50,100,200,400,1000 of Gamma$(5,1)$ and Beta$(5,9)$ data. 32
2.8 Ranges $R_1(\epsilon)$ and $R_2(\epsilon)$ of the right tail of $\hat{f}$ as described in Theorem 2.1. 46
3.1 Line transects of height recordings (Studentised) for new, treated and worn surfaces. 67
3.2 Short term dependencies and simulated Gaussian processes. 80
3.3 Examples of $\hat{f}(x)$ and $\hat{f}'(x)$ for simulated continuous-time Gaussian data ......................................................... 82
3.4 Log-linear relationship between optimal bandwidth and sample size . 100
3.5 Profiles of sample variances of $\hat{f}$ and $\hat{f}'$. ......................... 109
3.6 Sample and asymptotic integrated variances of $\hat{f}$ and $\hat{f}$ for samples sizes $T = 100, 200, 300, 400$. ........................................... 110
3.7 Estimates of density, derivatives and mode for a new surface ........... 115
3.8 Estimates of density, derivatives and mode for a treated surface ....... 116
3.9 Estimates of density, derivatives and mode for a worn surface ........ 117
Chapter 1

Introduction
1.1 The Setting for Kernel Density Estimation

Nonparametric density estimation is established as an important technique of statistical analysis and whilst it is an effective tool for description of a data set, it is often used as a component of more complex statistical models. The strength of the technique is that it does not require rigid assumptions about the data and is ideally suited to utilise computing power to extract detailed information from the data.

Rosenblatt [20] established the notion of kernel estimation of a univariate probability density, given a set of independent data. Parzen [17] extended these results by showing the consistency of the estimators, deriving their asymptotic distributions and showing how the kernel estimator could play a role in estimating other statistics; the mode in this case. The way was then clear for other researchers to find applications in diverse circumstances, necessitating research concerning consistency and convergence of the estimators in their diverse settings. A detailed account of theoretical developments is given by Prakasa Rao [18] and practical aspects have been given thorough coverage by Silverman [23]. The pace of research into the topic has not slackened and there has been substantial development of density estimation with dependent data such as time series and failure time data. Györfi, Härdle, Sarda and Vieu [12] have provided a valuable treatment of nonparametric curve estimation with dependent data. However, the subject area is extensive and numerous accounts are evident in the bibliographies of the above references.
1.2 Outline of Thesis

We provide statistical theory regarding the correction of negative density estimates that arise from high-order kernels (Chapter 2) and promote the use of kernel density estimators for continuous stationary stochastic processes (Chapter 3).

High-order kernels are well known and practitioners often apply ad hoc procedures to correct the negativity. Our theory will provide validity for four useful correction methods. We envisage that this will clear the way for the use of high-order kernels in statistical modelling such as the smoothed bootstrap and generalised additive models.

Our second topic regarding continuous processes is geared for data description at this stage. With large sets of data, we are in a position to extract more information than is achieved by traditional parametric models that use only the first two moments to summarise the data.

We shall give the basic concepts of kernel density estimation which establish our notation and serve as a reference point for the two topics we consider.

1.3 Basic Concepts of Kernel Density Estimation

The theory we present here is without rigour and is covered in detail by Silverman [23, Chapter 3].

For a sample of independent and identically distributed (i.i.d.) data $X_1, \ldots, X_n$. 
kernel function $K$ and bandwidth $h$, the kernel density estimator is

$$
\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} K \{ h^{-1}(x - X_i) \}. 
$$

(1.1)

The kernel function, $K(\cdot)$, is smooth, continuous, $\int_{-\infty}^{+\infty} K(t)dt = 1$ (normalization), $\sup_x |K(x)| < \infty$ (bounded) and $\lim_{|x| \to \infty} K(x) = 0$. The last condition is easily attained by choosing kernels with compact support. The density estimator, $\hat{f}$, will have the same continuity and differentiability properties as $K$.

The action of $K$ is to smooth the sample probability mass $(n^{-1})$ associated with each $X_i$ into the neighbourhood within radius $h$ of $X_i$. The choice of $h$ is of fundamental importance. If it is too large, the kernel function will oversmooth, obscure the detail supplied the data and induce excess bias. Conversely, $h$ being too small will undersmooth and the resulting estimator will give too much emphasis to the sample itself and have high variance.

The common approach to choosing a globally optimal bandwidth is to minimise the squared error risk function, integrated over the range of the density. This statistic is known as the mean integrated squared error (MISE) and by the familiar statistical association of bias and variance we have that

$$
\text{MISE} = E \int (\hat{f} - f)^2 = \int \text{var} \hat{f} + \int (\text{bias} \hat{f})^2.
$$

There is a trade-off between bias and variance as one can be reduced at the expense of increasing the other. The optimal bandwidth is one that gives matched orders of magnitude of the integrated variance and the integrated squared bias. The calculations for an exact MISE may be complex but a simpler approximate expression is
sufficient. We shall return to this issue shortly but first we need to consider the effect of the kernel function \( K \).

We shall define \( K \) to be and \( r \)th order kernel function if

\[
\int t^l K(t) dt = \begin{cases} 
1 & \text{if } l = 0 \\
0 & 0 < l < r \\
k_r \neq 0 & l = r
\end{cases}
\]  
(1.2)

The most straightforward case is when \( r = 2 \) but we shall retain the general form since in Chapter 2 we consider those kernel functions where \( r > 2 \).

We assume that the unknown density has continuous derivatives of at least order \( r \) and that \( h \to 0, nh \to \infty \) as \( n \to \infty \). Then straightforward calculations lead to expressions for the asymptotic bias and variance of \( \hat{f} \);

\[
\text{bias}\{\hat{f}\} = (-1)^r h^r f^{(r)} k_r / r! + o(h^r),
\]
(1.3)

\[
\text{var}\{\hat{f}\} = (nh)^{-1} f \int K(t)^2 dt + O(n^{-1}).
\]
(1.4)

Therefore for an \( r \)th order kernel we can write

\[
\text{MISE}\{\hat{f}(x)\} \approx (nh)^{-1} \int K(t)^2 dt + h^{2r} (k_r / r!)^2 \int \{f^{(r)}(x)\}^2 dx.
\]
(1.5)

By differentiating (1.5) with respect to \( h \), or by choosing \( h \) so that variance and squared bias are of the same magnitude, we get that

\[
h_{opt} \approx \left(\frac{\sqrt{2r}}{k_r / r!}\right)^{-2} \left\{\int f^{(r)} \right\}^{-1} \int K(t)^2 dt \right\}^{1/(2r+1)} n^{-1/(2r+1)}.
\]
(1.6)

Recall that \( f^{(r)} \) is unknown. An approximation recommended by Silverman [23. p48.f] is to substitute for \( f^{(r)} \) the \( r \)'th derivative of the normal density, standardized
to have the same scale and location as the data. Another proposal is to use least-squares cross validation for estimating $h_{opt}$. This process is attractive since it can be automated and it makes only weak assumptions about $f$ but the resampling involved can introduce extra noise and the chosen bandwidth may be too wide. Silverman [23] explains bandwidth choice by cross-validation and gives references to research on that topic.

Substitution of the result from (1.6) back into the approximate formula for MISE (1.5) gives that

$$\text{MISE} \approx \frac{2r + 1}{2r} C(K) \left[ \int \left\{ f^{(r)}(x) \right\}^2 dx \right]^{1/(2r+1)} n^{-2r/(2r+1)}$$

where

$$C(K) = \left\{ \sqrt{2\pi} \ k_r/r! \int K(t)^2 dt \right\}^{2r/(2r+1)}$$

$$= O \left\{ n^{-2r/(2r+1)} \right\} .$$

The role of the kernel function in deriving the estimator $\hat{f}$ is the subject matter of Chapter 2. The theory given in Chapter 1 is for the simple case of discrete i.i.d. data which makes the expression for variance of $\hat{f}$ quite simple. Dependencies amongst data are common and we often wish to exploit those dependence structures, at the same time requiring the flexibility that kernel estimators afford. In Chapter 3 we give theory for the variance of $\hat{f}$ and its derivatives when the data are continuous and correlated.
Chapter 2

Correcting the Negativity of High-Order Kernel Density Estimators
2.1 Introduction

The bias in kernel density estimation arises as we smooth the empirical probability from the sample points into the neighbourhoods surrounding them. The degree of smoothing, and hence bias, is determined by the bandwidth and by the nature of the kernel function. We can see from equation (1.3) that for an $r$th order kernel, the asymptotic bias has factors of $h^r/r!$ and $k_r = \int t^r K(t)dt$. In the Taylor series expansion for bias (see Silverman \[23, \text{pp39}\]), terms of $h^l/l! \int t^l K(t)dt$, $0 < l < r$, are zero due to properties (1.2) of $K(t)$ and the bias is $O(h^r)$. Since $h \to 0$ as $n \to \infty$, larger values of $r$ will give smaller orders of magnitude of bias.

High order kernels ($r > 2$) were proposed by Parzen \[17\] and Bartlett \[2\] as means to reduce the bias of density estimators. If $r = 2$, $K$ is itself a density function since it integrates to 1 and does not have negative values. For $r > 2$, $K$ must take some negative values in order that it integrates to 1. This leads to negative density estimates in places where the data are sparse, even though in other places the density estimates profit from the reduced bias.

This chapter deals with correcting that negativity so that high order kernels can be used to give bona fide density estimates. We propose four correction methods which retain the reduced bias for the non-negative estimates, and present theory to show that the corrections have negligible effects on the asymptotic squared error properties of the modified estimators for a wide range of distributions that are encountered in practice. This is confirmed in a simulation study.
Unless \( n \) is very large, it would seem that only a modest gain may be achieved by choosing a kernel of order greater than 2, given the negative values of \( \hat{f} \) that accompany it. For instance using equation (1.8), with \( n = 100 \) and \( r = 2 \), the MISE is \( O(0.025) \) and for \( r = 4 \), it is \( O(0.017) \). The advantages of high order kernels are found in higher level statistical procedures that use density estimators. Hall, DiCiccio and Romano [13] proved that the estimate of variance of a quantile, derived from bootstrap samples that are smoothed by a \( r \)'th order kernel, have a precise order of error equal to \( n^{-r/(2r+1)} \). However, bootstrap sampling demands that the density estimate be positive. If the gains of a high order kernel are to be realised, it is necessary to correct the negative values in the tails of \( \hat{f} \) whilst retaining the order of bias for the interior points.

The optimal kernel to minimise the asymptotically optimal MISE is found by minimising \( \int K(t)^2 \, dt \) subject to the constraints of equation (1.2). A generating function for optimal kernels has been given by Gasser, Müller and Mammitzsch [11] and they list several optimal kernels such as those in Table 2.1.

Table 2.1: Asymptotically optimal kernel functions for \( r = 2 \) and \( r = 4 \).

| \( r = 2 \) | \( K(t) = \frac{3}{4}(-t^2 + 1) \) | \( |t| \leq 1 \) |
| \( r = 4 \) | \( K(t) = \frac{15}{32}(7t^4 - 10t^2 + 3) \) | \( |t| \leq 1 \) |

Plots of these kernels in Figure 2.1 illustrate how bias is reduced with a high order kernel. With a fourth order kernel function, the density estimate at some data point \( X_i \) comes predominantly from evaluating the kernel function at points very close to
The second order kernel function smooths by using more information from the edges of the smoothing window so that distant points have more influence on the density estimate than is the case with a fourth order kernel function.

Figure 2.1: Asymptotically optimal kernel functions for \( r = 2 \) and \( r = 4 \).

\[
K(t) = \frac{3}{4}(-t^2 + 1), |t| \leq 1 \\
K(t) = \frac{15}{32}(7t^4 - 10t^2 + 3), |t| \leq 1
\]

\[ K(t) \]

\[ K(t) \]

\( t \)

\( t \)

\( -1.5 \)

\( 0 \)

\( 1.5 \)

\( 0 \)

\( 0.5 \)

\( 1.0 \)

\( -1.5 \)

\( 0 \)

\( 1.5 \)

\( 0 \)

\( 0.5 \)

\( 1.0 \)

\( -1.5 \)

\( 0 \)

\( 1.5 \)

\( 0 \)

\( 0.5 \)

\( 1.0 \)

2.2 Positive Estimators Derived from \( \hat{f} \)

Our general procedure is to use a density estimator, \( \hat{f} \), derived with a high order kernel and modify it so that it is entirely non-negative.

The first estimator is constructed by taking only the positive part of \( \hat{f} \) so that,

\[
\hat{f}_1 = \gamma_1 \hat{f} I(\hat{f} > 0),
\]

where \( I(\cdot) \) is the indicator function and \( \gamma_1 \) is a constant (a function of the data) that normalises \( \hat{f}_1 \), ie \( \int \hat{f}_1 = 1 \). Secondly we take the absolute value of \( \hat{f} \) to obtain.

\[
\hat{f}_2 = \gamma_2 |\hat{f}|, \text{ with } \gamma_2 \text{ chosen to normalise } \hat{f}_2.
\]
Where it is known that the negative regions of a higher order kernel density estimator occur only in the extreme tails, such as with many unimodal densities, another nonnegative estimator can be constructed. This estimator is obtained by truncating the estimator $\hat{f}$ outside the "central" range where it is nonnegative, and then renormalising. To do this, define $Y_1$ as the largest value in the lower tail and $Y_2$ as the smallest value in the upper tail, such that $\hat{f}(x) = 0$, and to take

$$\hat{f}_3(x) = \gamma_3 \hat{f}(x)I(Y_1 < x < Y_2), \quad \text{with } \gamma_3 = \left\{ \int \hat{f}(x)I(Y_1 < x < Y_2) \right\}^{-1}. \quad (2.3)$$

An example of $\hat{f}_2$ compared with $\hat{f}$ is shown in Figure 2.2 where $\hat{f}_2$ and $\hat{f}$ have been derived for a sample of size 50 simulated from a Normal distribution. From this figure, we can envisage the shape of $\hat{f}_1$ and $\hat{f}_3$.

Figure 2.2: The high-order kernel density estimators $\hat{f}$ and $\hat{f}_2$.

A fourth density estimator can be derived from $\hat{f}$ for densities with unbounded
support by tapering the tails within the ranges \([c_2 Y_i, c_1 Y_1]\) and \([c_1 Y_2, c_2 Y_2]\) where \(c_1, c_2\) are fixed constants satisfying \(0 < c_1 < 1 < c_2 < \infty\). The estimates between \(c_1 Y_1\) and \(c_1 Y_2\) are initially the same as for \(\hat{f}\) but are later modified by renormalising. Appropriate values for these constants would be \(c_2 = 2\) and \(c_1 = 0.5\). Monotone splines based on twelve knots are used to smooth the tails down to zero at \(c_2 Y_i\), \(i = 1, 2\). The procedure for the left tail is to break the interval \([c_2 Y_1, Y_1]\) into six equal, non-overlapping subintervals \([t_j, t_{j+1}], j = 1, \cdots, 6\) and another five similar subintervals for \([Y_1, c_1 Y_1]\). That is,

\[
[c_2 Y_1, Y_1] = \bigcup_{j=1}^{6} [t_j, t_{j+1}] \quad \text{and} \quad [Y_1, c_1 Y_1] = \bigcup_{j=7}^{12} [t_j, t_{j+1}]
\]

with \(t_1 = c_2 Y_1\), \(t_7 = Y_1\), \(t_{12} = c_1 Y_1\).

The first five knots are the coordinates \([\{t_j, \hat{g}(t_j)\}, j = 8, \cdots, 12]\) where \(\hat{g}(t_j)\) are the same as \(\hat{f}(t_j)\) and are positive since \(\max(|t_j|, j = 8, \cdots, 12) < |Y_1|\). The next six knots are determined in succession by using half the density estimate of the previous knot. ie. \([\{t_j, \frac{1}{2}\hat{g}(t_{j+1})\}, j = 2, \cdots, 7]\) and the final knot is \((t_1, 0)\). A similar procedure is used for the right tail. This estimator, \(\hat{g}\), is renormalised using

\[
\gamma_4 = \left[\int \hat{g}(x)dx\right]^{-1} \quad \text{so that} \quad \hat{f}_4 = \gamma_4 \hat{g}.
\]  \hspace{1cm} (2.4)

An example of \(\hat{f}_4\) is shown in Figures 2.3 and 2.4. This estimator is not usually constructed for densities with compact support.
Figure 2.3: The estimators $\hat{f}_3$ (with truncated tails) and $\hat{f}_4$ (with tapered tails).

The numbered points indicate the spline knots.

Figure 2.4: Close up views of the spline taper applied to $\hat{f}_3$ to get $\hat{f}_4$. 

Left tail

Right tail
We shall show that if the true density has moderately light tails and is unbounded (eg Normal, Gamma, Student's t), the estimators $\hat{f}_1$ and $\hat{f}_2$ have the same asymptotic integrated squared error as $\hat{f}$.

If the tails of the true density $f$ decrease like a power of $|x|^{-1}$, then the condition of finite $(1 + \epsilon)'th order moment, for some $\epsilon > 0$, is both necessary and sufficient for $\hat{f}$ to have the same asymptotic integrated squared error as $\hat{f}_3$ and $\hat{f}_4$. When the underlying distribution is compactly supported (eg Beta), asymptotic equivalence of $\hat{f}$ and ($\hat{f}_i, i = 1, 2, 3$) is available in a wide range of circumstances.

The principal conclusion to be drawn from this is that $\hat{f}_i$ may use the same bandwidth as $\hat{f}$ unless the tails of the density are inordinately large (eg Cauchy). A simulation study in Section 2.4 confirms these statements.

### 2.3 Main Results

Section 2.3.1 outlines integrated squared error properties of the estimators $\hat{f}_1$ and $\hat{f}_2$ and shows how they may be described very simply in terms of a single random variable $\gamma$. In Section 2.3.2 we investigate the cases where the underlying distribution is unbounded and the tails of the density vary regularly. In these cases, we examine how the removal of negative parts of the estimates affects the integrated squared error properties of high-order kernel density estimators. The influence of the correction on compactly supported densities is treated in Section 2.3.3. Proofs of major results and theorems are deferred to Section 2.5.
2.3.1 Integrated Squared Error Properties of the Main Estimators

The principal estimators, based on the positive part of \( \hat{f} \), are

\[
\hat{f}_1 = \gamma_1 \hat{f} I(\hat{f} > 0),
\]

\[
\hat{f}_2 = \gamma_2 |\hat{f}|,
\]

where the positive random variables \( \gamma_1, \gamma_2 \) are chosen to ensure that \( \hat{f}_1 \) and \( \hat{f}_2 \) both integrate to unity. The integral of the negative parts of \( \hat{f} \) is

\[
\gamma = \int |\hat{f}| I(\hat{f} < 0).
\]

In this notation,

\[
\int \hat{f} I(\hat{f} > 0) = \int \hat{f} - \int \hat{f} I(\hat{f} < 0)
= 1 - (-\gamma) = 1 + \gamma
\]

and

\[
\int |\hat{f}| = \int \hat{f} I(\hat{f} > 0) - \int \hat{f} I(\hat{f} < 0)
= 1 + 2\gamma.
\]

Therefore, \( \gamma_i = (1 + i\gamma)^{-1} \) for \( i = 1, 2 \).

Formulae for the ISE of \( \hat{f}_1 \) and \( \hat{f}_2 \) depend principally on properties of the random variable \( \gamma \) which converges to zero as \( n \to \infty \). For densities that have continuous derivatives of at least order \( r \), and assuming that \( h \to 0 \), \( nh \to \infty \) as \( n \to \infty \), we have for \( r \)th order kernel functions that.

\[
\int (\hat{f}_i - f)^2 = (nh)^{-1} A_1 + h^{2r} A_2 + (i\gamma)^2 A_3 + i\gamma h' A_4 + o_p \left\{ (nh)^{-1} + h^{2r} + \gamma^2 \right\},
\]

\[
(2.8)
\]
For each of these density estimators, we wish to compare its ISE with that of the uncorrected estimator \( \hat{f} \). We first substitute for \( \hat{f}_1 \) (using (2.5)) or \( \hat{f}_2 \) (using (2.6)) and then manipulate the expansion to get the contributions to the ISE from regions where \( \hat{f} > 0 \) and \( \hat{f} < 0 \). Subsequently we focus on components of the ISE that arise from the negative density estimates.
where $A_1 = \int K^2$, $A_2 = (k_r/r!)^2 \int \{f^{(r)}\}^2$ (with $r$ assumed to be even).

$A_3 = \int f^2$, $A_4 = 2(-1)^{-r/2+1}k_r/r! \int \{f^{(r)}\}^2$ and $i = 1, 2$.

(Similar formulae may be obtained for odd $r$, but kernels of odd order are seldom used in practice.) The conditions that we stipulated above are the same as those that led to the approximate formula for MISE, given at (1.5), and are available for a wide range of densities. This result permits a simple and direct comparison of the performance of $\hat{f}_i$ and $\hat{f}$ from the viewpoint of integrated squared error: the performances are asymptotically equivalent if and only if $\gamma^2 = o_p \{(nh)^{-1} + h^{2r}\}$.

To prove that the ISE's of $\hat{f}_1$ and $\hat{f}_2$ can be expressed in terms of the random variable, $\gamma$, via equation (2.8), we expand the formulae for the ISE's of $\hat{f}_1$ and $\hat{f}_2$ using the definitions (2.1) and (2.2). For the ISE of $\hat{f}_1$ we have,

$$
(\hat{f}_1 - f)^2 = \int \{\gamma_1 \hat{f} I(\hat{f} > 0) - f\}^2
= \int \{\gamma_1 (\hat{f} - f) I(\hat{f} > 0) - \gamma_1 f I(\hat{f} < 0) - (1 - \gamma_1) f\}^2
= (2\gamma_1 - 1) \int (\hat{f} - f)^2 I(\hat{f} > 0) + (\gamma_1 - 1)^2 \int \hat{f}^2 I(\hat{f} > 0)
+ 2(\gamma_1 - 1) \int (\hat{f} - f) f + \int f^2 I(\hat{f} < 0)
- 2(\gamma_1 - 1) \int (\hat{f} - f) f I(\hat{f} < 0).
$$

(2.9)

A similar expression for the ISE of $\hat{f}_2$ is,

$$
(\hat{f}_2 - f)^2 = \int \{\gamma_2 |\hat{f} | - f\}^2
= \int \{\gamma_2 \hat{f}^2 - 2\gamma_2 |\hat{f} | f + f^2\}
= (2\gamma_2 - 1) \int (\hat{f} - f)^2 + (\gamma_2 - 1)^2 \int \hat{f}^2
$$

14
The integrated squared error properties of \( \hat{f} \) are used to further modify equations (2.9) and (2.10). We assume that \( K \) is bounded, compactly supported, symmetric, Hölder continuous \(^1\) and of \( r \)th order; and \( f \) and its first \( r \) derivatives are bounded, continuous and integrable. Under these conditions, Marron and Härdle [15] show that

\[
\left| \frac{ISE(h) - MISE(h)}{MISE(h)} \right| \xrightarrow{a.s.} 0 \text{ as } n \to \infty.
\]

or

\[
\int (\hat{f} - f)^2 / \int E(\hat{f} - f)^2 \xrightarrow{a.s.} 1 \text{ as } n \to \infty.
\]  

The approximation for MISE, given in Silverman [23,p.39ff] and Prakasa Rao [18,Theorem 2.1.7] and previously stated in equation (1.5), is

\[
\int E(\hat{f} - f)^2 = (nh)^{-1} \int K^2 + h^{2r(k_r/r!)} \int \{f^{(r)}\}^2 + o \{ (nh)^{-1} + h^{2r} \}. \tag{2.12}
\]

To further reduce equations (2.9) and (2.10), we need the following results,

\[
\int \hat{f}^2 = \int f^2 + o_p \{ 1 \}, \tag{2.13}
\]

\[
\int \hat{f}^2 I(\hat{f} > 0) = \int f^2 + o_p \{ 1 \}, \tag{2.14}
\]

\[
\int (\hat{f} - f)I(\hat{f} < 0) = o_p \{ (nh)^{-1} + h^{2r} \}, \tag{2.15}
\]

\[
\int (\hat{f} - E\hat{f})f = o_p \{ (nh)^{-1/2} \}, \tag{2.16}
\]

and

\[
\int (E\hat{f} - f)f = -\frac{1}{2} A_4 h^r + o(h^r). \tag{2.17}
\]

The terms on the left side of (2.16) and (2.17) arise from \( \int (\hat{f} - f)f = \int (\hat{f} - E\hat{f})f + \)

\(^1g\) is Hölder continuous if \(|g(x) - g(y)| \leq c| x - y |^r \) \( \forall x, y \) and some \( c, c > 0 \).
\[ f(E\hat{f} - f)f. \] The results (2.13), (2.14), (2.15), (2.16) and (2.17) are proved in Section 2.5.

Substituting them into (2.9) we get,

\[
\int (\hat{f}_1 - f)^2
= (2\gamma_1 - 1) \left[ \int (\hat{f} - f)^2 \left\{ 1 - I(\hat{f} < 0) \right\} \right] + (\gamma_1 - 1)^2 \left[ \int f^2 + o_p(1) \right]
+ 2(\gamma_1 - 1) \left[ -\frac{1}{2} A_4 h^r + o(h^r) + o_p((nh)^{-1/2}) \right] + \left[ \int f^2 I(\hat{f} < 0) \right]
- 2(\gamma_1 - 1) \left[ o_p((nh)^{-1} + h^{2r}) \right]
= \{1 + o_p(1)\} \times
\left\{ \int (\hat{f} - f)^2 + (\gamma_1 - 1)^2 \int f^2 - (\gamma_1 - 1)A_4 h^r + \int f^2 I(\hat{f} < 0) \right\}. \tag{2.18}
\]

Equation (2.10) can be reduced to give

\[
\int (\hat{f}_2 - f)^2
= (2\gamma_2 - 1) \int (\hat{f} - f)^2 + (\gamma_2 - 1)^2 \left\{ \int f^2 + o_p(1) \right\}
+ 2(\gamma_2 - 1) \left[ -\frac{1}{2} A_4 h^r + o(h^r) + o_p((nh)^{-1/2}) \right] + 4\gamma_2 \int f \hat{f} I(\hat{f} < 0)
= \{1 + o_p(1)\} \times
\left\{ \int (\hat{f} - f)^2 + (\gamma_2 - 1)^2 \int f^2 - (\gamma_2 - 1)A_4 h^r + 4\gamma_2 \int f \hat{f} I(\hat{f} < 0) \right\}. \tag{2.19}
\]

During the proof of Theorem 2.3 it is shown that

\[
E \left\{ \int_0^\infty f |\hat{f}| I(\hat{f} < 0) \right\} = o\left\{ (nh)^{-1} \right\}
\text{ and } E \left\{ \int f^2 I(\hat{f} < 0) \right\} = o(h^{2r}).
\]

We examine this theorem in detail later but we require the results now so as to replace
terms in (2.18) and (2.19). We also use \((\gamma_1 - 1) \sim -\gamma\) and \((\gamma_2 - 1) \sim -2\gamma\), to give

\[
\int (\hat{f}_1 - f)^2 = \{1 + o_p(1)\} \left\{ (nh)^{-1}A_1 + h^{2r}A_2 + \gamma^2A_3 + \gamma A_4 h^r + o_p(h^{2r}) \right\}
\]

\[
= (nh)^{-1}A_1 + h^{2r}A_2 \gamma^2A_3 + \gamma A_4 h^r + o_p\left\{ h^{2r} + (nh)^{-1} + \gamma^2 \right\}
\]

and

\[
\int (\hat{f}_2 - f)^2 = \{1 + o_p(1)\} \times 
\]

\[
\left[ (nh)^{-1}A_1 + h^{2r}A_2 + (2\gamma)^2A_3 + 2\gamma A_4 h^r + 4\gamma_2 o\{(nh)^{-1}\} \right]
\]

\[
= (nh)^{-1}A_1 + h^{2r}A_2(2\gamma)^2A_3 + 2\gamma A_4 h^r + o_p\left\{ h^{2r} + (nh)^{-1} + \gamma^2 \right\},
\]

which proves the result (2.8).

2.3.2 Densities with Regularly Varying Tails

The aim of this section is to find properties of \(\gamma\) and then substitute into (2.8) to find the ISE of \(\hat{f}_1\) and \(\hat{f}_2\). We assume that the underlying density \(f\) has the following properties:

- tails that decrease as positive powers of \(|x|^{-1}\) as \(|x| \to \infty\),
- that \(f > 0\) on \((-\infty, +\infty)\) and that \(f\) has \(r\) continuous derivatives satisfying

\[
f^{(r)}(x) \sim \left( \frac{d}{dx} \right)^r c_1 x^{-\alpha_1}; \quad f^{(r)}(-x) \sim (-1)^r \left( \frac{d}{dx} \right)^r c_2 x^{-\alpha_2} \text{ as } n \to \infty,
\]

where \(c_1, c_2 > 0\) and \(\alpha_1, \alpha_2 > 1\) (so that \(f\) is integrable). (2.20)

The condition we assume of \(K\) is that

- for constants \(C_1, C_2 > 0\), \(K\) vanishes outside \((-C_1, C_1)\) and

\[
|K| \leq C_2; \text{ and } K \text{ is symmetric and of } r\text{'th order.} \quad (2.21)
\]
We make a transformation $x = (nh)^{1/\alpha_j}v$, $j = 1$ for the right tail and $j = 2$ for the left tail, and substitute for $x$ in the formula for $\hat{f}(x)$. The location of the negative regions of $\hat{f}$ is a function of $n$ and $h$. The transformation of $\hat{f}(x)$ to $Z(n,v)$ gives us a scale to investigate the convergence of an integral of the density estimator over a range dictated by $n$ and $h$, but free from the competing effect of the factor $(nh)^{-1}$ on the convergence of the integral of $\hat{f}$. It is shown in the proof that random variables $Z(n,v)$ converge to $Z_1(v)$ (or $Z_2(v)$).
Sometimes we shall strengthen this by requiring that $K$ be Hölder continuous. To investigate the properties of $\gamma$ (defined by (2.7)), we first define random variables $Z_1(v)$ associated with the right tail of $f$ and $Z_2(v)$ (left tail of $f$) that have characteristic functions,

$$
\psi_v(t, j) = \exp \left[ -c_j v^{-\alpha_j} \int \{1 - e^{itK(y)}\} dy \right], j = 1, 2
$$

$$
= \exp \left\{ -c_j v^{-\alpha_j} \beta(t) \right\}, \beta(t) = \int \{1 - e^{itK(y)}\} dy.
$$

(2.22)

The purpose of the random variables $Z_1(v)$ and $Z_2(v)$ becomes apparent during the proof of Theorem 2.1 (Section 2.5), where it is shown that random variables $Z(n, v) = nh \hat{f}(x)$ converge to $Z_1(v)$ (or $Z_2(v)$). We then define

$$
A_{s,j} = \int_0^\infty E [||Z_j(v)||I\{Z_j(v) < 0\}] dv, j = 1, 2,
$$

(2.23)

which we use to calculate $\gamma$. Under conditions (2.20) and (2.21), and assuming $K$ takes negative values on a set of positive measure, $A_{s,j}$ is absolutely convergent to a positive number. This follows from $Z_j(v)$ being the limit of a sequence of random variables which, with probability bounded away from zero, take values bounded below zero. Theorem 2.1 shows that $\gamma$ is very small so that the ISE of $\hat{f}$ is like the ISE of $\hat{f}$.

**Theorem 2.1** Assuming the conditions on $f$ (2.20) and $K$ (2.21) and that $h \to 0$, $nh \to \infty$ as $n \to \infty$, we have

$$
\gamma = \int_{-\infty}^{+\infty} |\hat{f}(x)|I\{\hat{f}(x) < 0\} \, dx
$$

$$
= \sum_{j=1}^2 (nh)^{-1+1/\alpha_j} A_{s,j} + o_p \left\{ \sum_{j=1}^2 (nh)^{-1+1/\alpha_j} \right\} \text{ as } n \to \infty.
$$

(2.24)
That is, the bandwidth which is asymptotically optimal for minimising the ISE of \( \hat{f} \) will also be optimal for minimising the ISE of \( \hat{f}_i, i = 1, 2 \), if and only if \( \alpha > 2 \).

From the viewpoint of the density, the condition \( \alpha > 2 \) implies the existence of finite means and when \( \alpha \leq 2 \), we note that \( \int_{-\infty}^{\infty} x f(x) dx \) is not finite. Thus our results will be relevant for those densities for which a moment higher than the mean is finite.

If this condition is not satisfied, there is too much information in the tails for our corrections to the negative estimates to be regarded as asymptotically negligible.
If one tail of the underlying density decreases slower than the other, it will exert
the dominant influence on $\gamma$ (and hence on $\hat{f}_i$) with the other tail having negligible
effect on $\gamma$. If both tails decrease at the same rate, both tails influence $\gamma$ equally and
we only need to calculate $\gamma$ for one tail as the other will be the same. To allow for
either balanced tails or one tail dominant, we define

$$
\alpha = \min(\alpha_1, \alpha_2), A_5 = A_{5,j} \text{ if } \alpha_1 \neq \alpha_2 \text{ and } \alpha = \alpha_j
$$

and $A_5 = A_{5,1} + A_{5,2}$ if $\alpha_1 = \alpha_2$. \hspace{1cm} (2.25)

Then equation (2.24) is equivalent to,

$$
\gamma = (nh)^{-1+1/\alpha}A_5 + o_p \{ (nh)^{-1+1/\alpha} \}. \hspace{1cm} (2.26)
$$

With the result from Theorem 2.1, we may substitute the right side of (2.26) for $\gamma$
into the formulae for ISE of $(\hat{f}_i, i = 1, 2)$ (see (2.8) on page 13) to get

$$
\int (\hat{f}_i - f)^2 = (nh)^{-1} A_i + h^2 R_2 + i^2 (nh)^{-2+2/\alpha} A_3 A_5^2
$$

$$
+ o_p \{ (nh)^{-1+1/\alpha} A_4 A_5 \}
$$

$$
+ o_p \{ (nh)^{-1} + h^2 + (nh)^{-2+2/\alpha} \}. \hspace{1cm} (2.27)
$$

If $h$ is chosen so that $(nh)^{-1}$ (a factor of the variance) and $h^2$ (a factor of the squared
bias) are of the same order, then

$$
\int (\hat{f}_i - f)^2 \sim \int (\hat{f} - f)^2 \text{ or } \int E(\hat{f} - f)^2 \text{ if and only if } \alpha > 2.
$$

Since $f(x) = e_1 x^{-\alpha}$, $\alpha \to 2$ implies the existence of finite means. In this sense, a
bandwidth which is asymptotically optimal in the sense of minimising ISE for $\hat{f}$ is
also optimal in the sense of minimising ISE for \( \hat{f}_i, i = 1, 2 \), if and only if a moment higher than the mean is finite.

The result from Theorem 2.1 is for densities given by (2.20). These conditions may be generalised by including slowly varying functions \( L_1, L_2 \) such that

\[
f^{(r)}(x) = x^{-(\alpha_1+r)} L_1(x), \quad f^{(r)}(-x) = x^{-(\alpha_2+r)} L_2(x)
\]

(2.28)

as \( x \to \infty \) and \( \alpha_1, \alpha_2 > 1 \). We need only make slight changes to the proof of Theorem 2.1 to accommodate these conditions and the only modifications from the previous result (2.27) are that terms in \((nh)^{-2+2/\alpha}\) and \((nh)^{-1+1/\alpha}\) are replaced by \((nh)^{-2+2/\alpha} L_3 \left\{ (nh)^{1/\alpha} \right\}^2\) and \((nh)^{-1+1/\alpha} L_3 \left\{ (nh)^{1/\alpha} \right\}\) respectively, where \( L_3 \) is another slowly varying function. The results due to the generalisations are elucidated in the proof of the theorem.

The former conclusion continues to hold. That is, the bandwidth which is asymptotically optimal for minimising the ISE for \( \hat{f} \) is also asymptotically optimal for minimising the ISE for \( \hat{f}_i, i = 1, 2 \), provided \( \alpha > 2 \), but not if \( 0 < \alpha < 1 \). The marginal case \( \alpha = 2 \) can go either way depending on the choice of \( L_1 \) and \( L_2 \).

Similar results are readily obtained in a wide range of other cases by adaption of Theorem 2.1. For example, consider densities with exponentially decreasing tails such that

\[
f^{(r)}(x) \sim \left( \frac{d}{dx} \right)^r c_{11} \exp \left( -c_{12} x^{\alpha_1} \right);
\]

\[
f^{(r)}(-x) \sim \left( \frac{d}{dx} \right)^r c_{21} \exp \left( -c_{22} x^{\alpha_2} \right)
\]

as \( x \to \infty \) where \( c_{ij}, \alpha_i \) are positive constants.

(2.29)
In such cases, $\gamma \to 0$ and

$$\int (\hat{f}_i - f)^2 = (nh)^{-1} A_1 + h^{2r} A_2 + o_p \left\{ (nh)^{-1} + h^{2r} \right\} .$$

We now examine the third density estimator given by

$$\hat{f}_3 = \gamma_3 \hat{f}(x) I(Y_1 \leq x \leq Y_2), \quad \text{where} \quad \gamma_3^{-1} = \int_{Y_1 \leq x \leq Y_2} \hat{f}(x) \, dx ,$$

$Y_1$ is the largest negative solution of $\hat{f}(x) = 0$ and $Y_2$ is the smallest positive solution of $\hat{f}(x) = 0$.

We continue to assume that $f$ has at least $r$ derivatives (conditions (2.20)) and strengthen the conditions on $K$ that were given at (2.21) by including Hölder continuity. Theorem 2.2 shows that $\alpha = \min(\alpha_1, \alpha_2) > 2$, which is a necessary and sufficient condition for $\hat{f}_1$ and $\hat{f}_2$ to have the same asymptotic ISE as $\hat{f}$, is also sufficient for $\hat{f}_3$ to share that formula.

**Theorem 2.2** The conditions (2.20), (2.21) and $K$ being Hölder continuous are assumed, with $\alpha > 2$ in (2.20). Also it is assumed that for some $0 < \epsilon < \frac{1}{2}$, $\epsilon^{-1+\epsilon} \leq n^{-1}$, $h \leq n^{-\epsilon}$. Then

$$\int (\hat{f}_3 - f)^2 = \{1 + o_p(1)\} \left\{ (nh)^{-1} A_1 + h^{2r} A_2 \right\} \quad \text{as} \quad n \to \infty .$$

The density estimate $\hat{f}_3$ is obtained by truncating $\hat{f}$ at a point where $\hat{f}$ goes negative and so the ISE of $\hat{f}_3$ will be similar to the ISE of $\hat{f}$ if the truncated part is negligible.

The density estimate $\hat{f}_4$ is similar to $\hat{f}_3$ but truncated at $2Y_2$ and we may conclude that

$$\int_{2Y_2}^{\infty} \left| \hat{f}(x) \right| = O_p \left\{ (nh)^{-1+1/\alpha} \right\} \quad \text{for} \quad \alpha > \alpha_1 .$$
This section concludes with Theorem 2.3 which gives the results that $\int f^2 I(\hat{f} < 0)$ and $\int f|\hat{f}|I(\hat{f} < 0)$ are smaller order than $(nh)^{-1}$. These integrals arise in the expansion of $f(\hat{f} - f)^2$ given at equations (2.9) and (2.10).

**Theorem 2.3** We assume conditions (2.20) regarding derivatives and the tails of $f$ and conditions (2.21) for the support, bound, symmetry, order and continuity of $K$, and that $h = h(n) \to 0$, $nh \to \infty$ as $n \to \infty$. Then,

$$\int f^2 I(\hat{f} < 0) + \int f|\hat{f}|I(\hat{f} < 0) = o_p\{ (nh)^{-1} \} \text{ as } n \to \infty .$$

### 2.3.3 Compactly Supported Densities

We describe the properties of $\gamma$ and of $\hat{f}_i$, $i = 1, 2, 3$, in the case where the true density $f$ is supported on a compact interval. Results analogous to those for unboundedly supported densities are derived to show that $\hat{f}_i$ and $\hat{f}$ are asymptotically equivalent.

Without loss of generality, we may take the support of $f$ to be $(0,1)$. However, the support of $\hat{f}(x)$ is close to $(-hC_1, 1 + hC_1)$, and to assess $\gamma$ we have to consider the contribution to $\hat{f}(x)$ outside of $(0,1)$. We assume that $f$ vanishes outside $(0,1)$, $f > 0$ and has $r$ continuous derivatives on $(0,1)$, and

$$f^{(r)}(x) \sim \left( \frac{d}{dx} \right)^r c_1 x^{\alpha_1} ,$$

$$f^{(r)}(1 - x) \sim (-1)^r \left( \frac{d}{dx} \right)^r c_2 x^{\alpha_2}$$

(2.30)

as $x \downarrow 0$, where $c_1, c_2 > 0$ and $\alpha_1, \alpha_2 \geq r$.

Beta densities represent examples of this type. The condition $\alpha_1, \alpha_2 \geq r$ ensures that $f$ has $r$ bounded derivatives on $(-\infty, \infty)$. Theorem 2.4 is an anologue of Theorem 2.4.
2.1 for densities satisfying (2.30) rather than (2.20). We define the function

\[ G_j(v) = c_j \int_{u<v} (v - u)^{\alpha_j} K(u) du, \]  

(2.31)

and put

\[ A_{s,j} = - \int G_j I(G_j < 0). \]

If \( K \) takes negative values then \( A_{s,j} \) is strictly positive. If we assume that \( K \) is compactly supported, then the integrand in the definition of \( A_{s,j} \) vanishes outside a compact set.

**Theorem 2.4** Assume conditions (2.21) for \( K \) and conditions (2.30) for \( f \), and that \( h = h(n) \to 0 \) and \( nh \to \infty \). Then,

\[ \gamma = \sum_{j=1}^{2} h^{\alpha_j+1} A_{s,j} + o_p \left( \sum_{j=1}^{2} h^{\alpha_j+1} \right) \] as \( n \to \infty . \]

(2.32)

As in Theorem 2.1, we assume that either one tail of the density provides most of the information about \( \gamma \) or that both tails operate equally. Substituting

\[ \gamma = h^{\alpha+1} A_5 + o_p \{ h^{\alpha+1} \} \]

into (2.8), we obtain the analogue of (2.27) for densities with compact support,

\[ \int (\hat{f}_i - f)^2 = (nh)^{-1} A_1 + h^{2r} A_2 + i^2 h^{2(\alpha+r)} A_3 A_5^2 + i h^{r+\alpha+1} A_4 A_5 \]

\[ + o_p \left\{ (nh)^{-1} + h^{2r} + h^{2(\alpha+1)} \right\} , i = 1, 2 . \]

(2.33)

However since \( \alpha \geq r \),

\[ \int (\hat{f}_i - f)^2 = \{ 1 + o_p(1) \} \left\{ (nh)^{-1} A_1 + h^{2r} A_2 \right\} , \]

(2.34)
and we may conclude that the integrated squared errors for $\hat{f}_i$, $i = 1, 2$, and $\hat{f}$ are asymptotically equivalent when the underlying density has compact support.

We now state an analogue of Theorem 2.2 which gives the ISE for $\hat{f}_3$ when the underlying density has compact support. Without loss of generality, we may assume that the support of $f$ is the interval $(0, 1)$. Let $Y_1$ and $Y_2$ denote respectively the largest solution less than $\frac{1}{2}$ and the smallest solution greater than $\frac{1}{2}$, of $\hat{f}(x) = 0$. Define $\hat{f}_3$ by (2.3). Theorem 2.5 shows that $\hat{f}_3$ has the same asymptotic ISE as $\hat{f}$.

**Theorem 2.5** Assume that $K$ has support $(-C_1, C_1)$, $|K| \leq C_2$, is symmetric of $r$th order, and is Hölder continuous. Also assume that for some $0 < \epsilon < \frac{1}{2}$, $n^{-1+\epsilon} \leq h \leq n^{-\epsilon}$. Then,

$$\int (\hat{f}_3 - f)^2 = (1 + o_p(1)) \{(nh)^{-\frac{1}{r}}A_1 + h^{-2r}A_2\}.$$

Theorem 2.6 is the analogue of Theorem 2.3 for compactly supported densities.

**Theorem 2.6** The conditions for $K$ and $f$ are the same as for theorem 2.5. Also, we assume that $h = h(n) \to 0$ and $nh \to \infty$. Then,

$$\int f^2I(\hat{f} < 0) + \int f|\hat{f}|I(\hat{f} < 0) = o_p \{(nh)^{-1} + h^{2r}\} \text{ as } n \to \infty.$$

The proofs of these two theorems are similar to their counterparts when the support of the underlying density is unbounded.
2.4 Simulation Study

We studied each of the estimators \( \hat{f}, \hat{f}_1, \hat{f}_2, \hat{f}_3 \) in the cases of data from Normal, Cauchy, Student's \( t \), Gamma and Beta distributions and \( \hat{f}_4 \) for Normal, Cauchy and Student's \( t \) distributions and compared the integrated squared errors of \( \hat{f} \) and \( \hat{f}_i \) as \( n \to \infty \).

The theory in Section 2.3, and its analogue for the Gamma case, predicts that for all but the Cauchy distribution, each type of nonnegative density estimator \( \hat{f}_i \), should have integrated squared error close to that of the basic estimator \( \hat{f} \), if the same bandwidth is used for all four estimators, and provided that in the Gamma and Beta densities the shape parameters are chosen so that the densities are sufficiently smooth.

2.4.1 Simulation Details

We took \( r = 4 \) throughout, and used the 4'th order kernel suggested by Gasser et al. [11],

\[
K(y) = (15/32)(7y^4 - 10y^2 + 3) I(|y| \leq 1) .
\] (2.35)

By ensuring that the shape parameters of Gamma and Beta distributions are greater than or equal to 5 we guarantee that \( f \) has \( r = 4 \) bounded and continuous derivatives on \((-\infty, \infty)\), except possibly at the origin (for a Gamma distribution with shape parameter 5) or at 0 or 1 (for a Beta distribution with shape parameter 5), where \( f^{(4)} \) might have a jump discontinuity.
Conditions (2.21) and Hölder continuity are satisfied by $K$; condition (2.20) is satisfied by the Cauchy distribution with $\alpha = \alpha_1 = \alpha_2 = 2$, and by Student’s t distribution with number of degrees of freedom equal to $\alpha - 1$; condition (2.30) is satisfied by the Beta distribution with shape parameters $(\beta_1, \beta_2) = (\alpha_1 + 1, \alpha_2 + 1)$; and it is straightforward to prove that for the Normal and Gamma distributions, the obvious analogues of our main results hold:

$$f(\hat{f} - f)^2 = \{1 + o_p(1)\} \{(nh)^{-1} A_1 + h^8 A_2\}, \quad i = 1, 2, 3,$$
$$\int f^2 I(\hat{f} < 0) + \int f|\hat{f}| I(\hat{f} < 0) = o_p\{(nh)^{-1} + h^8\}.$$  

We considered two empirical methods for bandwidth choice, cross-validation (e.g. Silverman [23,p.48ff] and reference to a standard distribution (e.g. Silverman [23,p.45ff]).

The latter technique may be developed by noting that by standard asymptotic theory for kernel estimators (Silverman [23,p.66ff]), the optimal bandwidth is given very nearly by

$$h = 72\frac{1}{5} \left\{ \int_{-1}^{1} y^2 K(y) dy \right\}^{\frac{1}{5}} \left\{ \int_{-\infty}^{\infty} f^{(4)}(x)^2 dx \right\}^{-\frac{1}{3}} n^{-\frac{1}{5}}. \quad (2.36)$$

If the underlying distribution were Normal $N(\mu, \sigma^2)$ then, for the particular kernel given at (2.35), formula (2.36) would reduce to

$$h = 2.58 \sigma n^{-\frac{1}{5}}.$$

Now, the interquartile range of the normal distribution equals 1.34 times its standard deviation. This observation motivates an empirical rule,

$$\hat{h} = 2.58 \min(\hat{\sigma}, \hat{\sigma}/1.34) n^{-\frac{1}{5}}. \quad (2.37)$$

26
where $\hat{\sigma}$ denotes sample standard deviation and $\hat{q}$ is sample interquartile range. By incorporating $\hat{q}$ into (2.37) we ensure that the bandwidth rule is practicable for heavy-tailed distributions like the Cauchy.

The cross-validation technique for choosing $h$ requires finding the minimum of a score function,

$$M_1(h) = n^{-2} \sum_i \sum_j K^* \left\{ h^{-1}(X_i - X_j) \right\} + 2n^{-1}h^{-1}K(0),$$

where $K^*(t) = (K * K)(t) - 2K(t)$ with $*$ being the convolution operator. The score function was initially calculated over a coarse grid of values of $h$ which were multiples $(0.25, 0.5, 1, 1.5, 2, 2.25, 2.5)$ of $\hat{h}$ as given by (2.37). Then the grid was refined stepwise by locating the current minimum of $M_1(h)$ and adding an extra grid point on either side of it. This continued until the relative change in $M_1(h)$ was less than 1%.

The computing of $K^* \{ h^{-1}(X_i - X_j) \}$ was done by first evaluating $K^*(t)$ for 401 values of $t \in [-2, 2]$ and recording these values in a table. For each $X_i$ and only values of $X_j$ within two bandwidths of $X_i$, we referred to the table (interpolating between the tabulated values if necessary) to get the value of $K^* \{ h^{-1}(X_i - X_j) \}$. By ordering the data and utilising the symmetry of $K^*(t)$, the score function was evaluated efficiently.

### 2.4.2 Results

One hundred data sets of sizes 50, 100, 200, 400 and 1,000 were simulated using NAG routines. Each density estimator was evaluated at 256 or more points, and numerical
integrations were performed on grids of at least 200 points.

As our index of the closeness of the integrated squared error of $\hat{f}$ to that of $f$ we took

$$R_i = \text{av} \frac{|\text{ISE}(\hat{f}_i) - \text{ISE}(\hat{f})|}{\{\text{av} \text{ISE}(\hat{f})\}} \tag{2.38}$$

where "av" denotes the average over all simulations. We also recorded the number of times that $\text{ISE}(\hat{f}_i) < \text{ISE}(\hat{f})$. These data are summarized in Tables 2.2 to 2.9, for the Normal, Cauchy, Student's t with 2, 4, 5, 8 degrees of freedom, Gamma (5, 1) and Beta (5, 9) distributions respectively, and for the sample sizes $n = 50, 100, 200, 400, 1000$. Relative errors for values of $R_i$, i.e. standard deviations divided by means, are between 7% and 12%.

Normal and Cauchy distributions are included because they represent opposite extremes of tail behaviour for distributions with unbounded support. In the case of Normal data, the tails of $f$ are so light that the ISE effect of rendering nonnegative the tails of $\hat{f}$ is very small indeed. In this case, we observe the predicted decrease in the values of $R_i$ as $n \to \infty$. The decrease occurs more rapidly when $h$ is estimated by the rule (2.37) than when cross-validation is used. By way of contrast, the Cauchy distribution has $\alpha_1 = \alpha_2 = 2$, and so comprises the critical case where the optimal bandwidths for $\hat{f}$, $\hat{f}_1$, $\hat{f}_2$, $\hat{f}_3$ and $\hat{f}_4$ are of the same order of magnitude but have different multiplicative constants. In consequence, the value of $R_i$ converges to a nonzero constant as $n \to \infty$. For Student's $t$ distribution with three or more degrees of freedom, the value of $R_i$ converges to zero as $n \to \infty$, and the rate of convergence increases with the number of degrees of freedom. The convergence of $R_i$ for Student's
t (2 df) distribution is quite slow and when using cross-validation to estimate \( h \), a sample size of 1000 is insufficient to detect the predicted decrease. This distribution has \( \alpha \) (defined at (2.20)) equal to 3. Whilst these tails are lighter than in the critical case of the Cauchy (\( \alpha = 2 \)), sufficient data can be found in the tails such that \( \gamma \) does not readily approach zero as \( n \to \infty \).

Data from the Beta distribution are presented because they illustrate the example discussed in Section 2.3.3. As expected, they exhibit similar behaviour to the Normal case, although in the case of Beta data the values of \( R_i \) converge to zero a little more rapidly. The convergence of \( R_i \) for the Gamma data is as predicted by the theory.

Plots illustrating these results are shown in Figures 2.5 - 2.7.
Figure 2.5: Values of relative difference ($R_i$) of $\hat{f}_i$ for sample sizes 50, 100, 200, 400, 1000 of Normal (0,1), Cauchy (0,1) and Student's t (2 df) data.

For graphs on the left, the bandwidth is estimated by cross validation and for those on the right, bandwidth is estimated by reference to the Normal distribution.
Figure 2.6: Values of relative difference ($R_i$) of $\hat{f}_i$ for sample sizes 50, 100, 200, 400, 1000 of Student's t (3 df), Student's t (5 df) and Student's t (8 df) data.
Figure 2.7: Values of relative difference ($R_i$) of $\hat{f}_i$ for sample sizes 50, 100, 200, 400, 1000 of Gamma(5,1) and Beta(5,9) data.
Table 2.2: Values of relative difference ($R_i$) of $\hat{f}$, for simulated Normal data. The numbers in parentheses in column $R_i$ represent the percentage of times that $\text{ISE}(\hat{f}) < \text{ISE}(\hat{f})$. The bandwidth methods are cross-validation (C.V.) and "reference to a standard distribution" (ref.).

<table>
<thead>
<tr>
<th>n</th>
<th>bandwidth method</th>
<th>mean ISE($\hat{f}$)</th>
<th>s.d. ISE($\hat{f}$)</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>C.V.</td>
<td>0.013</td>
<td>0.013</td>
<td>0.035</td>
<td>0.084</td>
<td>0.043</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(41)</td>
<td></td>
<td>(28)</td>
<td>(37)</td>
<td>(29)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.012</td>
<td>0.009</td>
<td>0.021</td>
<td>0.037</td>
<td>0.022</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(72)</td>
<td></td>
<td>(65)</td>
<td>(70)</td>
<td>(62)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>C.V.</td>
<td>0.0065</td>
<td>0.0052</td>
<td>0.031</td>
<td>0.070</td>
<td>0.031</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(47)</td>
<td></td>
<td>(35)</td>
<td>(46)</td>
<td>(38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0064</td>
<td>0.0045</td>
<td>0.014</td>
<td>0.027</td>
<td>0.013</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(58)</td>
<td></td>
<td>(54)</td>
<td>(53)</td>
<td>(55)</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>C.V.</td>
<td>0.0036</td>
<td>0.0029</td>
<td>0.023</td>
<td>0.057</td>
<td>0.022</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(41)</td>
<td></td>
<td>(30)</td>
<td>(41)</td>
<td>(32)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0036</td>
<td>0.0026</td>
<td>0.008</td>
<td>0.016</td>
<td>0.006</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(60)</td>
<td></td>
<td>(56)</td>
<td>(56)</td>
<td>(57)</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>C.V.</td>
<td>0.0022</td>
<td>0.0018</td>
<td>0.020</td>
<td>0.048</td>
<td>0.020</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(31)</td>
<td></td>
<td>(25)</td>
<td>(30)</td>
<td>(25)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0020</td>
<td>0.0016</td>
<td>0.0053</td>
<td>0.011</td>
<td>0.0053</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(49)</td>
<td></td>
<td>(46)</td>
<td>(49)</td>
<td>(48)</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>C.V.</td>
<td>0.0010</td>
<td>0.0007</td>
<td>0.013</td>
<td>0.029</td>
<td>0.012</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(32)</td>
<td></td>
<td>(28)</td>
<td>(30)</td>
<td>(27)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0009</td>
<td>0.00065</td>
<td>0.003</td>
<td>0.006</td>
<td>0.003</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(56)</td>
<td></td>
<td>(52)</td>
<td>(52)</td>
<td>(52)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.3: Values of relative difference ($R_i$) of $\hat{f}_i$ for simulated Cauchy data. The numbers in parentheses in column $R_i$ represent the percentage of times that $\text{ISE}(\hat{f}_i) < \text{ISE}(\hat{f})$. The bandwidth methods are cross-validation (C.V.) and "reference to a standard distribution" (ref.).

<table>
<thead>
<tr>
<th>n</th>
<th>bandwidth method</th>
<th>mean $\text{ISE}(\hat{f})$</th>
<th>s.d. $\text{ISE}(\hat{f})$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>C.V.</td>
<td>0.015</td>
<td>0.008</td>
<td>0.030</td>
<td>0.055</td>
<td>0.21</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(54)</td>
<td>(50)</td>
<td>(57)</td>
<td>(67)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.013</td>
<td>0.010</td>
<td>0.037</td>
<td>0.068</td>
<td>0.24</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(63)</td>
<td>(58)</td>
<td>(46)</td>
<td>(62)</td>
</tr>
<tr>
<td>100</td>
<td>C.V.</td>
<td>0.0088</td>
<td>0.0046</td>
<td>0.018</td>
<td>0.033</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(60)</td>
<td>(53)</td>
<td>(38)</td>
<td>(53)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0072</td>
<td>0.0036</td>
<td>0.023</td>
<td>0.044</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(62)</td>
<td>(56)</td>
<td>(34)</td>
<td>(51)</td>
</tr>
<tr>
<td>200</td>
<td>C.V.</td>
<td>0.0046</td>
<td>0.0023</td>
<td>0.018</td>
<td>0.033</td>
<td>0.23</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(54)</td>
<td>(51)</td>
<td>(32)</td>
<td>(42)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0039</td>
<td>0.0018</td>
<td>0.022</td>
<td>0.037</td>
<td>0.23</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(54)</td>
<td>(52)</td>
<td>(28)</td>
<td>(43)</td>
</tr>
<tr>
<td>400</td>
<td>C.V.</td>
<td>0.0027</td>
<td>0.0014</td>
<td>0.016</td>
<td>0.031</td>
<td>0.22</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(50)</td>
<td>(49)</td>
<td>(27)</td>
<td>(38)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0026</td>
<td>0.0013</td>
<td>0.016</td>
<td>0.033</td>
<td>0.21</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(49)</td>
<td>(46)</td>
<td>(34)</td>
<td>(43)</td>
</tr>
<tr>
<td>1000</td>
<td>C.V.</td>
<td>0.0012</td>
<td>0.0067</td>
<td>0.013</td>
<td>0.025</td>
<td>0.24</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(45)</td>
<td>(42)</td>
<td>(23)</td>
<td>(31)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0014</td>
<td>0.0006</td>
<td>0.011</td>
<td>0.022</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(37)</td>
<td>(35)</td>
<td>(28)</td>
<td>(45)</td>
</tr>
</tbody>
</table>
Table 2.4: Values of relative difference ($R_i$) of $\hat{f}$ for simulated Students’ $t$ (2 df.) data. The numbers in parentheses in column $R_i$ represent the percentage of times that $\text{ISE}(\hat{f}_i) < \text{ISE}(\hat{f})$. The bandwidth methods are cross-validation (C.V.) and “reference to a standard distribution” (ref.).

<table>
<thead>
<tr>
<th>n</th>
<th>bandwidth method</th>
<th>mean ISE($\hat{f}$)</th>
<th>s.d. ISE($\hat{f}$)</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>C.V.</td>
<td>0.012</td>
<td>0.008</td>
<td>0.042</td>
<td>0.074</td>
<td>0.097</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(98)</td>
<td>(97)</td>
<td>(42)</td>
<td>(71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.012</td>
<td>0.010</td>
<td>0.053</td>
<td>0.092</td>
<td>0.16</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)</td>
<td>(98)</td>
<td>(33)</td>
<td>(64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>C.V.</td>
<td>0.0075</td>
<td>0.0033</td>
<td>0.034</td>
<td>0.057</td>
<td>0.11</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(99)</td>
<td>(99)</td>
<td>(30)</td>
<td>(58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0066</td>
<td>0.0032</td>
<td>0.033</td>
<td>0.061</td>
<td>0.16</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)</td>
<td>(99)</td>
<td>(18)</td>
<td>(54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>C.V.</td>
<td>0.0053</td>
<td>0.0020</td>
<td>0.025</td>
<td>0.044</td>
<td>0.089</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)</td>
<td>(100)</td>
<td>(16)</td>
<td>(55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0047</td>
<td>0.0017</td>
<td>0.035</td>
<td>0.059</td>
<td>0.14</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)</td>
<td>(100)</td>
<td>(10)</td>
<td>(42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>C.V.</td>
<td>0.0040</td>
<td>0.0012</td>
<td>0.019</td>
<td>0.036</td>
<td>0.11</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)</td>
<td>(100)</td>
<td>(1)</td>
<td>(37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0037</td>
<td>0.0011</td>
<td>0.022</td>
<td>0.043</td>
<td>0.14</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)</td>
<td>(100)</td>
<td>(2)</td>
<td>(34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>C.V.</td>
<td>0.0030</td>
<td>0.0007</td>
<td>0.014</td>
<td>0.029</td>
<td>0.12</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)</td>
<td>(100)</td>
<td>(1)</td>
<td>(23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0029</td>
<td>0.0007</td>
<td>0.028</td>
<td>0.028</td>
<td>0.12</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(100)</td>
<td>(100)</td>
<td>(1)</td>
<td>(24)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

35
Table 2.5: Values of relative difference \( R_i \) of \( \hat{f} \) for simulated Students't (3 df.) data. The numbers in parentheses in column \( R_i \) represent the percentage of times that \( \text{ISE}(\hat{f}) < \text{ISE}(\hat{f}) \). The bandwidth methods are cross-validation (C.V.) and “reference to a standard distribution” (ref.).

<table>
<thead>
<tr>
<th>n</th>
<th>bandwidth method</th>
<th>mean ( \text{ISE}(\hat{f}) )</th>
<th>s.d. ( \text{ISE}(\hat{f}) )</th>
<th>( R_1 )</th>
<th>( R_2 )</th>
<th>( R_3 )</th>
<th>( R_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>C.V.</td>
<td>0.018</td>
<td>0.010</td>
<td>0.021</td>
<td>0.048</td>
<td>0.040</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(11)</td>
<td>(8)</td>
<td>(45)</td>
<td>(38)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.014</td>
<td>0.008</td>
<td>0.015</td>
<td>0.031</td>
<td>0.060</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(31)</td>
<td>(22)</td>
<td>(59)</td>
<td>(59)</td>
</tr>
<tr>
<td>100</td>
<td>C.V.</td>
<td>0.012</td>
<td>0.008</td>
<td>0.014</td>
<td>0.031</td>
<td>0.040</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4)</td>
<td>(3)</td>
<td>(58)</td>
<td>(45)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.009</td>
<td>0.0011</td>
<td>0.024</td>
<td>0.024</td>
<td>0.061</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(12)</td>
<td>(9)</td>
<td>(69)</td>
<td>(53)</td>
</tr>
<tr>
<td>200</td>
<td>C.V.</td>
<td>0.0088</td>
<td>0.0046</td>
<td>0.009</td>
<td>0.019</td>
<td>0.043</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(67)</td>
<td>(43)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.007</td>
<td>0.0034</td>
<td>0.010</td>
<td>0.022</td>
<td>0.057</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(77)</td>
<td>(48)</td>
</tr>
<tr>
<td>400</td>
<td>C.V.</td>
<td>0.0065</td>
<td>0.0028</td>
<td>0.007</td>
<td>0.014</td>
<td>0.036</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(82)</td>
<td>(41)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0055</td>
<td>0.0022</td>
<td>0.008</td>
<td>0.017</td>
<td>0.049</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(86)</td>
<td>(41)</td>
</tr>
<tr>
<td>1000</td>
<td>C.V.</td>
<td>0.0050</td>
<td>0.0014</td>
<td>0.005</td>
<td>0.010</td>
<td>0.029</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(88)</td>
<td>(36)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0045</td>
<td>0.0013</td>
<td>0.003</td>
<td>0.011</td>
<td>0.037</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(92)</td>
<td>(39)</td>
</tr>
</tbody>
</table>
Table 2.6: Values of relative difference ($R_i$) of $\hat{f}$ for simulated Students’ t (5 df.) data. The numbers in parentheses in column $R_i$ represent the percentage of times that $\text{ISE}(\hat{f}_i) < \text{ISE}(\hat{f})$. The bandwidth methods are cross-validation (C.V.) and “reference to a standard distribution” (ref.).

<table>
<thead>
<tr>
<th>n</th>
<th>bandwidth method</th>
<th>mean ISE($\hat{f}$)</th>
<th>s.d. ISE($\hat{f}$)</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>C.V.</td>
<td>0.016</td>
<td>0.019</td>
<td>0.028</td>
<td>0.057</td>
<td>0.090</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(31)</td>
<td>(25)</td>
<td>(38)</td>
<td>(36)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.011</td>
<td>0.009</td>
<td>0.020</td>
<td>0.038</td>
<td>0.035</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(51)</td>
<td>(44)</td>
<td>(54)</td>
<td>(55)</td>
</tr>
<tr>
<td>100</td>
<td>C.V.</td>
<td>0.008</td>
<td>0.005</td>
<td>0.016</td>
<td>0.034</td>
<td>0.021</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(24)</td>
<td>(21)</td>
<td>(34)</td>
<td>(31)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.007</td>
<td>0.004</td>
<td>0.011</td>
<td>0.020</td>
<td>0.027</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(43)</td>
<td>(35)</td>
<td>(46)</td>
<td>(44)</td>
</tr>
<tr>
<td>200</td>
<td>C.V.</td>
<td>0.005</td>
<td>0.003</td>
<td>0.009</td>
<td>0.019</td>
<td>0.014</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(23)</td>
<td>(19)</td>
<td>(46)</td>
<td>(26)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0042</td>
<td>0.0026</td>
<td>0.008</td>
<td>0.018</td>
<td>0.016</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(21)</td>
<td>(20)</td>
<td>(43)</td>
<td>(37)</td>
</tr>
<tr>
<td>400</td>
<td>C.V.</td>
<td>0.0032</td>
<td>0.0020</td>
<td>0.006</td>
<td>0.011</td>
<td>0.011</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(16)</td>
<td>(8)</td>
<td>(55)</td>
<td>(16)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0027</td>
<td>0.0014</td>
<td>0.005</td>
<td>0.009</td>
<td>0.016</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(25)</td>
<td>(12)</td>
<td>(72)</td>
<td>(26)</td>
</tr>
<tr>
<td>1000</td>
<td>C.V.</td>
<td>0.0019</td>
<td>0.0009</td>
<td>0.004</td>
<td>0.007</td>
<td>0.014</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(22)</td>
<td>(8)</td>
<td>(66)</td>
<td>(18)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0015</td>
<td>0.0006</td>
<td>0.004</td>
<td>0.009</td>
<td>0.022</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(30)</td>
<td>(11)</td>
<td>(11)</td>
<td>(20)</td>
</tr>
</tbody>
</table>
Table 2.7: Values of relative difference ($R_t$) of $\hat{f}$ for simulated Students't (8 df.) data. The numbers in parentheses in column $R_t$ represent the percentage of times that $\text{ISE}(\hat{f}) < \text{ISE}(\hat{f})$. The bandwidth methods are cross-validation (C.V.) and "reference to a standard distribution" (ref.).

<table>
<thead>
<tr>
<th>n</th>
<th>bandwidth</th>
<th>mean</th>
<th>s.d.</th>
<th>ISE(\hat{f})</th>
<th>ISE(\hat{f})</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(25)</td>
<td>(22)</td>
<td>(31)</td>
<td>(29)</td>
<td>(25)</td>
<td>(22)</td>
</tr>
<tr>
<td>50</td>
<td>C.V.</td>
<td>0.012</td>
<td>0.008</td>
<td>0.030</td>
<td>0.072</td>
<td>0.031</td>
<td>0.058</td>
<td>(25)</td>
<td>(22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(57)</td>
<td>(53)</td>
<td>(55)</td>
<td>(55)</td>
<td>(57)</td>
<td>(53)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.011</td>
<td>0.008</td>
<td>0.017</td>
<td>0.033</td>
<td>0.018</td>
<td>0.052</td>
<td>(57)</td>
<td>(53)</td>
</tr>
<tr>
<td>100</td>
<td>C.V.</td>
<td>0.008</td>
<td>0.005</td>
<td>0.021</td>
<td>0.048</td>
<td>0.021</td>
<td>0.047</td>
<td>(21)</td>
<td>(19)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(21)</td>
<td>(19)</td>
<td>(30)</td>
<td>(29)</td>
<td>(21)</td>
<td>(19)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.006</td>
<td>0.004</td>
<td>0.011</td>
<td>0.023</td>
<td>0.019</td>
<td>0.043</td>
<td>(40)</td>
<td>(36)</td>
</tr>
<tr>
<td>200</td>
<td>C.V.</td>
<td>0.0040</td>
<td>0.0030</td>
<td>0.012</td>
<td>0.027</td>
<td>0.015</td>
<td>0.035</td>
<td>(14)</td>
<td>(13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(14)</td>
<td>(13)</td>
<td>(38)</td>
<td>(20)</td>
<td>(14)</td>
<td>(13)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0035</td>
<td>0.0026</td>
<td>0.006</td>
<td>0.013</td>
<td>0.015</td>
<td>0.034</td>
<td>(24)</td>
<td>(21)</td>
</tr>
<tr>
<td>400</td>
<td>C.V.</td>
<td>0.0025</td>
<td>0.0016</td>
<td>0.009</td>
<td>0.019</td>
<td>0.010</td>
<td>0.035</td>
<td>(7)</td>
<td>(7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(7)</td>
<td>(7)</td>
<td>(28)</td>
<td>(11)</td>
<td>(7)</td>
<td>(7)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0020</td>
<td>0.0009</td>
<td>0.005</td>
<td>0.011</td>
<td>0.011</td>
<td>0.038</td>
<td>(9)</td>
<td>(8)</td>
</tr>
<tr>
<td>1000</td>
<td>C.V.</td>
<td>0.0014</td>
<td>0.0007</td>
<td>0.005</td>
<td>0.010</td>
<td>0.010</td>
<td>0.032</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(36)</td>
<td>(7)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0012</td>
<td>0.0006</td>
<td>0.004</td>
<td>0.009</td>
<td>0.011</td>
<td>0.037</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(41)</td>
<td>(3)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
</tbody>
</table>
Table 2.8: Values of relative difference \((R_i)\) of \(\hat{f}_i\) for simulated Gamma data. The numbers in parentheses in column \(R_i\) represent the percentage of times that \(\text{ISE}(\hat{f}_i) < \text{ISE}(\hat{f})\). The bandwidth methods are cross-validation (C.V.) and "reference to a standard distribution" (ref.).

<table>
<thead>
<tr>
<th>n</th>
<th>bandwidth method</th>
<th>mean ISE((\hat{f}))</th>
<th>s.d. ISE((\hat{f}))</th>
<th>(R_1)</th>
<th>(R_2)</th>
<th>(R_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>C.V.</td>
<td>0.007</td>
<td>0.005</td>
<td>0.009</td>
<td>0.018</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(39)</td>
<td>(30)</td>
<td>(70)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.005</td>
<td>0.003</td>
<td>0.010</td>
<td>0.017</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(71)</td>
<td>(64)</td>
<td>(71)</td>
</tr>
<tr>
<td>100</td>
<td>C.V.</td>
<td>0.004</td>
<td>0.003</td>
<td>0.006</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(36)</td>
<td>(30)</td>
<td>(62)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.003</td>
<td>0.002</td>
<td>0.006</td>
<td>0.011</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(58)</td>
<td>(51)</td>
<td>(56)</td>
</tr>
<tr>
<td>200</td>
<td>C.V.</td>
<td>0.002</td>
<td>0.001</td>
<td>0.003</td>
<td>0.006</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(33)</td>
<td>(32)</td>
<td>(43)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.002</td>
<td>0.0009</td>
<td>0.004</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(61)</td>
<td>(54)</td>
<td>(62)</td>
</tr>
<tr>
<td>400</td>
<td>C.V.</td>
<td>0.001</td>
<td>0.0009</td>
<td>0.002</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(33)</td>
<td>(30)</td>
<td>(33)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0009</td>
<td>0.0006</td>
<td>0.003</td>
<td>0.006</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(53)</td>
<td>(44)</td>
<td>(48)</td>
</tr>
<tr>
<td>1000</td>
<td>C.V.</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.002</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(42)</td>
<td>(27)</td>
<td>(43)</td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.003</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(46)</td>
<td>(37)</td>
<td>(45)</td>
</tr>
</tbody>
</table>
Table 2.9: Values of relative difference ($R_i$) of $\hat{f}$ for simulated Beta data. The numbers in parentheses in column $R_i$ represent the percentage of times that $\text{ISE}(\hat{f}) < \text{ISE}(\hat{f})$. The bandwidth methods are cross-validation (C.V.) and "reference to a standard distribution" (ref.).

<table>
<thead>
<tr>
<th>n</th>
<th>bandwidth method</th>
<th>mean s.d.</th>
<th>ISE((\hat{f})) ISE((\hat{f}))</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>C.V.</td>
<td>0.083 0.062</td>
<td>0.023 0.047 0.022 (44) (34) (48)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.090 0.076</td>
<td>0.026 0.045 0.026 (76) (71) (76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>C.V.</td>
<td>0.044 0.033</td>
<td>0.019 0.043 0.019 (40) (25) (41)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.043 0.029</td>
<td>0.014 0.025 0.014 (75) (66) (76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>C.V.</td>
<td>0.026 0.021</td>
<td>0.016 0.042 0.017 (33) (26) (35)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.022 0.013</td>
<td>0.010 0.018 0.010 (68) (57) (64)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>C.V.</td>
<td>0.014 0.011</td>
<td>0.015 0.035 0.015 (32) (20) (33)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.013 0.009</td>
<td>0.007 0.014 0.007 (61) (48) (65)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>C.V.</td>
<td>0.007 0.005</td>
<td>0.012 0.029 0.012 (29) (19) (29)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ref.</td>
<td>0.006 0.004</td>
<td>0.006 0.011 0.006 (59) (50) (60)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.5 Proofs of Results and Theorems

In this section we shall make frequent use of Bernstein's inequality (Serfling [22,p.95ff]) and we restate it here because of its importance.

Let $Y_1 \cdots Y_n$ be independent random variables satisfying $Pr \{|Y_i - E(Y_i)| \leq m\} = 1$, each i, where $m < \infty$. Then for $t > 0$, and for all $n = 1, 2 \cdots$,

$$Pr \left\{ \left| \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} E(Y_i) \right| \geq nt \right\} \leq 2 \exp \left\{ -\frac{n^2 t^2}{2 \sum_{i=1}^{n} \text{Var}(Y_i) + \frac{7}{3} mn t} \right\}. \quad (2.39)$$

2.5.1 Proof of \( \hat{f}^2 = f^2 + o_p(1) \)

This is statement (2.13) on page 15.

By the triangle inequality, with $\| \cdot \|$ denoting the $L^2$ metric for functions,

$$\| \hat{f} - f \| \leq \| \hat{f} - f \|.$$ Squaring both sides, and taking expectations, we see that

$$E \left( \| \hat{f} - f \| \right)^2 \leq \int E(\hat{f} - f)^2 \rightarrow 0,$$

which implies that $\| \hat{f} \| \rightarrow \| f \|$ in probability, or equivalently that $\int \hat{f}^2 \rightarrow \int f^2$ in probability.

2.5.2 Proof of \( \int \hat{f}^2 1(\hat{f} > 0) = \int f^2 + o_p(1) \)

This is statement (2.14) on page 15.

We use a result from Theorem 2.3 which is proved later. In that theorem, it is shown that $\int f^2 1(\hat{f} < 0) = o_p \{(nh)^{-1}\}$. We write the left side of (2.14) as

$$\int \hat{f}^2 1(\hat{f} > 0) = \int \hat{f}^2 - \int \hat{f}^2 1(\hat{f} < 0),$$
We now investigate the factor $Pr(\hat{f} < 0)$. We have that

$$\sup_j Pr(\hat{f} < 0) \leq \sup_j \left\{ Pr \left( |\hat{f} - E\hat{f}| > |E\hat{f}| \right) \right\}$$

and we can manipulate the RHS of this inequality using Bernstein’s inequality. To identify with the terms in our definition (2.39) on page 41, we write

$$Y_i(x) = (nh)^{-1}K\{h^{-1}(x - X_i)\}$$

so that $\hat{f}(x) = \sum_{i=1}^{n} Y_i(x)$ and $E\{\hat{f}(x)\} = E \{ \sum_{i=1}^{n} Y_i(x) \}$. Because the $X_i$’s are independent and identically distributed, so to are the $Y_i$’s and $E\{\hat{f}(x)\} = \sum_{i=1}^{n} E\{Y_i(x)\}$.

Under the i.i.d assumptions, we have that $E(\hat{f}) = nE(Y_1)$ which is identifiable with the term $nt$ in (2.39) but we retain $E(\hat{f})$ in the role of $nt$ for convenient evaluation of Bernstein’s inequality. In this notation,

$$\sup_j \left\{ Pr \left( |\hat{f} - E\hat{f}| > |E\hat{f}| \right) \right\} = \sup_j \left[ Pr \left\{ \left| \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} E(Y_i) \right| > E(\hat{f}) \right\} \right].$$

We require a bound for $|Y_i - E(Y_i)|$, designated $m$ in (2.39). Since $|K| \leq C_2$,

$$|Y_i - E(Y_i)| \leq |Y_i| + |E(Y_i)| \leq 2(nh)^{-1}C_2.$$

Thus by Bernstein’s inequality,

$$\sup_j \left\{ Pr \left( |\hat{f} - E\hat{f}| > |E\hat{f}| \right) \right\} \leq 2 \sup_j \left( \exp \left[ -\frac{|E\hat{f}|^2}{2 \sum_{i=1}^{n} \text{var}(Y_i) + \frac{2}{3}(2(nh)^{-1}C_2 |E\hat{f}|)} \right] \right).$$

Since the $Y_i$ are i.i.d., $\sum_{i=1}^{n} \text{var}(Y_i) = \text{var}(\hat{f})$. 

and by applying Theorem 2.3 and (2.13) we get (2.14).

2.5.3 Proof of \( f(\hat{f} - f)fI(\hat{f} < 0) = o_p\{(nh)^{-1} + h^{2r}\} \)

This is statement (2.15) on page 15.

Let \( c > 0 \) be any constant, and write \( J = \{x : f(x) > c\} \). We consider the integration in the left side of (2.15) for the ranges of \( x \) given by \( J \) and its complement, \( \tilde{J} \). By Hölder’s inequality, \(^2\)

\[
E\left| \int_J (\hat{f} - f)fI(\hat{f} < 0) \right| \leq \int_J E\left| (\hat{f} - f)fI(\hat{f} < 0) \right| \\
\leq \int_J \left[ E\left\{(\hat{f} - f)^2\right\}\right]^{1/2} fPr(\hat{f} < 0)^{1/2}.
\]

We assume (see (2.21)) that for kernel function \( K, |K| \leq C_2 \) (a constant) so that \( |\hat{f} - EF| \leq (nh)^{-1}C_2 \). Using Bernstein’s inequality (see (2.39) on page 44) we have that,

\[
\sup_J Pr(\hat{f} < 0) \leq \sup_{J} \left\{ \Pr\left( |\hat{f} - EF| > |EF| \right) \right\} \\
\leq 2\sup_{J} \left[ \exp\left\{ \frac{|EF|^2}{2\text{var} f + \frac{3}{2}(nh)^{-1}C_2|EF|} \right\} \right].
\]

Expressions for the expectation and variance of the kernel estimator are given by

\[
EF = f + (k_r/r!)f^{(r)}h^r + o(h^r) \quad \text{and}
\]
\[
\text{var}\hat{f} = (nh)^{-1}f \int K(t)^2dt + O(n^{-1})
\]

(see Silverman [23, p39]). Thus,

\[
\sup_J Pr(\hat{f} < 0) = o\{(nh)^{-1} + h^{2r}\}
\]

\(^2\)Hölder’s inequality: \( E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}, 1 < p < \infty, 1 < q < \infty, (1/p)+(1/q) = 1.\)
and with result (2.12) for the integrated mean square error, we have that

\[ E \left| \int (\hat{f} - f) fI(\hat{f} < 0) \right| = o\left\{ (nh)^{-1} + h^2 \right\} . \]

Since \( E| (\hat{f} - f) fI(\hat{f} < 0)| \leq E(\hat{f} - f)^2 \) then,

\[ E \left| \int (\hat{f} - f) fI(\hat{f} < 0) \right| \leq \int E(\hat{f} - f)^2 . \]

For any \( \epsilon > 0 \) the right-hand side may be made less than \( \epsilon\{ (nh)^{-1} + h^2 \} \), for all sufficiently large \( n \), by choosing \( c \) sufficiently small. The bounds for the integral over the ranges \( J \) and \( \tilde{J} \) give us the result (2.15).

### 2.5.4 Proof of \( f(\hat{f} - E\hat{f}) = o_p\{ (nh)^{-1/2} \} \)

This is statement (2.16) on page 15.

We first expand the left-hand side as a series of independent variables. That is,

\[
\int (\hat{f} - E\hat{f}) f = \int \left[ (nh)^{-1} \sum_{i=1}^{n} K \left\{ (x - X_i)/h \right\} - M \right] f \quad (M = E\hat{f})
\]

\[
= \frac{1}{n} \sum_{i} \left[ h^{-1} \int K \left\{ (x - X_i)/h \right\} f(x) dx - \int M f \right]
\]

\[
= \frac{1}{n} \sum_{i} \left\{ \int K(y)f(X_i - hy)dy - \int M f \right\} .
\]

By squaring both sides and taking expectations we have,

\[
E \left\{ \int (\hat{f} - E\hat{f}) f \right\}^2 \quad = \quad n^{-2} E \left[ \sum_{i} \left\{ \int K(y)f(X_i - hy)dy - \int M f \right\}^2 \right]
\]

\[
= \quad n^{-2} \sum_{i} \left[ E \left\{ \int K(y)f(X_i - hy)dy - \int M f \right\}^2 \right]
\]

(since the \( X_i \) are independent )

\[
= \quad n^{-1} E \left\{ \int K(y)f(X_1 - hy)dy - \int M f \right\}^2 .
\]

43
The kernel function is bounded with compact support and the expectation of the squared term gives a real non-negative number. Therefore,

\[ E \left\{ \int (\hat{f} - Ef)f \right\}^2 = n^{-1} B \quad \text{ (B is a constant)} \]

\[ = \left( nh \right)^{-1} Bh. \]

Since \( h \to 0 \) and \( nh \to \infty \) as \( n \to \infty \), we have

\[ \int (\hat{f} - Ef)f = o_p \left\{ (nh)^{-1/2} \right\}, \]

which is (2.16).

2.5.5 Proof of \( \int (Ef - f) f = -\frac{1}{2} A_4 h^r + o(h^r) \)

This is statement (2.17) on page 15. The bias, \( Ef - f \), has been shown (see (1.3)) to be \( k_r h^r f^{(r)} + o(h^r) \). (Recall that in (2.8) on page 13 we assume \( r \) to be even.) Since this is valid uniformly in the argument of \( \hat{f} \) and \( f \) then,

\[ \int (Ef - f)f = \int k_r h^r f^{(r)} f + o(h^r). \]

After integrating by parts \( r/2 \) times,

\[ \int (Ef - f)f = (-1)^{r/2} k_r \int f^{(r/2)} h^r + o(h^r) \]

\[ = -\frac{1}{2} A_4 h^r + o(h^r), \]

which proves (2.17).

In the following proofs, the symbols \( B_1, B_1^*, B_2, \ldots \) represent generic constants.
2.5.6 Proof of Theorem 2.1

Without loss of generality, we can assume that the right-hand tail of the density provides the essential information about $\gamma$ and prove that as $n \to \infty$,

$$\gamma = \int_0^{+\infty} \hat{f}(x) I \{ \hat{f}(x) < 0 \} \, dx = (nh)^{-1+1/\alpha_1} I(0) + o_p \{(nh)^{-1+1/\alpha_1}\} \quad (2.40)$$

where

$$I(\epsilon) = \int_{\epsilon}^{\epsilon^{-1}} E[Z(v)I\{Z(v) < 0\}] \, dv \quad 0 < \epsilon < 1. \quad (2.41)$$

The random variable $Z(v)$ is defined (see (2.22) on page 18) by having characteristic function $\psi_\epsilon(t) = \exp \{-cv^{-\alpha}\beta(t)\}$ and as $\epsilon \to 0$, $I(\epsilon) \to A_5$; $A_5$ being previously defined at (2.23) on page 18 and (2.25) on page 19.

To establish the result at (2.40), we consider in succession two ranges of $x$:

$$R_1(\epsilon) = \{x : 0 \leq x \leq \epsilon(nh)^{1/\alpha_1} \text{ or } x > \epsilon^{-1}(nh)^{1/\alpha_1}\},$$

$$R_2(\epsilon) = \{x : \epsilon(nh)^{1/\alpha_1} < x \leq \epsilon^{-1}(nh)^{1/\alpha_1}\}, \quad 0 \leq \epsilon < 1.$$

These ranges (illustrated in Figure (2.8)) are chosen because it isn’t at first clear what happens in the range $(0, \infty)$ so $R_1(\epsilon)$ and $R_2(\epsilon)$ are used to define limits where $\hat{f}$ can be described. Then $R_1(\epsilon)$ is allowed to shrink and $R_2(\epsilon)$ stretch until the limits of $(0, \infty)$ are obtained. In the first part of the proof, the values of $x$ are those where it can be expected that $\hat{f}(x) \geq 0$ and there is no contribution to $\gamma$. The theory of this part can be developed directly from the definition of $\gamma$ (defined at (2.7)) and does not involve $A_5$. This range will contain the bulk of the internal points and the extreme range of the tails.
It is proved that,
\[
\lim_{\epsilon \to 0} \limsup_{n \to \infty} (nh)^{1-1/\alpha_1} \int_{R_1(\epsilon)} E \left[ |\hat{f}(x)| I \{ \hat{f}(x) < 0 \} \right] dx = 0 .
\]

The second range of \( x \) contains those values where it can be expected that \( \hat{f}(x) < 0 \) and it is proved, using \( Z(v) \), that by taking \( \epsilon > 0 \) sufficiently small, the value of
\[
(nh)^{1-1/\alpha_1} \int_{R_2(\epsilon)} E \left[ |\hat{f}(x)| I \{ \hat{f}(x) < 0 \} \right] dx
\]
can be made arbitrarily close to \( I(0) \) for all sufficiently large \( n \).

**Range 1: \( R_1(\epsilon) \)**

We manipulate the expression for bias to to get useful terms for Bernstein's inequality which is then used to put a bound on \( Pr \{ f(x) < 0 \} \) for later use in bounding \( \gamma \). Put \( K_0 = K \) and \( K_j(u) = \int_{u>u} K_{j-1}(v) dv, \ j \geq 1 \). Using integration by parts,
\[
E \hat{f}(x) - f(x) = \int_{-\infty}^{\infty} \{ f(x - hy) - f(x) \} K(y) dy
\]
\[
= \int_{-\infty}^{\infty} f'(x - hy) hK_1(y) dy .
\]
Integrating by parts $r - 1$ more times,

$$E\hat{f}(x) - f(x) = h^r \int_{-\infty}^{\infty} f^{(r)}(x - hy)K_r(y)dy.$$ 

From the definition of the density at (2.20), for $x > C_1$ (a constant),

$$|f^{(r)}(x)| \leq B^r_1(1 + x)^{-(\alpha_1 + r)}.$$ 

Assuming that $h \leq 1$,

$$|E\hat{f}(x) - f(x)| \leq B^r_1 h^r \int_{-\infty}^{\infty} (1 + |x + hy|)^{-(\alpha_1 + r)} |K_r(y)|dy \text{, for } x > C_2$$

$$\leq B_1 h^r(1 + x)^{-(\alpha_1 + r)}. \quad (2.42)$$

(Note that $K_r$ vanishes outside a compact set.)

In the next section, $D_1$, $D_2$ and $D$ are constants.

Since $f(x) \sim c_1 x^{-\alpha_1}$ and $f$ is bounded, then $x^{\alpha_1}f(x) \leq D_1$.

For $x \geq 1$,

$$f(x) \leq \frac{D_1}{x^{\alpha_1}} \leq \frac{D_1}{\left(\frac{1}{2}(1 + x)\right)^{\alpha_1}} = \frac{D_1 2^{\alpha_1}}{(1 + x)^{\alpha_1}}.$$ 

For $x < 1$,

$$f(x) \leq D_2 \leq D_2 \left(\frac{2}{1 + x}\right)^{\alpha_1} = \frac{D_2 2^{\alpha_1}}{(1 + x)^{\alpha_1}}.$$ 

We choose $D = \max(D_1 2^{\alpha_1}, D_2 2^{\alpha_1})$, so that $f(x) \leq D(1 + x)^{-\alpha_1}$. We can now substitute these results into the relationship $E\hat{f}(x) \leq f(x) + |E\hat{f}(x) - f(x)|$ to give,

$$E\hat{f}(x) \leq D(1 + x)^{-\alpha_1} + B_1 h^r(1 + x)^{-(\alpha_1 + r)}$$

$$= (1 + x)^{-\alpha_1} \{D + B_1 h^r(1 + x)^{-r}\}$$

$$= B_2(1 + x)^{-\alpha_1}. \quad (2.43)$$

47
Also, 

\[ E\hat{f}(x) \geq f(x) - |E\hat{f}(x) - f(x)| \]

\[ \geq (1 + x)^{-\alpha_1} - B_1 h^r (1 + x)^{-(\alpha_1 + r)} \]

\[ \geq B_2^*(1 + x)^{-\alpha_1} \quad (2.44) \]

for all \( x \geq 0, \ 0 < h \leq h_0 \) (\( h_0 \) a constant) and where \( B_2^* < B_2 \). We restate an earlier result that, 

\[ \text{var} \hat{f}(x) = (nh)^{-1} \int f(x - hy)K(y)^2dy - n^{-1} \{f(x) + \text{bias}\}^2 \]

\[ \leq (nh)^{-1} \int f(x - hy)K(y)^2dy. \]

Since \( K \) vanishes outside \((-C_1, C_1)\), and \( f(x) \leq B_3^*(1 + |x|)^{-\alpha_1} \) for \( x > -1 \), then 

\[ \text{var} \{\hat{f}(x)\} \leq B_3(nh)^{-1}(1 + x)^{-\alpha_1}. \quad (2.45) \]

We have the identity, 

\[ Pr \{\hat{f}(x) < 0\} = Pr \{\hat{f}(x) - E\hat{f}(x) < -E\hat{f}(x)\} \]

\[ = Pr \{E\hat{f}(x) - \hat{f}(x) > E\hat{f}(x)\}, \]

to which we apply Bernstein's inequality (as in 2.5.3 on page 42) so that, 

\[ Pr \{\hat{f}(x) < 0\} \leq 2\exp \left[ -\frac{\{E\hat{f}(x)\}^2}{2\{\text{var}\hat{f}(x) + \frac{2}{3}E\hat{f}(x)(nh)^{-1}C_2\}} \right] \]

\[ \leq 2\exp \left[ -\frac{\{B_2(1 + x)^{-\alpha_1}\}^2}{2\{B_3(nh)^{-1}(1 + x)^{-\alpha_1} + \frac{2}{3}B_2(1 + x)^{-\alpha_1}(nh)^{-1}C_2\}} \right] \]

\[ = 2\exp \left[ -B_4(nh)(1 + x)^{-\alpha_1} \right], \quad (2.46) \]
where $B_4 = \frac{1}{2}B_2^2(B_3 + \frac{2}{3}C_2)^{-1}$. Observe that, if $K$ vanishes outside $(-C_1, C_1)$,

$$E[|\hat{f}(x)|I\{\hat{f}(x) < 0\}] = E\left((nh)^{-1} \sum_i K \{h^{-1}(x - X_i)\} I\{\hat{f}(x) < 0; |x - X_i| \leq C_1 h\}\right) \leq (nh)^{-1}C_2E\left(\sum_i I\{\hat{f}(x) < 0; |x - X_i| \leq C_1 h\}\right)$$

$$= (nh)^{-1}C_2 \sum_{k=1}^{n} k Pr\{\hat{f}(x) < 0; |x - X_i| \leq C_1 h, \text{ for precisely } k \text{ values of } i, 1 \leq i \leq n\}$$

$$\leq (nh)^{-1}C_2 \left[Pr\{\hat{f}(x) < 0\}\right]^{\frac{1}{2}} \sum_{k=1}^{n} k \left\{\binom{n}{k} p^k(1-p)^{n-k}\right\}^{\frac{1}{2}}, \quad (2.47)$$

where $p = Pr(|x - X_i| \leq C_1 h) \leq B_5 h(1 + x)^{-a_1}$. To estimate the series on the right-hand side, note that

$$\sum_{k=1}^{n} k \left\{\binom{n}{k} p^k(1-p)^{n-k}\right\}^{\frac{1}{2}} = \sum_{k=1}^{n} k \left\{\frac{n!}{(n-k)!k!} \left(\frac{p}{1-p}\right)^k (1-p)^n\right\}^{\frac{1}{2}},$$

and that by Stirling's formula,

$$(k!)^{\frac{1}{2}} \sim (2\pi)^{\frac{1}{4}} k^{(k+\frac{1}{2})} e^{-k} \sim B_6 \left(k^{(k+\frac{1}{2})} e^{-k}\right)^{\frac{1}{2}}$$

$$= B_6 \left(\frac{k}{2}\right)^{(\frac{k}{2}+\frac{1}{2})} e^{-\frac{k}{2}} \left(\frac{1}{2}\right)^{(\frac{k}{2}+\frac{1}{2})} k^{-\frac{1}{4}}$$

$$\geq B_7 \Gamma\left(\frac{k}{2} + 1\right)2^{\frac{3}{4}} k^{-\frac{1}{4}}.$$ 

Substituting $\lambda = np/(1 - p)$ and using $\binom{n}{k} = n!/(k!(n-k)!) \leq n^k/k!$ gives,

$$\binom{n}{k} p^k(1-p)^{n-k} \leq \frac{1}{k!} \lambda^k (1-p)^n \leq \frac{1}{k!} \lambda^k e^{-np} \quad \text{since } (1-p) \leq e^{-p}.$$
Therefore,

\[
\sum_{k=1}^{n} k \left\{ \binom{n}{k} p^k (1-p)^{n-k} \right\}^{\frac{1}{2}} \leq B_7^{-1} \sum_{k=1}^{n} k^{\frac{k}{2}} \frac{\left( \frac{1}{2} \lambda \right)^k}{\Gamma(k+1)} e^{-\frac{1}{2}np} \leq B_7^{-1}\left\{ \sum_{k=1}^{\infty} (2k)^{\frac{k}{2}} \frac{\left( \frac{1}{2} \lambda \right)^k}{\Gamma(k+1)} + \sum_{k=1}^{\infty} (2k-1)^{\frac{k}{2}} \frac{\left( \frac{1}{2} \lambda \right)^k}{k!} \right\} e^{-\frac{1}{2}np^2/(2(1-p))}
\]

Let \( V \) be Poisson with mean \( \lambda/2 \). Then \( E(V^2) = \sum_{k=1}^{\infty} \frac{k^2}{k!} \left( \frac{1}{2} \lambda \right)^k e^{-\frac{1}{2}k/2} \) and

\[
E(V^2) = \text{var}(V) + E(V)^2 = \frac{1}{2} \lambda + \left( \frac{1}{2} \lambda \right)^2 = \frac{1}{2} \lambda (1 + \frac{1}{2} \lambda).
\]

Therefore,

\[
\sum_{k=1}^{n} k \left\{ \binom{n}{k} p^k (1-p)^{n-k} \right\}^{\frac{1}{2}} \leq B_8 \lambda (1 + \lambda) e^{np^2/(2(1-p))}. \tag{2.48}
\]

Combining results (2.46), (2.47) and (2.48),

\[
E \left[ |f(x)| I \{ f(x) < 0 \} \right] \leq (nh)^{-1} C_2 \left[ \exp \left\{ -B_4 nh(1 + x)^{\alpha_1} \right\} \right]^{\frac{1}{2}} \left[ B_8 \lambda (1 + \lambda) \exp \left\{ \frac{1}{2} np^2/(1 - p) \right\} \right].
\]
We use the expressions for $\lambda$ and $p$ to write,

$$\lambda(1+\lambda) \leq 2\max(\lambda, \lambda^2)$$

$$\leq C_2 \max \left\{ nh(1+x)^{-\alpha_1}, (nh)^2(1+x)^{-2\alpha_1} \right\}$$

and $np^2/(1-p) = p\lambda \leq \sup(p)\sup(\lambda)$.

Therefore,

$$E \left[ |\hat{f}(x)| \mathbb{1} \{ \hat{f}(x) < 0 \} \right]$$

$$\leq (nh)^{-1} C_2 \exp \left\{ -\frac{1}{2} B_4 nh(1+x)^{-\alpha_1} + \frac{1}{2} C_1 B_5 nh^2(1+x)^{-2\alpha_1} \right\}$$

$$\times \max \left\{ nh(1+x)^{-\alpha_1}, (nh)^2(1+x)^{-2\alpha_1} \right\}$$

$$\leq B_6 (nh)^{-1} \exp \left\{ -\frac{1}{4} B_4 (nh)(1+x)^{-\alpha_1} \right\}$$

$$\times \max \left\{ nh(1+x)^{-\alpha_1}, (nh)^2(1+x)^{-2\alpha_1} \right\}, \quad (2.49)$$

provided $x > 0$ and $C_1 B_5 h \leq \frac{1}{4} B_4$ or that $h \leq \frac{1}{4} B_4/C_1 B_5$ (i.e. $h$ is small).

Integrating both sides of (2.49) with respect to $x$ over $x \in R_1(\epsilon)$,

$$\int_{R_1(\epsilon)} E \left[ |\hat{f}(x)| \mathbb{1} \{ \hat{f}(x) < 0 \} \right] dx$$

$$\leq \int_{R_1(\epsilon)} B_6 (nh)^{-1} \exp \left\{ -\frac{1}{4} B_4 (nh)(1+x)^{-\alpha_1} \right\}$$

$$\times \max \left\{ nh(1+x)^{-\alpha_1}, (nh)^2(1+x)^{-2\alpha_1} \right\} dx.$$

We now substitute $(\frac{1}{4} B_4 nh)^{1/\alpha_1} v = (1+x)$, absorb necessary constants into $B_6^\downarrow, B_6^+$ and let $\epsilon$ approach zero so that

$$\int_{R_1(\epsilon)} E \left[ |\hat{f}(x)| \mathbb{1} \{ \hat{f}(x) < 0 \} \right] dx$$

$$\leq B_6^\downarrow (nh)^{-1+1/\alpha_1} \left( \int_0^{B_6^{-1}\epsilon} + \int_{B_6^{\alpha_1}\epsilon}^{\infty} \right) \exp (-v^{-\alpha_1}) \max (v^{-\alpha_1}, v^{-2\alpha_1}) dv.$$
Hence,
\[
\lim_{\epsilon \to 0} \limsup_{n \to \infty} (nh)^{1-(1/\alpha_1)} \int_{R_1(\epsilon)} E \left[ |\hat{f}(x)| I \{ \hat{f}(x) < 0 \} \right] dx = 0. \tag{2.50}
\]

Range 2: \(R_2(\epsilon)\)

To investigate \(\int_{R_2(\epsilon)} |\hat{f}(x)| I \{ \hat{f}(x) < 0 \} dx\), we define a random variable \(Z(v)\) that has characteristic function \(\psi_\epsilon(t) = \exp \left\{ -c_1 v_1^{-\alpha_1} \beta(t) \right\}\), \(\beta(t) = \int (1 - e^{itK(y)}) dy\). We employ two transformations, \(x = (nh)^{1/\alpha_1} v\), \(v > 0\) and \(Z(n, v) = \sum_{j=1}^{n} K \{ h^{-1}(x - X_j) \}\). That is, \(Z(n, v(x)) = nh\hat{f}(x)\). With the change of variable we have that,
\[
\int_{\epsilon(nh)^{1/\alpha_1}}^{\epsilon^{-1}(nh)^{1/\alpha_1}} \hat{f}(x) I \{ \hat{f}(x) < 0 \} dx = (nh)^{-1+1/\alpha_1} \int_{\epsilon}^{\epsilon^{-1}} Z(n, v) I \{ Z(n, v) < 0 \} dv. \tag{2.51}
\]

Let
\[
I_n(\epsilon) = \int_{\epsilon}^{\epsilon^{-1}} Z(n, v) I \{ Z(n, v) < 0 \} dv,
\]
and
\[
I(\epsilon) = \int_{\epsilon}^{\epsilon^{-1}} E[Z(v) I \{ Z(v) < 0 \}] dv.
\]

We prove that \(I_n(\epsilon) \xrightarrow{p} I(\epsilon)\) as \(n \to \infty\) and then let \(\epsilon \to 0\) to achieve the desired result. The proof utilises the expected value of the squared difference of the integrals given by
\[
E \{ I_n(\epsilon) - I(\epsilon) \}^2 = \text{var} \{ I_n(\epsilon) \} + \{ E I_n(\epsilon) - I(\epsilon) \}^2, \tag{2.52}
\]
and proceeds by showing that:

(i) \(|E \{ I_n(\epsilon) \} - I(\epsilon)| \to 0\) as \(n \to \infty\),

52
(ii) \( \text{var} \{ I_n(\epsilon) \} \to 0 \) as \( n \to \infty \),

(iii) hence \( E \{ I_n(\epsilon) - I(\epsilon) \}^2 \to 0 \) as \( n \to \infty \) and

(iv) \( I_n(\epsilon) \to I(\epsilon) \) in probability as \( n \to \infty \).

The result (i) comes in part from showing the convergence of the characteristic function of \( Z(n, v) \) to the characteristic function of \( Z(v) \), which we do now. The characteristic function of \( Z(n, v) \) is

\[
\psi_{vn}(t) = E \left[ \exp \left\{ it \sum_{j=1}^{n} K \left( \frac{x - X_j}{h} \right) \right\} \right] \\
= \left( E \left[ \exp \left\{ it K \left( \frac{x - X_1}{h} \right) \right\} \right] \right)^n \\
= \{ \psi_{v1}(t) \}^n.
\]

Now,

\[
1 - E \left[ \exp \left\{ it K \left( \frac{x - X_1}{h} \right) \right\} \right] \\
= \int \left[ 1 - \exp \left\{ it K \left( \frac{x - z}{h} \right) \right\} \right] f(z)dz \\
= \int \left[ 1 - \exp \{ itK(y) \} \right] f(x + hz)hdz \quad (y = h^{-1}(x - z)) \\
\sim h c_1 x^{-\alpha_1} \int \left[ 1 - \exp \{ itK(y) \} \right] dy \quad (f(x) \sim c_1 x^{-\alpha_1}) \\
= h c_1 x^{-\alpha_1} \beta(t) \\
= n^{-1} c_1 v^{-\alpha_1} \beta(t) \quad (x = (nh)^{1/\alpha_1} v).
\]

Therefore, \( \psi_{v1}(t) \sim 1 - n^{-1} c_1 v^{-\alpha_1} \beta(t) \) and

\[
\psi_{vn}(t) \sim \left\{ 1 - n^{-1} c_1 v^{-\alpha_1} \beta(t) \right\}^n \to \exp \left\{- c_1 v^{-\alpha_1} \beta(t) \right\} = \psi_v(t).
\]
The convergence theorem for characteristic functions gives

\[ \psi_{un}(t) \to \psi_{v}(t) \Rightarrow Z(n,v) \overset{D}{\to} Z(v) \]

and

\[ Z(n,v)I\{Z(n,v) < 0\} \overset{D}{\to} Z(v)I\{Z(v) < 0\}. \quad (2.53) \]

By Rosenthal's inequality (Hall and Heyde [14, p.23]), for integers \( k \geq 2 \) and constants \( A_1, A_2, A_3 \) not depending on \( h, n \) or \( v \) (but depending on \( \epsilon \)),

\[
E |Z(n,v)|^k \leq A_1 \left[ |E Z(n,v)|^k + \{ \text{var} Z(n,v) \}^{k/2} \right] \\
\leq A_2 \left[ \left\{ nh(1+x)^{-a_1} \right\}^k + \left\{ nh(1+x)^{-a_1} \right\}^{k/2} \right] \\
\leq A_3. \quad (2.54)
\]

Chung [5, p95, Theorem 4.5.2] proves that if \( X_n \overset{D}{\to} X \), and for some \( p > 0 \), \( \sup_n E(|X|^p) = M < \infty \), then for each \( r < p \), \( \lim_{n \to \infty} E(|X_n|^r) = E(|X|^r) < \infty \).

With (2.53) and (2.54) we apply this theorem to give

\[
\lim_{n \to \infty} E[Z(n,v)I\{Z(n,v) < 0\}] = E[Z(v)I\{Z(v) < 0\}]. \quad (2.55)
\]

Furthermore, with the bound given at (2.54) and the convergence established by (2.55), we have via the dominated convergence theorem (see Chung [5, p42]) that \( |EI_n(\epsilon) - I(\epsilon)| \to 0 \) as \( n \to 0 \) and have established point (i) above.

To examine the variance of \( I_n(\epsilon) \) we consider

\[
Z(n,v_1) = \sum_{i=1}^{n} K \left\{ \frac{(nh)^{1/a_1} v_1 - X_i}{h} \right\} \quad \text{and} \\
Z(n,v_2) = \sum_{i=1}^{n} K \left\{ \frac{(nh)^{1/a_1} v_2 - X_i}{h} \right\}.
\]
Recalling that $K$ vanishes outside $(-C_1, C_1)$, we can see that they are independent provided only that $|v_1 - v_2| > 2C_1 h(nh)^{-1/\alpha_1}$. As $n \to \infty$, $h \to 0$, $nh \to \infty$ and for $\alpha_1 > 0$, we have that $2C_1 h(nh)^{-1/\alpha_1} \to 0$, implying that $Z_n(v_1)$ and $Z_n(v_2)$ are asymptotically independent. For a simpler notation, we put $Y_n(v) = Z(n, v) I \{Z(n, v) < 0\}$ and $Y(v) = Z(v) I \{Z(v) < 0\}$. In this notation,

$$\text{var} \{ \mathcal{I}_n(\epsilon) \} = \mathbb{E} \{ \mathcal{I}_n(\epsilon) - E \mathcal{I}_n(\epsilon) \}^2$$

$$= \int_\epsilon^{\epsilon} \int_\epsilon^{\epsilon} \mathbb{E} \left\{ Y_n(v_1) - EY_n(v_1) \right\} \{Y_n(v_2) - EY_n(v_2)\} dv_1 dv_2$$

$$= \int_\epsilon^{\epsilon} \mathbb{E} \left\{ Y_n(v_1) - EY_n(v_1) \right\} dv_1 \int_\epsilon^{\epsilon} \mathbb{E} \left\{ Y_n(v_2) - EY_n(v_2)\right\} dv_2 + o(1)$$

($Z_n(v_1)$ and $Z_n(v_2)$ are asymptotically independent)

$$= o(1).$$

Both parts of the right side of (2.52) converge to zero as $n \to \infty$, so that

$$E \{ \mathcal{I}_n(\epsilon) - \mathcal{I}(\epsilon) \}^2 \to 0 \text{ and } \mathcal{I}_n(\epsilon) \overset{p}{\to} \mathcal{I}(\epsilon) \text{ as } n \to \infty.$$

From (2.51),

$$(nh)^{1-(1/\alpha_1)} \int_{R_2(\epsilon)} \hat{f}(x) I \left\{ \hat{f}(x) < 0 \right\} dx \to \mathcal{I}(\epsilon) \text{ as } n \to \infty.$$

By the same arguments leading to (2.50) on page 52, $\mathcal{I}(0) < \infty$ and

$$\lim_{\epsilon \to 0} |\mathcal{I}(0) - \mathcal{I}(\epsilon)| = 0.$$

Therefore,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \left| (nh)^{1-(1/\alpha_1)} \int_{R_2(\epsilon)} \hat{f}(x) I \left\{ \hat{f}(x) < 0 \right\} dx - \mathcal{I}(0) \right| = 0. \quad (2.57)$$

From results (2.50) and (2.57), we have the result (2.40) and have completed the proof of the theorem.

For the more general case where the right tail of the density is described by $f(x) = L_1(x)x^{-\alpha_1}, \alpha_1 > 0$ as $x \to \infty$ and $L_1$ is a slowly varying function, we require
a transformation so that \( \psi_n(t) \sim \{1 - hL_1(x)x^{-\alpha_1}\beta(t)\}^n \) can be transformed to give 
\[
\psi_n(t) \sim \{1 - n^{-1}v^{-\alpha_1}\beta(t)\}^n.
\]
Let 
\[
x = nhL_1\left\{(nh)^{1/\alpha_1}\right\}^{1/\alpha_1}v.
\]
Then 
\[
hL_1(x)x^{-\alpha_1} = n^{-1}v^{-\alpha_1}\left[L_1\left\{(nh)^{1/\alpha_1}\right\}\right]^{-1}L_1\left[\left\{(nh)^{1/\alpha_1}\right\}\right]^{1/\alpha_1}v
\approx n^{-1}v^{-\alpha_1}.
\]

The slowly varying nature of \( L_1 \) allows the cancellation of 
\[
L_1\left[\left\{(nh)^{1/\alpha_1}\right\}\right]^{1/\alpha_1}v \text{ by } \left[L_1\left\{(nh)^{1/\alpha_1}\right\}\right]^{-1}.
\]
If we use the transformation 
\[
x = (nh)^{1/\alpha_1}L_3\{(nh)^{1/\alpha_1}\}v,
\]
with \( L_3 \) being another slowly varying function and equal to \( L_1^{1/\alpha_1} \), and we redefine the limits for \( R_2(\epsilon) \) as \( \epsilon(nh)^{1/\alpha_1}L_3\{(nh)^{1/\alpha_1}\} \) and \( \epsilon^{-1}(nh)^{1/\alpha_1}L_3\{(nh)^{1/\alpha_1}\} \), the proof for this case follows along the same lines as previously to give 
\[
(nh)^{1-(1/\alpha_1)}\left[L_3\{(nh)^{1/\alpha_1}\}\right]^{-1}\int_{R_2(\epsilon)}\hat{f}(x)I\{\hat{f}(x) < 0\}dx \to A_3 \text{ as } \epsilon \to 0.
\]
The limits for \( R_1(\epsilon) \) would be adjusted accordingly and that part of the proof unaltered.

### 2.5.7 Proof of Theorem 2.2

Suppose we can prove that for each \( \epsilon > 0 \),
\[
\int_{x > Y_2} \left|\hat{f}(x)\right|dx = O_p\{(nh)^{-1+(1/\alpha_1)+\epsilon}\},
\]
and
\[
\int_{x < Y_1} \left|\hat{f}(x)\right|dx = O_p\{(nh)^{-1+(1/\alpha_2)+\epsilon}\}.
\]
(It suffices to derive the former result as a proof of the latter is similar). Then, with
\( \alpha_0 = \min(\alpha_1, \alpha_2) \) and \( \varepsilon > 0 \),
\[
\gamma_3^{-1} = \int_{Y_1 \leq z \leq Y_2} \hat{f}(x) dx = \int_{-\infty}^{\infty} \hat{f} + O_p \left\{ (nh)^{-1+1/(\alpha_0)+\varepsilon} \right\}
= 1 + O_p \left\{ (nh)^{-1+1/(\alpha_0)+\varepsilon} \right\}.
\]

Let \( \alpha^{(i)} > \alpha_i \) for \( i = 1, 2 \). We designate the range of \( x \) given by \( x < -(nh)^{1/\alpha^{(i)}} \)
or \( x > (nh)^{1/\alpha^{(2)}} \) as \( \mathcal{R} \). It is straightforward to show that
\[
\int_{\mathcal{R}} \hat{f}(x)^2 dx = O \left[ \sum_{i=1}^{2} \left\{ (nh)^{1/\alpha^{(i)}} \right\}^{1-2\alpha_i} \right] = o \left\{ (nh)^{-1} \right\},
\]
provided each \( \alpha^{(i)} \) is chosen sufficiently close to \( \alpha_i \) (and exceeds 1). Therefore, for
such values of \( \alpha^{(1)} \) and \( \alpha^{(2)} \),
\[
\int_{\mathcal{R}} E \left\{ \hat{f}(x)^2 \right\} dx \leq 2 \int_{\mathcal{R}} \left[ E \left\{ \hat{f}(x) - f(x) \right\}^2 + f(x)^2 \right] dx
= o_p \left\{ (nh)^{-1} + h^{2\gamma} \right\}.
\]
Hence, if we prove in addition that for each \( \alpha^{(i)} > \alpha_i \), \( Pr \left\{ Y_i \leq (nh)^{1/\alpha^{(i)}} \right\} \to 0 \), then
\[
\left\{ \int (\hat{f}_3 - f)^2 \right\}^{1/2} - \left\{ \int (\hat{f} - f)^2 \right\}^{1/2} \leq \int (\hat{f}_3 - \hat{f})^2
= (1 - \gamma_3)^2 \int_{Y_1 \leq z \leq Y_2} \hat{f}(x)^2 dx + \left( \int_{x < Y_1} + \int_{x > Y_2} \right) \hat{f}(x)^2 dx
= O_p \left[ (1 - \gamma_3)^2 + \int_{\mathcal{R}} E \left\{ \hat{f}(x)^2 \right\} dx \right]
= O_p \left\{ (nh)^{-1+1/(\alpha_0)+\varepsilon} \right\}^2 + o_p \left\{ (nh)^{-1} + h^{2\gamma} \right\}
= o_p \left\{ (nh)^{-1} + h^{2\gamma} \right\},
\]
provided \( \varepsilon > 0 \) is sufficiently small. (Recall that \( \alpha_0 > 2 \), by assumption.) Theorem
2.2 follows from this result.
It remains to prove that if \( Y \) denotes the smallest positive \( x \) such that \( \hat{f}(x) \leq 0 \), and \( \epsilon > 0 \), then

\[
\int_{Y}^{\infty} |\hat{f}(x)| \, dx = O_p\{(nh)^{-1+1/(\alpha_1)+\epsilon}\},
\]

and that if \( \alpha > \alpha_1 \),

\[
Pr\{Y \leq (nh)^{1/\alpha}\} \to 0. \tag{2.58}
\]

In view of the Hölder continuity of \( K \), given any \( B_{11} > 0 \), we can construct a sequence \( \mathcal{X}_n = \{x_{ni}, 1 \leq i \leq m_n\} \) of equally spaced points such that

(i) \( 0 = x_{n1} < \cdots < x_{n,m_n} = (nh)^{1/\alpha_1} \)

(ii) \( m_n = O(n^{B_{12}}) \) for some \( B_{12} > 0 \), depending on \( B_{11} \) but not on \( n \)

(iii) with probability one, \( \max_{1 \leq i \leq m_n-1} \sup_{x_{ni} \leq x \leq x_{n,i+1}} |\hat{f}(x) - \hat{f}(x_{ni})| \leq n^{-B_{11}}. \)

For \( D(x) = \hat{f}(x) - Ef(x) \),

\[
|D(x) - D(x_{ni})| \leq |\hat{f}(x) - \hat{f}(x_{ni})| + E|\hat{f}(x) - \hat{f}(x_{ni})|. \tag{2.59}
\]

If we take the supremum of the differences (2.59) within each interval \((x_{ni}, x_{n,i+1})\) and then take the maximum over the intervals, we have

\[
\max_{1 \leq i \leq m_n-1} \sup_{x_{ni} \leq x \leq x_{n,i+1}} \{|D(x) - D(x_{ni})|\}
\leq \max_{1 \leq i \leq m_n-1} \sup_{x_{ni} \leq x \leq x_{n,i+1}} \{|\hat{f}(x) - \hat{f}(x_{ni})| + E|\hat{f}(x) - \hat{f}(x_{ni})|\}
\leq \max_{1 \leq i \leq m_n-1} \sup_{x_{ni} \leq x \leq x_{n,i+1}} \{|\hat{f}(x) - \hat{f}(x_{ni})|\}
+ \max_{1 \leq i \leq m_n-1} \sup_{x_{ni} \leq x \leq x_{n,i+1}} \{E|\hat{f}(x) - \hat{f}(x_{ni})|\}
\leq n^{-B_{11}} + n^{-B_{11}} \quad \text{(from (iii) above)}
\]
\[
= 2n^{-B_{11}}. 
\]
In Theorem 2.1, we obtained the result (2.44) that \( E\hat{f}(x) \geq B_2^*(1 + x)^{-\alpha_1} \) uniformly in \( x \geq 0 \). We choose \( B_{11} \) such that \((nh)^{-1}\) is a strictly larger order of magnitude than \( n^{-B_{11}} \). Then for all \( x \geq 0 \) and \( n \geq n_0 \) say,

\[
\inf_{x_{n_i} \leq x \leq x_{n_i+1}} E\hat{f}(x) - 2n^{-B_{11}} \geq \frac{1}{2} B_2^*(1 + x_{n_i})^{-\alpha_1}.
\]

Let \( i_0 \) denote the largest \( i \) such that \( x_{i_0} \leq (nh)^{1/\alpha} \).

By the definition of \( Y \), we can say that if \( Y \leq (nh)^{1/\alpha} \), then \( \hat{f}(x) < 0 \) for some \( 0 \leq x \leq (nh)^{1/\alpha} \). The inequality \( \hat{f}(x) < 0 \) can be rewritten as \( E\hat{f}(x) - \hat{f}(x) > E\hat{f}(x) \) so that

\[
Pr \{Y \leq (nh)^{1/\alpha}\} = Pr \{-D(x) > E\hat{f}(x) \text{ for some } 0 \leq x \leq (nh)^{1/\alpha}\}
\]

\[
\leq Pr \{|D(x_{ni})| > \frac{1}{2} B_2^*(1 + x_{ni})^{-\alpha_1} \text{ for some } 1 \leq i \leq i_0\}
\]

\[
\leq 2\sum_{i=1}^{i_0} \exp \{-B_{13}nh(1 + x_{ni})^{-\alpha_1}\}
\]

(using Bernstein's inequality)

\[
\leq 2m_n \max_i \{\exp \{-B_{13}nh(1 + x_{ni})^{-\alpha_1}\}\}
\]

\[
\leq 2m_n \exp \{-B_{14}nh(nh)^{-\alpha_1/\alpha}\}
\]

\[
= O(n^{B_{12}}) \exp \{-B_{14}(nh)^{1-(\alpha_1/\alpha)}\}.
\]

For each \( \alpha > \alpha_1, 1 - \alpha_1/\alpha > 0 \) and \( \exp \{-B_{14}(nh)^{1-\alpha_1/\alpha}\} \to 0 \) as \( nh \to \infty \). With this result,

\[
Pr \{Y \leq (nh)^{1/\alpha}\} \to 0 \text{ as } n \to \infty,
\]

which proves (2.58). Having showed that it is unlikely that \( Y \leq (nh)^{1/\alpha} \), the integral

\[
J_\beta \equiv \int_{(nh)^{1/\beta}}^{\infty} E \{|\hat{f}(x)|\} \, dx \quad \alpha_1 < \beta < \alpha_1 + \delta, \ \delta \text{ sufficiently small},
\]

59
will be adequate to measure the amount of the density lost by disregarding the truncated tail area. We show that \( J_\beta = O \left\{ (nh)^{-1+1/\alpha} \right\} \) for each \( \alpha > \alpha_1 \).

As in Theorem 2.1, \( p = Pr (|x - X| \leq C_1 h) \leq B_3 h (1 + x)^{-\alpha_1}, \lambda = np/(1 - p) \) and

\[
E \left\{ \hat{f}(x) \right\} \leq (nh)^{-1} C_2 \sum_{k=1}^{n} k Pr (|x - X_i| \leq C_1 h),
\]

for precisely \( k \) values of \( i, 1 \leq i \leq n \)

\[
= (nh)^{-1} C_2 \sum_{k=1}^{n} k \left( \frac{n}{p} \right) p^k (1 - p)^{n-k}
\]

\[
\leq (nh)^{-1} B_{15} \sum_{k=1}^{n} \frac{\lambda^k}{k!} e^{-np}
\]

\[
= (nh)^{-1} B_{15} \sum_{k=1}^{n} \frac{\lambda^k}{k!} e^{-\lambda} e^{\lambda p}
\]

\[
\leq B_{15} (nh)^{-1} \lambda \exp \{ \sup(p) \sup(\lambda) \}
\]

\[
\leq B_{16} (1 + x)^{-\alpha_1} \exp \left\{ B_{17} h (1 + x)^{-\alpha_1} \right\} n B_7 h (1 + x)^{-\alpha_1}
\]

\[
\leq B_{16} (1 + x)^{-\alpha_1} \exp \left[ B_{17} \left\{ (nh(1 + x)^{-\alpha_1}) \right\}^2 \right].
\]

After substituting \( v = (nhB_{17})^{-1/\alpha_1} (1 + x) \) and integrating,

\[
J_\beta \leq B_{16} (nh)^{-1 + 1/\alpha_1} \int_{(nh)^{1/\beta - 1/\alpha_1}}^{\infty} u^{-\alpha_1} \exp \left( u^{-2\alpha_1} \right) du
\]

\[
\leq B_{16} (nh)^{-1 + 1/\alpha_1} \int_{(nh)^{1/\beta - 1/\alpha_1}}^{\infty} u^{-\alpha_1} du
\]

\[
= B_{16} (nh)^{-1 + 1/\alpha_1} \left[ -u^{-1-\alpha_1} \right]_{(nh)^{1/\beta - 1/\alpha_1}}^{\infty}
\]

Since \( \alpha_1 > 1, 1 - \alpha_1 < 0, v^{1-\alpha_1} \to 0 \) as \( v \to \infty \). This allows us to say

\[
J_\beta = O \left\{ (nh)^{(1-\alpha_1)/\beta} \right\} = O \left\{ (nh)^{-1+1/\alpha} \right\}
\]

if \( \alpha > \alpha_1 \) and if \( \beta > \alpha_1 \) is chosen sufficiently close to \( \alpha_1 \). This result leads to

\[
\int_{(nh)^{1/\beta}}^{\infty} |\hat{f}(x)| dx = O \left\{ (nh)^{-1+1/\alpha} \right\}.
\]
2.5.8 Proof of Theorem 2.3

The expectation of the incidence of the negative regions of \( \hat{f} \) is given by,

\[
E\{I(\hat{f} < 0)\} = \text{Pr}(\hat{f} < 0)
\]

\[
\leq \exp\{-B_4 nh(1 + x)^{-\alpha_1}\} \quad \text{(from (2.46) in Theorem 2.1)}.
\]

Therefore,

\[
E\left\{ \int_0^\infty f^2 I(\hat{f} < 0) \right\} \leq B_{19} \int_0^\infty (1 + x)^{-2\alpha_1} \exp\{-B_4 nh(1 + x)^{-\alpha_1}\} \, dx
\]

\[
= O\{(nh)^{-2+1/\alpha_1}\}.
\]

Since \( \alpha_1 > 1 \), the right hand side equals \( o\{(nh)^{-1}\} \) so that

\[
\int f^2 I(\hat{f} < 0) = o\{(nh)^{-1}\}. \quad (2.60)
\]

Using (2.49) from Theorem 2.1, we have

\[
E \left[ \int_0^\infty f(x)|\hat{f}(x)|I\{\hat{f}(x) < 0\} \right] \leq B_9(nh)^{-1} \times
\]

\[
\int_0^\infty f(x) \exp\left\{-\frac{1}{4} B_4 nh(1 + x)^{-\alpha_1}\right\} \max\{nh(1 + x)^{-\alpha_1},(nh)^2(1 + x)^{-2\alpha_1}\} \, dx.
\]

By substituting \( (1 + x) = (\frac{1}{4} B_4 nh)^{1/\alpha_1} v \) and \( f(x) \sim c_1 x^{-\alpha_1} \),

\[
E \left[ \int_0^\infty f(x)|\hat{f}(x)|I\{\hat{f}(x) < 0\} \right]
\]

\[
\sim B_9(nh)^{-2+1/\alpha_1} \int_0^\infty v^{-\alpha_1} \exp(-v^{-\alpha_1}) \times \max(v^{-\alpha_1},v^{-2\alpha_1}) \, dv
\]

\[
= O\{(nh)^{-2+1/\alpha_1}\}
\]

\[
= o\{(nh)^{-1}\}. \quad (2.61)
\]

With (2.60) and (2.61) and similar results over \((-\infty, 0)\), Theorem 2.3 is proved.
2.5.9 Proof of Theorem 2.4

We assume that the left tail provides the essential information about $\gamma$ and prove that as $n \to \infty$,

$$\int_{-\infty}^{\frac{1}{2}} \hat{f}(x)I(\hat{f}(x) < 0)dx = -h^{a_1+1} \int G_1I(G_1 < 0) + o_p \{h^{a_1+1} + (nh)^{-1}\},$$

(2.62)

where $G_1 = c_1 \int_{u<v} (v-u)^{a_1} K(u)du$.

We treat separately two ranges of $x$:

$R_1(B) = \{x : Bh < x < \frac{1}{2}\}$, $R_2(B) = \{x : -\infty < x < Bh\}$,

where $B$ is a large but fixed constant. The first range is the “central” part of the density where we do not anticipate much contribution to $\gamma$. The second range contains the “tails” of the density where negative density estimates are likely to arise.

Range(1): $R_1(B)$

With $K_r(y)$ defined in theorem 2.1, we may write

$$E|\hat{f}(x) - f(x)| = |h^r \int_0^{\infty} \{f^{(r)}(x + hy) + f^{(r)}(x - hy)\} K_r(y)dy|.$$

For reasons similar to those that lead to (2.42) in the proof of Theorem 2.1,

$$|E\hat{f}(x) - f(x)|$$

$$\leq B_1 h^r \left\{ \int_0^{B_2} (h + x + hy)^{a_1-r} dy + \int_0^{\min(B_2, h^{-1}-x)} (h + x - hy)^{a_1-r} dy \right\}$$

$$\leq B_3 h^r (x + h)^{a_1-r} \text{ uniformly in } 0 < x \leq \frac{1}{2}.$$

If $Bh < x \leq \frac{1}{2}$ and $B$ is sufficiently large,

$$E\hat{f}(x) \geq f(x) - |E\hat{f}(x) - f(x)|$$
\[ \begin{align*}
\geq \ & B_3 x^{a_1} - B_3 h^r (x + h)^{(a_1 - r)} \\
\geq \ & B_4 x^{a_1}.
\end{align*} \]

For reasons similar to those that gave (2.45) in Theorem 2.1, we have \( \text{var} \hat{f}(x) \leq B_5 (nh)^{-1} x^{a_1} \) uniformly in \( Bh < x \leq \frac{1}{2} \). The probability of a negative density estimate in \( R_1(B) \) is given by

\[
Pr \left\{ \hat{f}(x) < 0 \right\} = Pr \left\{ \hat{f}(x) - E \hat{f}(x) < -E \hat{f}(x) \right\}
\]

\[
= Pr \left\{ E \hat{f}(x) - \hat{f}(x) > E \hat{f}(x) \right\}
\]

\[
\leq 2 \exp \left[ \frac{-E \left\{ \hat{f}(x) \right\}^2}{2 \left\{ \text{var} \hat{f}(x) + \frac{1}{3} E \hat{f}(x)(nh)^{-1} C_2 \right\}} \right]
\]

(by Bernstein's inequality)

\[
\leq 2 \exp \left[ \frac{-\left\{ \frac{1}{2} B_4 x^{a_1} \right\}^2}{2 \left\{ B_5 (nh)^{-1} x^{a_1} + \frac{1}{6} B_4 x^{a_1} (nh)^{-1} C_2 \right\}} \right]
\]

\[
\leq 2 \exp \left\{ -(nh) B_6 x^{a_1} \right\}
\]

uniformly in \( Bh < x \leq \frac{1}{2} \). In the same range,

\[
E \hat{f}(x)^2 = \text{var} \hat{f}(x) + E \left\{ \hat{f}(x) \right\}^2
\]

\[
\leq B_7 \left\{ (nh)^{-1} x^{a_1} + x^{2a_1} \right\}.
\]

By the Cauchy-Schwarz inequality,

\[
\begin{align*}
\int_{B_h} \frac{1}{2} E \left[ |\hat{f}(x)| I \{ \hat{f}(x) < 0 \} \right] dx \\
\leq \int_{B_h} \frac{1}{2} E \left\{ \hat{f}(x)^2 \right\}^{\frac{1}{2}} Pr \left\{ \hat{f}(x) < 0 \right\}^{\frac{1}{2}} dx
\end{align*}
\]

\[
\leq (B_7)^\frac{1}{2} \int_{0}^{\infty} \left\{ (nh)^{-\frac{1}{2}} x^{a_1/2} + x^{a_1} \right\} \exp (-B_9 nh x^{a_1}/2) dx. \quad (2.63)
\]
We evaluate these integrals by

\[
\int_0^\infty (nh)^{-\frac{1}{2}} x^{\alpha_1/2} \exp (-B_8 nh x^{\alpha_1/2}) \, dx \\
= (nh)^{-\frac{1}{2}} \frac{2}{\alpha_1} \int_0^\infty u^{2/\alpha_1} \exp (-B_8 nh u^2/2) \, du \\
= (nh)^{-\frac{1}{2}} \frac{1}{\alpha_1} \Gamma \left(1/\alpha_1 + 1/2\right) (B_8/2)^{-(1/\alpha_1+1/2)} (nh)^{-(1/\alpha_1+1/2)} \\
= O \left\{ nh^{-1-(1/\alpha_1)} \right\},
\]

(2.64) and

\[
\int_0^\infty x^{\alpha_1} \exp (-B_8 nh x^{\alpha_1/2}) \, dx \\
= \int_0^\infty u \exp (-B_8 nh u/2) (1/\alpha_1) u^{(1/\alpha_1-1)} \, du \\
= \frac{1}{\alpha_1} \Gamma \left(1/\alpha_1 + 1\right) (B_8 nh/2)^{-(1/\alpha_1+1)} \\
= O \left\{ nh^{-1-(1/\alpha_1)} \right\}.
\]

(2.65)

By substituting (2.64) and (2.65) into (2.63),

\[
\int_{B_h} \hat{f}(x) I \{ \hat{f}(x) < 0 \} \, dx \\
= O_p \left\{ (nh)^{-1-(1/\alpha_1)} \right\} \\
= o_p \left\{ (nh)^{-1} \right\} \text{ since } \alpha_1 > 1.
\]

(2.66)

Range (2): \( R_2(B) \)

Recall that \( \hat{f}(x) = (nh)^{-1} \sum_j K \{ h^{-1}(x-X_j) \} \), \( K \) vanishes outside \( (-C_1, C_1) \) and \( |K| \leq C_2 \) which implies that,

\[
E \hat{f}(x) \leq (nh)^{-1} C_2 n Pr(|x-X| \leq C_1 h).
\]

With this inequality we have,

\[
\int_0^{B_h} E \left[ |\hat{f}(x)| I \{ \hat{f}(x) < 0 \} \right] \, dx \\
\leq (nh)^{-1} C_2 n \int_0^{B_h} Pr(|x-X| \leq C_1 h) \, dx
\]

64
\[
\leq B_s \int_0^{B_h} x_{\alpha_1} dx = O\left(h^{\alpha_1+1}\right).
\]

If \(nh^{\alpha_1+2} \to 0\) as \(n \to \infty\), then

\[
\int_0^{B_h} E \left[ |\hat{f}(x)| I \{\hat{f}(x) < 0\} \right] dx = o\left((nh)^{-1}\right). \tag{2.67}
\]

Next, assume that \((nh)^{\alpha_1+2}\) is bounded away from zero. In this case, we calculate \(\gamma\) as a function of \(G_1(v)\) defined at the beginning of the theorem. If \(v_0\) is sufficiently large then \(G_1(v) \geq 0\) for all \(v \geq v_0\). Choose \(B \geq v_0\). If \(x = hv\), then

\[
E \{\hat{f}(x)\} = \int_{-\infty}^{v} f(hv - hy)K(y)dy = h^{\alpha_1}G_1(v) + o(h^{\alpha_1}),
\]

\[
\text{var}\{\hat{f}(x)\} \leq (nh)^{-1} \int_{-\infty}^{v} f(hv - hy)K(y)^2dy
\]

\[
= O(n^{-1}h^{\alpha_1-1}) = o(h^{2\alpha_1}).
\]

Hence by Markov's inequality\(^3\) (Ross [21,p.243ff]), \(h^{-\alpha_1}\hat{f}(hv) \to G_1(v)\) in probability. Thus, it is to be expected that

\[
h^{-\alpha_1+1} \int_{-\infty}^{B_h} \hat{f}(x) I \{\hat{f}(x) < 0\} dx = h^{-\alpha_1} \int_{-\infty}^{B} \hat{f}(hv) I \{\hat{f}(hv) < 0\} dv
\]

\[
\to \int_{-\infty}^{B} G_1(v) I \{G_1(v) < 0\} dv = \int_{-\infty}^{\infty} G_1(v) I \{G_1(v) < 0\} dv, \tag{2.68}
\]

with convergence in probability. The veracity of this result is readily checked by establishing \(L^2\) convergence. The result (2.62) follows from (2.66), (2.67) and (2.68), and Theorem 2.4 is proved.

\(^3\)Markov's inequality : \(Pr(x \geq a) \leq E(x)/a\) \(\forall a > 0\).
Chapter 3

Estimation of the Density and Mode of a Continuous-Time Stationary Stochastic Process
3.1 Introduction

We are motivated to consider kernel density estimation of continuous stationary processes by the increasing prevalence of data of this type. Modern data recording instruments allow data to be recorded on extremely fine sampling grids so that for all intents and purposes the data may be considered continuous. These instruments have the capability of amassing large tracts of data and while they provide much information, they reveal statistical properties that are not apparent in smaller samples. Improved statistical techniques are required to fully exploit the information contained in such data.

It is to be expected that sample values taken closer together (temporally or spatially) will be more alike than widely separated sample values, and statistical analysis needs to account for the correlations amongst the data. We demonstrate the suitability of kernel density estimation for the analysis of continuous stationary stochastic processes.

A convenient data set which we use as a focus for this chapter is the height of a metal surface, measured under powerful magnification. The data were collected as continuous traces recorded over line transects where height was measured from electrical impulses induced in a fine stylus drawn across the surface. The height above a given level may be thought of as the value of a stationary stochastic process $X$. Since height is recorded over line transects, then for our purpose $X$ is indexed by a single variable, $t$, representing distance along the transect. Figure 3.1 illustrates typical
traces of the studentized electrical readings from the height measuring instrument. Part (a) depicts a new surface, part (b) depicts a treated (i.e. chrome plated) but still new surface and part (c) a worn surface. The skewness of the distribution of surface height for a worn surface is clear from this comparison.

Figure 3.1: Line transects of height recordings (Studentised) for new, treated and worn surfaces.

(a) new

(b) treated

(c) worn
With data such as measurements of heights of the metal surface, we must consider the competing merits of parametric or nonparametric estimation of the marginal density. If the assumption of a parametric distribution could be relied upon, parametric estimation may be a viable procedure. However, departures in the data away from the parametric distribution may be quite subtle and not noticed in large data sets. The reliance upon the mathematical properties of an assumed, but not necessarily correct, parametric distribution may lead to unsound inference about the data. Furthermore, the estimation of the parameters that describe the dependence is often a complex procedure that relies heavily upon the parametric assumptions.

Discrete height measurements for the new surface pass several tests for normality, providing the sample is not too extensive and the data are gathered at widely separated points. Very large data sets can be relatively inexpensive to obtain in this problem. Therefore, parametric modelling of the marginal density is a possibility. Surface height for the worn surface may be modelled by a chi-squared process, ie by a sum of squares of independently and identically distributed Gaussian processes. However, the chi-squared model is not readily motivated by physical considerations.

Thus we are led to consider nonparametric methods for estimating both the marginal height density and the modal height.

Kernel estimation of the marginal density is quite flexible and appropriate for data such as the metal surface where wearing is gradually inducing skewness into the distribution of height measurements. We shall show that under favourable conditions,
We show that the nonparametric estimator is at least as efficient as its parametric counterpart, even when the distribution of the data is given.
nonparametric estimation of the marginal density is as effective as parametric estimation and under general conditions it is still quite satisfactory. Thus we may feel quite confident in using nonparametric methods for inference about the data, safe in the knowledge that we shall not be tricked by misspecification of the parametric distribution.

To demonstrate the worth of kernel density estimators, relative to parametric estimators, we develop theory for the case where $X(t)$ is a Gaussian process over the interval $(0, T)$ with covariance function $\rho(t) = \text{cov}\{X(s), X(s + t)\}$. It might be thought initially that this situation is relatively uninteresting since the marginal density of $X$ depends on only two unknown parameters, the mean and standard deviation, and these may be generally estimated at a rate $O_p(\lambda^{1/2}_T)$, where $\lambda_T = T^{-1} \int_0^T |\rho(t)| dt$. For example, the sample mean $\bar{X} = T^{-1} \int_0^T X(t) dt$, has this convergence rate. Our methods may be readily extended to processes which are functions of Gaussian processes, e.g. $X = g(Y)$, where $Y$ is Gaussian and $g$ is a polynomial. We do not have a convincing theory for the case where the function $g$ is determined nonparametrically, and anyway, the case of a simple Gaussian process $X$ provides on its own a wealth of intricate and fascinating possibilities. Furthermore, one of our major sources of real data, the height of the surface of an unworn roller used to mill sheet metal, is strikingly close to being Gaussian. For all these reasons we have chosen to study the Gaussian process in detail, so as to illustrate the unusual properties that may be expected from kernel estimators for other sources of continuous data. We shall also investigate the effect of long term correlations which may be evident in large
3.2 Previous Developments in Density and Mode Estimation for Stationary Processes

3.2.1 Density Estimation

Whilst there has been considerable work done on nonparametric density estimation for discrete samples from a continuous stochastic process, there has been little written on estimation when the data are truly continuous. We shall present a brief review of some of the previous findings for discrete data so that we can see the effect of the extra information that continuous data provides.

Chanda and Ruymgaart [4] present theory for the convergence rates of estimates of density and mode when the data are discrete samples of a general linear process defined by $X_i = \sum_{k=-\infty}^{\infty} A_k Z_{i-k}$ for $i = 1, \ldots, n$, where the $A_k$ are given $d \times d$ matrices and the $Z_j$ are independent and identically distributed $d$-dimensional random vectors. The authors study the density and mode based on a sample $X_1, \ldots, X_n$ from the above process. Their approach is a decomposition of $X_i$ into a $m$-dependent part, $X_{i,m} = \sum_{k \leq \mid m \mid} A_k Z_{i-k}$, and a remainder, $\tilde{X}_{i,m} = \sum_{k > \mid m \mid} A_k Z_{i-k}$, which is asymptotically (for $m \to \infty$) negligible. The degree of dependence is quantified by $m = O(n^\beta)$ where $0 < \beta < \frac{1}{2}$. They consider the
naive estimator given by

$$
\hat{f}(x) = (2nh)^{-1} \left[ \text{number of } X_1, \ldots, X_n \text{ falling in } (x-h, x+h) \right]
$$

(see Silverman [23, p12]). For the case of \( d = 1 \), with \( |f''| \leq M \in (0, \infty) \),

$$
n^\xi \sup_x |\hat{f}(x) - f(x)| \xrightarrow{a.s.} 0 \text{ as } n \to \infty \text{ for any } 0 < \xi < 2(1 - \beta)/5
$$

when \( h \) is chosen as \( \propto n^{-(1-\beta)/5} \). The analogue of this result for independent data is given by Devroye and Györfi [8, p79] who show that \( E \int |\hat{f}_n(x) - f(x)| \, dx \) decreases at the rate \( n^{-3/2} \) if \( h \propto n^{-1/3} \). Thus we can see how the dependence amongst the data can slow the rate of convergence in the discrete case.

Masry [16] established the consistency, asymptotic bias and bound for the covariance of kernel probability density estimators (\( \hat{f} \)) of continuous-time stationary processes where the data had been sampled at discrete time instances. Widely spaced samples of the process were assumed asymptotically independent and the correlations of near observations were defined in four ways; strong mixing, uniform mixing, maximal correlation condition and asymptotically uncorrelated.

The bias of \( \hat{f} \) is the same as for independent samples (see (1.3)) and does not depend upon the covariance of the stochastic process or the number of sampling points, other than for the influence of sample size on bandwidth.

The bound on the covariance of \( \hat{f} \) derived from discrete samples of a continuous-time process is a function of the covariance of the data and the strength of the assumptions regarding the data influences the bound of the covariance of \( \hat{f} \). We first present the definitions used to describe the dependence amongst the data.
Let $\mathcal{F}_a^b = \sigma\{X(t), a \leq t \leq b\}$ be the $\sigma$-algebra of events in $\mathcal{F}$ generated by the random variables $\{X(t), a \leq t \leq b, -\infty < a < b < \infty\}$. We denote $L_2(\mathcal{F}_a^b)$ as the collection of all second-order random variables measurable with respect to $\mathcal{F}_a^b$. The stationary process is strong mixing if for $\tau \geq 0$,

$$\sup_{A \in \mathcal{F}_a^b, B \in \mathcal{F}_c^d} \left| \Pr[AB] - \Pr[A] \Pr[B] \right| = \alpha(\tau) \downarrow 0 \text{ as } \tau \to \infty,$$

and is uniform mixing if

$$\sup_{A \in \mathcal{F}_a^b, B \in \mathcal{F}_c^d, P[A] > 0} \left| \Pr[B|A] - \Pr[B] \right| = \phi(\tau) \downarrow 0 \text{ as } \tau \to \infty.$$

The terms $\alpha(\tau)$ and $\phi(\tau)$ are known as the strong mixing and uniform mixing coefficients respectively. The uniform mixing condition is a stronger assumption about the correlation than the strong mixing condition.

The stationary process, $X$, is asymptotically uncorrelated if for every $\tau \geq 0$,

$$\left| \text{cov}\{g_1(X_\tau), g_2(X_0)\} \right| \leq a(\tau) \{E[g_1^2(X_0)]E[g_2^2(X_0)]\}^{1/2}$$

for some function $a(\tau) \downarrow 0$ as $\tau \to \infty$ for all functions $g_i(x) \in L_2[f(x)dx]$.

It is assumed that the sampling instants $\{t_k\}_{k=1,N}$ are non-negative real numbers such that $\inf_{k \geq 1} |t_{k+1} - t_k| = 1/\lambda > 0$. Under the assumption of uniform mixing with $\int_0^\infty \phi(\tau)^{1/2}d\tau < \infty$, the bound for the covariance of $\hat{f}$ is given by

$$\limsup_{n \to \infty} n h \text{cov}\{\hat{f}(x), \hat{f}(y)\} \leq \left\{ f(x)\delta_{y,x} + 4\lambda \sqrt{f(x)f(y)} \int_0^\infty \phi(\tau)^{1/2}d\tau \right\} \int_{-\infty}^\infty K^2(v)dv , \quad (3.1)$$

where $\delta_{y,x}$ is the Kronecker delta and $K$ is the kernel function.
This expression reduces to the classical asymptotic result for independent data when \( \int_0^\infty \phi(t)^{1/2} dt \) is asymptotically negligible. The bound on the covariance is proportional to \( \lambda \), reflecting that observations become more dependent as their sampling intervals decrease. Under the weaker assumption of strong mixing with \( \int_0^\infty [\alpha(t)]^q dt < \infty, 0 < q < 1 \), the convergence rate for the bound is \( O\{(nh^{1+q})^{-1}\} \) compared with \( O\{(nh)^{-1}\} \) for uniform mixing where more stringent conditions were imposed on the mixing coefficient. The effect of the sampling rate, \( \lambda \), is similar to that in the case of uniform mixing. For asymptotically uncorrelated processes, we have

\[
\lim_{n \to \infty} \sup \ n h \ \text{cov}\{\hat{f}(x), \hat{f}(y)\} \leq \{f(x)\delta_{y,x} + 2\lambda \sqrt{f(x)f(y)} \int_0^\infty \phi(t)^{1/2} dt\} \int_{-\infty}^\infty K^2(u) du . \tag{3.2}
\]

In his paper, Masry compares the performance of discrete-time density estimators with continuous-time estimators based on the observations of \( X \) over the interval \([0,T]\). For asymptotically uncorrelated processes he shows that if

\[
\int_0^\infty |f_r(x,y) - f(x)f(y)| dt \leq M < \infty \ \forall x, y \tag{3.3}
\]

where \( f_r(x,y) \) is the bivariate density of \( x \) and \( y \) that are \( r \) apart, then

\[
|\text{cov}\{\hat{f}(x), \hat{f}(y)\}| \leq \frac{2M}{T} \left( \int_{-\infty}^\infty |K(u)| du \right)^2 .
\]

The convergence rate of this bound is \( O(T^{-1}) \) which is much faster than its discrete time analogue \( O\{(Th)^{-1}\} \) but requiring condition (3.3) for it to hold and Masry demonstrates that this condition does not hold for smooth Gaussian processes.

Castellana and Leadbetter [3] investigate kernel density estimation for continuous processes with discrete and continuous samples. Their density estimators are of
'smoothing function type', derived using smoothing functions $\delta_n$ (for discrete samples $X_1, \cdots, X_n$) and $\delta_T$ (for continuous samples $X_t, \ t \in (0,T)$). Kernel functions are examples of these smoothing functions with $\delta_n(x) = h_n^{-1}K(h_n^{-1}x)$ and $\delta_T(x) = h_T^{-1}K(h_T^{-1}x)$ where $K$ is the kernel function and $h_n$ and $h_T$ are the bandwidths for the discrete and continuous cases respectively. The constraints on $\delta_n$ and $\delta_T$ are the same as given for $K$ in (1.2) and we define the quantities $\alpha_n = \int \delta_n^2(x)dx < \infty$ for each $n$ and $\alpha_T = \int \delta_T^2(x)dx < \infty$. These terms are required later for discussing the convergence rates of the variance of density estimators from discrete and continuous samples.

The density estimator based on a discrete sample is

$$\hat{f}_n(x) = n^{-1} \sum_{i=1}^{n} \delta_n(x - X_i) ,$$

and for a continuous sample it is

$$\hat{f}_T(x) = T^{-1} \int_{0}^{T} \delta_T(x - X_t)dt .$$

The variances of these estimators depend on the correlations amongst the data, the conditions on the smoothing function and the assumptions about the underlying true density.

The measure of dependence for the discrete sequence $\{X_j, j = 1, n\}$ is the dependence index sequence defined by

$$\beta_n = \sup_{x,y} \sum_{i=1}^{n} |f_i(x, y) - f(x)f(y)|$$

$(n \geq 1)$

where $f_i(x, y)$ is the joint density of $X_i$ and $X_{1+i}$. For i.i.d. sequences, $\beta_n \equiv 0$ for all $n$, for sequences with high long range dependence $\beta_n$ may tend to infinity, and in
between $\beta_n$ may converge to a finite limit at various rates. Castellana and Leadbetter
[3] give two results for the convergence of $\text{var}(\hat{f}_n)$. The first, Theorem 3.3, is for the
situation where the dependence is restricted so that $\beta_n = o(\alpha_n)$ as $n \to \infty$. For kernel
estimators, this condition may be interpreted as requiring that $h\beta_n \to 0$ as $n \to \infty$.
In these circumstances,

$$(n/\alpha_n)\text{var}\{\hat{f}_n(x)\} \to f(x) \text{ as } n \to \infty.$$  \hfill (3.4)

By imposing extra but general conditions on the density and smoothing function,
a more precise result for the convergence of $\text{var}(\hat{f}_n)$ is obtained in their Theorem 3.5. We restate this theorem now so that we can discuss differences between the
convergence rates of $\text{var}(\hat{f})$ when the data are sampled discretely and continuously.

Let the density $f$ have bounded second derivative $f''$. Let $\{\delta_n(x); n \geq 1\}$ be a
non-negative $\delta$-sequence with each $\delta_n$ even, and such that $\alpha_n = \int \delta_n^2(x)dx < \infty$,
$\theta_n = \int x^2 \delta_n^2(x)dx < \infty$ (where $\delta_n^2(x) = \delta_n(x)/\alpha_n$) and $\theta_n^{-1} \int_{|x|>\lambda} x^2 \delta_n^2(x)dx \to 0$ for
each $\lambda > 0$. Then

$$(n/\alpha_n)\text{var}\{\hat{f}_n(x)\} = f(x) + \frac{1}{2} \theta_n f''(x) \{1 + o(1)\} - \alpha_n^{-1} f^2(x) \{1 + o(1)\} + O(\beta_n/\alpha_n).$$  \hfill (3.5)

Thus we can see that for discrete samples, the variance of the density estimator
typically is asymptotic to $C\alpha_n/n$, where $C$ is a nonzero constant.

For continuous processes $(X_t, t = (0,T))$, where condition (3.3) could be assumed,
Castellana and Leadbetter [3] prove (Theorem 5.2) that

$$\lim_{T \to \infty} T\text{var}\{\hat{f}_T(x)\} = 2 \int_0^\infty \{f_r(x,x) - f^2(x)\} \, dx,$$  \hfill (3.6)
so the convergence rate is precisely $O(T^{-1})$. Condition (3.3) allows us to specify the local dependence between $X_t$ and $X_{t+\tau}$ as $\tau \to 0$ and this restriction on the local dependence of $X_t$ and $X_{t+\tau}$ leads to a 'full rate' $1/T$ of convergence of the variance to zero. This contrasts sharply with the sequence case where the condition $\tau \to 0$ is not a feature. The authors' intuitive explanation of this phenomenon is that the continuous sampling collects a whole continuum of 'somewhat independent' random variables.

Whilst there is potential for fast ($O(T^{-1})$) convergence rates of $\text{var} (\hat{f}_T)$, this may only be realised when the restrictive conditions (3.3) apply. A more general result is given by their Theorem 5.4 which provides lower bounds for the convergence rate. To obtain this they define the dependence index function to be the function of $0 < \gamma < T$ given by

$$\beta_T(\gamma) = \sup_{x,y} \int_{\gamma}^{T} |f_T(x,y) - f(x)f(y)| \, d\tau$$

where $f_T(x,y)$ is the joint density of $X_0$ and $X_T$. This dependence index function is assumed finite for all $0 < \gamma < T$. Theorem 5.4 states that if \{ $\gamma_T; T \geq 0$ \} are positive constants such that $\gamma_T \to 0$ as $T \to \infty$, and if

$$\gamma_T^{-1} \beta_T(\gamma_T) = o(\alpha_T) \text{ as } T \to \infty,$$

then

$$\limsup_{T \to \infty} T \gamma_T^{-1} \alpha_T^{-1} \text{var} \{ \hat{f}_T(x) \} \leq 2f(x).$$

Thus for cases where local dependence is greater than required in their Theorem 5.2, the convergence rate is slower than $T^{-1}$ but not as slow as $\alpha_T/T$, which would be the
discrete-time analogue.

The consistency of kernel density estimators for a continuous parameter and stationary process was also proved by Delecroix [7].

3.2.2 Mode Estimation

The problem of estimation of the mode of independent data via kernel density estimators has been addressed by Parzen [17], Eddy [9, 10] and Romano [19]. With the sample mode, \( \hat{\theta} \), defined as the location such that \( \hat{f}(\hat{\theta}) = \max_x \hat{f}(x) \), Parzen proved that \( \hat{\theta} \) is a consistent estimator of the population mode, \( \theta \), and that the sample mode is asymptotically normal with a variance that depends upon \( f''(\theta) \). Eddy [9] proved that the same result holds when less stringent assumptions are imposed on the kernel than those assumed by Parzen. For second order kernels, the asymptotic bias of \( \hat{\theta} \) depends on \( f'''(\theta) \).

Eddy [9] showed that with second order kernels, \( E(\hat{\theta} - \theta)^2 \) is minimised by choosing \( h_n \propto n^{-1/7} \) whereas the fixed bandwidth rule for density estimation based on the mean integrated squared error gives \( h \propto n^{-1/5} \). Romano expanded this subject by emphasizing that the mode is precisely a location where choice of bandwidth is most sensitive and that estimation of it, using kernel density estimators, should be by a data dependent, random bandwidth. The choice of a bandwidth proportional to \( n^{-1/7} \), results in \( \hat{\theta} \), converging to \( \theta \) at a rate of \( (nh^3)^{1/2} \) which is \( \propto n^{2/7} \); that is \( \hat{\theta} \) is an \( n^{2/7} \)-consistent estimator of \( \theta \).

We would not expect these results to flow through exactly for the situation where...
the data are continuous yet they give a guide for bandwidth selection and subsequent estimation for the mode of continuous data. Since the squared error properties of the mode are related to $f''(\theta)$ and $f'''(\theta)$, we should look to these functions for determining a bandwidth rule for mode estimation. With densities such as we are studying, the existence of $f''(\theta)$ and $f'''(\theta)$ is guaranteed and the theory that we develop for $\hat{f}^{(i)}$ gives us a foundation for mode estimation. Analogous results are of course true for density estimators based on kernels of higher orders, and in fact the estimators that we use are of this form.

Chanda and Ruymgaart [4] prove that for a general linear process, described in the previous sub-section, $n^{\xi/2}(\hat{\theta} - \theta) \xrightarrow{a.s.} 0$ as $n \to \infty$, for any $0 < \xi < 2(1 - \beta)/5$, but this is so when $h \propto n^{-(1-\beta)/5}$. (Recall that $0 < \beta < \frac{1}{2}$.) They use the same bandwidth that is required for optimal convergence of the density estimate, rather than using a wider bandwidth as suggested in the case of independent data. We can take these results as a guide to consistency but we are unaware of theory for the convergence rates of the kernel estimator of the mode of continuous dependent data.

3.3 Estimation of the Marginal Density and its Derivatives of a Continuous-Time Process

In this section, we describe kernel estimation of the marginal density for a continuous-time process, concentrating on Gaussian processes. The $L^2$ convergence of $\hat{f}^{(i)}$ to $f^{(i)}$ will depend upon the bias and variance of $\hat{f}^{(i)}$ and we provide theory for these
components. Before building our theory about the density estimators, we establish the Gaussian process to which our theory applies.

### 3.3.1 The Covariance Function of a Gaussian Process

Let $X_t$, $t > 0$, denote a Gaussian process observed on the interval $(0, T)$. We suppose that the covariance function, $\rho(\tau) = \text{cov}\{X_t, X_{t+\tau}\}$, behaves like

$$\rho(\tau) = \rho(0) - (c_1|\tau|)^\alpha + o(|\tau|^\alpha) \text{ as } \tau \to 0 ,$$  \hspace{1cm} (3.7)

and like

$$\rho(\tau) \sim (c_2|\tau|)^{-\beta} \text{ as } \tau \to \infty ,$$  \hspace{1cm} (3.8)

where $\alpha, \beta > 0$ and $c_1, c_2 > 0$. The constants $c_1$ and $c_2$ are scaling constants to account for the units of measurement of $t$. Necessarily, $0 < \alpha \leq 2$ and the relationship (3.7) describes the short term dependence amongst the data. The relationship (3.8) represents long term dependence if $\beta \leq 1$, since $\int_0^\infty |\rho(\tau)|d\tau = \infty$. For $\beta > 1$, the integral will be finite and relation (3.8) will have relatively little impact.

In Figure 3.2, we illustrate with simulated data how roughness of the data influences the short term covariance function. The simulation details are described later and this example uses relationships like (3.7), $\alpha = 0.6, 1, 1.95$, and (3.8), $\beta = 1$, in the proportions of 0.9 and 0.1 respectively. The scaling constants $c_1, c_2$ are 1 and the plots here are sections from surfaces with $E(X_t) = 0$, $\text{var}(X_t) = 1$ and $T = 400$. 

79
Figure 3.2: Short term dependencies and simulated Gaussian processes.

\[ \alpha = 0.6, \beta = 1 \]

\[ \alpha = 1.0, \beta = 1 \]

\[ \alpha = 1.95, \beta = 1 \]
3.3.2 Kernel Estimators for Continuous-Time Data

A kernel estimator of the marginal density of $X_t$, $f(x)$, is given by

$$\hat{f}(x) = (Th)^{-1} \int_0^T \{K(x - X_t)/h\} dt, \quad (3.9)$$

where $h$ denotes the smoothing parameter (bandwidth) and the kernel function $K$ is symmetric and of order at least $r \geq 1$ (see (1.2)). If $K$ has $j$ derivatives, the $j$th derivative of $f$ may be estimated by the $j$th derivative of $\hat{f}$,

$$\hat{f}^{(j)}(x) = (Th^{j+1})^{-1} \int_0^T K^{(j)} \{(x - X_t)/h\} dt. \quad (3.10)$$

Consistency demands that $h \to 0$ as $T \to \infty$ but it is not essential that $Th \to \infty$ as $T \to \infty$. For the analogue of this latter condition with discrete data of sample size $n$, $nh \to \infty$ as $n \to \infty$ is a necessary constraint.

We give an example in Figure 3.3 of an estimate of a density and its first derivative, derived from a simulated Gaussian process. The data, $X_t$, $t \in (0, 400)$, were simulated so that $E(X_t) = 0$, $\text{var}(X_t) = 1$, with covariance function $\rho(\tau) = \exp(-\tau)$ and the bandwidths for estimating the density and first derivative were 0.27 and 0.36 respectively. The solid lines represent $f(x)$ and $f'(x)$ and the dotted lines represent $\hat{f}(x)$ and $\hat{f}'(x)$. 

81
3.3.3 The Bias of $\hat{f}^{(j)}$

The expected value and bias of $\hat{f}^{(j)}$ may be deduced by traditional arguments for kernel type estimators. From (3.10) we have,

$$E\{\hat{f}^{(j)}(x)\} = (Th^{j+1})^{-1} \int_0^T \int_{-\infty}^{+\infty} K^{(j)}((x-y)/h) f(y) dy dt$$

$$= (Th^{j+1})^{-1} h \int_0^T \left\{ \int_{-\infty}^{+\infty} K^{(j)}(u)f(x-hu)du \right\} dt .$$

(3.11)

Evaluating the inner integral by parts once gives,

$$h^{-j} \left[ \left\{ K^{(j-1)}(u)f'(x-hu) \right\}_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} K^{(j-1)}(u)f'(x-hu)(-h)du \right]$$

$$= h^{-j+1} \int_{-\infty}^{+\infty} K^{(j-1)}(u)f'(x-hu)du .$$

After integrating by parts $j$ times and integrating with respect $t$, we have

$$E\{\hat{f}^{(j)}(x)\} = \int_{-\infty}^{+\infty} K(u)f^{(j)}(x-hu)du .$$
By expanding \( f^{(j)}(x - hu) \) in a Taylor series about \( f^{(j)}(x) \) and using the properties of \( K \) (see (1.2)), a simple, approximate expression for the bias is given by

\[
\text{bias} \left\{ \hat{f}^{(j)}(x) \right\} = (-1)^r h^{r} f^{(j+r)}(x) k_r / r! + o(h^r). 
\]  

(3.12)

The bias formula does not involve the dependence amongst the data and is of the same form as for a kernel density estimator for discrete data.

### 3.3.4 The Variance of \( \hat{f}^{(j)} \)

The dependence amongst the data has substantial influence on the variance of the kernel estimator, with results quite different from those in discrete data cases. An unexpected result is the way in which the convergence rate depends on the fractal dimension of the continuous-time process, \( X_t \). For Gaussian processes such as we have described, the fractal dimension is given by \( D = 2 - \frac{1}{2} \alpha \) (\( \alpha \) is the parameter shown in (3.7)); see Adler [1, Chapter 8]. The rate of convergence of \( \text{var}(\hat{f}) \) will be faster for rough sample paths of the stochastic process (corresponding to \( \alpha \) being 'small') than for smooth sample paths where \( \alpha \) may be close to 2. Paths that are rough have less correlation than smooth paths and there is more information in the measurement of \( X_t \), leading to a faster rate of convergence of the variance of \( \hat{f}^{(j)} \).

The expression for the variance of \( \hat{f}^{(j)} \) is complex, involving a triple integral, but we can reduce this analytically using asymptotic expressions for the covariance function and leading to expressions for the asymptotic variance of \( \hat{f}^{(j)} \).
3.3.5 **Asymptotic Variance of $\hat{f}^{(j)}$ for Gaussian Processes**

There are separate components to the variance of $\hat{f}^{(j)}$ due to the short term and long term dependence amongst the data. Within each of these components, we can identify contributions which are

(i) functions of the underlying true marginal density and

(ii) functions of the bandwidth and length of the stochastic process.

The "roughness" of the surface influences the short term and long term covariances and the type of the function for each of the four components of the asymptotic variance of $\hat{f}^{(j)}$ changes depending upon the covariance function (or the "roughness") of the stochastic process. When we develop theory for this, we find that there are critical values for the parameters $\alpha$ and $\beta$ of the covariance function. These are stated here and the reasons for these critical values are made apparent during the proof of Theorem 3.1. For consideration of the effect of short term dependency, the nature of the functions depends on whether $\alpha$ is less than, equal to, or greater than $(j + \frac{1}{2})^{-1}$. The function for the long term component will depend whether $0 < \beta < 1$ or $\beta = 1$. To illustrate this, we briefly describe some results now, with a more detailed explanation of these results to be given later.

If $\alpha < (j + \frac{1}{2})^{-1}$, then one contribution to $\text{var}(\hat{f}^{(j)})$ is of size $T^{-1}$. When $\alpha = (j + \frac{1}{2})^{-1}$, this changes to $T^{-1} |\log h|$, and when $\alpha > (j + \frac{1}{2})^{-1}$ it becomes $T^{-1} h^{(2/\alpha)-(2j+1)}$. If $\beta < 1$, there is an additional contribution of $T^{-\beta}$, which changes to $T^{-1} \log T$ if $\beta = 1$.  

84
To appreciate the importance of these results, let us assume for the sake of simplicity that there is effectively no long term dependence (i.e. $\beta > 1$) and so the asymptotic variance of $\hat{f}^{(j)}$ does not depend upon $\beta$. For the sake of simplicity we consider here the standard case where the kernel is second order, $r = 2$. The mean integrated squared error is the sum of integrated squared bias and integrated variance which can be written as

$$MISE\{\hat{f}^{(j)}(x)\} \approx h^4(k_2/2)^2 \int f^{(j+2)}(x)^2 dx + \int \text{var}\{\hat{f}^{(j)}(x)\} dx.$$ 

When $\alpha < (j + \frac{1}{2})^{-1}$, the variance and squared bias are of sizes $T^{-1}$ and $h^4$ respectively. Therefore, bias may rendered negligible by choosing $h = o(T^{-1/4})$, to give an $L^1$ convergence rate for $\hat{f}^{(j)}$ of $O_p(T^{-1/2})$. The choice of bandwidth is of relatively little importance. However, if $\alpha \geq (j + \frac{1}{2})^{-1}$ then the convergence rate is necessarily slower than $O_p(T^{-1/2})$ and depends on the choice of bandwidth. For example, if $\alpha > (j + \frac{1}{2})^{-1}$ then the optimal bandwidth is of size $T^{-1/(2j+1+4-2/\alpha)}$ and produces an $L^1$ convergence rate of $T^{-1/(j+1/2+2-1/\alpha)}$. We see from the argument above that, apart from the point $\alpha = (j + \frac{1}{2})^{-1}$ where a discontinuity occurs, convergence results are generally slower for larger values of $\alpha$ which correspond to smoother processes $X$. Convergence rates are also slower for smaller values of $\beta < 1$, representing longer ranges of dependence.

A formal statement of our results concerning the variance of $\hat{f}^{(j)}$ is now provided. The theoretical results that describe how the "roughness" of the Gaussian process influences the rate of convergence of $\hat{f}^{(j)}$ to $f^{(j)}$ are stated and the proof of the
main theorem is deferred until Section 3.6. We take $X$ to be a Gaussian process over the interval $(0, T)$ and for simplicity we consider the case where $E(X) = 0$ and $\text{var}(X) = 1$. We are interested in estimating $\hat{f}^{(j)}$ and so we require appropriate regularity conditions such that the $j$th derivative is estimable.

Let $\kappa$ denote the Fourier transform of the kernel function $K$,

$$\kappa(u) = \int_{-\infty}^{+\infty} e^{iuv} K(v) dv.$$  \hfill (3.13)

This term is not scaled by $(2\pi)^{-1/2}$ but the scale factor will be absorbed into other terms. We assume that $K$ is symmetric and that for some $\eta > 0$,

$$\int (1 + |u|^{2j+1+\eta}) \kappa(u)^2 du < \infty.$$  \hfill (3.14)

This condition ensures that $K^{(j)}$ is well defined and continuous.

We assume that the covariance function, $\rho(\tau)$, satisfies the following conditions:

as $\tau \downarrow 0$, either $\{1 - \rho(\tau)\}^{-1} = O(\tau^{-\alpha})$ for $0 < \alpha < (j + \frac{1}{2})^{-1}$

or $\{1 - \rho(\tau)\} \sim (c_1 \tau)^{\alpha}$ for $(j + \frac{1}{2})^{-1} \leq \alpha \leq 2$, $c_1 > 0$ \hfill (3.15)

and

as $\tau \to \infty$, either $\int |\rho(\tau)| d\tau < \infty$,

or $\rho(\tau) \sim (c_2 \tau)^{-\beta}$ for some $0 < \beta \leq 1$, $c_2 > 0$. \hfill (3.16)

The contributions to the asymptotic variance of $\hat{f}^{(j)}$ due to short term dependence are designated $A_j(x)$ which is a function of the true underlying marginal density, and $\delta$ which is a function of the bandwidth, $h$, and length of the stochastic process, $T$. Similar terms arising from long term dependence are designated $B_j(x)$ and $\epsilon$. We
define $f_r$ to be the bivariate Normal density when both marginal densities are standard Normal and the correlation coefficient equals $\rho(\tau)$. The forms of the components of the variance of $\hat{f}^{(i)}$ are given in Tables 3.1 and 3.2.

Table 3.1: Asymptotic variance components of $\hat{f}^{(i)}$ due to short term dependence.

<table>
<thead>
<tr>
<th>Conditions on $\alpha$ and $\rho$</th>
<th>$A_j(x)$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 &lt; \alpha &lt; (j + \frac{1}{2})^{-1}, f</td>
<td>\rho</td>
<td>&lt; \infty$</td>
</tr>
<tr>
<td>$0 &lt; \alpha &lt; (j + \frac{1}{2})^{-1}, f</td>
<td>\rho</td>
<td>= \infty$</td>
</tr>
<tr>
<td>$\alpha = (j + \frac{1}{2})^{-1}$</td>
<td>$\pi^{-1} \left( \int_0^\infty \exp\left{ -(c_1 \tau)^a \right} f(x) \right)$</td>
<td>$(c_1 T)^{-1}</td>
</tr>
<tr>
<td>$(j + \frac{1}{2})^{-1} &lt; \alpha \leq 2$</td>
<td>$\pi^{-1} \left( \int_0^\infty \exp\left{ -(c_1 \tau)^a \right} f(x) \right) \times \left{ \int_0^\infty u^{2(j-1/\alpha)} \kappa(u)^2 du \right}$</td>
<td>$(c_1 T)^{-1} \times \frac{h^{2(\alpha-2)(j+1/2)}}{\alpha}$</td>
</tr>
</tbody>
</table>

Table 3.2: Asymptotic variance components of $\hat{f}^{(i)}$ due to long term dependence.

<table>
<thead>
<tr>
<th>Conditions on $\beta$ and $\rho$</th>
<th>$B_j(x)$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f</td>
<td>\rho</td>
<td>= \infty, 0 &lt; \beta &lt; 1$</td>
</tr>
<tr>
<td>$f</td>
<td>\rho</td>
<td>= \infty, \beta = 1$</td>
</tr>
<tr>
<td>$f</td>
<td>\rho</td>
<td>&lt; \infty, \beta &gt; 1$</td>
</tr>
</tbody>
</table>

**Theorem 3.1** Let $X$ be a stationary Gaussian process satisfying (3.15) and (3.16) and assume that $h = h(T) \rightarrow 0$ as $T \rightarrow \infty$. Then

$$\text{var}\left\{ \hat{f}^{(j)}(x) \right\} = A_j(x) \delta + B_j(x) \epsilon + o(\delta + \epsilon) \quad \text{as} \quad T \rightarrow \infty,$$

(3.17)

where $A_j(x)$, $B_j(x)$, $\delta$ and $\epsilon$ are the variance components defined in Tables 3.1 and 3.2.
Results that arise from this theorem shed light on how the dependence amongst the data of the stochastic process influence the rate at which \( \hat{f}^{(j)} \to f^{(j)} \). We discuss these results now by considering how different situations of short and long range dependence affect the asymptotic variance of \( \hat{f}^{(j)} \) through the variance components listed in Tables 3.1 and 3.2.

I. The quantity \( \lambda_T = T^{-1} \int_0^T |\rho(\tau)| d\tau \) converges to zero at precisely the rate \( T^{-1} + \epsilon \) which can be demonstrated with an example. The covariance function \( \rho(\tau) \) may be thought of as including separate components that describe the short term and long term dependence. Suppose that the short term dependence is modelled by

\[
\rho_s(\tau) = \exp(-\tau^\alpha), \quad 0 < \alpha \leq 2
\]

and the long term dependence by

\[
\rho_l(\tau) = (1 + \tau^2)^{-\beta/2}, \quad 0 < \beta \leq 1.
\]

A covariance function which is a combination of \( \rho_s(\tau) \) and \( \rho_l(\tau) \) such as,

\[
\rho(\tau) = \rho_s(\tau) + \rho_l(\tau),
\]

exhibits the asymptotic conditions (3.15) and (3.16) (with \( c_1 = c_2 = 1 \)) that we require for \( \rho(\tau) \).

For this example,

\[
\lambda_T = T^{-1} \int_0^T |\rho_s(\tau) + \rho_l(\tau)| d\tau \\
\leq T^{-1} \left\{ \int_0^T |\rho_s(\tau)| d\tau + \int_0^T |\rho_l(\tau)| d\tau \right\}.
\]

88
We evaluate the contribution from the short term component by

\[ \int_0^T |\rho_s(\tau)| d\tau = \int_0^T \exp(-\tau^\alpha) d\tau < \int_0^\infty \exp(-\tau^\alpha) d\tau = O(T^{-1}) \]

since the integral over the range \((0, \infty)\) is a constant. The contribution from the long term component is

\[ T^{-1} \int_0^T |\rho_l(\tau)| d\tau \leq T^{-1} \int_0^T (1 + \tau^2)^{-\beta/2} d\tau + T^{-1} \int_1^\infty \tau^{-\beta} d\tau \]

\[ = O(T^{-1}) + \begin{cases} O(T^{-\beta}) & \text{if } 0 < \beta < 1 \\ O(T^{-1} \log T) & \text{if } \beta = 1 \end{cases} \]

\[ = O(T^{-1}) + O(\epsilon). \]

Therefore \( \lambda_T = O(T^{-1} + \epsilon) \) in this case.

We now establish conditions for which \( \text{var} \{ \hat{f}^{(j)}(x) \} = O(\lambda_T) \) when the parameters are from the categories represented in Tables 3.1 and 3.2. The functions \( A_j(x) \) and \( B_j(x) \) are evaluated at fixed \( x \).

(a) When \( 0 < \alpha < (j + \frac{1}{2})^{-1} \) and either \( 0 < \beta < 1 \) or \( \beta = 1 \),

\[ \text{var} \{ \hat{f}^{(j)}(x) \} = A_j(x)T^{-1} + B_j(x)\epsilon + o(\delta + \epsilon) \]

\[ \leq \max \{ A_j(x), B_j(x) \} (T^{-1} + \epsilon) + o(\delta + \epsilon) \]

\[ = O(T^{-1} + \epsilon). \]

In this case the variance does not depend upon bandwidth and the same result applies whether \( 0 < \beta < 1 \) or \( \beta = 1 \) and \( \epsilon \) is either \( T^{-\beta} \) or \( T^{-1} \log T \) depending on the value of \( \beta \), see Table 3.2.
(b) (i) When $\alpha = (j + \frac{1}{2})^{-1}$ and $0 < \beta < 1$,

$$\text{var } \{ \hat{f}^{(j)}(x) \} = A_j(x)T^{-1}|\log h| + B_j(x)T^{-\beta} + o(\delta + \epsilon).$$

The components due to short term and long dependence have the same orders of magnitude if $h^{-1} = O\{\exp(CT^{1-\beta})\}$, where $C > 0$ is a constant. Then,

$$\text{var } \{ \hat{f}^{(j)}(x) \} = O(T^{-\beta}) = O(\epsilon).$$

(ii) When $\alpha = (j + \frac{1}{2})^{-1}$ and $\beta = 1$,

$$\text{var } \{ \hat{f}^{(j)}(x) \} = A_j(x)T^{-1}|\log h| + B_j(x)T^{-1}\log T + o(\delta + \epsilon)
= O(T^{-1}\log T) \quad \text{if } h^{-1} = O(T^{C_1}) \quad (C_1 > 0)
= O(\epsilon).$$

(c) When $(j + \frac{1}{2})^{-1} < \alpha \leq 2$ and $0 < \beta < 1$,

$$\text{var } \{ \hat{f}^{(j)}(x) \} = A_j(x)T^{-1}h^{2/\alpha-(2j+1)} + B_j(x)T^{-\beta} + o(\delta + \epsilon)
= O(T^{-\beta}) \text{ if } h^{-1} = O(T^{(1-\beta)/(2j+1)-(2/\alpha)})
= O(\epsilon).$$

We summarize these statements with the following remark. The quantity $\lambda_T$ converges to zero at precisely the rate $T^{-1} + \epsilon$, where $\epsilon$ is the variance component due to long dependence given in Table 3.2. Therefore $\text{var } \{ \hat{f}^{(j)}(x) \} = O(\lambda_T)$ if and only if one or other of the following holds:

(a) $0 < \alpha < (j + \frac{1}{2})^{-1}$;
(b) $\alpha = (j + \frac{1}{2})^{-1}$ and

(i) either $0 < \beta < 1$ and $h^{-1} = O\{\exp(CT^{1-\beta})\}$ for $C > 0$

(ii) or $\beta = 1$ and $h^{-1} = O(T^{C_1})$ for $C_1 > 0$；

(c) $(j + \frac{1}{2})^{-1} < \alpha \leq 2$, $0 < \beta < 1$ and $h^{-1} = O\{T^{(1-\beta)/(2j+1-2/\alpha)}\}$. 

II. The bandwidth, $h$, has no effect on $\epsilon$, nor on $\delta$ when $\alpha < (j + \frac{1}{2})^{-1}$. It only influences the asymptotic variance of $\hat{f}^{(j)}(x)$, through $\delta$, when $\alpha \geq (j + \frac{1}{2})^{-1}$. The situation when $\alpha < (j + \frac{1}{2})^{-1}$ gives a result for the first-order asymptotic variance of $\hat{f}^{(j)}(x)$ which is in striking contrast to the result for a kernel estimator for discrete data, in that the asymptotic variance of the estimator derived from continuous data does not depend on $h$, whereas its counterpart for discrete data does. The results proved by Masry [16] and presented as equations (3.1) and (3.2) show this latter result.

From equation (3.12) we see that the squared bias of $\hat{f}^{(j)}(x)$ is $O(h^{2r})$ and in practice, the kernel function, $K$, will always be of order $r \geq 2$. If $\alpha < (j + \frac{1}{2})^{-1}$, the asymptotic variance of $\hat{f}^{(j)}(x)$ is $O(T^{-1} + \epsilon)$ and the orders of magnitude of the mean squared error (MSE) are given by

$$\text{MSE}\{\hat{f}^{(j)}(x)\} = O(T^{-1} + \epsilon) + O(h^{2r}).$$

If $h$ is chosen so that it is $o(T^{-1/4})$, squared bias is $o(T^{-r/2})$ and so for $r \geq 2$, squared bias is of smaller order of magnitude than the variance and the value of $h$ has no effect on the first-order asymptotic properties of $\text{MSE}\{\hat{f}^{(j)}(x)\}$. 91
When \( \alpha < 2 \), either \( 0 < \beta < 1 \) or \( \beta = 1 \) and \( h = o(T^{-1/4}) \), the convergence rate of \( \hat{f} = \hat{f}^{(0)} \) is identical to that of the parametric estimator under the Normal model. For a Normal distribution, the mean \( (\mu) \) and standard deviation \( (\sigma) \) may generally be estimated at a rate of \( O_p(\lambda_T^{1/2}) \) and so the parametric estimator, \( \hat{f}(x; \bar{x}, s) \), converges to \( f(x; \mu, \sigma) \) at a rate \( O_p(\lambda_T^{1/2}) \). For the nonparametric estimator with \( h = o(T^{-1/4}) \), the bias contribution is of smaller order of magnitude than the variance and the rate of convergence of \( \hat{f} \) to \( f \) is determined by the variance. We previously showed that for \( \alpha < 2 \), \( \text{var}\{\hat{f}\} = O(\lambda_T) \). Therefore,

\[
E\{(\hat{f} - f)^2\} = O(\lambda_T),
\]

\[
\left[E\{(\hat{f} - f)^2\}\right]^{\frac{1}{2}} = O(\lambda_T^{1/2}) \text{ and}
\]

\( \hat{f} \to f \) at rate \( O_p(\lambda_T^{1/2}) \).

III. When \( \alpha \geq (j + \frac{1}{2})^{-1} \), inconsistency can result if \( h \) is chosen too small. In this instance,

\[
\text{var}\{\hat{f}(x)\} = A_j(x)T^{-1\alpha^{-2(j+2/3)}} + B_j(x)e + o(e + e).
\]

Then \( \hat{f}(x) \to f(x) \) in mean square if and only if \( h \to 0 \) and \( T^{-2j+1-2/\alpha} \to \infty \) as \( T \to \infty \).

IV. Under the conditions of Theorem 3.1, formula (3.17) admits the obvious extension to the integral of the variance,

\[
\int_{-\infty}^{+\infty} \text{var}\{\hat{f}(x)\} = \delta \int A_j(x) + e \int B_j(x) + o(e + e) \text{ as } T \to \infty.
\]

Variants of the results can be proved for a variety of stochastic processes related to Gaussian processes. Examples include sums of powers of Gaussian processes and
log-Gaussian processes. In the first of these cases, the pertinent values of $\alpha$ and $\beta$ are respectively the maximum and minimum of those quantities for the component Gaussian processes.

We give a result for stationary Gaussian processes, $X$, where the local dependence of $X_s$ and $X_t$ is sufficiently restricted when $s \neq t$, and for such Gaussian processes, the convergence rate of the variance of $\hat{f}^{(i)}$ is precisely $O(T^{-1})$. For the purpose of stating this result, let $f_r$ denote the joint bivariate Normal density of $X_t$ and $X_{t+r}$, and let $\psi_r$ be the corresponding characteristic function. Let $\psi$ be the characteristic function corresponding to the marginal Normal density $f$ and assume that $f_r, f$ are continuous at $(x, x), x$ respectively. We assume that either

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_0^{+\infty} |\theta_1 \theta_2|^j |\psi_r(\theta_1, \theta_2) - \psi(\theta_1)\psi(\theta_1)| d\tau d\theta_1 d\theta_2 < \infty$$

(3.18)

or that

$$\int_0^{+\infty} \left\{ \sup_{u, v} |f_r(u, v) - f(u)f(v)| \right\} d\tau < \infty .$$

(3.19)

**Theorem 3.2** Let $X$ be a stationary Gaussian process satisfying (3.18) or (3.19).

Then

$$\text{var}\{\hat{f}^{(i)}(x)\} = 2T^{-1} \int_0^{\infty} \left\{ f_r^{(j,j)}(x) - f^{(i)}(x)^2 \right\} d\tau + o(T^{-1}) .$$

Masry [16] explains that for a Gaussian process with $\rho(\tau) = 1 - |\tau|^\alpha + o(|\tau|^\alpha)$ as $\tau \downarrow 0$, condition (3.19) can only be obtained if $0 < \alpha < 2$. We elaborate this point by the example of a Gaussian process with

$$f_r(u, v) = (2\pi)^{-1} \{1 - \rho(\tau)^2\}^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \{1 - \rho(\tau)^2\}^{-1} \{u^2 + v^2 - 2uv\rho(\tau)\} \right] .$$
Condition (3.19) requires that $f_T(u,v)$ be integrable in $\tau$ over $[0,\nu]$ for some $\nu > 0$, or that

$$\int_0^\nu \{1 - \rho(\tau)^2\}^{-1/2} d\tau < \infty .$$

After substituting for $\rho(\tau)$, we would need that

$$\int_0^\nu \tau^{-\alpha/2} (2 - \tau^\alpha)^{-1/2} d\tau < \infty ,$$

and this does hold for $0 < \alpha < 2$ but cannot hold for $\alpha = 2$. Thus in the Gaussian case, Theorem 3.2 would not hold for smooth processes where closely spaced samples are highly correlated.

Conditions (3.18) and (3.19) restrict the class of Gaussian processes to those where the covariance function decays sufficiently fast so that there is no effect of long term dependence on the variance of the estimator.

This theorem is quite like Theorem 5.2 of Castellana and Leadbetter [3] but goes further in that it also applies to estimators of the derivatives of the density. Theorem 3.2 is stated for Gaussian processes but the same rate of convergence is possible for other stationary stochastic processes where the above conditions on the density and local dependence also apply.
3.4 Simulation Study of Kernel Density Estimators of a Gaussian Process

The aim of this study is to investigate how the accuracy of the estimators is influenced by the degree of roughness of the Gaussian process and numerically elucidate the asymptotic theory. The simulation study encompasses a wide range of situations where the parameters of the covariance function, $\alpha$ and $\beta$, are chosen to represent situations that give rise to the different forms of the variances of the estimators given in Tables 3.1 and 3.2. We restrict the study to processes $X$ where $E(X) = 0$ and $\text{var}(X) = 1$ so that we can focus on the effect of parameters $\alpha$ and $\beta$.

We develop rules for choosing bandwidths that are satisfactory for estimating the density, the first and second derivatives of the density and the mode and then compare the sample variances and squared errors of the kernel estimators with the asymptotic variances and squared errors established in Theorem 3.1 and the expression for bias given by (3.12).

3.4.1 Generating the Data for a Gaussian Process

We generate data with a covariance function

$$\rho(\tau) = b_s \rho_s(\tau) + b_l \rho_l(\tau), \text{ where}$$

(3.20)

$$\rho_s(\tau) = \exp\{-c_1 \tau^2\}, \text{ representing the short term dependence and}$$

(3.21)

$$\rho_l(\tau) = \{1 + (c_2 \tau)^2\}^{-\beta/2}, \text{ representing the long term dependence.}$$

(3.22)
This discreteness will not invalidate our simulations. In practice, all measurements are discrete if the measurement scale is fine enough and the measurement instrument for the surface roughness data also records the data as discrete entries on a fine sampling grid. In our simulations, we do not pick up extra noise from a sampling process and the points are so close so as to define a dependence structure that is consistent with that of a continuous process.
The constants $b_1, b_t (b_t = 1 - b_1)$ give the proportions of short and long term dependence in the covariance function and $c_1, c_2$ have been described before as scaling constants to account for the units of measurement of $t$. We set $c_1 = c_2 = 1$ for the purposes of these simulations but in practice, the scale on which short term dependence is measured may be quite different to that of long term dependence. In that case $c_1 \neq c_2$.

Whilst we cannot simulate data that are entirely continuous, we can simulate discrete data on a fine grid with the appropriate covariances that will reflect the properties of continuous data. To do this, we construct a variance-covariance matrix $A$ of dimension $d$ where the diagonal elements are 1 (or $\rho(0)$) and the elements of the $j$th super and sub diagonals are $\rho(\tau_j+1)$ where $\{\tau_j\}_{j=1,d}$ are values from the fine grid of points on the interval $(0, T)$.

The computer had 16 megabytes of memory and this limited the dimensions of $A$ to $(1200, 1200)$. With the grid of $\tau$ values at intervals of $\frac{1}{3}$ and with $c_1 = c_2 = 1$, we are able to consider stochastic processes up to lengths $T = 400$ on an arbitrary scale. Thus for given values of $b_1, b_t, \alpha$ and $\beta$, we derive $\rho(\tau)$ for $\tau = 0, 0.33, 0.67, \ldots 400$; that is, $d = 1201$.

In Section 3.5, we describe the analysis of the heights of surfaces where 1150 readings were recorded over a length of $4\text{mm}$ and so the sample sizes in our simulations are commensurate with those used in practice and we have merely changed our units of measurement with the constants $c_1, c_2$.

We define $Z$ to be a vector of length $d$ of independent normal random variables.
with \( E(Z) = 0 \) and \( \text{var}(Z) = \sigma^2 \) that is generated by standard simulation techniques. The stochastic process with covariance function \( \sigma^2 \rho \) can be generated by \( X = L'Z \) where is \( L \) such that \( L'L = A \) and is calculated by the Cholesky decomposition of \( A \).

This may be verified by considering the expected value and variance of \( X \):

\[
E(X) = E(L'Z) = L'E(Z) = 0
\]

\[
\text{cov}(X) = E(L'ZZ'L) = L'E(ZZ')L = \sigma^2 L'L = \sigma^2 A,
\]

which is the required Gaussian process.

By choosing different values of the parameters \( \alpha \) and \( \beta \) we can generate a wide range of stochastic processes representing varying degrees of roughness. Variance functions with values of \( \alpha = 0.6, \frac{2}{3}, 1.0, 1.95 \) and \( \beta = 0.6, 0.8, 1.0 \) provide examples of the different situations for the variance of estimators of the marginal density and its derivatives. Ideally we would also like to simulate data from a covariance function with \( \alpha = 2 \) but the numerical rounding by the computer often gives \( A \) as not positive definite. The choice of \( \alpha = 1.95 \) leads to data that are smooth enough to approximate the case when \( \alpha = 2 \) and allow comparison of the convergence rates of \( \hat{f} \) for most of the range of possible values of \( \alpha \).

We consider combinations of short term and long term dependence where the proportions of short term to long term are \((1, 0), (0.9, 0.1), (0.8, 0.2)\).

Shorter versions of \( X_t \) are obtained by first generating \( X_t, t \in (0, 400), \) and then selecting a subsequence of points from these data such as the data over the interval \((0, 200)\). From the simulated data, we estimate the marginal density \( \hat{f}(x) \), its first
derivative, \( \hat{f}'(x) \), and its second derivative \( \hat{f}''(x) \), at each point \( x \) from a grid of 201 equally spaced values from the interval \((-3.5\sigma, 3.5\sigma)\).

### 3.4.2 Choosing the Bandwidth

One of our aims in this study is to derive rules that give satisfactory bandwidths to estimate the density, its derivative and the mode by kernel methods. For the kernel density estimator defined at (3.10), we use the standard normal density kernel, 

\[
K(u) = \left(2\pi\right)^{-\frac{1}{2}} e^{-\frac{1}{2}u^2},
\]

which is a second order kernel function that is smooth enough to enable us to construct any \( \hat{f}^{(j)} \).

The mean squared error, \( E\{\hat{f}^{(j)}(x) - f^{(j)}(x)\}^2 \), is an expression of the average closeness of \( \hat{f}^{(j)} \) to \( f^{(j)} \) and an appropriate bandwidth is one such that the mean squared error is close to its minimum value.

The curvature of the density will be influential in mode estimation and \( f'' \) conveys a measure of curvature. The estimate of the mode can be determined as \( \hat{\theta} \) such that \( \hat{f}''(\hat{\theta}) = \min_x \hat{f}''(x) \) and the bandwidth for mode estimation will be the \( h \) such that 

\[
E\{\hat{f}''(\hat{\theta}) - f''(\theta)\}^2
\]

is minimised.

Monte-Carlo methods provide means of determining appropriate bandwidths and our procedure is as follows.

(i) We define a grid of candidate values for \( h \) ranging from \( h = 0.1 \) to \( h = 0.6 \).

(ii) We simulate a Gaussian process of length 400 with fixed \( b_i, c_i, \alpha \) and \( \beta \).

(iii) For subsamples of lengths \( T = 100, 200, 300, 400 \) we calculate \( \hat{f}(x) \), \( \hat{f}'(x) \) and
\( \hat{f}''(x) \) with \( h = 0.1 \) and then repeat the estimation for each of these subsets of data using the other values of \( h \) from the grid.

(iv) We estimate the mode, \( \hat{\theta} \), from \( \hat{f}''(x) \) and hence \( \hat{f}''(\hat{\theta}) \).

(v) As \( f, f' \) and \( f'' \) are known, it is possible to calculate the integrated squared error (ISE),

\[
ISE_j = \int \left\{ \hat{f}^{(j)}(x) - f^{(j)}(x) \right\}^2 dx \quad , j = 0, 1, 2,
\]

and note for which value of \( h \) the ISE was least. We also calculate the squared error of \( \hat{f}'' \) at the sample mode and record the grid value of \( h \) where that squared error was least.

(vi) Steps (iii), (iv) and (v) are repeated for one hundred independent simulated data sets and the frequency table for the optimum \( h \) is constructed.

(vii) The mean value of \( h \) from the frequency table (\( h_{\text{min}} \)) is taken as the best for the estimator of marginal density (or its derivatives) for that Gaussian process.

For a given set of parameters \( h, c_i, \alpha \) and \( \beta \), we have a profile of how the value of \( T \) influences the choice of \( h \). An example of results from one set of Monte-Carlo simulations with \( \alpha = 0.6, \beta = 1, b_s = 0.9 \) is presented in Table 3.3.
Table 3.3: Frequencies of $h$ values that are optimal for simulated data.

| $h$ | $T$ | 100 | 200 | 300 | 400 | $T$ | 100 | 200 | 300 | 400 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.2 | 4   | 13  | .26 | 38  | 0   | 0   | 0   | 0   | 0   | 0   |
| 0.3 | 39  | 53  | 54  | 45  | 0   | 2   | 13  | 32  | 0   | 0   |
| 0.4 | 38  | 22  | 14  | 17  | 35  | 67  | 72  | 60  | 0   | 0   |
| 0.5 | 15  | 11  | 6   | 0   | 54  | 29  | 14  | 8   | 0   | 0   |
| 0.6 | 4   | 1   | 0   | 0   | 11  | 2   | 1   | 0   | 0   | 0   |

$\bar{h}_{\text{min}}$ 0.38 0.33 0.30 0.28 

A plot of $\log(\bar{h}_{\text{min}})$ against $\log(T)$ and the least squares fits for the above data are shown in Figure 3.4 and reveals a linear relationship between the log of optimal bandwidth and $\log T$ for the estimators $\hat{f}$, $\hat{f}'$.

Figure 3.4: Log-linear relationship between optimal bandwidth and sample size.

The linear relationship was confirmed in the Monte-Carlo simulations of a diverse
set of Gaussian processes, including those where \( \text{var}(X) = \sigma^2, \sigma^2 \neq 1 \). If we propose a rule that \( \hat{h} = \sigma DT^{-p} \), where \( D \) is a constant, then the slope from least squares fits such as done in Figure 3.4 provides an estimate of \( p \) and an estimate of \( D \) can be obtained by \( \hat{D} = \exp\{\text{intercept} - \log(\sigma)\} \).

The MSE is often minimized when \( h \) is chosen such that the squared bias and variance are of the same order and in this regard, we consider our rules for bandwidth selection.

For \( \hat{f} \), \( MSE(\hat{f}) \approx O\{(c_1T)^{-1}\} + O(h^4) \) and the choice of \( h = DT^{-1.25} \) makes the squared bias and variance terms approximately equal.

We consider the three situations for \( \alpha \) that occur with \( \hat{f} \). If \( 0 < \alpha < \frac{2}{3} \), we have the same result as above that \( h = DT^{-1.25} \) would be appropriate. For \( \alpha = \frac{2}{3} \), \( MSE(\hat{f}) = O\{(c_1T)^{-1} \log h\} + O(h^4) \) and squared bias will be of the same order of magnitude as variance if \( h \approx (c_1T)^{-0.25} \{\log(c_1T)\}^{-2.25} \). When \( \frac{2}{3} < \alpha \leq 2 \), \( MSE(\hat{f}) = O\{(c_1T)^{-1} h^{(2/\alpha)-3}\} + O(h^4) \), suggesting that \( h \) be chosen as \( D_1 T^{-1/(7-2/\alpha)} \), or in the range \( D_1^\star T^{-1.17} \) to \( D_1^\dagger T^{-1.15} \), with \( D_1, D_1^\star, D_1^\dagger \) being constants.

For bandwidth choice for \( \hat{f}'' \), consider the situation when \( 0.4 < \alpha \leq 2 \), \( MSE(\hat{f}'') = O\{(c_1T)^{2/\alpha-5}\} + O(h^4) \) and an appropriate bandwidth would be given by \( h = D_2 T^{-0.12} \) (\( D_2 \) a constant).

The optimum bandwidth for mode estimation is that which minimises the squared error of the second derivative at the sample mode. We do not present any theoretical results for this but our simulations suggest that this bandwidth should be wider than the bandwidth for estimating the density or first derivative but narrower than that.
for the second derivative.

Our results from the simulations to estimate bandwidth, given in Tables 3.4 and 3.5, pages 103, 104, are consistent with the above principles. We propose that for the estimator \( \hat{f} \), we use \( h_0 = 1.2\sigma T^{-0.25} \); for \( \hat{f}' \), \( h_1 = \sigma T^{-0.17} \); for \( \hat{f}'' \), \( h_2 = \sigma T^{-0.12} \); and for the mode \( h_3 = 0.7\sigma T^{-0.09} \). For estimation of \( f' \) when \( \alpha \leq \frac{2}{3} \) and \( f' \) when \( \alpha \leq \frac{2}{5} \), our bandwidth rule will produce a bandwidth that is slightly wider than the optimum bandwidth but there is little loss of precision in these instances.

In Table 3.6 on page 105, the optimal bandwidths from our simulations are listed and we can see that the rules are good approximations in most cases.

We do not take a fixed position on the use of these rules in that we may prefer to take \( h = o(T^{-\frac{1}{4}}) \) as suggested in Section 3.3.5. Then the error in estimating the density can be ascribed to the variance of \( \hat{f} \), the bias being negligible (even though the ISE is not minimum). However our rules will be useful when analysing data where \( \alpha \) is unknown and the nature of the variance function for \( \hat{f}' \) and \( \hat{f}'' \) cannot be specified.

By taking \( h = o(T^{-\frac{1}{4}}) \), we get that \( \hat{f} \to f \) at the same rate as the parametric estimator. Yet the choice of \( h = O(T^{-\frac{1}{4}}) \), where squared bias and variance are of equal orders of magnitude, leads to a smaller ISE and bandwidths based on this rule may provide even faster convergence than the parametric estimator. This is not surprising since kernel density estimators use all the information coded in the data whereas parametric estimators are approximations based on the first two moments.
Table 3.4: Least squares fits of the linear relationship between log(\(T\)) and log(\(\hat{h}_{min}\)) for different proportions of short and long term dependence.

The constant \(\hat{D}\) equals \(\exp\{\text{intercept} - \log(\sigma)\}\)

and \(p\) is the slope of the line such that \(\hat{h} = \sigma DT^p\).

The proportion of short term dependence is \(b_s\).

<table>
<thead>
<tr>
<th>(b_s)</th>
<th>(\beta)</th>
<th>(\alpha)</th>
<th>(\hat{D})</th>
<th>(p)</th>
<th>(R^2)</th>
<th>(\hat{D}')</th>
<th>(p')</th>
<th>(R'^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>1.14</td>
<td>-0.22</td>
<td>0.99</td>
<td>1.0</td>
<td>-0.16</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.17</td>
<td>-0.23</td>
<td>0.98</td>
<td>1.02</td>
<td>-0.16</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.09</td>
<td>-0.21</td>
<td>1.0</td>
<td>1.03</td>
<td>-0.16</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.11</td>
<td>-0.20</td>
<td>1.0</td>
<td>1.01</td>
<td>-0.14</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>1.20</td>
<td>-0.24</td>
<td>0.95</td>
<td>1.02</td>
<td>-0.17</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.17</td>
<td>-0.24</td>
<td>0.97</td>
<td>1.03</td>
<td>-0.17</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.02</td>
<td>-0.21</td>
<td>0.99</td>
<td>0.97</td>
<td>-0.15</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.99</td>
<td>-0.19</td>
<td>0.99</td>
<td>0.96</td>
<td>-0.14</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1.0</td>
<td>0.6</td>
<td>1.19</td>
<td>-0.24</td>
<td>0.99</td>
<td>1.05</td>
<td>-0.17</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.22</td>
<td>-0.25</td>
<td>0.99</td>
<td>1.06</td>
<td>-0.17</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.07</td>
<td>-0.23</td>
<td>0.98</td>
<td>1.03</td>
<td>-0.17</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.98</td>
<td>-0.19</td>
<td>0.99</td>
<td>0.98</td>
<td>-0.14</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.6</td>
<td>0.6</td>
<td>1.17</td>
<td>-0.24</td>
<td>0.99</td>
<td>1.01</td>
<td>-0.16</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.17</td>
<td>-0.24</td>
<td>0.98</td>
<td>1.02</td>
<td>-0.17</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.12</td>
<td>-0.23</td>
<td>1.0</td>
<td>1.06</td>
<td>-0.17</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.10</td>
<td>-0.21</td>
<td>0.99</td>
<td>1.06</td>
<td>-0.16</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>0.6</td>
<td>1.18</td>
<td>-0.24</td>
<td>0.98</td>
<td>1.02</td>
<td>-0.17</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.19</td>
<td>-0.25</td>
<td>0.99</td>
<td>1.10</td>
<td>-0.18</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.11</td>
<td>-0.23</td>
<td>0.99</td>
<td>1.10</td>
<td>-0.18</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.16</td>
<td>-0.23</td>
<td>0.99</td>
<td>1.08</td>
<td>-0.16</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>1.0</td>
<td>0.6</td>
<td>1.19</td>
<td>-0.25</td>
<td>0.99</td>
<td>1.05</td>
<td>-0.17</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.21</td>
<td>-0.25</td>
<td>0.99</td>
<td>1.11</td>
<td>-0.19</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.11</td>
<td>-0.23</td>
<td>0.99</td>
<td>1.12</td>
<td>-0.19</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.17</td>
<td>-0.23</td>
<td>0.99</td>
<td>1.10</td>
<td>-0.17</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>0.6</td>
<td>1.15</td>
<td>-0.25</td>
<td>0.98</td>
<td>1.09</td>
<td>-0.19</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.19</td>
<td>-0.25</td>
<td>0.99</td>
<td>1.07</td>
<td>-0.18</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.14</td>
<td>-0.25</td>
<td>0.99</td>
<td>1.10</td>
<td>-0.19</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.21</td>
<td>-0.24</td>
<td>0.99</td>
<td>1.10</td>
<td>-0.17</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.5: Least squares fits of the linear relationship between $\log(T)$ and $\log(h_{\text{min}})$ for different proportions of short and long term dependence.

The constant $D$ equals $\exp\{\text{intercept} - \log(\sigma)\}$ and $p$ is the slope of the line such that $\tilde{h} = \sigma DT^p$.

The proportion of short term dependence is $b_s$.

<table>
<thead>
<tr>
<th>$b_s$</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$D$</th>
<th>$p$</th>
<th>$R^2$</th>
<th>$\tilde{h}$</th>
<th>$D$</th>
<th>$p$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>0.88</td>
<td>-0.10</td>
<td>0.98</td>
<td>0.66</td>
<td>-0.09</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.6</td>
<td>0.92</td>
<td>-0.11</td>
<td>1.0</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.6</td>
<td>0.96</td>
<td>-0.12</td>
<td>0.99</td>
<td>0.67</td>
<td>-0.07</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.6</td>
<td>0.81</td>
<td>-0.08</td>
<td>0.96</td>
<td>0.73</td>
<td>-0.09</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.87</td>
<td>-0.10</td>
<td>0.97</td>
<td>0.71</td>
<td>-0.10</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.8</td>
<td>0.87</td>
<td>-0.10</td>
<td>0.95</td>
<td>0.71</td>
<td>-0.10</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.8</td>
<td>0.85</td>
<td>-0.09</td>
<td>0.97</td>
<td>0.66</td>
<td>-0.08</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.8</td>
<td>0.78</td>
<td>-0.07</td>
<td>1.0</td>
<td>0.74</td>
<td>-0.09</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1.0</td>
<td>0.6</td>
<td>0.91</td>
<td>-0.11</td>
<td>0.99</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.0</td>
<td>0.92</td>
<td>-0.11</td>
<td>0.97</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.8</td>
<td>0.85</td>
<td>-0.09</td>
<td>0.96</td>
<td>0.69</td>
<td>-0.09</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.0</td>
<td>0.79</td>
<td>-0.07</td>
<td>1.0</td>
<td>0.73</td>
<td>-0.09</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.6</td>
<td>0.6</td>
<td>0.92</td>
<td>-0.11</td>
<td>0.97</td>
<td>0.68</td>
<td>-0.09</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.6</td>
<td>0.94</td>
<td>-0.12</td>
<td>0.99</td>
<td>0.67</td>
<td>-0.08</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.9</td>
<td>0.97</td>
<td>-0.12</td>
<td>0.99</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.9</td>
<td>0.84</td>
<td>-0.09</td>
<td>0.99</td>
<td>0.71</td>
<td>-0.09</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>0.6</td>
<td>0.93</td>
<td>-0.11</td>
<td>0.98</td>
<td>0.68</td>
<td>-0.09</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.6</td>
<td>0.97</td>
<td>-0.12</td>
<td>1.0</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.9</td>
<td>0.99</td>
<td>-0.13</td>
<td>0.99</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.9</td>
<td>0.84</td>
<td>-0.08</td>
<td>1.0</td>
<td>0.69</td>
<td>-0.08</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>1.0</td>
<td>0.6</td>
<td>0.95</td>
<td>-0.12</td>
<td>0.99</td>
<td>0.70</td>
<td>-0.10</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>1.0</td>
<td>0.91</td>
<td>-0.11</td>
<td>1.0</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.9</td>
<td>0.99</td>
<td>-0.13</td>
<td>0.97</td>
<td>0.65</td>
<td>-0.08</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.0</td>
<td>0.86</td>
<td>-0.09</td>
<td>1.0</td>
<td>0.69</td>
<td>-0.09</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>0.9</td>
<td>0.89</td>
<td>-0.11</td>
<td>0.99</td>
<td>0.69</td>
<td>-0.10</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.9</td>
<td>0.86</td>
<td>-0.10</td>
<td>0.96</td>
<td>0.70</td>
<td>-0.10</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.9</td>
<td>1.0</td>
<td>-0.13</td>
<td>0.98</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.9</td>
<td>0.87</td>
<td>-0.09</td>
<td>0.99</td>
<td>0.72</td>
<td>-0.10</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.6: Comparisons of bandwidths that minimise the ISE and the bandwidths given by the rules for $T = 400$ and $\sigma = 1$.

<table>
<thead>
<tr>
<th>$b_s$</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$\tilde{h}_0 = 0.27$</th>
<th>$\tilde{h}_1 = 0.36$</th>
<th>$\tilde{h}_2 = 0.48$</th>
<th>$\tilde{h}_3 = 0.41$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\hat{f}_{opt}$</td>
<td>$\hat{f}_{opt}$</td>
<td>$\hat{f}_{opt}$</td>
<td>$\hat{f}_{opt}$</td>
</tr>
<tr>
<td>0.8</td>
<td>0.60</td>
<td>0.60</td>
<td>0.31</td>
<td>0.40</td>
<td>0.48</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.67</td>
<td>0.31</td>
<td>0.40</td>
<td>0.48</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>0.31</td>
<td>0.39</td>
<td>0.48</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.95</td>
<td>0.33</td>
<td>0.42</td>
<td>0.51</td>
<td>0.42</td>
</tr>
<tr>
<td>0.8</td>
<td>0.60</td>
<td>0.60</td>
<td>0.30</td>
<td>0.39</td>
<td>0.48</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.67</td>
<td>0.30</td>
<td>0.39</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>0.29</td>
<td>0.39</td>
<td>0.47</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.95</td>
<td>0.32</td>
<td>0.41</td>
<td>0.49</td>
<td>0.42</td>
</tr>
<tr>
<td>1.0</td>
<td>0.60</td>
<td>0.60</td>
<td>0.29</td>
<td>0.38</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.67</td>
<td>0.29</td>
<td>0.38</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>0.29</td>
<td>0.38</td>
<td>0.47</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.95</td>
<td>0.31</td>
<td>0.41</td>
<td>0.51</td>
<td>0.41</td>
</tr>
<tr>
<td>0.9</td>
<td>0.60</td>
<td>0.60</td>
<td>0.29</td>
<td>0.39</td>
<td>0.47</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.67</td>
<td>0.29</td>
<td>0.38</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>0.29</td>
<td>0.38</td>
<td>0.46</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.95</td>
<td>0.31</td>
<td>0.41</td>
<td>0.50</td>
<td>0.41</td>
</tr>
<tr>
<td>0.8</td>
<td>0.60</td>
<td>0.60</td>
<td>0.29</td>
<td>0.38</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.67</td>
<td>0.28</td>
<td>0.38</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>0.28</td>
<td>0.38</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.95</td>
<td>0.31</td>
<td>0.41</td>
<td>0.50</td>
<td>0.41</td>
</tr>
<tr>
<td>1.0</td>
<td>0.60</td>
<td>0.60</td>
<td>0.28</td>
<td>0.38</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.67</td>
<td>0.28</td>
<td>0.38</td>
<td>0.48</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>0.27</td>
<td>0.38</td>
<td>0.47</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.95</td>
<td>0.31</td>
<td>0.40</td>
<td>0.50</td>
<td>0.41</td>
</tr>
<tr>
<td>1.0</td>
<td>0.60</td>
<td>0.60</td>
<td>0.28</td>
<td>0.38</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.67</td>
<td>0.27</td>
<td>0.38</td>
<td>0.48</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>0.27</td>
<td>0.37</td>
<td>0.47</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>1.95</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.41</td>
</tr>
</tbody>
</table>
3.4.3 Numerical Study of $L^2$ Convergence of Kernel Estimators

For a particular Gaussian process with $\rho(\tau)$ specified, 100 data sets with $E(X) = 0$, $\text{var}(X) = 1$, $T = 100, 200, 300, 400$ were simulated and for each data set the estimators $\hat{f}(x)$, $\hat{f}'(x)$ and $\hat{f}''(x)$ were calculated at every point in the grid of 201 values over the interval $(-3.5, 3.5)$. The bandwidth rules discussed in the previous section were used for these estimators. From these simulated data sets, the sample variance of $\hat{f}^{(i)}$ is calculated by

\[
\text{var}\{\hat{f}^{(i)}(x)\} = \left[ \sum_{s=1}^{NS} \left( \hat{f}^{(i)}(x) - \sum_{s=1}^{NS} \hat{f}^{(i)}(x) / NS \right)^2 / (NS - 1) \right] / (NS - 1)
\]

where $\hat{f}^{(i)}(x)$ is the estimate at $x$ for simulation number $s$ out of $NS = 100$.

The comparisons of sample integrated variances and integrated squared errors with the corresponding asymptotic results for a sample of size $T = 400$ are presented in Table 3.8 (page 111, integrated variances) and Table 3.9 (page 112, integrated squared errors).

Provided that the influence of the long term dependence is not too strong, the sample integrated variance of $\hat{f}$ is close to the asymptotic integrated variance. As the long term dependence is increased, we notice increasing disparity between the sample and asymptotic variances with the sample variances being less than their asymptotic counterparts. However, there is closer agreement between the sample and asymptotic integrated squared errors. Since asymptotic biases are negligible, we may conclude that with a sample much larger than $T = 400$ and with the corresponding reduction in the bandwidth and bias, the sample variances would reflect the asymptotic variances.
An arbitrary reduction in the bandwidth without increase in sample size may improve the agreement between sample and asymptotic variances but this is not optimal and would lead to a degradation of the estimator $\hat{f}$.

Our simulations show that samples of size $T = 400$ are not sufficiently large to detect the effect of long term dependence on the variance of $\hat{f}$. The surface data that we analyse in Section 3.5 is of similar sample size and so our analysis will be determined by the short term dependence.

The accuracy of estimates of derivatives of the density decreases as the order of differentiation increases. The sample variances of $\hat{f}'$ and $\hat{f}''$ are not particularly close to the asymptotic variances yet their ISE's compare well with their asymptotic MISE's.

These results confirm the asymptotic result that the rate at which $\hat{f}^{(j)} \rightarrow f^{(j)}$ becomes slower as the process becomes smoother.

We have investigated the performance of $\hat{f}'$ when the covariance function has parameter $\alpha = \frac{2}{3}$ but the process of simulating data is too coarse to give data that would reflect the theoretical results at this critical point.

We can illustrate the general results from the simulations with examples of three Gaussian processes where $\alpha = 0.6$, $\beta = 1$, $c_1 = c_2 = 1$ and $T = 400$ in each case but differing by the proportion of short term dependence; $b_s = 1.0, 0.9$ and 0.8. In the case of $b_s = 1.0$, there is no long term dependence and hence $\beta$ does not figure. In Figure 3.5 (page 109) we give examples of estimates of the variances of $\hat{f}(x)$ and $\hat{f}'(x)$, $x \in (-3.5, 3.5)$ for the three Gaussian processes based on 100 simulations. Figures
3.5 (a),(c),(e) are for the variance of \( \hat{f} \) when \( b_s = 1.0, 0.9, 0.8 \) respectively and Figures 3.5 (b),(d),(e) are the corresponding plots for the variance of \( \hat{f}' \). It is clear that the data do not detect the influence of long term dependence in estimating the variance of \( \hat{f} \) and the asymptotic nature of \( \text{var}(\hat{f}') \) is not apparent with \( T = 400 \).

The effect of sample size on the estimation of the integrated variances of \( \hat{f} \) and \( \hat{f}' \) can be gauged from the plots of integrated variances for Gaussian processes (\( \alpha = 0.6, \beta = 1, b_s = 0.8, 0.9, 1.0 \)) calculated at \( T = 100, 200, 300, 400 \) and plotted in Figure 3.6 on page 110.

Our simulations do not reveal any obvious pattern for the convergence of the sample mode (\( \hat{\theta} \)) to the population mode (\( \theta \)). The squared errors ((\( \hat{\theta} - \theta \))^2) are shown in Table 3.7.

Table 3.7: Squared errors for the sample mode of Gaussian processes.

<table>
<thead>
<tr>
<th>( b_s )</th>
<th>( \alpha )</th>
<th>( \beta = ) 0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.6</td>
<td>0.11</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.11</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.10</td>
<td>0.10</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>0.9</td>
<td>0.6</td>
<td>0.12</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.09</td>
<td>0.10</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.11</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>0.11</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.5: Profiles of sample variances of $\hat{f}$ and $\hat{f}'$.

\[ \alpha = 0.6, \beta = 1.0, T = 400. \]

(a) \( \text{var}(\hat{f}) \), \( b_s = 1.0 \), (b) \( \text{var}(\hat{f}') \), \( b_s = 1.0 \),
(c) \( \text{var}(\hat{f}) \), \( b_s = 0.9 \), (d) \( \text{var}(\hat{f}') \), \( b_s = 0.9 \),
(e) \( \text{var}(\hat{f}) \), \( b_s = 0.8 \), (f) \( \text{var}(\hat{f}') \), \( b_s = 0.8 \)

The unbroken lines represent the asymptotic variances and the dotted lines represent the sample variances.
Figure 3.6: Sample and asymptotic integrated variances of \( \hat{f} \) and \( \hat{f} \) for samples sizes \( T = 100, 200, 300, 400 \).

\[ \alpha = 0.6, \quad \beta = 1.0. \]

(a) \( \text{var}(\hat{f}), b_s = 1.0 \), (b) \( \text{var}(\hat{f}'), b_s = 1.0 \),

(c) \( \text{var}(\hat{f}), b_s = 0.9 \), (d) \( \text{var}(\hat{f}'), b_s = 0.9 \),

(e) \( \text{var}(\hat{f}), b_s = 0.8 \), (f) \( \text{var}(\hat{f}'), b_s = 0.8 \)

The unbroken lines represent the integrated asymptotic variances and the dotted lines represent the integrated sample variances.
Table 3.8: Asymptotic and sample integrated variances ($\times 10^4$) of kernel estimators for Gaussian processes of length $T=400$.

The asymptotic variances are shown in bold type.

<table>
<thead>
<tr>
<th>$b_1$</th>
<th>$\alpha$</th>
<th>$f \text{ var}(f)$ for $\beta = 0.6$</th>
<th>$f \text{ var}(f)$ for $\beta = 0.8$</th>
<th>$f \text{ var}(f)$ for $\beta = 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.6</td>
<td>71 54 32 114 106 94</td>
<td>85 73 56</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>38 28 22 63 53 46</td>
<td>257 220 193</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.8</td>
<td>70 53 31 42 35 24</td>
<td>86 74 57</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>39 26 19 66 56 45</td>
<td>265 240 223</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>68 52 30 51 41 33</td>
<td>106 94 77</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>42 27 22 74 45 51</td>
<td>306 251 247</td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td>0.8</td>
<td>69 53 31 49 40 29</td>
<td>180 168 151</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>36 29 20 58 51 40</td>
<td>338 324 300</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.6</td>
<td>44 36 25 104 100 94</td>
<td>57 51 42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>28 28 18 49 59 40</td>
<td>222 248 194</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.8</td>
<td>43 35 24 26 22 17</td>
<td>58 52 44</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>28 23 21 59 45 47</td>
<td>247 209 199</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>42 34 23 36 31 26</td>
<td>78 72 64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 25 18 50 49 49</td>
<td>217 225 230</td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td>0.8</td>
<td>44 36 25 39 30 27</td>
<td>151 145 137</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>28 23 18 61 42 49</td>
<td>320 285 284</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>19</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.8</td>
<td>17</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>16</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td>0.8</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>123</td>
<td></td>
</tr>
</tbody>
</table>

111
Table 3.9: Asymptotic and sample integrated squared errors ($\times 10^4$) of kernel estimators for Gaussian processes of length $T=400$.

The asymptotic squared errors are shown in bold type.

<table>
<thead>
<tr>
<th>$b_*$</th>
<th>$\alpha$</th>
<th>$ISE(f)$ for $\beta = 0.6$</th>
<th>$ISE(f)$ for $\beta = 0.8$</th>
<th>$ISE(f)$ for $\beta = 1.0$</th>
<th>$ISE(f'')$ for $\beta = 0.6$</th>
<th>$ISE(f'')$ for $\beta = 0.8$</th>
<th>$ISE(f'')$ for $\beta = 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.6</td>
<td>73 57 35</td>
<td>136 128 117</td>
<td>195 183 167</td>
<td>343 303 295</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>42 30 26</td>
<td>83 68 66</td>
<td>343 303 295</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.6</td>
<td>72 56 34</td>
<td>65 57 46</td>
<td>197 185 168</td>
<td>357 338 330</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>43 30 22</td>
<td>87 75 65</td>
<td>357 338 330</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>72 54 32</td>
<td>75 66 55</td>
<td>223 210 194</td>
<td>411 340 335</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>47 29 24</td>
<td>99 71 68</td>
<td>411 340 335</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td>0.6</td>
<td>72 55 32</td>
<td>80 72 61</td>
<td>315 302 286</td>
<td>437 427 406</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>41 33 27</td>
<td>93 85 79</td>
<td>437 427 406</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.6</td>
<td>47 39 28</td>
<td>126 122 116</td>
<td>166 160 152</td>
<td>322 348 300</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>31 318 21</td>
<td>72 78 61</td>
<td>322 348 300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.6</td>
<td>46 38 27</td>
<td>49 45 39</td>
<td>168 162 154</td>
<td>329 328 323</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>30 27 25</td>
<td>74 71 70</td>
<td>329 328 323</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>45 36 25</td>
<td>58 54 48</td>
<td>194 187 179</td>
<td>319 346 304</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>27 29 20</td>
<td>70 75 61</td>
<td>319 346 304</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td>0.6</td>
<td>47 39 28</td>
<td>64 60 55</td>
<td>284 278 270</td>
<td>406 379 386</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>32 27 22</td>
<td>82 74 72</td>
<td>406 379 386</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>21</td>
<td>117</td>
<td>139</td>
<td>302</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>61</td>
<td>302</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.6</td>
<td>20</td>
<td>33</td>
<td>142</td>
<td>292</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>58</td>
<td>292</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.6</td>
<td>19</td>
<td>42</td>
<td>166</td>
<td>315</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>62</td>
<td>315</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td>0.6</td>
<td>25</td>
<td>49</td>
<td>255</td>
<td>403</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>70</td>
<td>403</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

112
3.5 The Surface Roughness Data

We return to the data that motivated the study of kernel density estimators for continuous stochastic processes. These data are the electrical readings from the instrument that measures surface height above a reference level and we estimate the density and mode of the heights measured on a new, a treated and a worn surface. Plots of sections of these data were presented in Figure 3.1, page 67. Each data set comprises 1150 measurements of height at equally spaced points along a transect of length 4 millimetres.

We estimate the density, its derivatives and the mode using kernel estimators described previously by equation (3.10) and using the bandwidth rules suggested by our simulations in Section 3.4.2. These rules are that for data of length 4 millimetres (or 400 × 10^-5 metres) with sample standard deviation \( \hat{\sigma} \), we take \( \hat{h}_0 = 0.27 \hat{\sigma} \) to estimate the density, \( \hat{h}_1 = 0.36 \hat{\sigma} \) to estimate the first derivative, \( \hat{h}_2 = 0.48 \hat{\sigma} \) to estimate the second derivative and \( \hat{h}_3 = 0.41 \hat{\sigma} \) to estimate the mode by way of the second derivative.

Plots of these estimates are shown on pages 115 (new surface), 116 (treated surface) and 117 (worn surface). We also show the estimates that would be obtained parametrically by assuming that the data for the new and treated surfaces were Gaussian processes and the data for the worn surface were a chi-square process. The heights of the different surfaces are not directly comparable due to the data being recalibrated, either at the instrument level or after the initial measurements were
made.

We can see that for new and treated surfaces there is close agreement between the parametric and nonparametric estimators.

The marked skewness in the data for the worn surface is easily detected by the kernel estimators. The parametric density estimator for these data that we have plotted is a negative chi-square density function given by:

\[ f_d(x; \mu, \sigma) = \frac{1}{\sigma 2^{d/2} \Gamma(d/2)} e^{-x^2/(2\sigma^2)} \left(\frac{d}{2}\right)^{-d/2} e^{-d/2} \left(\frac{q}{2}\right)^{d-1} ; q = \frac{x + \mu}{\sigma}, \]

where \( d \) denotes the degrees of freedom. It is not readily obvious what values to assign to the parameters and this model with \( d = 20, \mu = 5000 \) and \( \sigma = 100 \), was deduced by changing the parameters until there was close visual agreement between the parametric and nonparametric density estimates. We can see the strength of the nonparametric estimators for situations such as this.

The estimation of the mode is done using a bandwidth that is smaller than that required for estimating the second derivative, leading to a deeper trough of the second derivative at the mode but at the expense of reliable estimation in the tails. However, this is of no consequence since the objective is to estimate the mode.
Figure 3.7: Estimates of density, derivatives and mode for a new surface.

Estimates for new surface

density

1st deriv

2nd deriv

mode

mode = 29420.
Figure 3.8: Estimates of density, derivatives and mode for a treated surface.

Estimates for treated surface

density

2nd deriv

1st deriv

mode

mode = 14969.
Figure 3.9: Estimates of density, derivatives and mode for a worn surface.

Estimates for worn surface

- **Density**
  - Chi-square
  - Nonparam.

- **1st deriv**

- **2nd deriv**

- **Mode**

mode = 3155.
3.6 Proofs

3.6.1 Proof of Theorem 3.1.

We get a general result for \( \text{var}\{\hat{f}^{(j)}\} \) and then consider the separate effects of short and long term dependence on the asymptotic variance. For short term dependence, \( \int |\rho(\tau)| d\tau < \infty \) and for long term dependence, \( \int |\rho(\tau)| d\tau = \infty \).

From (3.11),

\[
\begin{align*}
\left[ E\left\{ \hat{f}^{(j)}(x) \right\} \right]^2 & = \\
(Th^{j+1})^{-2} & \int_{0}^{T} \int_{0}^{T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^{(j)}(u_1)K^{(j)}(u_2)f(x-hu_1)f(x-hu_2)du_1du_2dt_1dt_2,
\end{align*}
\]

with \( f \) being the standard normal density.

For random integrals of the form \( L = \int g(t)\xi(t)dt \), where \( \xi(t) \) is a stationary process with covariance function \( \rho(t,u) \) and \( g(t) \) is a deterministic function,

\[
E(|L|^2) = \int \int g(t_1)g(t_2)\rho(t_1,t_2)dt_1dt_2 \quad \text{(see Cramér and Leadbetter [6, p.86ff])}.
\]

Using this result,

\[
\begin{align*}
E\left\{ \hat{f}^{(j)}(x)^2 \right\} & = \\
(Th^{j+1})^{-2} & \int_{0}^{T} \int_{0}^{T} \text{cov}\left(K^{(j)}\left[x-X(t_1)\right]/h\right)K^{(j)}\left[x-X(t_2)\right]/h) dt_1dt_2 \\
& = (Th^{j+1})^{-2}h^2 \int_{0}^{T} \int_{0}^{T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^{(j)}((x-y_1)/h)K^{(j)}((x-y_2)/h) \times \\
& \quad f_{t_1-t_2}(y_1,y_2)dy_1dy_2dt_1dt_2 \\
& = (Th^{j})^{-2} \int_{0}^{T} \int_{0}^{T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^{(j)}(u_1)K^{(j)}(u_2) \times \\
& \quad f_{t_1-t_2}(x-hu_1,x-hu_2)du_1du_2dt_1dt_2,
\end{align*}
\]
where \( f_T \) denotes the bivariate normal density when the marginal distributions are standard normal and the coefficient of correlation equals \( \rho(\tau) \). It follows that,

\[
\text{var} \{ \hat{f}^{(j)}(x) \} = (T h')^{-2} \int_0^T \int_0^T \int_{-\infty}^\infty \int_{-\infty}^\infty K^{(j)}(u_1) K^{(j)}(u_2) \times \\
\{ f_{1-\tau_2}(x - hu_1, x - hu_2) - f(x - hu_1)f(x - hu_2) \} du_1 du_2 dt_1 dt_2 . \tag{3.23}
\]

In the following sections, the underscored symbols represent vector quantities, (e.g. \( x = (x_1, x_2)^T \)).

Define the components of \( \text{var} \{ \hat{f}^{(j)}(x) \} \) in equation (3.23) as

\[
g(u) = (2\pi)^{-1} f_{1-\tau_2}(x - hu) \quad \text{and} \quad \\
h(u) = (2\pi)^{-1} K^{(j)}(u_1) K^{(j)}(u_2) .
\]

Write \( \theta = (\theta_1, \theta_2) \) and let \( G(\theta) \) and \( H(\theta) \) be the Fourier transforms of \( g(u) \) and \( h(u) \) respectively. A bivariate form of Parseval's identity gives that

\[
\int_{-\infty}^\infty \int_{-\infty}^\infty g(u)h(u)du_1 du_2 = \int_{-\infty}^\infty \int_{-\infty}^\infty G(\theta)H(\theta)d\theta_1 d\theta_2 . \tag{3.24}
\]

We now derive the quantities \( G(\theta) \) and \( H(\theta) \) and use the spectral representation, given by the right side of (3.24), to evaluate the integrals with respect to \( u_1 \) and \( u_2 \).

Let \( \Sigma \) denote the variance-covariance matrix of the distribution with density \( f \) and \( \mu = (x, x)^T \); that is \( f(u_1, u_2) \) is a \( N(\mu, \Sigma) \) distribution. Substituting \( y = x - hu \) (\( u = h^{-1}x - h^{-1}y \)), the bivariate density is given by

\[
f(y) = h^{-2} f_u(h^{-1}x - h^{-1}y) ,
\]

with corresponding characteristic function

\[
G(\theta) = h^{-2} \exp(- \frac{1}{2} h^{-2} \theta' \Sigma \theta + ih^{-1} \theta' \mu) . \tag{3.25}
\]

119
The Fourier transform of \( K^{(j)}(u_1)K^{(j)}(u_2) \) can be found by writing the required integral as the product of the two separate Fourier integrals,

\[
I = \int_{-\infty}^{\infty} e^{i\theta_1 u_1} K^{(j)}(u_1) du_1 \int_{-\infty}^{\infty} e^{i\theta_2 u_2} K^{(j)}(u_2) du_2 = I_1 \times I_2.
\]

Integration by parts of \( I_1 \) gives

\[
I = \left\{ e^{i\theta_1 u_1} K^{(j-1)}(u_1) I_2 \right\}_{-\infty}^{\infty} - (i\theta_1) \int_{-\infty}^{\infty} e^{i\theta_1 u_1} K^{(j-1)}(u_1) du_1 I_2.
\]

For the normal kernel, \( K(u_1) \to 0 \) as \( u_1 \to \infty \), \( K^{(j-1)}(u_1) \to 0 \) as \( u_1 \to \infty \), and after integrating by parts \( j \) times,

\[
I = (-1)^j (i\theta_1)^j \int_{-\infty}^{\infty} e^{i\theta_1 u_1} K(u_1) du_1 I_2.
\]

Repeating this process for the second integral, \( I_2 \), we get

\[
I = (-i\theta_1)^j \kappa(\theta_1)(-i\theta_2)^j \kappa(\theta_2) = (-\theta_1 \theta_2)^j \kappa(\theta_1) \kappa(\theta_2), \tag{3.26}
\]

where \( \kappa(\theta_1) \) and \( \kappa(\theta_2) \) are the Fourier transforms of \( K(u_1) \) and \( K(u_2) \) defined by (3.13). We temporarily replace \( \theta_1 \) and \( \theta_2 \) by \( \theta_1^* \) and \( \theta_2^* \) respectively and note that by Parseval's identity (see (3.24)),

\[
(2\pi)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^{(j)}(u_1)K^{(j)}(u_2)f_{x_1-x_2}(x_1-u_1,x_2-u_2)du_1du_2
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (-1)^j (\theta_1^* \theta_2^*)^j \kappa(\theta_1^*) \kappa(\theta_2^*) h^{-2} \exp(i h^{-1} \theta^* \mu) \exp(-\frac{1}{2} h^{-2} \theta^* \Sigma \theta^*) \theta_1^* d\theta_1^* d\theta_2^*.
\]

Now substitute \( \theta_1 = h^{-1} \theta_1^* \) and \( \theta_2 = h^{-1} \theta_2^* \) and replace \( \exp(i \theta^* \mu) \) by its cosine form to give the right side of the above equation equal to

\[
(-1)^j (h^2)^j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 \theta_2)^j \kappa(h \theta_1) \kappa(h \theta_2) \cos\{x(\theta_1 + \theta_2)\} \exp\left(-\frac{1}{2} \theta^* \Sigma \theta^* \right) d\theta_1 d\theta_1. \tag{3.27}
\]

120
A similar transform of the univariate function gives

$$2\pi \int_{-\infty}^{\infty} K^{(j)}(u_k) f(x_k - h u_k) du_k = (3.28)$$

$$(ih)^j \int_{-\infty}^{\infty} \theta_k^j \kappa(h \theta_k) \exp(-\frac{1}{2} \theta_k^2) \cos(x_k \theta_k) d\theta_k .$$

With results (3.27) and (3.29) we can rewrite the double integral with respect to $u_1, u_2$ in (3.23) as its Fourier equivalent,

$$V(t_1 - t_2) = (-1)^j h^{2j}(2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 \theta_2)^j \kappa(h \theta_1) \kappa(h \theta_2) \cos(x(\theta_1 + \theta_2))$$

$$\times \left\{ \exp(-\frac{1}{2} \theta' \Sigma \theta) - \exp(-\frac{1}{2} \|\theta\|^2) \right\} d\theta \quad \text{where } \|\theta\|^2 = \theta_1^2 + \theta_2^2 .$$

We now substitute this form of the double integral with respect to $u_1, u_2$ into (3.23) to give,

$$T^2 \text{var} \{ \hat{f}^{(j)}(x) \}$$

$$= h^{-2j} \int_{0}^{T} \int_{0}^{T} V(t_1 - t_2) dt_1 dt_2$$

$$= 2h^{-2j} \int_{0}^{T} \int_{0}^{T-\tau} V(\tau) dt_2 d\tau , \quad (\text{substituting } \tau = t_1 - t_2)$$

$$= 2h^{-2j} \int_{0}^{T} (T - \tau) V(\tau) d\tau$$

$$= 2(2\pi)^{-2} (-1)^j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 \theta_2)^j \kappa(h \theta_1) \kappa(h \theta_2) \cos(x(\theta_1 + \theta_2)) \exp(-\frac{1}{2} \|\theta\|^2)$$

$$\times \int_{0}^{T} (T - \tau) \left[ \exp\{-\rho(\tau) \theta_1 \theta_2\} - 1 \right] d\tau d\theta_1 d\theta_2 . \quad (3.29)$$

Equation (3.29) is simplified by replacing $\rho(\tau)$ with its asymptotic functions for short and long term dependence and the integration leads to results for the asymptotic variance of $\hat{f}^{(j)}$.

To examine the effect of short term dependence ($\int |\rho(\tau)| d\tau < \infty$), we rearrange
the integral with respect to $r$ so that

$$T^2 \text{var} \left\{ \hat{f}^{(j)}(x) \right\} =$$

$$2(2\pi)^{-2}(-1)^j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 \theta_2)^j \kappa(h_1) \kappa(h_2) \cos\{x(\theta_1 + \theta_2)\} \exp\left(-\frac{1}{2}||\theta||^2\right)$$

$$\times \int_0^\infty T \left[ \exp\{-\rho(\tau)\theta_1 \theta_2\} - 1 \right] d\tau \: d\theta_1 d\theta_2$$

$$+ 2(2\pi)^{-2}(-1)^j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 \theta_2)^j \kappa(h_1) \kappa(h_2) \cos\{x(\theta_1 + \theta_2)\} \exp\left(-\frac{1}{2}||\theta||^2\right)$$

$$\times \left( \int_0^T (T - \tau) \left[ \exp\{-\rho(\tau)\theta_1 \theta_2\} - 1 \right] d\tau - T \int_0^\infty \left[ \exp\{-\rho(\tau)\theta_1 \theta_2\} - 1 \right] d\tau \right) d\theta_1 d\theta_2 .$$

(3.30)

(3.31)

We define the second component of the sum, (3.31), as $A(T)$. Our objective will be to show that $T^{-1}|A(T)|$ is bounded by a constant so that

$$T^{-2}|A(T)| \to 0 \text{ as } T \to \infty$$

(3.32)

and that this term may be disregarded in the asymptotic expression for $\text{var} \left\{ \hat{f}^{(j)}(x) \right\}$.

We manipulate the ranges of integration so that

$$\left\{ \int_0^T (T - \tau) - T \int_0^\infty \right\} \left[ \exp\{-\rho(\tau)\theta_1 \theta_2\} - 1 \right] d\tau$$

$$= \left\{ T \int_0^\infty - T \int_0^T - \int_0^T \tau - T \int_0^\infty \right\} \left[ \exp\{-\rho(\tau)\theta_1 \theta_2\} - 1 \right] d\tau$$

$$= - \left\{ T \int_T^\infty + \int_0^T \tau \right\} \left[ \exp\{-\rho(\tau)\theta_1 \theta_2\} - 1 \right] d\tau .$$

Since $\kappa$ is a bounded function and $\sup |\cos| = 1,$

$$T^{-1}|A(T)|$$

$$\leq 2(2\pi)^{-2} \left\{ \int \kappa(h \theta) d\theta \right\}^2 \left( \int |\theta_1 \theta_2|^j \exp\left(-\frac{1}{2}||\theta||^2\right) \times$$

122
\[ \left[ \int_{T}^{\infty} \exp \left\{ -\rho(\tau)\theta_1\theta_2 \right\} - 1 \right] d\tau + \int_{0}^{T} \frac{\tau}{T} \left[ \exp \left\{ -\rho(\tau)\theta_1\theta_2 \right\} - 1 \right] d\tau \right] d\theta_1 d\theta_2 \]
\[ \leq 2(2\pi)^{-2} \left( \int \kappa(h\theta) d\theta \right)^2 \left( \int \int |\theta_1\theta_2|^i \exp \left( -\frac{1}{2} \|\theta\|^2 \right) \times \right. \]
\[ \int_{0}^{\infty} \min \left( \frac{T}{T}, 1 \right) \left[ \exp \left\{ -\rho(\tau)\theta_1\theta_2 \right\} - 1 \right] d\tau \right) d\theta_1 d\theta_2 . \]

Now collect the terms involving \( \theta_1 \) and \( \theta_2 \), apart from \( \{ \int \kappa(h\theta) d\theta \}^2 \) which is bounded, and define

\[ B(\theta_1, \theta_2, \tau) = |\theta_1\theta_2|^i \exp \left( -\frac{1}{2} \|\theta\|^2 \right) \left[ \exp \left\{ -\rho(\tau)\theta_1\theta_2 \right\} - 1 \right] . \]

The above integrand is dominated by \( B(\theta_1, \theta_2, \tau) \) and we show that this term is integrable and bounded. The result (3.32) follows from the Dominated Convergence Theorem.

We rearrange integral with respect to \( \tau \) so that it is in a form where we can consider the effect of the sign of \( \theta_1\theta_2 \). That is,

\[ \int_{0}^{\infty} \left[ \exp \left\{ -\rho(\tau)\theta_1\theta_2 \right\} - 1 \right] d\tau = \exp(-\theta_1\theta_2) \int_{0}^{\infty} \left[ \exp \left\{ \{1 - \rho(\tau)\} \theta_1\theta_2 \right\} - \exp(\theta_1\theta_2) \right] d\tau . \]

We have proposed that the short term covariance can be modelled by \( \rho(\tau) = 1 - (c_1|\tau|^\alpha) + o(|\tau|^\alpha) \) as \( \tau \downarrow 0 \), \( c_1 > 0 \), see (3.7). After substituting for \( \rho(\tau) \) we evaluate

\[ \int_{0}^{\infty} \exp\left\{ -(c_1\tau)^\alpha \theta_1\theta_2 \right\} d\tau = \alpha^{-1} \Gamma(\alpha)c_1^{-1}(\theta_1\theta_2)^{-1/\alpha} , \]

leading to integrals with respect to \( \theta_1, \theta_2 \) that have components \( (\theta_1\theta_2)^{-1/\alpha} \). By definition, \( \alpha \leq 2 \) so the nature of \( \operatorname{var} \left\{ \hat{f}(j)(x) \right\} \) will depend whether \( 0 < \alpha < (j + \frac{1}{2})^{-1} \), \( \alpha = (j + \frac{1}{2})^{-1} \) or \( (j + \frac{1}{2})^{-1} < \alpha \leq 2 \).
With this regard, we examine $B(\theta_1, \theta_2, \tau)$ and \( \text{var}\{\hat{f}(x)\} \) for the following modes of behaviour of $\rho(\tau)$;

(i) \( \{1 - \rho(\tau)\}^{-1} = O(\tau^{-\alpha}) \) as $\tau \downarrow 0$, where $0 < \alpha < (j + \frac{1}{2})^{-1}$, \hspace{1cm} (3.33)

(ii) \( 1 - \rho(\tau) \sim (c_1 \tau)^{\alpha} \) as $\tau \downarrow 0$, where $0 < \alpha = (j + \frac{1}{2})^{-1}$ and \hspace{1cm} (3.34)

(iii) \( 1 - \rho(\tau) \sim (c_1 \tau)^{\alpha} \) as $\tau \downarrow 0$, where $(j + \frac{1}{2})^{-1} < \alpha \leq 2$, \hspace{1cm} (3.35)

In the following work, $C, C_1, C_2 \ldots$ represent constants.

In determining a bound for $B(\theta_1, \theta_2, \tau)$, we must examine the cases where $\theta_1 \theta_2 < 0$, $\theta_1 \theta_2 \geq 0$ and with consideration of the influence of $\rho(\tau)$ over the range $[0, \infty)$.

For $\theta_1 \theta_2 \geq 0$ we only require the property that $\int |\rho(\tau)|d\tau < \infty$ to show that $B(\theta_1, \theta_2, \tau)$ is bounded. For $\theta_1 \theta_2 < 0$, the integral of $B(\theta_1, \theta_2, \tau)$ will be dominated by the behaviour of $\rho(\tau)$ when $\tau < 1$ so the range of integration is separated into $[0, 1]$ and $[1, \infty)$. In this case, we write \( \int_0^\infty B(\theta_1, \theta_2, \tau)d\tau = I_1 + I_2 \) where

\[
I_1 = \int_0^1 |\exp\{-\rho(\tau)\theta_1 \theta_2\} - 1|d\tau \quad \text{and} \quad (3.36)
\]

\[
I_2 = \int_1^\infty |\exp\{-\rho(\tau)\theta_1 \theta_2\} - 1|d\tau . \quad (3.37)
\]

To put a bound on $I_2$, it is only necessary to use that $\int |\rho(\tau)|d\tau < \infty$. The bound for $I_1$ will depend upon the conditions (3.33), (3.34) and (3.35) and we treat in succession these three different modes of behaviour of $\rho$.

Suppose first that $\alpha < (j + \frac{1}{2})^{-1}$ and $\rho$ is as defined in (3.33). We examine separately the cases $\theta_1 \theta_2 \geq 0$ and $\theta_1 \theta_2 < 0$ and show that for some $0 < \epsilon < 1$,

\[
\int_0^\infty |\exp\{-\rho(\tau)\theta_1 \theta_2\} - 1|d\tau \leq C\{(1 + |\theta_1 \theta_2|)^{-1/\alpha}\exp(-\theta_1 \theta_2) + \exp(\epsilon \theta_1 \theta_2)\}
\]

124
uniformly in \( \theta_1 \theta_2 \).

When \( \theta_1 \theta_2 < 0 \), put \( \omega = -\theta_1 \theta_2 \) so that

\[
|I_1| \leq \int_0^1 |\exp \{\rho(\tau)\omega\}| \, d\tau + \int_0^1 d\tau = e^\omega \int_0^1 |\exp [-\omega \{1 - \rho(\tau)\}]| \, d\tau + 1.
\]

Hence, when (3.33) holds,

\[
|I_1| \leq e^\omega \int_0^1 \exp \{-C_1 \gamma \omega\} \, d\tau + C_2
\]

\[
\leq e^\omega \int_0^\infty \exp \{-C_1 \gamma \omega\} \, d\tau + C_2
\]

\[
= e^\omega C_3 \omega^{-1/\alpha} + C_2.
\]

For \( \omega > 1 \), \( C_3 \omega^{-1/\alpha} \leq C_3 \left\{ \frac{1}{2} (1 + \omega) \right\}^{-1/\alpha} = 2^{1/\alpha} C_3 (1 + \omega)^{-1/\alpha} \) and for \( 0 < \omega \leq 1 \),

\[
\int_0^1 \exp \{-C_1 \gamma \omega\} \, d\tau \leq 1, (1 + \omega)^{-1/\alpha} \geq 2^{-1/\alpha},
\]

giving that \( \int_0^1 \exp \{-C_1 \gamma \omega\} \, d\tau \leq 2^{1/\alpha}(1 + \omega)^{-1/\alpha} \). Hence,

\[
\int_0^1 \exp \{-C_1 \gamma \omega\} \, d\tau \leq 2^{1/\alpha} \max(C_3, 1)(1 + \omega)^{-1/\alpha} \ \forall \omega > 0, \text{ and}
\]

\[
|I_1| \leq C_4 e^\omega (1 + \omega)^{-1/\alpha}
\]

\[
= C_4 e^{\theta_1 \theta_2} (1 + |\theta_1 \theta_2|)^{-1/\alpha}.
\]  

For the integral over the range \((1, \infty)\) when \( \theta_1 \theta_2 < 0 \), \( \omega = -\theta_1 \theta_2 \),

\[
|I_2| = \int_1^\infty |\exp \{\rho(\tau)\omega\} - 1| \, d\tau
\]

\[
\leq \int_1^\infty |\rho(\tau)\omega| \exp \{|\rho(\tau)\omega|\} \, d\tau
\]

\[
\leq \omega \exp \left\{ \omega \sup_{\tau > 1} |\rho(\tau)| \right\} \int_1^\infty |\rho(\tau)| \, d\tau
\]

\[
= \omega \exp \{\omega(1 - \delta)\} \int_1^\infty |\rho(\tau)| \, d\tau \quad \left( \sup_{\tau > 1} |\rho(\tau)| = (1 - \delta) < 1 \right)
\]

\[
= (1 + \omega)^{-1/\alpha} e^\omega \left\{ (1 + \omega)^{1/\alpha} \omega \exp(-\delta\omega) \int_1^\infty |\rho(\tau)| \, d\tau \right\}.
\]
Since the terms within the braces give a bounded constant in \( \omega > 0 \), then for \( \theta_1 \theta_2 < 0 \),

\[
|I_2| \leq C_5 \left(1 + |\theta_1 \theta_2|\right)^{-1/\alpha} e^{-\delta_1 \theta_2} . \tag{3.39}
\]

When \( \theta_1 \theta_2 \geq 0 \), we can examine the integral over the entire range recalling that

\[
C_6 = \int_0^\infty |\rho(\tau)| \, d\tau < \infty \quad \text{and} \quad \int_0^\infty |\exp \{-\rho(\tau)\theta_1 \theta_2\} - 1| \, d\tau \leq \int_0^\infty |\theta_1 \theta_2 \rho(\tau)| \exp\{|\theta_1 \theta_2 \rho(\tau)|\} \, d\tau 
\]

\[
\leq \theta_1 \theta_2 \int_0^\infty \exp \{\theta_1 \theta_2 |\rho(\tau)|\} |\rho(\tau)| \, d\tau 
\]

\[
\leq \theta_1 \theta_2 \exp \{\theta_1 \theta_2 \sup_{\tau} |\rho(\tau)|\} \int_0^\infty |\rho(\tau)| \, d\tau 
\]

\[
= \theta_1 \theta_2 \exp \{\theta_1 \theta_2 (1 - \delta)\} C_6 
\]

\[
\left(\sup_\tau |\rho(\tau)| = 1 - \delta < 1\right) 
\]

\[
= C_6 \left\{\theta_1 \theta_2 \exp\left(-\frac{1}{2} \delta \theta_1 \theta_2\right)\right\} \exp \left\{\theta_1 \theta_2 (1 - \frac{1}{2} \delta)\right\} 
\]

\[
\leq C_7 \exp(\epsilon \theta_1 \theta_2) \quad (\epsilon = 1 - \frac{1}{2} \delta < 1) \tag{3.40}
\]

Combining (3.38), (3.39) and (3.40), we have that

\[
\int_0^\infty |\exp \{-\rho(\tau)\theta_1 \theta_2\} - 1| \, d\tau \leq C \left\{(1 + |\theta_1 \theta_2|)^{-1/\alpha} \exp(-\delta_1 \theta_2) + \exp(\epsilon \theta_1 \theta_2)\right\} 
\]

\[
\int_0^\infty \int_0^\infty \int_0^\infty B(\theta_1, \theta_2, \tau) \, d\tau \, d\theta_1 \, d\theta_2 \leq \int_0^\infty \int_0^\infty \int_0^\infty |\theta_1 \theta_2|^i \exp\left(-\frac{1}{2} \|\theta\|^2\right) 
\]

\[
\times \left\{(1 + |\theta_1 \theta_2|)^{-1/\alpha} \exp(-\delta_1 \theta_2) + \exp(\epsilon \theta_1 \theta_2)\right\} \, d\theta_1 \, d\theta_2 
\]

\[
= C \int_0^\infty \int_0^\infty \int_0^\infty |\theta_1 \theta_2|^i (1 + |\theta_1 \theta_2|)^{-1/\alpha} \exp\left(-\frac{1}{2} \|\theta\|^2 - \delta_1 \theta_2\right) \, d\theta_1 \, d\theta_2 
\]

\[
+ C \int_0^\infty \int_0^\infty |\theta_1 \theta_2|^i \exp\left(-\frac{1}{2} \|\theta\|^2 + \epsilon \theta_1 \theta_2\right) \, d\theta_1 \, d\theta_2 .
\]

126
In the second double integral, \(-\frac{1}{2}||\theta||^2\) overrides the influence of \(\epsilon \theta_1 \theta_2\) when \(\theta_1 \theta_2 \geq 0\) and we can replace this double integral by a constant, \(C_8\).

In the first double integral, we substitute \(\omega_1 = \theta_1 + \theta_2, \omega_2 = \theta_1 - \theta_2\) and \(\gamma = \alpha^{-1} - j (\gamma > \frac{1}{2})\), then consider the integration over the regions where \(\omega_1 > \omega_2\) and \(\omega_2 > \omega_1\).

This gives,

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(\theta_1, \theta_2, \tau) d\tau d\theta_1 d\theta_2 \leq 2C \int_{0}^{\infty} \left\{ \int_{\omega_1}^{\infty} \left( 1 + \frac{1}{4} |\omega_2 - \omega_2^2| \right)^{-\gamma} \exp\left( -\frac{1}{2} \omega_1^2 \right) d\omega_2 \right\} d\omega_1
\]

\[
+ 2C \int_{0}^{\infty} \left\{ \int_{\omega_1}^{\infty} \left( 1 + \frac{1}{4} |\omega_2 - \omega_2^2| \right)^{-\gamma} \exp\left( -\frac{1}{2} \omega_1^2 \right) d\omega_2 \right\} d\omega_1 + C_8.
\]

In the first integral on the right side, put \(\omega = \omega_2 - \omega_1 (\omega > 0)\) to give,

\[
\int_{0}^{\infty} \exp\left( -\frac{1}{2} \omega_1^2 \right) d\omega \int_{0}^{\infty} \left\{ \left( 1 + \frac{1}{4} (\omega^2 + 2\omega_1) \right)^{-\gamma} \right\} d\omega
\]

\[
\leq \int_{0}^{\infty} \exp\left( -\frac{1}{2} \omega_1^2 \right) d\omega \int_{0}^{\infty} \left( 1 + \frac{1}{4} \omega^2 \right)^{-\gamma} d\omega
\]

\[
\leq \int_{0}^{\infty} \exp\left( -\frac{1}{2} \omega_1^2 \right) d\omega \int_{0}^{\infty} (1 + \omega^2)^{-\gamma} d\omega \quad \text{since } \gamma > \frac{1}{2},
\]

\[
\leq C_9 4^\gamma \sqrt{\pi}.
\]

In the second integral, \((1 + \frac{1}{4} |\omega_2 - \omega_2^2|)^{-\gamma} \leq 1\) so that integral is less than

\[
2C_{10} \int_{0}^{\infty} \omega_1 \exp\left( -\frac{1}{2} \omega_1^2 \right) d\omega_1 = C_{11}.
\]

This establishes the integrability of \(B(\theta_1, \theta_2, \tau)\). Therefore

\[
T^{-1}|A(T)| \leq C_{11} \left( \int |\xi| \right)^2 \left( 4^\gamma \pi^{1/2} + 1 \right).
\]

By the Dominated Convergence Theorem, \(T^{-2}|A(T)| \to 0\) as \(T \to \infty\).
We now consider $\kappa(h\theta_1)\kappa(h\theta_2)$ in (3.30). Since $\kappa(t) = \int e^{itx}K(x)dx$, $\kappa(0) = \int K(x)dx = 1$ and as $h \to 0$, $\kappa(h\theta) \to 1$. Therefore,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\phi_{1,2}| |\kappa(h\theta_1)\kappa(h\theta_2) - 1| \exp\left(-\frac{1}{2}||\theta||^2\right)d\theta \int_{0}^{\infty} |\exp\left\{-\rho(\tau)\theta_1\theta_2\right\} - 1| d\tau \to 0$$

as $h \to 0$ by the Dominated Convergence Theorem.

Summing up for the situation where $f|\rho| < \infty$ and $\alpha < (j + \frac{1}{2})^{-1}$,

$$\text{var}\{f^{(j)}(x)\} \sim 2T^{-1}(2\pi)^{-\frac{1}{2}}(-1)^j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (-\theta_1\theta_2)^j \cos\{x(\theta_1 + \theta_2)\} \exp\left(-\frac{1}{2}||\theta||^2\right)$$

$$\times \int_{0}^{\infty} [\exp\left\{-\rho(\tau)\theta_1\theta_2\right\} - 1] d\tau d\theta_1 d\theta_2$$

$$= 2T^{-1} \int_{0}^{\infty} \left\{(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos\{x(\theta_1 + \theta_2)\}(-\theta_1\theta_2)^j \times \right.$$  
$$\left. \left[\exp\left\{-\frac{1}{2}(\theta_1^2 + \theta_2^2 - 2\rho(\tau)\theta_1\theta_2)\right\} - \exp\left\{-\frac{1}{2}(\theta_1^2 + \theta_2^2)\right\}\right]\right] d\theta_1 d\theta_2 \right\} d\tau$$

$$= 2T^{-1} \int_{0}^{\infty} \left\{f^{(j,j)}(x,x) - f^{(j)}(x)^2\right\} d\tau$$  

(3.42) via Parseval's identity.

We now investigate var\{f^{(j)}\} for the situation where $0 < \alpha = (j + \frac{1}{2})^{-1} \leq 2$ and $1 - \rho(\tau) \sim (c_\tau)^\alpha$ as $\tau \downarrow 0$ (see (3.34)). As for the previous case where $\alpha < (j + \frac{1}{2})^{-1}$, we show that $T^{-2}|A(T)| \to 0$ as $T \to \infty$ using similiar theory regarding the bounds for $\int_{0}^{\infty} |\exp\left\{-\rho(\tau)\theta_1\theta_2\right\} - 1| d\tau$. Since the bound for that integral when $\theta_1\theta_2 \geq 0$, (3.40), only depends upon $\int |\rho(\tau)| d\tau < \infty$ and does not depend upon the nature of $\rho(\tau)$, it also applies in this case. We are required to show that the integral is bounded for the situation where $\theta_1\theta_2 < 0$. To do this, we focus on the interval $\tau \in [0,1]$, since the bound given by (3.39) applies to the component of the integral for the range $\tau \in [1,\infty)$.
We put $\theta_1 \theta_2 = -\omega$ and define $C_{12} = \sup_{r > 1} \rho(\tau) < 1$, $C(\alpha) = \int_0^\infty \exp\{-C(\tau)^\alpha\} d\tau$.

Then as $\omega \to +\infty$,

$$\int_0^\infty [\exp \{\rho(\tau)\omega\} - 1] d\tau = \int_0^1 \exp \{\rho(\tau)\omega\} d\tau + O\{\exp(C_{12}\omega) + \omega\}$$

$$= e^\omega \int_0^1 \exp \{-1 - \rho(\tau)\} \omega d\tau + O\{\exp(C_{12}\omega) + \omega\}$$

$$\sim e^\omega \int_0^\infty \exp \{-C(\tau)^\alpha\} d\tau .$$

Substituting $\tau = u \omega^{-1/\alpha}$,

$$e^\omega \int_0^\infty \exp \{-C(\tau)^\alpha\} d\tau = e^\omega \int_0^\infty \exp \{-C(u \omega^{-1/\alpha})^\alpha\} \omega^{-1/\alpha} du$$

$$= e^\omega \omega^{-1/\alpha} \int_0^\infty \exp \{-C(u)^\alpha\} du$$

$$= e^\omega \omega^{-1/\alpha} C(\alpha) . \quad (3.43)$$

Therefore,

$$\int_0^\infty [\exp \{\rho(\tau)\omega\} - 1] d\tau \sim e^\omega \omega^{-1/\alpha} C(\alpha) = \exp(|\theta_1 \theta_1|)|\theta_1 \theta_1|^{-1/\alpha} C(\alpha) . \quad (3.44)$$

Combining results (3.39), (3.40) and (3.44), we have a result akin to (3.41) which gives that $B(\theta_1, \theta_2, \tau)$ is integrable and bounded. We may also conclude that $T^{-2}|A(T)| \to 0$ as $T \to \infty$ for $0 < \alpha = (j + \frac{1}{2})^{-1}$ and now proceed to evaluate

$$\frac{1}{2}(2\pi)^2 T \text{Var}\{\hat{f}^{(j)}\} = \int_{-\infty}^\infty \int_{-\infty}^\infty (-\theta_1 \theta_2)^j \cos\{x(\theta_1 + \theta_2)\}$$

$$\times \kappa(h\theta_1) \kappa(h\theta_1) \exp\{-\frac{1}{2}\|\theta\|^2\} \int_0^\infty [\exp \{-\rho(\tau)\theta_1 \theta_2\} - 1] d\tau d\theta_1 d\theta_2 . \quad (3.45)$$

For the integral over $\tau$, we can use the result (3.43) so that

$$\frac{1}{2}(2\pi)^2 C(\alpha)^{-1} T \text{Var}\{\hat{f}^{(j)}\} \sim W(h)$$
\[
\int_{\theta_1, \theta_2 < 0} \left(1 + |\theta_1 \theta_2|\right)^{(j-1/\alpha)} \cos \{x(\theta_1 + \theta_2)\} \kappa(h \theta_1) \kappa(h \theta_1) \\
\exp\left(-\frac{1}{2} \|\theta\|^2 + \theta_1 \theta_2\right) d\theta_1 d\theta_2 \\
= \int_{\theta_1, \theta_2 < 0} \left(1 + |\theta_1 \theta_2|\right)^{-1/2} \cos \{x(\theta_1 + \theta_2)\} \kappa(h \theta_1) \kappa(h \theta_1) \times \\
\exp\left\{\frac{1}{2} (\theta_1 + \theta_2)^2\right\} d\theta_1 d\theta_2 \quad (3.46)
\]
\[
= 2 \int_{0 < \theta_1 < \theta_2 < \infty} \left(1 + \theta_1 \theta_2\right)^{-1/2} \cos \{x(\theta_2 - \theta_1)\} \kappa(h \theta_1) \kappa(h \theta_1) \times \\
\exp\left\{\frac{1}{2} (\theta_2 - \theta_1)^2\right\} d\theta_1 d\theta_2 .
\]

We make the substitutions \(u = 2(\theta_1 \theta_2)^{1/2}\) and \(v = \theta_2 - \theta_1; \theta_1 = \frac{1}{2} \{u^2 + v^2\}^{1/2} - v\) and \(\theta_2 = \frac{1}{2} \{u^2 + v^2\}^{1/2} + v\). The Jacobian for this transformation is

\[
\begin{vmatrix}
\frac{\partial \theta_1}{\partial u} & \frac{\partial \theta_1}{\partial v} \\
\frac{\partial \theta_2}{\partial u} & \frac{\partial \theta_2}{\partial v}
\end{vmatrix} = \begin{vmatrix}
\frac{1}{2} u(u^2 + v^2)^{-1/2} & \frac{1}{2} v(u^2 + v^2)^{-1/2} - \frac{1}{2} \\
\frac{1}{2} u(u^2 + v^2)^{-1/2} & \frac{1}{2} v(u^2 + v^2)^{-1/2} + \frac{1}{2}
\end{vmatrix}
\]
\[
= \frac{1}{4} uv(u^2 + v^2)^{-1} + \frac{1}{4} u(u^2 + v^2)^{-1/2} - \frac{1}{4} uv(u^2 + v^2)^{-1} + \frac{1}{4} u(u^2 + v^2)^{-1/2}
\]
\[
= \frac{1}{2} u(u^2 + v^2)^{-1/2}
\]
\[
= \frac{1}{2} u \left\{ (u^2)^{-1/2} - \frac{1}{2} (u^2)^{-3/2} v^2 + \ldots \right\}
\]
\[
= \frac{1}{2} u \left\{ u^{-1} + O(u^{-3} v^2) \right\}
\]
\[
= \frac{1}{2} + O(u^{-2} v^2) .
\]

In this notation,

\[
W(h) = 2 \int_0^\infty \int_0^\infty \left(1 + \frac{1}{4} u^2\right)^{-1/2} \cos(xv) \kappa(h \theta_1(u,v)) \kappa(h \theta_2(u,v)) \\
\times \exp\left(-\frac{1}{2} v^2\right) \times \left\{ \frac{1}{2} + O(u^{-2} v^2) \right\} dudv .
\]

Since \(\theta_1 = \frac{1}{2} (u - v) + O(u^{-1} v^2)\), \(\theta_1 = \frac{1}{2} (u + v) + O(u^{-1} v^2)\), then
\[ \kappa\{h\theta_1(u,v)\}\kappa\{h\theta_2(u,v)\} \sim \kappa\left\{\frac{1}{2}h(u-v)\right\}\kappa\left\{\frac{1}{2}h(u+v)\right\} \text{ and,} \]

\[ W(h) \sim \int_0^\infty \int_0^\infty \cos(xv) \exp\left(-\frac{1}{2}v^2\right)(1 + u^2/4)^{-\frac{1}{2}} \times \kappa\left\{\frac{1}{2}h(u-v)\right\}\kappa\left\{\frac{1}{2}h(u+v)\right\} \, du \, dv \]

\[ \sim \int_0^\infty \cos(xv) \exp\left(-\frac{1}{2}v^2\right)dv \int_1^\infty (1 + \frac{1}{4}u^2)^{-\frac{1}{2}} \kappa\left\{\frac{1}{2}h(u-v)\right\}\kappa\left\{\frac{1}{2}h(u+v)\right\} \, du . \]

Substitute \( w = hu/2, \ (u = 2h^{-1}w, \ du = 2h^{-1}dw \) and when \( u = 1, \ w = h/2 \), then

\[ W(h) \sim \int_0^\infty \cos(xv) \exp\left(-\frac{1}{2}v^2\right)dv \int_{h/2}^\infty \left(1 + \frac{1}{4}(2h^{-1}w)^2\right)^{-\frac{1}{2}} \kappa(w)^2 2h^{-1}dw \]

\[ \sim 2 \int_0^\infty \cos(xv) \exp\left(-\frac{1}{2}v^2\right)dv \int_{h/2}^\infty w^{-1}\kappa(w)^2dw . \]

Because \( \kappa(w) \) is a bounded function, \( w^{-1}\kappa(w)^2 \to 0 \) as \( w \to \infty \) and

\[ \int_{h/2}^\infty w^{-1}\kappa(w)^2dw \leq \sup_w \kappa(w)^2 \left\{\left|\log h\right|^{-1} - \log 2\right\} \sim \left|\log h\right|^{-1} . \]

Now

\[ \int_0^\infty \cos(xv)e^{-\frac{1}{2}v^2}dv = \sqrt{\frac{\pi}{2}} e^{-\frac{1}{2}x^2} = \pi f(x) , \]

giving \( W(h) \sim 2\pi f(x)|\log h| \) and

\[ \text{var} \{\tilde{f}(j)(x)\} \sim 2(2\pi)^{-2}C(\alpha)2\pi T^{-1}|\log h| = \pi^{-1}C(\alpha)f(x)|\log h| . \quad (3.47) \]

For the third situation where \( (j + \frac{1}{2})^{-1} < \alpha \leq 2 \) and \( \rho \) is as described by (3.35), the bound and integrability of \( B(\theta_1, \theta_2, \tau) \) is the same as for the previous case. We define \( \gamma = j - \alpha^{-1} > -\frac{1}{2} \) and by reasons similar to those leading to (3.45) and (3.46) we have that

\[ \frac{1}{2}(2\pi)^2C(\alpha)^{-1}T\text{var}\{\tilde{f}(j)\} \]

131
The integral in equation (3.48) is finite under condition (3.14) on page 86.
\[
\sim \int \int_{\theta_1 \theta_2 < 0} (1 + |\theta_1 \theta_2|)^\gamma \cos\{x(\theta_1 + \theta_2)\} \kappa(\theta_1 \theta_2) \exp\{(\frac{-1}{2}(\theta_1 + \theta_2)^2) d\theta_1 d\theta_2
\]

\[
\sim \int_0^\infty \int_0^\infty \left(\frac{1}{4} u^2\right)^\gamma \cos(xv) \kappa\left(\frac{1}{2} hu - v\right) \kappa\left(\frac{1}{2} hu + v\right) \exp\left(\frac{-1}{2} v^2\right) dudv
\]

\[
\sim 2^{-2\gamma} \int_0^\infty \int_0^\infty u^{2\gamma} \cos(xv) \kappa\left(\frac{1}{2} hu\right)^2 \exp\left(\frac{-1}{2} v^2\right) dudv
\]

\[
= 2^{-2\gamma} \int_0^\infty \cos(xv) \exp\left(\frac{-1}{2} v^2\right) dv \int_0^\infty u^{2\gamma} \kappa\left(\frac{1}{2} hu\right)^2 du
\]

\[
= 2^{-2\gamma} \pi f(x) 2^{1+2\gamma} h^{-2\gamma - 1} \int_0^\infty w^{2\gamma} \kappa(w)^2 dw \quad (w = \frac{1}{2} hu)
\]

\[
= 2 h^{-2\gamma - 1} \pi f(x) \int_0^\infty w^{2\gamma} \kappa(w)^2 dw .
\]

Therefore in the final case for the short term dependence,

\[
\text{var}\{f^{(i)}\} \sim 2(2\pi)^{-2} C(\alpha) T^{-1} 2 h^{-2\gamma - 1} \pi f(x) \int_0^\infty w^{2\gamma} \kappa(w)^2 dw
\]

\[
= \pi^{-1} C(\alpha) T^{-1} h^{2 \left\{1/2 - (i+\frac{1}{2})\right\}} f(x) \int_0^\infty w^{2\gamma} \kappa(w)^2 dw . \quad (3.48)
\]

When there is long term dependence, we require a further adjustment to the variance of \(f^{(i)}\). In these cases, we consider \(\rho(\tau) \sim (c_2 |\tau|)^{-\beta} \) as \(|\tau| \to \infty\) with \(0 < \beta \leq 1\). In the expression given by (3.29), the integration over \(\tau\) can be done in two parts; the first on the interval \((0,1)\) and the second on \((1,T)\). Over the interval \((0,1)\), we have components such as those previously discussed and this integral produces a contribution equal to \(\delta + o(\delta) + O(T^{-1})\). For the integral over \((1,T)\), we first write

\[
\exp\{ -\rho(\tau) \theta_1 \theta_2 \} - 1 \sim - \theta_1 \theta_2 \rho(\tau) ,
\]

and then that integral over the range \((1,T)\) is asymptotic to

\[
2(2\pi)^{-2} (-1)^{i+1} \int_{-\infty}^\infty \int_{-\infty}^\infty (\theta_1 \theta_2)^{i+1} \cos \{x(\theta_1 + \theta_2)\} \exp\{-||\theta||^2/2\} d\theta_1 d\theta_2
\]

\[
\times \int_1^T (T - \tau) \rho(\tau) d\tau .
\]
The integral with respect to $\theta_1, \theta_2$ is equal to

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 \theta_2)^{i+1} \{ \cos(x \theta_1) \cos(x \theta_2) - \sin(x \theta_1) \sin(x \theta_2) \} \exp(-\|\theta\|^2/2) d\theta_1 d\theta_2
\]

\[
= \int_{-\infty}^{\infty} \theta_1^{i+1} \cos(x \theta_1) e^{-\frac{1}{2} \theta_1^2} d\theta_1 \int_{-\infty}^{\infty} \theta_2^{i+1} \cos(x \theta_2) e^{-\frac{1}{2} \theta_2^2} d\theta_2
\]

\[
- \int_{-\infty}^{\infty} \theta_1^{i+1} \sin(x \theta_1) e^{-\frac{1}{2} \theta_1^2} d\theta_1 \int_{-\infty}^{\infty} \theta_2^{i+1} \sin(x \theta_2) e^{-\frac{1}{2} \theta_2^2} d\theta_2
\]

\[
= \left\{ \int_{-\infty}^{\infty} (u)^{i+1} \cos(xu) e^{-\frac{1}{2} u^2} du \right\}^2 - \left\{ \int_{-\infty}^{\infty} (u)^{i+1} \sin(xu) e^{-\frac{1}{2} u^2} du \right\}^2
\]

\[
= (2\pi)^2 (-1)^{i+1} f^{(i)}(x)^2.
\]

(3.49)

(3.50)

To evaluate the integral over $\tau$, we use $\rho(\tau) \sim (c_2 \tau)^{-\beta}$ as $\tau \to \infty$.

If $0 < \beta < 1$,

\[
(c_2)^{-\beta} \int_1^T (T - t)^{-\beta} dt
\]

\[
= (c_2)^{-\beta} \left\{ (1 - \beta)^{-1}(T^{2-\beta} - T) - (2 - \beta)^{-1}(T^{2-\beta} - 1) \right\}
\]

\[
= (c_2)^{-\beta} \left\{ (1 - \beta)^{-1}(2 - \beta)^{-1} \{ (2 - \beta) - (1 - \beta) \} T^{2-\beta} \times \right.
\]

\[
- (1 - \beta)^{-1} T + (2 - \beta)^{-1} \left. \right\}
\]

\[
= (c_2)^{-\beta} \left\{ (1 - \beta)^{-1}(2 - \beta)^{-1} T^{2-\beta} \right\} + (c_2)^{-\beta} \left\{ (2 - \beta)^{-1} - (1 - \beta)^{-1} T \right\}
\]

\[
\sim (c_2)^{-\beta} (1 - \beta)^{-1} (2 - \beta)^{-1} T^{2-\beta} \quad \text{since } T^{2-\beta} > T.
\]

(3.51)

If $\beta = 1$, $\int_1^T (T - \tau)^{-\beta} d\tau
\]

\[
= (c_2)^{-\beta} [T \log \tau]_1^T - (c_2)^{-\beta} [\tau]_1^T
\]

\[
= (c_2)^{-\beta} T \log T - T + 1
\]

\[
\sim (c_2)^{-\beta} T \log T.
\]

(3.52)

Combining results (3.50), (3.51) and (3.52), the contribution to the variance of $\hat{f}^{(i)}$
due to long term dependence is asymptotic to
\[
2(c_2)^{-\beta} f^{(j)}(x)^2 (1 - \beta)^{-1}(2 - \beta)^{-1} T^{-\beta} \quad \text{if } 0 < \beta < 1 \text{ and }
\]
\[
2(c_2)^{-\beta} f^{(j)}(x)^2 T^{-1} \log T \quad \text{if } \beta = 1. \quad (3.53)
\]
Thus with (3.42), (3.47), (3.48) and (3.53), we have proved Theorem 3.1.

3.6.2 Proof of Theorem 3.2.

From (3.23) and (3.29) in Theorem 3.1, we have that
\[
T \text{var} \{f^{(j)}(x)\} = 2h^{-2j} \int_0^T \left(1 - \frac{\tau}{T}\right) V(\tau) d\tau
\]
where
\[
V(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^{(j)}(x - u)K^{(j)}(x - v) \{f_r(u,v) - f(u)f(v)\} dudv
\]
or in the Fourier form it is given by
\[
V(\tau) = (-h)^{2j} (2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1, \theta_2)^j \{\psi_r(\theta_1, \theta_2) - \psi(\theta_1)\psi(\theta_2)\} d\theta_1 d\theta_2.
\]
With condition (3.18), the proof of this theorem is similar to that component of Theorem 3.1 leading to (3.42). With condition (3.19), the proof is similar to that given by Castellana and Leadbetter [3]. We first state Lemma 2.3 from that paper (since it is required for the theorem) and then proceed to give the proof.

Lemma 2.3 (Castellana and Leadbetter). Let \(g(u,v)\) be a bounded measurable function which is continuous at the point \((x,y)\), and let \(\{\delta_T(x); \ T \geq 0\}\) be a \(\delta\)-family. Then,
\[
\int \int \delta_T(u-x)\delta_T(v-y)g(u,v)dudv \to g(x,y) \text{ as } n \to \infty.
\]
Kernel estimators are part of the $\delta$-family of smoothing functions (satisfying the axioms of boundedness, continuity and integration to 1, see (1.2)) and with our assumption of $K^{(j)}$ being well defined and continuous,

$$\int \int \left\{ h^{-j}K^{(j)}(u-x) \right\} \left\{ h^{-j}K^{(j)}(v-y) \right\} g(u,v) dudv \to g^{(j,j)}(x,y)$$

as $n \to \infty$.

We commence the proof by utilising (3.23) and (3.29) to write,

$$T \text{var} \{ \hat{f}^{(j)}(x) \} =$$

$$2h^{-2j} \int \int K^{(j)}(x-u)K^{(j)}(x-v) \int_0^\infty \left( 1 - \frac{\tau}{T} \right) \{ f_T(u,v) - f(u)f(v) \} d\tau dudv .$$

Define $\zeta(\tau) = \sup_{u,v} |f_T(u,v) - f(u)f(v)|$. The inner integral is bounded above in absolute value by $\int_0^\infty \zeta(\tau) d\tau$ for all $u, v$ and differs from $\int_0^\infty \{ f_T(u,v) - f(u)f(v) \} dudv$ by no more than $\int_0^T (\tau/T) \zeta(\tau) d\tau + \int_T^\infty \zeta(\tau) d\tau$ which converges to zero by dominated convergence. It follows that,

$$T \text{var} \{ \hat{f}^{(j)}(x) \} =$$

$$2h^{-2j} \int \int K^{(j)}(x-u)K^{(j)}(x-v) \int_0^\infty \{ f_T(u,v) - f(u)f(v) \} d\tau dudv + o(1)$$

as $T \to \infty$. The function $g(u,v) = \int_0^\infty \{ f_T(u,v) - f(u)f(v) \} d\tau$ is continuous at $(x, x)$ since if $(u_m, v_m)$ is any sequence converging to $(x, x)$, $g(u_m, v_m) \to g(x, x)$ by dominated convergence. Thus with the above lemma, we have that

$$T \text{var} \{ \hat{f}^{(j)}(x) \} = 2 \int_0^\infty \{ f_T^{(j,j)}(x,x) - f^{(j)}(x)^2 \} d\tau .$$
References


