12-2006

A BAYESIAN APPROACH FOR BANDWIDTH SELECTION IN KERNEL DENSITY ESTIMATION WITH CENSORED DATA

Chinthaka Kuruwita
Clemson University, ckuruwi@clemson.edu

A BAYESIAN APPROACH FOR BANDWIDTH SELECTION IN KERNEL DENSITY ESTIMATION WITH CENSORED DATA

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Mathematical Sciences

by
Chinthaka Kuruwita
December 2006

Accepted by :
Dr.K.B.Kulasekera, Committee Chair
Dr.W.J.Padgett
Dr.C.Park

ABSTRACT

Estimating an unknown probability density function is a common problem arising frequently in many scientific disciplines. Among many density estimation methods, the kernel density estimators are widely used. However, the classical kernel density estimators suffer from an intrinsic problem as they assign positive values outside the support of the target density. This problem is commonly known as the 'Spill over' effect. A modification to the regular kernel estimator is proposed to circumvent this problem. The proposed method uses a lognormal kernel and can be used even in the presence of censoring to estimate any density with a positive support without any spill over at the origin. Strong consistency of this estimator is established under suitable conditions.

A Bayesian approach using as inverted gamma prior density is used in the computation of local bandwidths. These bandwidths yield better density estimates. It was shown that these bandwidths converge to zero for suitable choices of prior parameters and as a result the density estimator achieved its asymptotic unbiasedness.

A simulation study was carried out to compare the performance of the proposed method with two competing estimators. The proposed estimator was shown to be superior to both competitors under pointwise and global error criteria.

DEDICATION

I dedicate this work to my loving wife Pujitha. It is her love and support that made this work a complete one.

ACKNOWLEDGMENTS

I would like to express my heartiest gratitude to my advisor Dr.K.B.Kulasekera for his guidance and support. His mathematical insights inspired me and helped me immensely in making this a success. My sincere thanks to Dr.W.J.Padgett, for sharing his world of experience during the course of this work and also many thanks to Dr.C.Park for his insightful suggestions.

A special thank to Dr.J.Peterson for motivating me to study 'math' through his fascinating lectures on Measure Theory that benefited me in many ways in completing this text.

Loads of love to my wife Pujitha, for having an infinite patience and putting up sleepless nights with me through out the completion of this work.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF FIGURES

Chapter 1

Introduction

Like many other branches of statistics, nonparametric statistical methods have advanced in many directions from its inception. With the advent of high speed computing power in the recent past at a low cost, applications of nonparametric statistical methods grew many fold. These technological advancements and the vast array of problems which came to the fore as a result of these developments brought about a new era in nonparametric statistics.

Nonparametric density estimation is one of the major branches in nonparametric statistics that has been developed since the late 1950's. Its extensive application both in theoretical and practical settings has opened up new avenues in the field of statistics, particularly in data analysis.

1.1 Density Estimation

The theory and methods of density estimation focus on obtaining an accurate, and a robust estimator of an unknown probability density function. Typically, the functional form of the probability density function (pdf) $f(x)$ is unknown. Even if it is known, or assumed to be of a certain type, it will usually depend on some unknown parameters; i.e $f(x, \theta) = f_o(x|\theta)$

for a known function f_o. In such situations we may estimate the density by a 'plug in' method: that is plugging in an estimator for θ to come up with a density estimator of the form $f_o(\cdot|\hat{\theta})$. This is referred to as the parametric approach for density estimation as we only estimate unknown parameters. In this setting, the prior assumption of the functional form of the density function restricts our search to a small class of functions, within which the estimator is chosen. As a result, the density estimator may not adequately represent important features of the underlying probability density. Wand & Jones (1995) provide a classic example of the perils of this approach, by showing how important features such as multi-modes will be undetected when we restrict our density estimator to be of a predetermined parametric family.

In contrast, nonparametric density estimation procedures do not make any assumptions of the functional form of the target density. Rather, they attempt to uncover the underlying density, guided primarily by the data. This usually is known as *"letting the data speak for themselves"*.

One of the main advantages that this approach has over the parametric approach is the flexibility to choose an estimator from a very large class of functions, e.g. all the nonnegative continuous functions that integrate to unity. Consequently, nonparametric density estimators can and often will detect important features of the data which could otherwise be undetected.

However, if a parametric density estimator is justifiable, then the use of a nonparametric density estimator instead will reduce the precision of the inferences that one makes based on such an estimator. Another drawback of nonparametric density estimators is their heavy usage of computing power. However, at present powerful computing resources are readily available and as a result, nonparametric density estimators are used extensively in many areas of research.

1.2 Kernel Density Estimation

Among many available techniques, kernel density estimation (KDE) is probably the most widely used nonparametric density estimation method because of its simplicity. This estimator is a generalization of a naive estimator which is based on the definition of a pdf. A naive density estimator can be defined based on the definition of a probability density function. We know for an absolutely continuous probability distribution function $F(\cdot)$ of a random variable X, the pdf is defined as

$$f(x) = \frac{d}{dx}F(x) = \lim_{h\to 0}\frac{F(x+h)-F(x-h)}{2h} = \lim_{h\to 0}\frac{P(x-h<X<x+h)}{2h}.$$

Given a random sample from F, $P(x-h<X<x+h)$ could be estimated by the relative frequency of the sample and therefore, for small h values a naive estimator of $f(x)$ is formed as

$$\hat{f}(x) = \frac{1}{2h}\frac{Number\ of\ observations\ falling\ in\ (x-h, x+h)}{n}.$$

By defining a weighting function $W(t)$, this can be expressed as

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}W\left(x - X_i\right)$$

where

$$W(t) = \begin{cases} \frac{1}{2} & -h < t < h \\ 0 & otherwise \end{cases} .$$

The kernel density estimator is a generalization of the naive estimator above. Given a random sample from a density f, the kernel density estimator at a point x in the support of f is defined as a weighted sum of the observations,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right),$$

where $K(\cdot)$ is a weighting function called the kernel and h is a user defined quantity which is called the smoothing parameter or the bandwidth. This estimator is based on the intuitive notion of denseness and sparseness of the observations around the point of estimation x. If there are many observations (dense) around x , we would expect the true density to be high at x. On the other hand, the lesser the concentration of observations around x , the smaller the true density. The kernel $K(\cdot)$ assures that the density estimate $\hat{f}(x)$ adapts to sparse regions of the data.

Typically, the kernel function is a symmetric probability density function. It is desirable to have the kernel to satisfy the following conditions:

$$\begin{aligned} \int_{-\infty}^{\infty} K(u)\, du &= 1, \\ \int_{-\infty}^{\infty} uK(u)\, du &= 0, \\ \int_{-\infty}^{\infty} u^2K(u)\, du &= k_2 < \infty. \end{aligned} \tag{1.1}$$

These conditions are useful in the analysis of the performance of the kernel density estimator for finite sample sizes as well as in deriving their asymptotic properties. Some of the commonly used kernels are given in Table 1.1.

Kernel	K(t)
Epanechnikov	$\frac{\frac{3}{4}(1-\frac{t^2}{5})}{\sqrt{5}}$ for $\mid t \mid < \sqrt{5}$ and 0 otherwise
Biweight	$\frac{15}{16}(1-t^2)^2$ for $\mid t \mid < 1$ and 0 otherwise
Triangular	$1- \mid t \mid$ for $\mid t \mid < 1$ and 0 otherwise
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$
Rectangular	$\frac{1}{2}$ for $\mid t \mid < 1$ and 0 otherwise

Table 1.1: *Commonly Used Kernels*

1.3 Properties of Kernel Estimators

Assessing the performance of the estimators is an essential part of any density estimation problem. For density estimators, a widely used measure of performance is the Mean Integrated Squared Error (MISE) criterion is defined as

$$MISE\ \hat{f}(x) = E \int [\hat{f}(x) - f(x)]^2\ dx. \tag{1.2}$$

This is a measure of global accuracy of the density estimator and it accounts for the sampling variability in the data by taking the expectation of the integrated squared error across all possible samples . The MISE criterion is a widely used measure of performance of density estimators due to its mathematical tractability. Performance measures such as mean integrated absolute error (MIASE) which is defined as $E \int |\hat{f}(x) - f(x)|\ dx$, although more intuitive is much harder to compute than the MISE. Furthermore, MISE of $\hat{f}(x)$ can be decomposed to give an alternative representation in terms of its bias and variance resembling the classical MSE of a parametric estimator $\hat{\theta}$:

$$\begin{aligned} MISE\ \hat{f}(x) &= E \int [\hat{f}(x) - f(x)]^2\ dx \\ &= \int E[\hat{f}(x) - f(x)]^2\ dx \\ &= \int MSE[\hat{f}(x)]\ dx \\ &= \int V[\hat{f}(x)]\ dx + \int [Bias \hat{f}(x)]^2\ dx. \end{aligned}$$

As stated in Silverman (1986) one can show that the bias and the variance of $\hat{f}(x)$ depends on the smoothing parameter h in a complicated way. Except in very special cases, the formulae for bias and variance become intractable and have very little intuitive meaning. However, more appealing formulae can be derived for asymptotic bias and variance of kernel density estimators.

1.4 Large Sample Properties of Kernel Density Estimators

Using Taylor series expansion one can show

$$
\begin{aligned}
\int [Bias_h \hat{f}(x)]^2 \; dx &\approx \frac{1}{4} \; h^4 k_2 \int f''(x) \; dx, \\
\int V_h[\hat{f}(x)] \; dx &\approx \frac{1}{nh} \int K(x)^2 \; dx.
\end{aligned}
$$

Note that the only quantity, other than the kernel $K(\cdot)$, which is at the control of the experimenter is the smoothing parameter h. Any attempt to decrease either the bias or the variance with respect to h will result in an increase of the other. This is a fundamental problem in kernel density estimation. Further, the quantity $\int f''(x) \; dx$ is commonly known as the 'curvature' and it measures how 'wiggly' the density is. The bias of $\hat{f}$ will be substantial for densities with high 'curvature' even with large samples .

1.5 Background and Overview of the Study

Although symmetric kernels are widely used in KDE, for densities with a bounded support such as $[0, \infty)$, the resulting density estimate will pose problems at the boundary of the support. When estimating a density with a bounded support, we would want our estimate $\hat{f}(x)$ to be zero for all x outside the support of the underlying density. However, typical kernel estimators will assign positive weights for x values outside the support. This problem is known as the 'Spill Over Effect'. In this study we will be looking at a remedy for the spill over problem at the origin when estimating a lifetime density with a positive support $[0, \infty)$.

Moreover, in lifetime data analysis problems it is of particular interest to estimate certain percentiles and other features of the underlying lifetime density related to one or few x values in the domain. Hence, locally optimal bandwidths are preferred than global bandwidths. We will examine how to compute locally optimal bandwidths for density estimators using Bayesian methods.

Further, we propose a methodology for computing bandwidth values for kernel estimators entirely based on the data at hand. In other words, devising a method that will eliminate the role of the experimenter in bandwidth selection by utilizing the data to govern the bandwidth. These automatic bandwidth selection techniques will serve as guides to develop bandwidth selection routines in statistical software packages. Moreover, these methods could be used as preliminary analytical tools to provide insights for sophisticated analysis.

We will explore these issues in a Bayesian framework in relation to estimating a lifetime density. Therefore, we would inevitably work under the additional constraint of censored data. Spill over of the density estimate at the origin is avoided by using an asymmetric kernel with a bounded support, and bandwidth selection is automated by using improper priors.

1.6 Literature Review

Available literature on nonparametric density estimation is vast. Pioneering work on nonparametric density estimation was initiated by Rosenblatt (1956) and Parzen (1962). Since then numerous studies have been done on various aspects of nonparametric density estimation. Wegman (1972) provides a survey of some of the earliest nonparametric density estimation methods, and more recent developments are discussed in Izenman (1991).

Studies on nonparametric density estimation with censored data are comparatively less. Literature on this area did not appear until the 1980's and Blum & Susarla (1980) constructed the first density estimator and failure rate estimator based on censored data. Padgett & McNichols (1984) have compiled a comprehensive survey of the earlier nonparametric density estimation methods designed for censored observations.

Bandwidth selection on its own has generated an extensive amount of literature. Marron (1988) gives an excellent exposition about various bandwidth selection procedures and Jones *et al.* (1996) have surveyed some of the recent advancements in bandwidth selection methods including data driven bandwidths.

Bayesian methodologies in nonparametric density estimation began to be developed in 1970's. Ferguson & Phadia (1979) discuss a nonparametric method based on censored data for estimating a distribution function using Bayesian methodologies. A recent study was carried out by Gangopadhyay & Cheung (2002) on bandwidth selection using a Bayesian approach. Chen (1999) and Chen (2000) and Scalliet (2004) have proposed the use of asymmetric kernels to circumvent the spill over effect at the origin for estimating densities with bounded supports. By combining the Bayesian concept in bandwidth selection and asymmetric kernel method, Kulasekera & Padgett (2006) have developed a novel methodology for estimating probability densities with bounded support, with the presence of random censoring.

Chapter 2

Methodology

The notation and some existing results that are being used in this study will be introduced in this section. General theories are stated as references and will not be presented here.

2.1 Notation for Randomly Right Censored Data

Let $X_1, \ldots, X_n$ be independent and identically distributed lifetimes of n individuals or items that are censored from the right by a sequence of random variables $U_1, \ldots, U_n$ which are independent from the X_i's. Let F be the unknown distribution function of the X_i's with density f and G be the distribution function of the censoring variables U_i's. The observed data will be denoted by the pairs (Z_i, δ_i) where

$$Z_i = \min\{X_i, U_i\} \qquad \text{and}$$

$$\delta_i = \begin{cases} 1 & , \; X_i \le U_i \\ 0 & , \; X_i > U_i. \end{cases} \tag{2.1}$$

Then Z_i's will be independent and identically distributed random variables with distribution function H satisfying $[1 - H(x)] = [1 - F(x)][1 - G(x)]$. These observations are considered as randomly right censored data. When all the U_i's are equal to a constant 'c' then they are called Type I censored observations. If all the U_i's are equal to the r^{th} order statistic $X_{(r)}$ then we call them Type II censored observations.

2.2 Kaplan-Meier Product Limit Estimator

This is the most widely used estimator of an unknown survival function $(1 - F)$ and is defined as follows:

$$\hat{S}_n(t) = \begin{cases} \prod_{j:Z_j \le t} \left(\frac{n-j}{n-j+1} \right)^{\delta_j} & , \; t < Z_n \\ \\ 0, & t \ge Z_n \end{cases} \tag{2.2}$$

where Z_j is the j^{th} order statistic of the sample and $\delta_j = \begin{cases} 1 & if \; Z_j \; uncensored \\ 0 & if \; Z_j \; censored \end{cases}$.

By reversing the role of the indicator variable δ_i, we get the following estimator of the survival function $1 - G$ of the censoring variable :

$$\hat{S}^*_n(t) = \begin{cases} \prod_{j:Z_j \le t} \left(\frac{n-j}{n-j+1} \right)^{1-\delta_j} & t < Z_n \; . \\ \\ 0 & t \ge Z_n \end{cases} \tag{2.3}$$

2.3 Least Squares Cross Validation Bandwidth Selector for Censored Data

This estimator was developed by Marron & Padgett (1987). The optimal data based bandwidth estimator $\hat{h}_c$ is defined as the minimizer of the least squares cross validation criterion

$$CV(h) = \int \left(\hat{f}(x)\right)^2 w(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_i(X_i)\frac{w(X_i)}{S_n^{**}(X_i)} I_{[\delta_i=1]}$$

and for a suitable weight function $w(\cdot)$. Here $\hat{f}_i$ is the "leave out one" version of a density estimator $\hat{f}$ which is defined as

$$\hat{f}_i(x) = \sum_{j\neq i} \frac{1}{(n-1)S_n^{**}(X_j)h} K\left(\frac{x-X_j}{h}\right) I_{[\delta_i=1]}\ ,$$

where the kernel function $K(\cdot)$ is as defined in (1.1) and S_n^{**} is a modified version of the Kaplan-Meier estimator which was proposed by Blum & Susarla (1980),

$$S_n^{**}(t) = \begin{cases} 1 & 0 \le t \le Z_1 \\ \prod_{i=1}^{k-1}\left(\frac{n-i+1}{n-i+2}\right)^{1-\delta_i} & Z_{k-1} < t \le Z_k \quad k = 2, ..., n \\ \prod_{i=1}^{n}\left(\frac{n-i+1}{n-i+2}\right)^{1-\delta_i} & Z_n < t\ . \end{cases}$$

Chapter 3

Bandwidth Selection

One of the main challenges in density estimation problems is to choose the smoothing parameter or the bandwidth appropriately so that the estimator $\hat{f}$ neither will contain any unwarranted noise due to undersmoothing nor will it not detect important features of the density due to oversmoothing. For complete samples there are several bandwidth selection methods available for researchers. These methods have been developed to satisfy various optimality criteria, such as minimizing pointwise mean squared error. Jones *et.al* [1996] provide a survey of existing bandwidth selection procedures that are frequently used in practice.

We propose a local bandwidth selection method under a Bayesian framework and it is specifically designed to compute local bandwidths that can be used to estimate densities arising in reliability and lifetime data analysis, i.e. with support over $[0, \infty)$. The local nature of these bandwidths are expected to provide more reliable estimates at desired points of the support than global bandwidths. In addition we propose to use a lognormal kernel to avoid the spill over effect at the origin.

3.1 Derivation of the Bayesian Bandwidth

We derive the Bayesian bandwidth formula for a kernel estimator using the lognormal kernel which could be used to estimate densities with positive support, in particular lifetime densities. We develop our methodology for randomly right censored data.

We can define a function $f_h(x)$ associated with an unknown probability density $f(x)$ at a point x using a convolution of kernel weights as $f_h(x) = \int k(x,y,h)dF(y)$ where, $k(x,y,h)$ is a kernel function centered at y with a scale parameter h. Since F is unknown we use a suitable estimator $\hat{F}$ of the associated probability distribution function F and estimate f_h by

$$\hat{f}_h(x) = \int k(x,y,h)d\hat{F}(y) \ .$$

Note that $\int \hat{f}_h(x)\ dx = 1$ and $\hat{f}_h(x) \geq 0$, making it a proper pdf.

In this work, we use the Kaplan-Meier product limit estimator $\hat{S}_n$ defined in (2.2) to get the estimator $\hat{F}$ of F, by $\hat{F} = 1 - \hat{S}_n$. This will then lead to the following estimator of f_h:

$$\hat{f}_h(x) = \sum_{i=1}^{n} s_j k(x, \ln Z_j, h) \tag{3.1}$$

where $s_j = \hat{S}_n(Z_j) - \hat{S}_n(Z_j^-)$ are the jump sizes at each observation Z_j s defined in (2.1) and $k(x,\mu,\sigma)$ is a lognormal kernel with parameters μ and σ which is defined as

$$k(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}} \frac{1}{x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}$$

and h is the smoothing parameter or the bandwidth associated with the estimation process.

In the Bayesian framework, we would treat the smoothing parameter h as a random quantity with a prior distribution. Suppose that the bandwidth h follows an inverted gamma prior distribution with parameters α and β given by

$$\xi(h) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)h^{\alpha+1}} e^{-\frac{1}{\beta h}} \quad , \ h > 0.$$

Then the posterior density of h given the data $\underline{\boldsymbol{Z}} = \{Z_i, \delta_i\}, i = 1, 2, ...n$ is given by

$$P(h|x, \underline{\boldsymbol{Z}}) \quad = \quad \frac{f_h(x)\xi(h)}{\int f_h(x)\xi(h)dh} \ .$$

Since f_h is unknown, we use $\hat{f}_h$ in (3.1) as our estimator of f_h, leading to

$$\hat{P}(h|x, \underline{\boldsymbol{Z}}) \quad = \quad \frac{\hat{f}_h(x)\xi(h)}{\int \hat{f}_h(x)\xi(h)dh} \ . \tag{3.2}$$

Consider the denominator of (3.2). It simplifies to

$$\begin{aligned} \int \hat{f}_h(x)\xi(h)dh \quad &= \quad \int_0^{\infty} \sum_{j=1}^{n} s_j k(x, \ln Z_j, h)\xi(h) \ dh \\ &= \quad \int_0^{\infty} \sum_{j=1}^{n} s_j \frac{1}{\sqrt{2\pi}} \frac{1}{xh} e^{-\frac{1}{2}\left(\frac{\ln x - \ln Z_j}{h}\right)^2} \frac{1}{\beta^{\alpha}\Gamma(\alpha)h^{\alpha+1}} e^{-\frac{1}{\beta h}} \ dh. \end{aligned}$$

Finding a closed form for the posterior density of h is difficult under this parameterization. Therefore, let $\delta = h^2$ and assign the same prior to δ instead of h. Then, the prior distribution of h could be calculated using the square root transformation. For example suppose that the random variable has an inverted gamma (α, β) density. Let Y=g(X)=$\sqrt{X}$. Therefore, the inverse mapping $g^{-1}(y) = y^2$. Now by using the transformation technique the density of Y could be derived as follows:

$$
\begin{aligned}
f_Y(y) &= f_X(g^{-1}(y)).\left|\frac{d}{dy}g^{-1}(y)\right| \\
&= f_X(y^2).2y \\
&= \frac{1}{\beta^\alpha\Gamma(\alpha)(y^2)^{\alpha+1}}\,e^{-\frac{1}{\beta y^2}}.2y \\
&= \frac{2}{\beta^\alpha\Gamma(\alpha)y^{2\alpha+1}}\,e^{-\frac{1}{\beta y^2}} \quad .
\end{aligned}
$$

Therefore, if $\delta = h^2 \sim$ Inverted Gamma(α, β), then $h = \sqrt{\delta}$ has a pdf given by

$$
\xi(h) = \frac{2}{\beta^\alpha\Gamma(\alpha)h^{2\alpha+1}}\,e^{-\frac{1}{\beta h^2}} \ , \qquad h > 0 \ .
$$

Then the denominator of (3.2) can be written as

$$
\begin{aligned}
\int \hat{f}_h(x)\xi(h)dh &= \int_0^\infty \sum_{j=1}^n s_j \frac{1}{\sqrt{2\pi}}\frac{1}{xh}\,e^{-\frac{1}{2}\left(\frac{\ln x - \ln Z_j}{h}\right)^2}\frac{2}{\beta^\alpha\Gamma(\alpha)h^{2\alpha+1}}\,e^{-\frac{1}{\beta h^2}}\ dh \\
&= \int_0^\infty \sum_{j=1}^n s_j \frac{1}{\sqrt{2\pi}}\frac{1}{x}\frac{2}{\beta^\alpha\Gamma(\alpha)(h^2)^{\alpha+1}}\,e^{-\frac{1}{h^2}\left[\frac{1}{\beta}+\frac{1}{2}(\ln x - \ln Z_j)^2\right]}\ dh \ .
\end{aligned}
$$

Let

$$
\beta_j^* = \left[\frac{1}{\beta} + \frac{1}{2}(\ln x - \ln Z_j)^2\right]^{-1} \quad \mathit{and} \quad \alpha^* = \alpha + \frac{1}{2} \tag{3.3}
$$

Now changing variables by letting $h^2 = t$ we obtain

$$\int \hat{f}_h(x)\xi(h)dh = \int_0^\infty \sum_{j=1}^n s_j \frac{1}{\sqrt{2\pi}} \frac{1}{x} \frac{1}{\beta^\alpha \Gamma(\alpha) t^{\alpha+1}} e^{-\frac{1}{t\beta_j^*}} . \frac{1}{\sqrt{t}}\, dt$$

$$= \frac{1}{x} \sum_{j=1}^n s_j \frac{(\beta_j^*)^{\alpha^*} \Gamma(\alpha^*)}{\sqrt{2\pi}\beta^\alpha \Gamma(\alpha)} \int_0^\infty \frac{1}{(\beta_j^*)^{\alpha^*} \Gamma(\alpha^*) t^{\alpha^*+1}} e^{-\frac{1}{t\beta_j^*}}\, dt\ .$$

The integral in the last expression above is just unity as the integrand is the pdf of an inverted gamma random variable with parameters α^* and β_j^*. Therefore the denominator of equation (3.2) becomes

$$\frac{\Gamma(\alpha^*)}{\Gamma(\alpha)x} \sum_{j=1}^n s_j \frac{(\beta_j^*)^{\alpha^*}}{\sqrt{2\pi}\beta^\alpha}\ .$$

Likewise, the numerator of equation (3.2) simplifies to

$$\hat{f}_h(x)\xi(h) = \sum_{j=1}^n s_j \frac{1}{\sqrt{2\pi}} \frac{1}{x} \frac{2}{\beta^\alpha \Gamma(\alpha)(h^2)^{\alpha+1}} e^{-\frac{1}{h^2\beta_j^*}}\ .$$

Therefore, the estimated posterior density of h given the data $\underline{\boldsymbol{Z}}$ becomes

$$\hat{P}(h|x, \underline{\boldsymbol{Z}}) = \frac{\displaystyle\sum_{j=1}^n s_j \frac{1}{\sqrt{2\pi}} \frac{1}{x} \frac{2}{\beta^\alpha \Gamma(\alpha)(h^2)^{\alpha+1}} e^{-\frac{1}{h^2\beta_j^*}}}{\displaystyle\frac{\Gamma(\alpha^*)}{\Gamma(\alpha)x} \sum_{j=1}^n s_j \frac{(\beta_j^*)^{\alpha^*}}{\sqrt{2\pi}\beta^\alpha}}$$

$$= \frac{\displaystyle\sum_{j=1}^n s_j \frac{2}{(h^2)^{\alpha+1}} e^{-\frac{1}{h^2\beta_j^*}}}{\displaystyle\Gamma(\alpha^*) \sum_{j=1}^n s_j (\beta_j^*)^{\alpha^*}}\ .$$

Under the squared error loss, the Bayes estimator of the smoothing parameter h is the posterior mean, i.e.

$$\begin{aligned}
\tilde{h}(x) &= \int_0^\infty h.\hat{P}(h|x, \underline{\boldsymbol{Z}})\, dh \\
&= \int_0^\infty h.\frac{\sum_{j=1}^n s_j \frac{2}{(h^2)^{\alpha+1}} e^{-\frac{1}{h^2\beta_j^*}}}{\Gamma(\alpha^*)\sum_{j=1}^n s_j(\beta_j^*)^{\alpha^*}}\, dh \\
&= \frac{1}{\Gamma(\alpha^*)\sum_{j=1}^n s_j(\beta_j^*)^{\alpha^*}} \int_0^\infty h.\sum_{j=1}^n s_j \frac{2}{(h^2)^{\alpha+1}} e^{-\frac{1}{h^2\beta_j^*}}\, dh \\
&= \frac{\sum_{j=1}^n s_j \int_0^\infty h.\frac{2}{(h^2)^{\alpha+1}} e^{-\frac{1}{h^2\beta_j^*}}\, dh}{\Gamma(\alpha^*)\sum_{j=1}^n s_j(\beta_j^*)^{\alpha^*}}
\end{aligned}$$

Let $h^2 = t$ in the above expression to get

$$\tilde{h}(x) = \frac{\sum_{j=1}^n s_j \int_0^\infty \frac{1}{(t)^{\alpha+1}} e^{-\frac{1}{t\beta_j^*}}\, dt}{\Gamma(\alpha^*)\sum_{j=1}^n s_j(\beta_j^*)^{\alpha^*}} .$$

We can simplify the integral in the above expression by making the integrand into an inverted gamma density with parameters α and β_j^* as follows

$$\tilde{h}(x) = \frac{\Gamma(\alpha)\sum_{j=1}^n s_j(\beta_j^*)^{\alpha}}{\Gamma(\alpha^*)\sum_{j=1}^n s_j(\beta_j^*)^{\alpha^*}} . \int_0^\infty \frac{1}{(\beta_j^*)^{\alpha}\Gamma(\alpha)(t)^{\alpha+1}} e^{-\frac{1}{t\beta_j^*}}\, dt .$$

Thus, the Bayesian local bandwidth h for estimating the pdf at x is given by

$$\tilde{h}(x) = \frac{\Gamma(\alpha)\sum_{j=1}^n s_j(\beta_j^*)^{\alpha}}{\Gamma(\alpha^*)\sum_{j=1}^n s_j(\beta_j^*)^{\alpha^*}} , \tag{3.4}$$

where β_j^* and α^* are defined in (3.3) .

3.2 Data Based Bandwidths Using Improper Priors

Suppose we assume the prior distribution is improper where $\xi(h) \propto \dfrac{1}{h^r}$, where $r \in \mathbb{R}$, and $r \geq 0$. Then, the denominator of (3.2) becomes

$$
\begin{aligned}
\int_0^\infty \hat{f}_h(x)\xi(h)dh \quad &\propto \quad \int_0^\infty \sum_{j=1}^{n} s_j k(x, \ln Z_j, h)\frac{1}{h^r}dh \\
&= \quad \int_0^\infty \sum_{j=1}^{n} s_j \frac{1}{\sqrt{2\pi}}\frac{1}{xh} e^{-\frac{1}{2}\left(\frac{\ln x - \ln Z_j}{h}\right)^2} \frac{c}{h^r}\; dh \; .
\end{aligned}
$$

Note that c is the constant of proportionality for the improper prior. Now, changing variables using $h^2 = v$ and letting $\phi_j^* = \left[\frac{1}{2}\left(\ln x - \ln Z_j\right)^2\right]^{-1}$ and $r^* = \frac{r}{2}$, we get,

$$
\begin{aligned}
\int_0^\infty \hat{f}_h(x)\xi(h)dh \quad &= \quad \int_0^\infty \sum_{j=1}^{n} s_j \frac{1}{\sqrt{2\pi}} \frac{1}{x(v^{1/2})^{r+1}} e^{-\frac{1}{v\phi_j^*}} \frac{c}{2v^{1/2}} dv \\
&= \quad \sum_{j=1}^{n} s_j \frac{c}{\sqrt{2\pi}} \frac{(\phi_j^*)^{r^*}\Gamma(r^*)}{2x} \int_0^\infty \frac{1}{(\phi_j^*)^{r^*}\Gamma(r^*)v^{(r^*+1)}} e^{-\frac{1}{v\phi_j^*}} dv \; .
\end{aligned}
$$

As the value of the integral in the above expression is unity, the denominator reduces to

$$
\int_0^\infty \hat{f}_h(x)\xi(h)dh \quad \propto \quad \sum_{j=1}^{n} s_j \frac{1}{\sqrt{2\pi}} \frac{(\phi_j^*)^{r^*}\Gamma(r^*)}{2x} \; .
$$

Similarly, the numerator of (3.2) $\quad \propto \quad \displaystyle\sum_{j=1}^{n} s_j \frac{1}{\sqrt{2\pi}} \frac{1}{xh^{(r+1)}} e^{-\frac{1}{h^2\phi_j^*}} \; .$

Hence, the posterior density of the smoothing parameter h at a point of estimation x with data $\underline{\boldsymbol{Z}}$ can be written as,

$$\hat{P}(h,x,\underline{\boldsymbol{Z}}) = \frac{\sum_{j=1}^{n} s_j \frac{1}{\sqrt{2\pi}} \frac{c}{xh^{(r+1)}} e^{-\frac{1}{h^2\phi_j^*}}}{\sum_{j=1}^{n} s_j \frac{c}{\sqrt{2\pi}} \frac{(\phi_j^*)^{r^*}\Gamma(r^*)}{2x}} \; .$$

Under squared error loss, the Bayes estimator of h is found by computing the posterior mean in the following manner:

$$\begin{aligned}
\tilde{h}(x) &= \int_0^\infty h \; \hat{P}(h,x,\underline{\boldsymbol{Z}})dh \\
&= \int_0^\infty h \frac{\sum_{j=1}^{n} s_j \frac{1}{\sqrt{2\pi}} \frac{1}{xh^{(r+1)}} e^{-\frac{1}{h^2\phi_j^*}}}{\sum_{j=1}^{n} s_j \frac{1}{\sqrt{2\pi}} \frac{(\phi_j^*)^{r^*}\Gamma(r^*)}{2x}} dh \\
&= \frac{1}{\frac{\Gamma(r^*)}{2} \sum_{j=1}^{n} s_j (\phi_j^*)^{r^*}} \int_0^\infty \sum_{j=1}^{n} s_j \frac{1}{h^r} e^{-\frac{1}{h^2\phi_j^*}} dh \; .
\end{aligned}$$

By changing variables as $h^2 = v$ we get,

$$\begin{aligned}
\tilde{h}(x) &= \frac{1}{\frac{\Gamma(r^*)}{2} \sum_{j=1}^{n} s_j (\phi_j^*)^{r^*}} \sum_{j=1}^{n} s_j \int_0^\infty \frac{1}{(\sqrt{v})^r} e^{-\frac{1}{v\phi_j^*}} \frac{1}{2\sqrt{v}} dv \\
&= \frac{\Gamma(r^* - \frac{1}{2}) \sum_{j=1}^{n} s_j (\phi_j^*)^{(r^*-\frac{1}{2})}}{\Gamma(r^*) \sum_{j=1}^{n} s_j (\phi_j^*)^{r^*}} \int_0^\infty \frac{1}{(\phi_j^*)^{(r^*-\frac{1}{2})}\Gamma(r^*-\frac{1}{2}) v^{(r^*-\frac{1}{2})+1}} e^{-\frac{1}{v\phi_j^*}} dv.
\end{aligned}$$

As before, the integral on the right hand side is unity since the integrand is an inverted gamma density with parameters $(r^* - \frac{1}{2}, \phi_j^*)$.

Hence,

$$\tilde{h}(x) = \frac{\Gamma(\frac{r-1}{2}) \sum_{j=1}^{n} s_j (\phi_j^*)^{(\frac{r-1}{2})}}{\Gamma(\frac{r}{2}) \sum_{j=1}^{n} s_j (\phi_j^*)^{\frac{r}{2}}} \tag{3.5}$$

where $\phi_j^* = \left[\frac{1}{2} (\ln x - \ln Z_j)^2\right]^{-1}$.

In particular, when $r = 1$, i.e. $\xi(h) \propto \dfrac{1}{h}$, the Bayes estimator of the smoothing parameter h at a point of estimation x reduces to,

$$\tilde{h}(x) = \frac{1}{\sqrt{2\pi}} \left[\sum_{j=1}^{n} \frac{s_j}{[\ln x - \ln Z_j]} \right]^{-1} .$$

Chapter 4

Asymptotic Properties

Now we explore the large sample properties of the proposed kernel density estimator using the lognormal kernel with Bayesian bandwidths. In particular, we will establish that the proposed density estimator converges almost surely to the underlying pdf and show that the Bayesian bandwidths converge to zero as $n \to \infty$ under suitable conditioins.

4.1 Convergence of the Bayesian Estimator

Theorem 1

Let f be a bounded density with distribution function F and let G be any censoring distribution satisfying $G(\tau_F) < 1$, where $\tau_F = sup\{t : F(t) < 1\}$. Then the Bayesian estimator (3.1) defined at a point of estimation x by $\hat{f}_h(x) = \int k(x, y, h) d\hat{F}(y)$ satisfies $|\hat{f}_h(x) - f(x)| \quad \to \quad 0 \quad a.s.$, whenever $h = h_n(x) \to 0$ as $n \to \infty$ at a rate slower than $\sqrt{\frac{\log \log n}{n}}$.

Proof :

The Bayesian estimator of the lifetime density $f(x)$ is given by

$$\hat{f}_h(x) = \sum_{i=1}^{n} s_j k(x, \ln z_j, h)$$

where $k(x, \mu, \sigma)$ is a lognormal kernel with parameters μ and σ . Consider $\hat{f}_h(x)$ for a fixed x and a particular h.

Then,

$$\hat{f}_h(x) = \sum_{i=1}^{n} s_j k(x, \ln Z_j, h) = \int_0^\infty k(x, \ln u, h)\ d\hat{F}(u)$$

where s_j is the jump size of the Kaplan-Meier survival function at the observation value Z_j. Therefore we can write,

$$\hat{f}_h(x) = \int_0^\infty k(x, \ln u, h)\ d[1 - \hat{S}_n(u)] = -\int_0^\infty k(x, \ln u, h)\ d\hat{S}_n(u)$$

where $\hat{S}_n(u)$ is the Kaplan-Meier estimator of the survival function. Consider $f_h(x)$ defined in Chapter 3,

$$f_h(x) = -\int_0^\infty k(x, \ln u, h)\ dS(u)\ .$$

Then,

$$\begin{aligned} |\hat{f}_h(x) - f(x)| &= |\hat{f}_h(x) - f_h(x) + f_h(x) - f(x)| \\ &\leq |\hat{f}_h(x) - f_h(x)| + |f_h(x) - f(x)|\ . \end{aligned} \tag{4.1}$$

Consider the first term on the right hand side of (4.1),

$$\left|\hat{f}_h(x) - f_h(x)\right| = \left|\int_0^\infty k(x, \ln u, h)\ d\hat{S}_n(u) - \int_0^\infty k(x, \ln u, h)\ dS(u)\right|$$

$$= \left|\int_0^\infty k(x, \ln u, h)\ d[\hat{S}_n(u) - S(u)]\right|.$$

Integration by parts yields,

$$\begin{aligned}
\left|\hat{f}_h(x) - f_h(x)\right| &= \left| k(x, \ln u, h)[\hat{S}_n(u) - S(u)]\Big|_0^\infty - \int_0^\infty [\hat{S}_n(u) - S(u)] d_u k(x, \ln u, h)\right| \\
&= \left|\int_0^\infty [\hat{S}_n(u) - S(u)] d_u k(x, \ln u, h)\right| \\
&\leq \int_0^\infty \left|\hat{S}_n(u) - S(u)\right| \left| d_u k(x, \ln u, h)\right|.
\end{aligned}$$

Then we can write,

$$\begin{aligned}
|\ \hat{f}_h(x) - f_h(x)\ | &\leq \int_0^\infty \sup_{0<u<\infty} |\hat{S}_n(u) - S(u)|\ \left| d_u k(x, \ln u, h)\right| \\
&\leq \sup_{0<u<\infty} |\hat{S}_n(u) - S(u)| \int_0^\infty \left| d_u k(x, \ln u, h)\right|.
\end{aligned}$$

Now consider the integral on the right hand side of the above inequality

$$= \int_0^{\infty} \left| \frac{d}{du}\{\frac{1}{\sqrt{2\pi}}\frac{1}{xh}e^{-\frac{1}{2}(\frac{\ln x - \ln u}{h})^2}\}\ du \right|$$

$$= \int_0^{\infty} \left| \frac{1}{\sqrt{2\pi}}\frac{1}{xh}e^{-\frac{1}{2}(\frac{\ln x - \ln u}{h})^2}\left(\frac{\ln x - \ln u}{h}\right)\frac{1}{hu}\ du \right|$$

$$= \int_0^{\infty} \left| \frac{1}{(2\pi)^{\frac{1}{4}}}\frac{1}{xh^{\frac{1}{2}}u^{\frac{1}{2}}}e^{-\frac{1}{4}(\frac{\ln x - \ln u}{h})^2}\frac{1}{(2\pi)^{\frac{1}{4}}}\frac{1}{h^{\frac{1}{2}}u^{\frac{1}{2}}}e^{-\frac{1}{4}(\frac{\ln x - \ln u}{h})^2}\left(\frac{\ln x - \ln u}{h^2}\right) \right| du$$

$$= \int_0^{\infty} \left| \psi(u).\psi^*(u) \right| du$$

where

$$\begin{aligned} \psi(u) &= \frac{1}{(2\pi)^{\frac{1}{4}}}\frac{1}{xh^{\frac{1}{2}}u^{\frac{1}{2}}}e^{-\frac{1}{4}(\frac{\ln x - \ln u}{h})^2} \\ \psi^*(u) &= \frac{1}{(2\pi)^{\frac{1}{4}}}\frac{1}{h^{\frac{1}{2}}u^{\frac{1}{2}}}e^{-\frac{1}{4}(\frac{\ln x - \ln u}{h})^2}\left(\frac{\ln x - \ln u}{h^2}\right) \ . \end{aligned}$$

Then by Hölder's inequality we obtain

$$\int_0^{\infty} |\ \psi(u).\psi^*(u)|\ du \quad \leq \quad \left[\int_0^{\infty} |\psi(u)|^2\ du\right]^{\frac{1}{2}} \left[\int_0^{\infty} |\psi^*(u)|^2\ du\right]^{\frac{1}{2}} \ .$$

Therefore the integral $\int_0^\infty \mid d_u k(x, \ln u, h) \mid$

$$\leq \left[\int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{1}{x^2 h u} e^{-\frac{1}{2}(\frac{\ln x - \ln u}{h})^2} du \right]^{\frac{1}{2}} \left[\int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{1}{hu} e^{-\frac{1}{2}(\frac{\ln x - \ln u}{h})^2} \left(\frac{\ln x - \ln u}{h^2} \right)^2 du \right]^{\frac{1}{2}}$$

$$= \left[\frac{1}{x^2} \int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{1}{hu} e^{-\frac{1}{2}(\frac{\ln u - \ln x}{h})^2} du \right]^{\frac{1}{2}} \left[\frac{1}{h^4} \int_0^\infty (\ln u - \ln x)^2 \frac{1}{\sqrt{2\pi}} \frac{1}{hu} e^{-\frac{1}{2}(\frac{\ln u - \ln x}{h})^2} du \right]^{\frac{1}{2}} .$$

The integral in the first term of the above inequality is just unity because the integrand is nothing but the density of a lognormal random variable. The integral of the second term is the expectation of $[\ln U - \ln x]^2$ with U being a lognormal random variable with parameters $\mu = \ln x$ and $\sigma = h$.

Note that

$$U \sim LogNormal(\ \ln x, h\ ) \quad \Rightarrow \quad \ln U \sim N(\ \ln x, h^2\ ) .$$

Therefore,

$$E_u[\ln U - \ln x]^2 \ = \ Var[\ln U] \ = \ h^2 .$$

This results in

$$\int_0^{\infty} \left| \, d_u k(x, \ln u, h) \, \right| \;\leq\; \left[\frac{1}{x^2} . 1 \right]^{\frac{1}{2}} \left[\frac{1}{h^4} . h^2 \right]^{\frac{1}{2}}$$

$$= \frac{1}{xh}.$$

Finally, we get

$$\left| \, \hat{f}_h(x) - f_h(x) \, \right| \;=\; \left| \int_0^{\infty} [\hat{S}_n(u) - S(u)] d_u k(x, \ln u, h) \right| \leq \sup_{0<u<\infty} |\hat{S}_n(u) - S(u)| \, . \frac{1}{xh}$$

where $\hat{S}_n(\cdot)$ is the Kaplan-Meier product limit estimator of the true survival function $S(\cdot)$ associated with the underlying lifetime distribution $F(\cdot)$. Földes & Rejtő (1981) have shown that $\hat{S}_n(t)$ is almost sure consistent with rate $O\left(\sqrt{\frac{\log\log n}{n}}\right)$ if $G(\tau_F) < 1$ where $\tau_F = \sup_x\{x \ : \ F(x) < 1\}$

i.e.

$$P\left[\sup_{-\infty<t<\infty} |\hat{S}_n(t) - S(t)| = O\left(\sqrt{\frac{\log\log n}{n}} \right) \right] = 1 \qquad as \qquad n \rightarrow \infty$$

where F and G are the distribution functions of the lifetime and censoring random variables respectively.

Using this result we see that for bandwidth sequences $h = h_n$, that converge to zero at a rate slower than $\sqrt{\frac{\log\log n}{n}}$, we can achieve strong convergence of $\left|\hat{f}_h(x) - f_h(x)\right|$ to zero in (4.1) provided the distribution function G of the censoring random variable satisfies the condition, $G(\tau_F) < 1$ where, $\tau_F = \sup_x\{x \ : \ F(x) < 1\}$.

Now consider the second term $|\, f_h(x) - f(x)\,|$ in equation (4.1). Noting that

$$
\begin{aligned}
f_h(x) &= \int_0^\infty k(x, \ln u, h)\ dF(u) \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{1}{xh} e^{-\frac{1}{2}(\frac{\ln x - \ln u}{h})^2}\ f(u)\ du\ ,
\end{aligned}
$$

we make the substitution $\left(\frac{\ln x - \ln u}{h}\right) = v$ to get $\ln u = \ln x - hv$. This gives us $u = e^{\ \ln x - hv}$ and $du = xe^{-hv}(-h)dv$. Therefore $f_h(x)$ becomes,

$$
\begin{aligned}
f_h(x) &= \int_\infty^{-\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{xh} e^{-\frac{1}{2}v^2}\ f(xe^{-hv})xe^{-hv}(-h)\ \ dv \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}\ f(xe^{-hv})e^{-hv}\ \ dv\ .
\end{aligned}
$$

Now we examine $\lim_{n\to\infty} f_h(x)$.

$$\begin{aligned}
\lim_{n\to\infty} f_h(x) &= \lim_{h\to 0} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} \; f(xe^{-hv}) e^{-hv} \;\; dv \\
&= \left[\lim_{h\to 0} e^{\frac{1}{2}h^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v+h)^2} \; f(xe^{-hv}) \;\; dv \right] \\
&= \left[\lim_{h\to 0} e^{\frac{1}{2}h^2} \right] \cdot \left[\lim_{h\to 0} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v+h)^2} \; f(xe^{-hv}) \;\; dv \right]
\end{aligned}$$

provided both limits exists. It is clear that the first limit exists and equals to one. The second limit also exists as the terms in the integrand are bounded in the following manner. The first term, $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v+h)^2}$ is bounded on $(-\infty, \infty)$ regardless of h. The second term $f(xe^{-hv})$ is bounded on $(0, \infty)$ as f is a bounded lifetime pdf on $(0, \infty)$, and hence is bounded on $(-\infty, \infty)$.

Also note that

$$\lim_{h\to 0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v+h)^2} \; f(xe^{-hv}) \quad = \quad \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} \; f(x) \; .$$

Therefore by the bounded convergence theorem, we get

$$\begin{aligned}
\lim_{h\to 0}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(v+h)^2}\ f(xe^{-hv})\ \ dv &= \int_{-\infty}^{\infty}\lim_{h\to 0}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(v+h)^2}\ f(xe^{-hv})\ dv \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}v^2}\ f(x)\ dv \\
&= f(x)
\end{aligned}$$

Hence $\mid f_h(x) - f(x) \mid\ \rightarrow 0$ as $n \rightarrow \infty$. We have now shown that the two terms in (4.1) converge to zero almost surely as $n \rightarrow \infty$ for a suitably chosen bandwidth sequence h_n that converges to zero at a rate slower than $\sqrt{\frac{\log\log n}{n}}$. This proves the strong convergence of $\left| \hat{f}_h(x) - f(x) \right|$ to zero.

4.2 Convergence of the Bayesian Bandwidths

We now establish the convergence of the Bayesian bandwidths to zero as n$\rightarrow \infty$. This is a highly desirable property for any bandwidth estimator as it will ensure that the window width of the kernel estimator will shrink as more and more data is available.

Theorem 2

The Bayesian bandwidth estimator $\tilde{h}_n(x)$ at a point of estimation x given in (3.4) will converge to zero almost surely as $n \rightarrow \infty$, for prior parameter sequences satisfying $\beta_n \rightarrow 0$ as $n \rightarrow \infty$ for fixed $\alpha \in \mathbb{N}$ and $\alpha > 2$.

Proof :

First, note that (3.4) can be written as,

$$\tilde{h}_n(x) = \frac{\Gamma(\alpha)\displaystyle\int \left[\frac{1}{\beta} + \frac{1}{2}(\ln x - \ln u)^2\right]^{-\alpha} d\hat{S}_n(u)}{\Gamma(\alpha^*)\displaystyle\int \left[\frac{1}{\beta} + \frac{1}{2}(\ln x - \ln u)^2\right]^{-\alpha^*} d\hat{S}_n(u)}$$

where α and β are the prior parameters of the inverted gamma distribution and integration is over $[0, \infty)$, the support of the density . We would pick β as a function of the sample size so that we could make the prior to be concentrated at zero as $n \to \infty$.

Let β_n be a sequence of real numbers that diverges. Then, for fixed $\alpha > 2$, the mean of the prior distribution $\frac{1}{\beta_n(\alpha-1)} \to 0$ *as* $n \to \infty$ and the variance $\frac{1}{\beta_n^2(\alpha-1)^2(\alpha-2)} \to 0$ *as* $n \to \infty$. Now rewriting (3.4) by letting $c = \frac{\Gamma(\alpha)}{\Gamma(\alpha^*)}$ and adding and subtracting the true survival function S of the density we get

$$\begin{aligned}
\tilde{h}_n(x) &= \frac{c\displaystyle\int \left[\frac{1}{\beta_n} + \frac{1}{2}(\ln x - \ln u)^2\right]^{-\alpha} d[\hat{S}_n(u) - S(u) + S(u)]}{\displaystyle\int \left[\frac{1}{\beta_n} + \frac{1}{2}(\ln x - \ln u)^2\right]^{-\alpha^*} d[\hat{S}_n(u) - S(u) + S(u)]} \\
&= \frac{c\displaystyle\int \left[\frac{1}{\beta_n} + \frac{1}{2}(\ln x - \ln u)^2\right]^{-\alpha} d[\hat{S}_n(u) - S(u)] + \ c\displaystyle\int \left[\frac{1}{\beta_n} + \frac{1}{2}(\ln x - \ln u)^2\right]^{-\alpha} dS(u)}{\displaystyle\int \left[\frac{1}{\beta_n} + \frac{1}{2}(\ln x - \ln u)^2\right]^{-\alpha^*} d[\hat{S}_n(u) - S(u)] + \displaystyle\int \left[\frac{1}{\beta_n} + \frac{1}{2}(\ln x - \ln u)^2\right]^{-\alpha^*} dS(u)}.
\end{aligned}$$

By the consistency result of the Product Limit Estimator $\hat{S}_n$ by Földes & Rejtő (1981), we see that the first integral in both the numerator and the denominator of the above expression converges to zero almost surely as $n \to \infty$. Hence, to prove that $\tilde{h}_n(x)$ converge to zero, it suffices to show that the ratio

$$R_n(x) \quad = \quad \frac{c}{\sqrt{\beta_n}} \frac{\displaystyle\int \left[1 + \frac{\beta_n}{2}(\ln x - \ln u)^2\right]^{-\alpha} dS(u)}{\displaystyle\int \left[1 + \frac{\beta_n}{2}(\ln x - \ln u)^2\right]^{-\alpha^*} dS(u)}$$

goes to zero as $n \to \infty$. Now let $\phi_n(u) = \left[1 + \frac{\beta_n}{2}(\ln x - \ln u)^2\right]$ and $\epsilon = \epsilon_n(x)$ be a sequence such that $0 < \epsilon_n(x) < x$ and $\epsilon_n(x) \to 0$ as $n \to \infty$. Then

$$\begin{aligned}
R_n(x) \quad &= \quad \frac{c}{\sqrt{\beta_n}} \frac{\displaystyle\int_0^\infty \phi_n(u)^{-\alpha} dS(u)}{\displaystyle\int_0^\infty \phi_n(u)^{-\alpha^*} dS(u)} \\
&= \quad \frac{c}{\sqrt{\beta_n}} \frac{\displaystyle\int_0^{x-\epsilon} \phi_n(u)^{-\alpha} dS(u) + \int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha} dS(u) + \int_{x+\epsilon}^{\infty} \phi_n(u)^{-\alpha} dS(u)}{\displaystyle\int_0^{x-\epsilon} \phi_n(u)^{-\alpha^*} dS(u) + \int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha^*} dS(u) + \int_{x+\epsilon}^{\infty} \phi_n(u)^{-\alpha^*} dS(u)} \\
&\leq \quad \frac{c}{\sqrt{\beta_n}} \frac{\displaystyle\int_0^{x-\epsilon} \phi_n(u)^{-\alpha} dS(u) + \int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha} dS(u) + \int_{x+\epsilon}^{\infty} \phi_n(u)^{-\alpha} dS(u)}{\displaystyle\int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha^*} dS(u)} \\
&= \quad \frac{c}{\sqrt{\beta_n}} \frac{\displaystyle\int_0^{x-\epsilon} \phi_n(u)^{-\alpha} dS(u) + \int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha} f(u) du + \int_{x+\epsilon}^{\infty} \phi_n(u)^{-\alpha} dS(u)}{\displaystyle\int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha^*} f(u) du} \qquad (4.2) \\
&\leq \quad \frac{c}{\sqrt{\beta_n}} \frac{\displaystyle\int_0^{x-\epsilon} \phi_n(u)^{-\alpha} dS(u) + \sup_{x-\epsilon<u<x+\epsilon} f(u) \int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha} du + \int_{x+\epsilon}^{\infty} \phi_n(u)^{-\alpha} dS(u)}{\displaystyle\inf_{x-\epsilon<u<x+\epsilon} f(u) \int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha^*} du} .
\end{aligned}$$

Now, consider the Taylor series expansion of $g(u) = (\ln x - \ln u)$ of order 1 around the point x in the neighbourhood of $(x-\epsilon, x+\epsilon)$ with $\epsilon = \epsilon_n \to 0$ as $n \to \infty$. Then,

$$
\begin{aligned}
g(u) &\approx g(x) + (u-x)g^{'}(u)|_{u=x} \\
&= \ln x - \ln x + (u-x)\frac{-1}{x} \\
g(u) &\approx \frac{(x-u)}{x}
\end{aligned}
$$

For fixed β_n , consider the two integrals $\int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha} du$ and $\int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha^*} du$ in (4.2),

$$
\begin{aligned}
\int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha} du &\approx \int_{x-\epsilon}^{x+\epsilon} \frac{1}{\left[1 + \frac{\beta_n}{2}(\frac{x-u}{x})^2\right]^{\alpha}} du \\
&= \int_{x-\epsilon}^{x+\epsilon} \frac{1}{\left[1 + \frac{\beta_n}{2}(\frac{x-u}{x})^2\right]^{\alpha}} du \ .
\end{aligned}
$$

Let $\sqrt{\frac{\beta_n}{2x^2}}(x-u) = w$. Then we can rewrite the above integral as

$$
\int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha} du \approx \sqrt{\frac{2x^2}{\beta_n}} \int_{-\delta}^{\delta} \frac{1}{(1+w^2)^{\alpha}} dw
$$

where $\delta = \epsilon\sqrt{\frac{\beta_n}{2x^2}}$ that converge to ∞ as $n \to \infty$ by propoperly choosing $\epsilon_n \to 0$ and $\beta_n \to \infty$.

Now repeatedly using the fact that for $\alpha \in \mathbb{N}$

$$\int \frac{1}{(a^2+u^2)^{\alpha}} du = \frac{1}{2a^2(\alpha-1)} \left(\frac{u}{(a^2+u^2)^{\alpha-1}} + (2\alpha-3) \int \frac{1}{(a^2+u^2)^{\alpha-1}} \right)$$

we get

$$\int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha} du \approx \sqrt{\frac{2x^2}{\beta_n}} \left(\sum_{i=1}^{\alpha-1} \frac{2\delta}{[1+\delta^2]^i} + K(\alpha)\ tan^{-1}(\delta) \right) ,$$

where

$$K(\alpha) = \frac{(2\alpha-3)(2\alpha-5)(2\alpha-7)...1}{2^{\alpha-1}(\alpha-1)(\alpha-2)(\alpha-3)...1}$$

and

$$\int_{x-\epsilon}^{x+\epsilon} \phi_n(u)^{-\alpha^*} du \approx \sqrt{\frac{2x^2}{\beta_n}} \left(\sum_{i=1}^{\alpha-1} \frac{2\delta}{[1+\delta^2]^{i+\frac{1}{2}}} + K(\alpha^*)\ \frac{2\delta}{\sqrt{1+\delta^2}} \right)$$

where

$$K^*(\alpha) = \frac{(2\alpha^*-3)(2\alpha^*-5)(2\alpha^*-7)...1}{2^{\alpha^*-1}(\alpha^*-1)(\alpha^*-2)(\alpha^*-3)...1}$$

with $\alpha^* = \alpha + \frac{1}{2}$.

Using these approximations in (4.2) we get

$$R_n(x) \quad \leq \quad \frac{c}{\sqrt{\beta_n}} \frac{\int_A + \sup_{x-\epsilon<u<x+\epsilon} f(u)\sqrt{\frac{2x^2}{\beta_n}}\left(\sum_{i=1}^{\alpha-1} \frac{2\delta}{[1+\delta^2]^i} + K(\alpha)\ tan^{-1}(\delta)\right) + c_n + \int_B}{\inf_{x-\epsilon<u<x+\epsilon} f(u)\sqrt{\frac{2x^2}{\beta_n}}\left(\sum_{i=1}^{\alpha-1} \frac{2\delta}{[1+\delta^2]^{i+\frac{1}{2}}} + K(\alpha^*)\ \frac{2\delta}{\sqrt{1+\delta^2}}\right) + c_n^*}$$

where c_n and c_n^* are the first order approximation errors in the Taylor expansion and

$$\int_A \quad = \quad \int_0^{x-\epsilon} \phi_n(u)^{-\alpha} dS(u) \ ,$$

$$\int_B \quad = \quad \int_{x+\epsilon}^{\infty} \phi_n(u)^{-\alpha} dS(u) \ .$$

We note that $\phi_n(u)^{-\alpha}$ is a sequence of functions bounded by 1 and converging to zero and so is the sequence $\phi_n(u)^{-\alpha^*}$. Therefore, by the bounded convergence theorem we obtain,

$$\lim_{n\to\infty} \int_0^{x-\epsilon} \phi_n(u)^{-\alpha} dS(u) = \quad 0 \quad a.e$$

and

$$\lim_{n\to\infty} \int_{x+\epsilon}^{\infty} \phi_n(u)^{-\alpha^*} dS(u) = 0 \quad a.e \ .$$

Hence,

$$\lim_{n\to\infty} R_n(x) \quad \leq \quad c \; \frac{0 + \sup\limits_{x-\epsilon<u<x+\epsilon} f(u) \lim\limits_{n\to\infty} \left[\sqrt{\frac{2x^2}{\beta_n}} \left(\sum_{i=1}^{\alpha-1} \frac{2\delta}{[1+\delta^2]^i} + K(\alpha) \;\; tan^{-1}(\delta) \right) + c_n \right] + 0}{\inf\limits_{x-\epsilon<u<x+\epsilon} f(u) \lim\limits_{n\to\infty} \left[\sqrt{2x^2} \left(\sum_{i=1}^{\alpha-1} \frac{2\delta}{[1+\delta^2]^{i+\frac{1}{2}}} + K(\alpha^*) \;\; \frac{2\delta}{\sqrt{1+\delta^2}} \right) + c_n^* \right]}$$

$$= \quad \frac{0}{\inf\limits_{x-\epsilon<u<x+\epsilon} 2f(u) \; \sqrt{2x^2} K(\alpha^*)}$$

Therefore, when $\beta_n \to \infty$ as $n \to \infty$ we have $h_n(x) \to 0$ almost surely as desired.

Chapter 5

Simulation Study

We will now discuss the results obtained from the simulation study carried out to assess the performance of the proposed lognormal kernel density estimator (KDE). We study the proposed estimator in two perspectives. First, we will examine the effect of the kernel on the proposed estimator. Then, we investigate how the Bayesian bandwidths have affected the performance of the said estimator.

5.1 Overview

We use both pointwise and mean integrated squared error criteria as our measure of closeness of the proposed density estimator to a target density. In particular, we will compare the proposed lognormal KDE with another KDE which uses an inverse Gaussian kernel that was shown to be promising in Kulasekera & Padgett (2006), with both estimators using Bayesian local bandwidths associated with their respective kernels. This comparison is expected to reflect the effect of the kernel on the KDE. Then we will assess the performance of the proposed lognormal KDE with two different choices of bandwidth selection methods, namely the Bayesian local bandwidths and the Least Squares Cross Validation (LSCV) bandwidths proposed by Marron & Padgett (1987) to demonstrate the superiority of the Bayesian local bandwidths.

We define the pointwise estimated mean squared error (EMSE) of a density estimator $\hat{f}(t)$ at a point of estimation t , by

$$EMSE(\ \hat{f}(t)\ ) = \frac{\sum_{i=1}^{N} \left[\hat{f}_i(t) - f(t) \right]^2}{N} \tag{5.1}$$

where N is the number of simulations which was chosen to be 1000 and all simulations were carried out using R (2004).

Then, we will examine the ratio

$$R_{\hat{f}_1,\hat{f}_2}(t) = \frac{EMSE(\ \hat{f}_1(t)\ )}{EMSE(\ \hat{f}_2(t)\ )} \tag{5.2}$$

over a grid of t values in the domain of the underlying density, where $\hat{f}_1$ and $\hat{f}_2$ are any two density estimators of a target density f. We plot these ratios against t to assess the pointwise performance of the two density estimators $\hat{f}_1$ and $\hat{f}_2$. Furthermore, we will use the mean integrated squared error (MISE) criterion defined in (1.2) as a global measure of performance of the density estimators.

5.2 Comparison of the Two Kernels

The performance of the two kernels, namely the lognormal and the inverse Gaussian, is compared with 3 different sample sizes n=20,n=40 and n=100. Simulated data from Weibull(θ,1) densities with pdf defined as

$$f(t) = \theta t^{\theta-1} e^{-t^{\theta}} \ \ , t > 0 \ \ , \theta > 0$$

were used and performance was assessed under 3 different failure rate models by changing the parameter θ in the Weibull density where $\theta = 0.5, 1, 1.5$ corresponding to decreasing, constant and increasing failure rates respectively. These data are randomly right censored by an exponential(λ) variate with pdf

$$g(x) = \lambda e^{-\lambda x} \ \ , x > 0 \ \ , \lambda > 0$$

where λ was chosen to achieve three levels of censoring namely 10% , 20% and 50%. All comparisons were made with both KDE s using Bayesian local bandwidths associated with their respective kernels.

5.2.1 Decreasing Failure Rate Data

Density estimates of a Weibull(0.5,1) density were computed using the the two KDE's (LN and IG) and the ratio $R_{\hat{f}_{IG},\hat{f}_{LN}}(t)$ was plotted against the domain values t of the underlying density. As shown in Figure 5.1 the lognormal KDE clearly outperformed the inverse Gaussian KDE in the neighborhood of the origin. However, the decreasing ratio values indicate that the pointwise estimates for the lognormal KDE are not as close as the ones we get from the inverse Gaussian KDE toward the end of the support. The effect of the censoring fraction was only observable at the tail of the support. In particular, within that region, the higher the censoring, the better the performance of the inverse Gaussian KDE when compared with the lognormal KDE. The increasing sample size shows no significant effect on the performance of the KDE's.

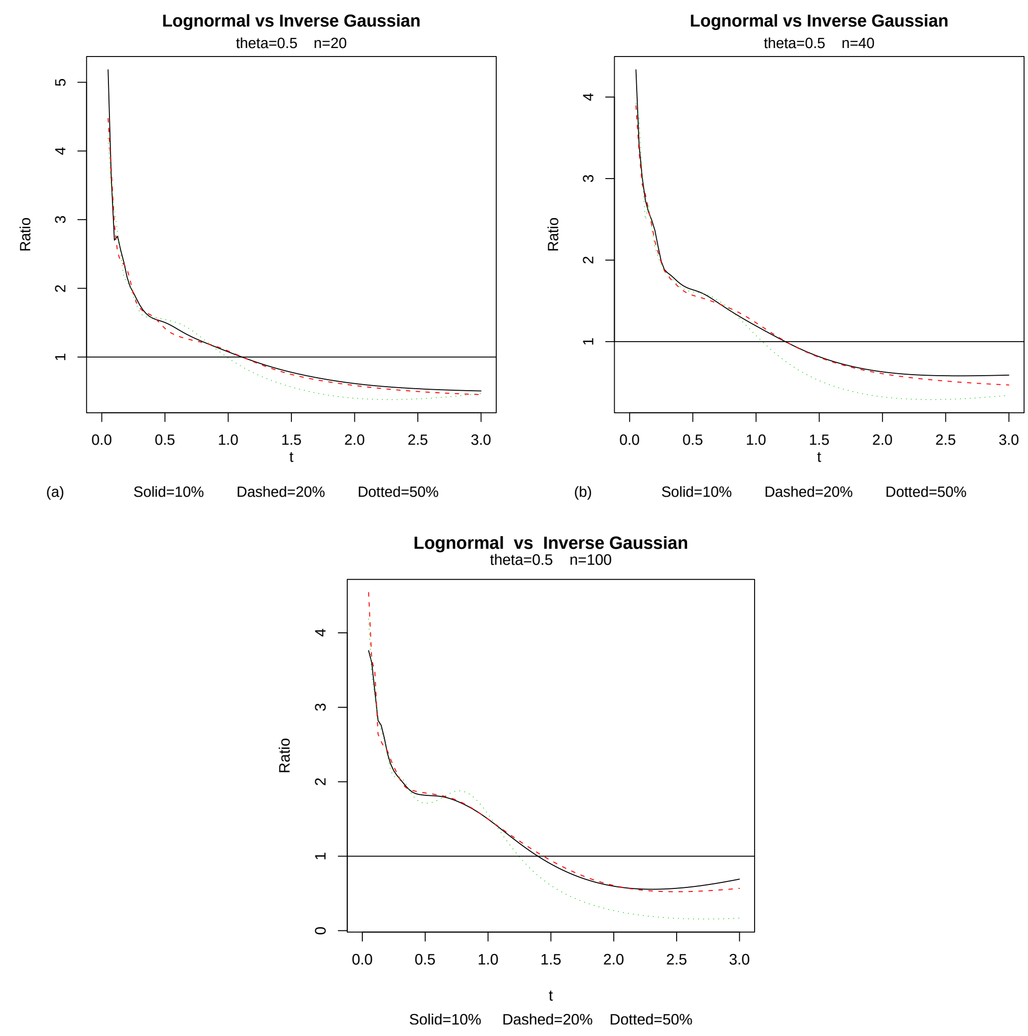

Figure 5.1: *Comparison of inverse Gaussian and lognormal kernels using pointwise error ratios with DFR data. (a) n=20 (b) n=40 (c) n=100.*

5.2.2 Constant Failure Rate Data

In this setting too, we observe a similar pattern in the performances of the two density estimators as with the decreasing failure rate data. The lognormal KDE is far superior than the inverse Gaussian KDE in the neighborhood of the origin. Further, the proportion of domain values in which the lognormal KDE is superior have increased with the increment of the sample size as indicated by Figure 5.2.

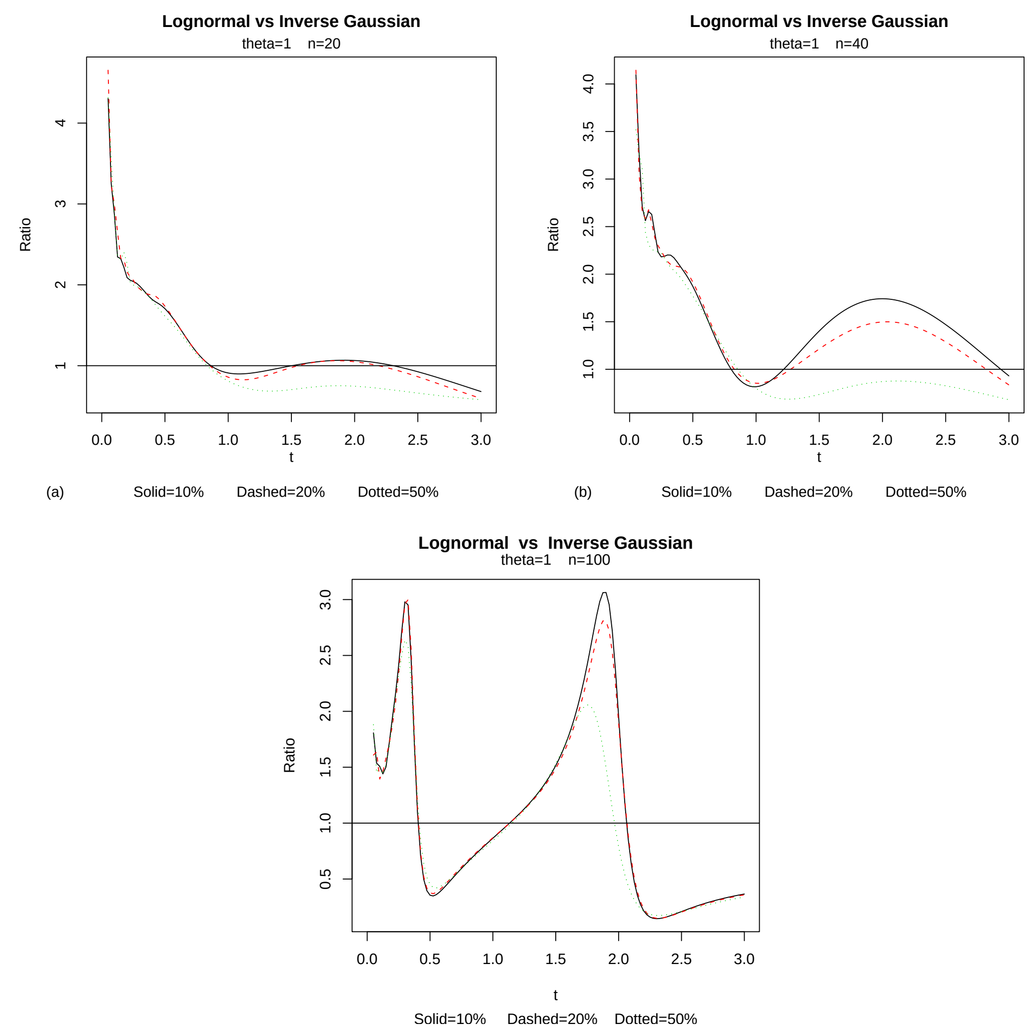

Figure 5.2: *Comparison of inverse Gaussian and lognormal kernels using pointwise error ratios with CFR data. (a) n=20 (b) n=40 (c) n=100.*

5.2.3 Increasing Failure Rate Data

Almost all the features that we observed in the previous two settings can be seen with increasing failure rate data as well. The most noticeable feature in this case is that the lognormal KDE performed better than the inverse Gaussian KDE over a large proportion of the support with the increment of the sample size. As before, the effect of the censoring was only observed toward the tail of support as shown in Figure 5.3.

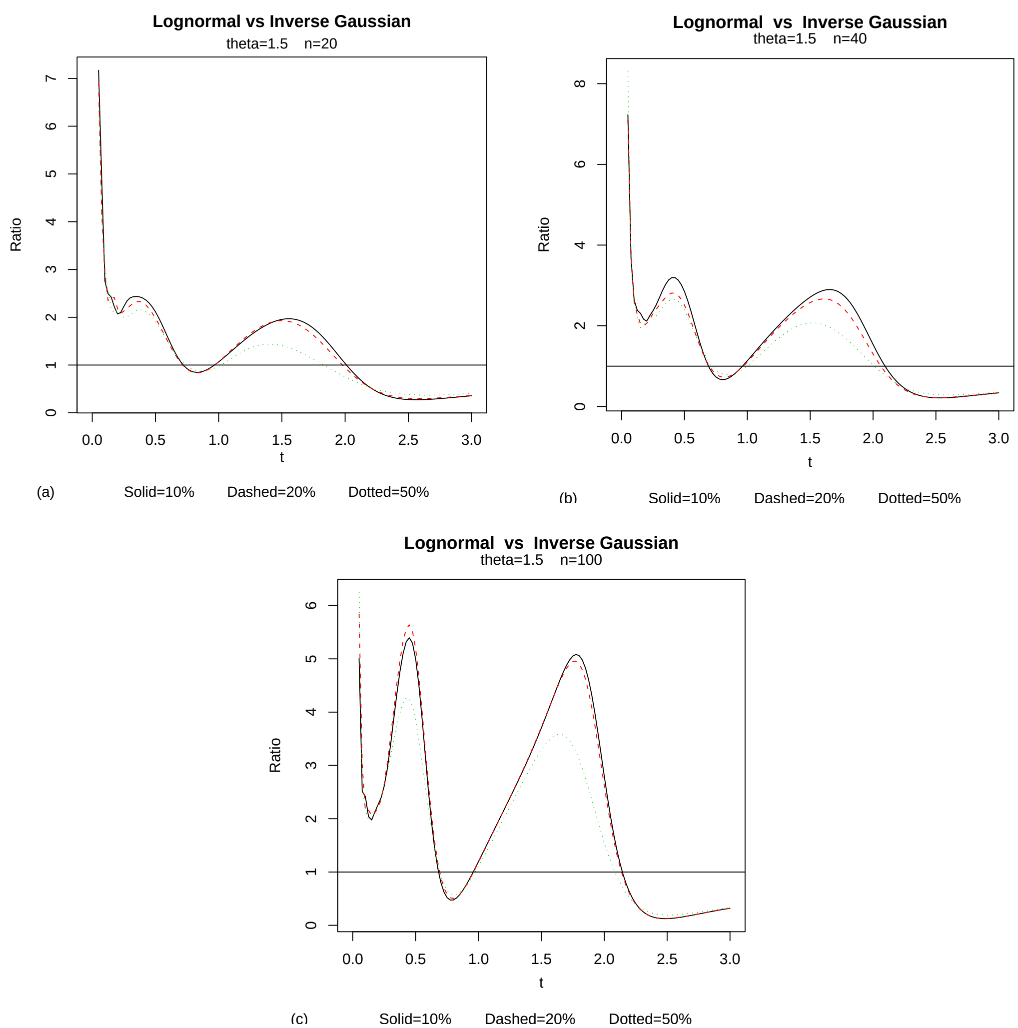

Figure 5.3: *Comparison of inverse Gaussian and lognormal kernels using pointwise error ratios with IFR data. (a) n=20 (b) n=40 (c) n=100.*

5.3 Comparison of the Two Bandwidth Selection Methods

Smoothing parameter (bandwidth) selection is an extremely important step in any density estimation problem. Numerous studies have been done on this issue and still there is no 'ideal' method that one can use. However, there are several well established methods available for experimenters for bandwidth selection. We will now compare the proposed Bayesian local bandwidth selection method with a well known bandwidth selection procedure, namely the least squares cross validation (LSCV) where both methods use a lognormal kernel.

5.3.1 Decreasing Failure Rate Data

Simulated data were generated from a Weibull(0.5,1) density and density estimates were computed using the Bayesian and the LSCV bandwidth selection. Then, point wise error ratios,

$$R_{\hat{f}_{LSCV},\hat{f}_{Bayes}}(t) = \frac{EMSE(\hat{f}_{LSCV}(t))}{EMSE(\hat{f}_{Bayes}(t))}$$

were plotted against the values of the support of the underlying density. Figure 5.4 clearly indicates that the Bayesian bandwidth selection method is far superior than the LSCV method. Moreover, neither the sample size nor the censoring level has any appreciable effect on the pointwise error ratio, although under 50% censoring the LSCV method appears to be have a lower pointwise MSE than 10% and 20% censoring levels, at the first half of the support.

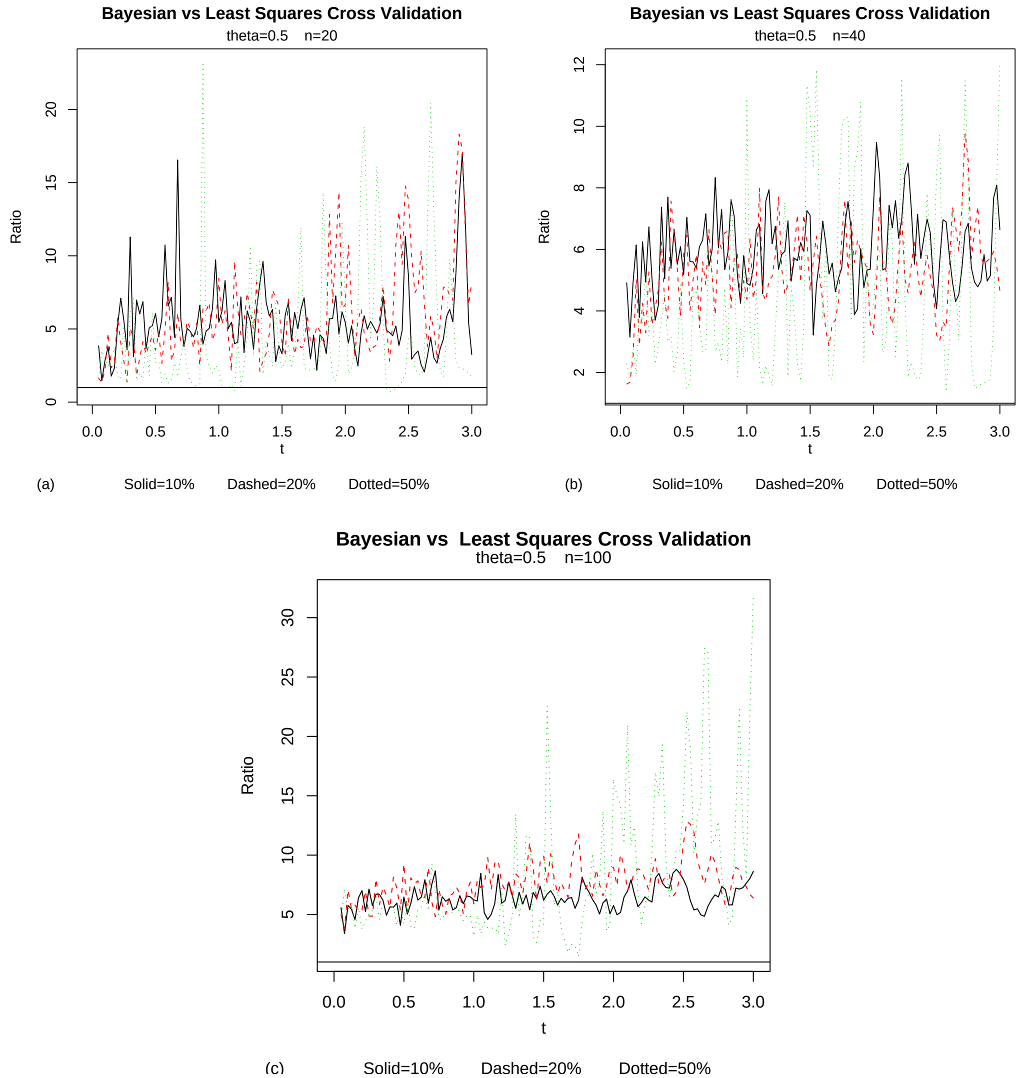

Figure 5.4: *Comparison of Bayesian and LSCV bandwidths using pointwise error ratios with DFR data. (a) n=20 (b) n=40 (c) n=100.*

5.3.2 Constant Failure Rate Data

Density estimates were computed using the Bayesian and the LSCV bandwidths based on data generated from a Weibull(1,1) density and then, pointwise error ratios were plotted against the values of the support as before. We observe a uniform dominance

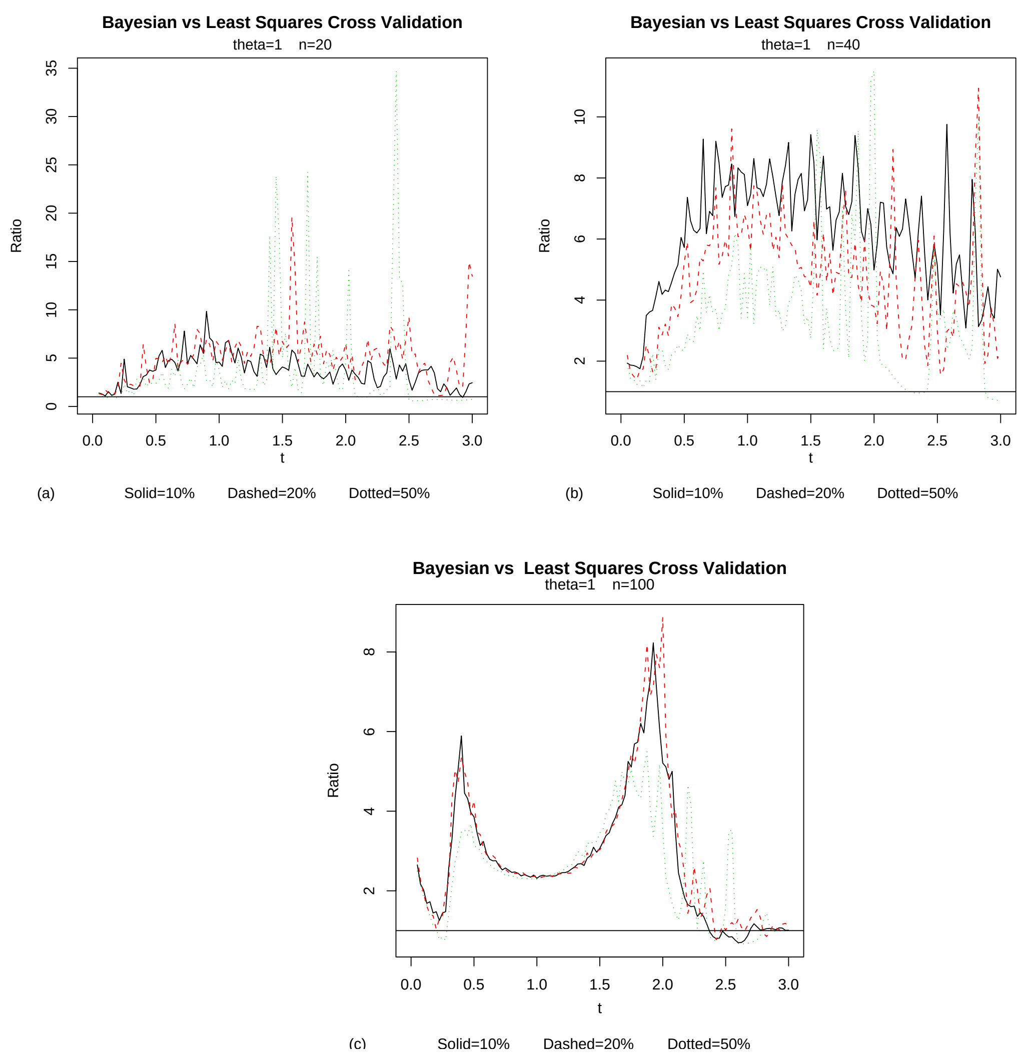

Figure 5.5: *Comparison of Bayesian and LSCV bandwidths using pointwise error ratios with CFR data. (a) n=20 (b) n=40 (c) n=100.*

of the Bayesian bandwidths over the LSCV bandwidths except at the tail of the support. Toward the tail, both methods appears to have equal pointwise MSEs.

5.3.3 Increasing Failure Rate Data

The behavior of the pointwise error ratios in this setting is almost similar to the ones we observe with the DFR and CFR data. The Bayesian bandwidths clearly outperformed the LSCV bandwidths in terms of pointwise MSE. No significant impact can be seen with the increasing sample size or within the 3 levels of censoring.

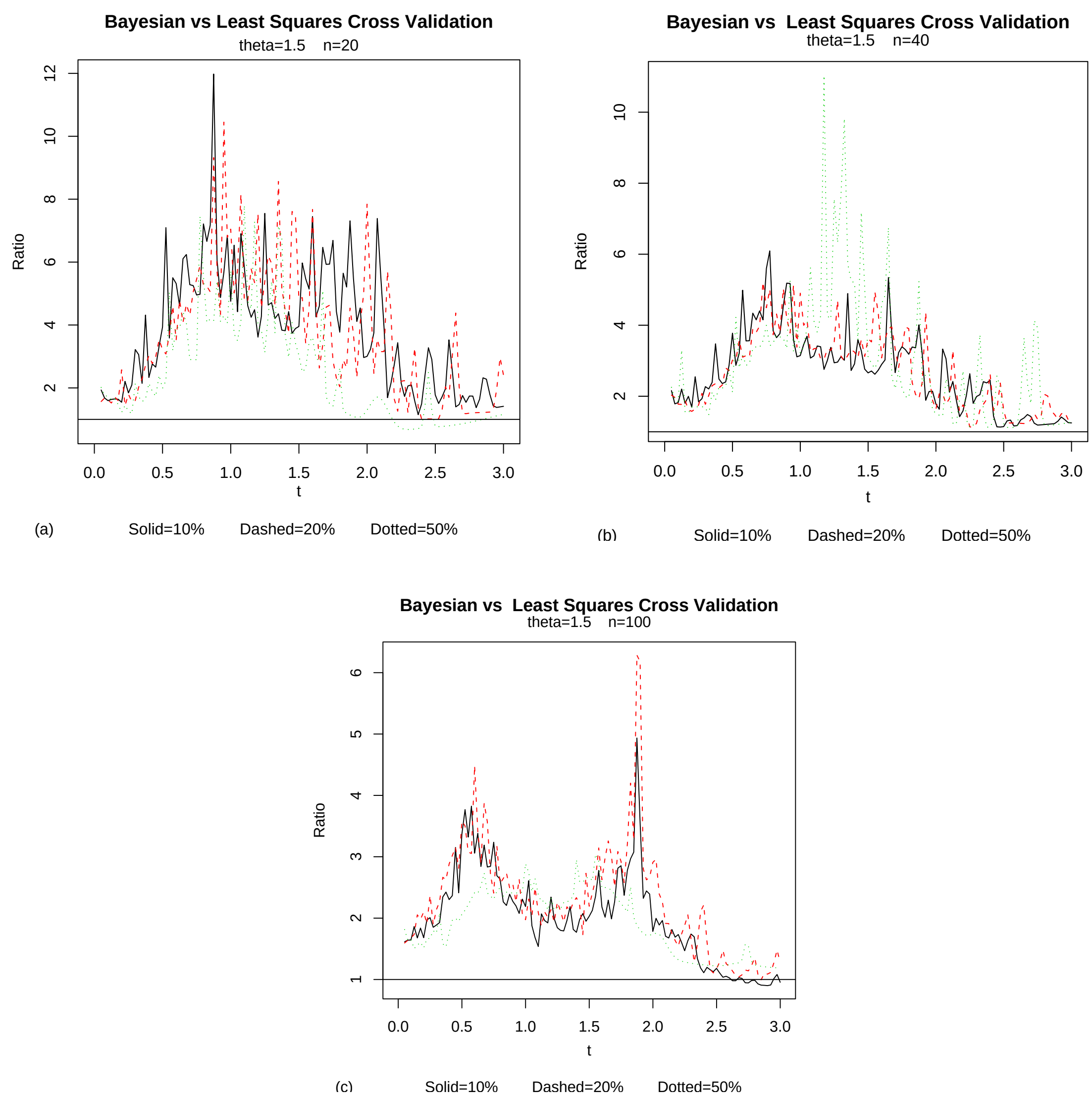

Figure 5.6: *Comparison of Bayesian and LSCV bandwidths using pointwise error ratios with IFR data. (a) n=20 (b) n=40 (c) n=100.*

5.4 Performance under Varying Scale Parameters in the Prior

It is highly desirable for any bandwidth estimator to converge to zero as $n \to \infty$, to achieve unbiasedness in the density estimator for which the bandwidths are computed. As discussed in section 4.2, to achieve this convergence in the proposed Bayesian bandwidths, we need to pick the scale parameter β of the prior density as a diverging sequence as $n \to \infty$. Therefore, we are interested in assessing the performance of the proposed lognormal density estimator in this setting.

To study the effect of the scale parameter of the inverted gamma prior density on the Bayesian bandwidths and hence on the lognormal KDE for large samples, we generated data from a Weibull(1.5,1) with a sample size n=100 with 20% censoring and then density estimates under different β values were computed. Five different scale parameter values were chosen in increasing order of magnitude for the β parameter ($\beta = 3, 5, 7, 10, 20$) and performance of the proposed lognormal KDE is compared with the inverse Gaussian KDE with both kernels using their associated Bayesian bandwidths. Further, we also compare the lognormal KDE with Bayesian and LSCV bandwidths. As before all comparisons are assessed in terms of pointwise error ratios.

5.4.1 Comparison of the two Kernels

As shown in the following figures the lognormal KDE is clearly superior to the inverse Gaussian KDE at the origin and remains so in most part of the support.

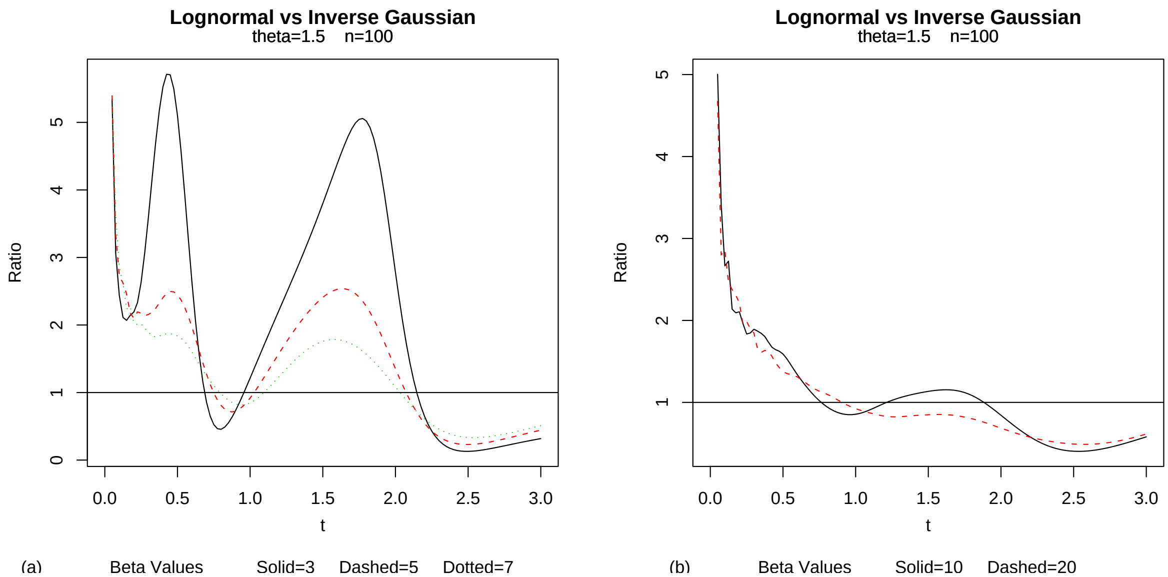

Figure 5.7: *Comparison of pointwise error ratios of lognormal and inverse Gaussian KDEs with increasing values of* β. *(a)* $\beta = 3, 5, 7$ *(b)* $\beta = 10, 20$

As β increases the pointwise errors exhibit a more stable behavior. This is in line with our argument because, in order to achieve better performance in our proposed KDE for large sample sizes, we need to make the bandwidth smaller and therefore, larger β values would naturally give better estimates and hence stabilizing pointwise MSEs throughout the support of the underlying density, as indicated in Figure 5.7(b).

5.4.2 Comparison of the two Bandwidth Selection Methods

Pointwise error ratios for the two bandwidth selection methods exhibit a similar behavior as we observed with the two kernels in section 5.4.1. The Bayesian bandwidth selection method clearly outperformed the LSCV method.

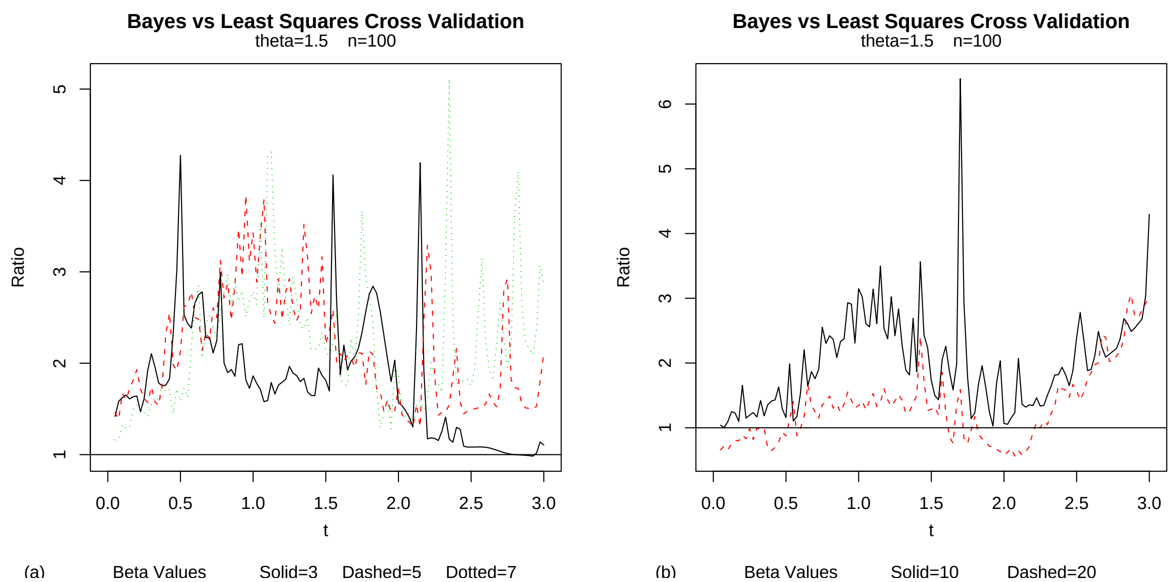

Figure 5.8: *Comparison of pointwise error ratios of the lognormal KDE using Bayesian and LSCV bandwiths with increasing values of* β. *(a)* $\beta = 3, 5, 7$ *(b)* $\beta = 10, 20$

A noticeable feature in Figure 5.8 is that when $\beta = 20$ the LSCV method seems to be yielding better pointwise estimates over some parts of the support, notably at the origin. However, as a whole, the Bayesian bandwidths resulted in smaller pointwise MSEs than the MSEs generated by the LSCV bandwidths.

5.5 Assessment of Overall Performance

In preceding sections we examined the pointwise MSEs of the proposed lognormal KDE, the inverse Gaussian KDE with both using Bayesian local bandwidths and the lognormal KDE with global cross validated bandwidt. Although the proposed lognormal KDE consistently outperformed the other two, it was difficult to conclude that the proposed KDE is uniformly superior than the other two methods. To overcome this, we looked at the MISE values of the density estimates generated by the 3 density estimators. The MISE criterion defined in (1.2) is a global measure of performance of density estimators and is a useful tool that is commonly used to compare several density estimators.

Sample	Censoring		$\theta = 0.5$			$\theta = 1$			$\theta = 1.5$	
Size	Level	LN	IG	LSCV	LN	IG	LSCV	LN	IG	LSCV
n=20	10%	0.475	0.683	0.638	0.445	0.636	0.517	0.440	0.607	0.521
	20%	0.485	0.696	0.591	0.436	0.632	0.533	0.445	0.615	0.525
	50%	0.513	0.748	0.787	0.471	0.688	0.564	0.467	0.643	0.536
n=40	10%	0.439	0.544	0.569	0.404	0.526	0.477	0.415	0.536	0.451
	20%	0.439	0.542	0.504	0.406	0.528	0.462	0.412	0.532	0.444
	50%	0.441	0.569	0.618	0.417	0.552	0.478	0.428	0.558	0.477
n=100	10%	0.420	0.471	0.479	0.377	0.459	0.428	0.399	0.495	0.415
	20%	0.416	0.471	0.481	0.382	0.464	0.436	0.398	0.495	0.416
	50%	0.404	0.469	0.573	0.382	0.469	0.430	0.402	0.500	0.416

Table 5.1: *Estimated mean integrated squared error values*

Table 5.1 shows the MISE values for the 3 density estimators considered, namely the proposed lognomal KDE, the inverse Gaussian KDE and the lognormal KDE with global cross validated bandwidth. We observe that the MISE of the lognormal KDE (LN) is always less than the other two competing estimators in all experiment settings. Further, as the sample size gets larger, the MISE have become smaller with all 3 estimators as expected. However, the level of censoring has had only a minor effect on the MISE in all 3 estimators, suggesting that all three of them are capable of utilizing censored observations effectively.

5.6 Application to Real Data

We now give an example of a density estimation problem with censored data and compare the performance of the proposed density estimator with the other 2 methods discussed earlier. In an experiment by Harwell Harwell (1995), to determine debond strength of carbon fibers, the stress at debonding for specimens were recorded after placing under a tensile load. Some specimens were broke before debonding, resulting in right censoring. Due to the complexity of the experiment, the data consisted of only 12 observations out of which 3 are censored.

The following figure shows the density estimates computed under the three methods, together with the histogram of the data.

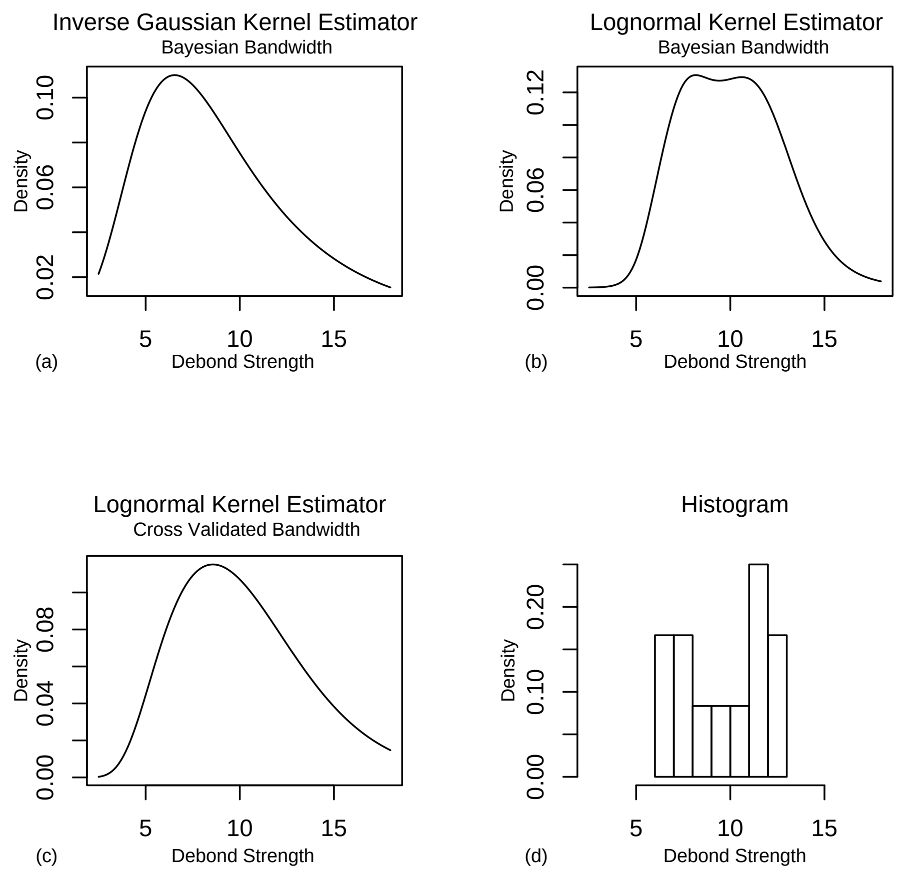

Figure 5.9: *Density estimates of the debond strength of carbon fibers using four estimation methods. (a) Inverse Gaussian kernel with Bayesian bandwidths. (b) Lognormal kernel with Bayesian bandwidths. (c) Lognormal kernel with LSCV bandwidths. (d) Histogram estimate.*

Figure 5.9 (b) shows clearly how the proposed lognormal KDE with the Bayesian bandwidths was able to capture the two apparent modes in the data which invariably gives more insight about the debond strength distribution. No such information was uncovered with the other two density estimators as shown in Figure 5.9 (a) and (c).

5.7 Conclusion and Future Work

The simulation study provided compelling evidence with regard to the potential of the lognormal KDE with the Bayesian bandwidths. In particular, it is the Bayesian local bandwidths that made the key contribution in achieving this superiority of the lognormal KDE over the other two KDEs. The performance of the lognormal KDE near the origin is undisputedly better the other two density estimators. Although pointwise MSEs seems to be high toward the tail of the support, the global performance of the proposed lognormal KDE as quantified by the MISE is consistently low in all simulation settings.

More extensive simulations with comparisons with other types of kernels, e.g. gamma, beta, reciprocal inverse Gaussian, etc. is needed to establish concrete evidence of the performance of the proposed estimator. Further, a close examination of the boundary effect on the right of the support, i.e. for densities with finite support of the form $[0, \tau], \;\; \tau < \infty$, would be a fruitful exercise as an extension of this study for the future.

Bibliography

Blum, J.R, & Susarla, V. 1980. *Maximal deviation theory of density and faliure rate function estimates based on censored data*. North-Holland, Amsterdam.

Chen, S.X. 1999. Beta kernel estimators for density functions. *Computational statistics and data analysis*, **31**, 131–145.

Chen, S.X. 2000. Probability density function estimation using gamma kernels. *Annals of the institute of statistical mathematics*, **52**, 471–480.

Ferguson, T.S., & Phadia, E.G. 1979. Bayesian nonparametric estimation based on censored data. *The annals of statistics*, **7**, 163–186.

Földes, A., & Rejtő, L. 1981. A LIL type result for the product limit estimator. *Z. wahrsheinlichkeitstheori verw. gebiete*, **56**, 75–86.

Gangopadhyay, A.K., & Cheung, K.N. 2002. Bayesian approach to the choice of smoothing parameter in kernel density estimation. *Nonparametric statistics*, **14**, 655–664.

Harwell, M. 1995. *Microbond tests for ribbon fibers*. M.S. thesis, Clemson University.

Izenman, A.J. 1991. Recent developments in nonparametric density estimation. *Journal of the american statistical association.*, **86**, 205–224.

Jones, M.C., Marron, J.S, & Sheather, S.J. 1996. A brief survey of bandwidth selection for density estimation. *Journal of the american statistical association.*, **91**, 401–407.

Kulasekera, K.B., & Padgett, W.J. 2006. Bayes bandwidth selection in kernel density estimation with censored data. *Nonparametric statistics*, **18**, 129–143.

Marron, J.S. 1988. Automatic smoothing parameter selection : A survey. *Empirical economics*, **13**, 187–208.

Marron, J.S., & Padgett, W.J. 1987. Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *The annals of statistics*, **15**, 1520–1535.

Padgett, W.J., & McNichols, D.T. 1984. Nonparametric density estimation from censored data. *Communications in statistics : Theory and methods*, **13**, 1581–1611.

Parzen, E. 1962. On estimation of a probability density function and mode. *Ann. math. stat.*, **33**, 1065–1076.

R Development Core Team. 2004. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *Ann. math. stat.*, **27**, 832–837.

Scalliet, O. 2004. Density estimation using inverse and reciprocal inverse gaussian kernels. *Nonparametric statistics*, **16**, 217–226.

Silverman, B.W. 1986. *Density estimation for statistics and data analysis.* Chapman and Hall.

Wand, M.P., & Jones, M.C. 1995. *Kernel smoothing.* Chapman and Hall.

Wegman, E.J. 1972. Nonparametric probability density estimation:i a summary of available methods. *Technometrics.*, **14**, 533–546.