WAVELET ESTIMATION OF DENSITY AND HAZARD RATE FOR RANDOMLY RIGHT CENSORED DATA

JAHIDA GULSHAN

Wavelet Estimation of Density and Hazard Rate for Randomly Right Censored Data

by

©Jahida Gulshan

A thesis submitted to the School of Graduate Studies

in partial fulfilment of the requirements for degree of Master of Science

Department of Mathematics and Statistics

Memorial University of Newfoundland

August 2005 Submitted

St. John's Newfoundland Canada

Your file Votre référence
ISBN: 978-0-494-19364-8
Our file Notre référence
ISBN: 978-0-494-19364-8

Abstract

In this study, different estimators of probability density functions and hazard rates are constructed under randomly right censored data. Nonparametric approaches are adopted under the assumption that the density and hazard rate has no specific parametric form. Some currently available methods of density and hazard rate estimation are compared to a modified approach. It is shown that wavelet estimators are competitive with the other available methods, and that no specific method can be uniquely used for all subdensities.

Acknowledgement

I would like to express my sincere gratitude to Dr. Alwell J. Oyet for his constant encouragement, valuable comments and helpful suggestions in completing this thesis. It was indeed a great pleasure to work on this problem area of wavelet applications in density and hazard estimation, which was suggested by Dr. Oyet.

I sincerely acknowledge the financial support provided by the School of Graduate Studies and Department of Mathematics and Statistics. Further, I wish to thank my professors who have shared with me their wealth of statistical knowledge at various points of my program. I also thank the Department for providing all necessary facilities and a very friendly atmosphere to complete my program.

Last but not the least, my sincere thanks to my family, friends and well-wishers whose direct or indirect encouragement and support have seen me to the completion of my program.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Wavelet means small waves that can be put together to make bigger ones, or varying ones. The objective is to use just a few basic waves, stretch them infinitely many ways, and move those in infinitely many ways to produce the wavelet system which can make an exact model of any wave. Although, wavelet theory has a long history, it has drawn much attention in the last two decades and has developed now into a methodology with applications in several disciplines including mathematics, geophysics, astronomy, signal processing, numerical analysis and statistics. Application of wavelets range from speech, music, to signal or image processing and fast algorithm in numerical analysis were developed using wavelet bases.

Recently, wavelet shrinkage curve estimation has become a well-known and mathemat-

ically sound technique for adaptively estimating functions. Software for fast wavelets smoothing is effectively implemented in many popular packages (Nason and Silverman, 1994 and Buckheit and Donoho, 1995). Most current wavelet methods often focus on density estimation or on ordinary regression (Donoho and Johnstone,1994, 1995; Donoho et.al., 1995 and Nason, 1996). In this study, wavelet estimators of probability density functions, cumulative distribution functions and hazard rates are constructed under randomly right censored data.

Censored observations are often encountered in medical follow-up, survival analysis, reliability and other studies. In such studies, interest usually focuses on estimating two functions: the underlying distribution density and the derivative of the log-survival probability known as the hazard rate. The estimation of density function and hazard rate has been studied extensively and many estimation methods are proposed including kernel and nearest neighbor smoothing method on time axis (Beran, 1981; Tanner and Wong, 1983; Dabrowska, 1987; Gray, 1992). Hazard rate estimation in the uncensored situation is discussed in Watson and Leadbatter (1964). Földes et.al. (1981), McNichols and Padgett (1985) proposed estimating the hazard from censored data using the density estimation and nonparametric apporaches. Tanner and Wong (1983) approach the problem by smoothing the empirical hazard directly. Tanner (1983) discussed the variable kernel estimator of the hazard function from

censored data. Yandell (1983), Ramlau-Hansen (1983) discuss a kernel estimator. A practical difficulty that arises with these estimators is the selection of the smoothing parameter. Penalized likelihood methods have been described by O'Sullivan (1988), Antoniadis (1989), Antoniadis, Gregoire and McKeague (1990). In all these methods, the programming to implement reasonably fast algorithm is not trivial. Kooperberg and Stone(1992) introduced an approach based on multivariate adaptive regression spline models. A major limitation of their implementation is that their method tends to be computationally intensive. Another traditional approach to density and hazard rate estimation is by orthogonal series (Kronmal and Tarter, 1968; Tanner and Wong, 1984). Wavelet smoothing methods have been applied with success in density estimation (Hall and Patil, 1996; David et.al., 1996, Patil, 1997).

Antoniadis et.al. (1999) explored the possibility of applying an ordinary nonparametric wavelet smoother to the problem of estimating the density and hazard function of right censored data. The goal is to take advantage of fast wavelet methods and software for nonparametric regression and to simplify the task of implementing software for the more complex problem of hazard smoothing. Xue (2004) constructed a random weighted statistic of a wavelet density estimator under random right censored data. The distribution of wavelet estimator is simulated by the distribution of random weighted statistic and the confidence interval of $f(x)$ is obtained by the

quantile of the distribution of random weighted statistic which is claimed to produce confidence intervals with greater coverage accuracy than those obtained by bootstrap method (Wang, 1997; Sun and Zhu, 1999).

The objective of this study is to compare different estimates of density and hazard rate available in the literature for randomly right censored data. We examined the density estimates suggested by Antoniadis et.al.(1999), Xue (2004) and compared them with that of nearest neighbor approach. To find the hazard rate we adopted the Antoniadis approach but suggested a different estimator for the CDF in the denominator of the hazard function which seem to perform better than the CDF of Antoniadis et.al.

In Section 1.1, we provide some background on wavelets. Section 1.2 describes the wavelet system construction and Section 1.3 introduces some important wavelet bases.

1.1 Some Background on Wavelets

In this section, we give a brief description on some background on wavelets. Some definitions and theories related to this study are also discussed briefly. The fundamental idea behind wavelets is to perform analysis according to scale. A wavelet system is formed by dilation and translation of two functions, $\phi(x)$, a scaling function and $\psi(x)$, a primary wavelet. The dilated and translated versions of the functions are defined

by

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k) \tag{1.1}$$

and

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \;\; j,k \in \mathbb{Z} \tag{1.2}$$

where $\mathbb{Z}$ is the set of all integers.

For a sequence of constants $\{h_r\}$ called the filter coefficients, the functions $\phi(x)$ and $\psi(x)$ are chosen to satisfy the equations

$$\phi(x) = \sum_{p\in\mathbb{Z}} h_p \phi(2x - p) \tag{1.3}$$

$$\psi(x) = \sum_{r\in\mathbb{Z}} g_r \phi(2x - r) \tag{1.4}$$

$$g_r = (-1)^r h_{-r+1} \tag{1.5}$$

and

$$\int \phi(x)dx = 1, \;\; \int \psi(x)dx = 0, \;\; \int \phi^2(x)dx = 1. \tag{1.6}$$

The condition

$$\sum_{p\in\mathbb{Z}} h_p = 2 \tag{1.7}$$

ensures the existence of a unique solution to equations (1.3) and (1.4). Orthogonality of the translates of the scaling function $\phi(x)$ is ensured by the following condition

$$\sum_{p\in\mathcal{Z}} h_p h_{p-2j} = \delta_j, j \in \mathbb{Z}. \tag{1.8}$$

In the theory of wavelets, the space of square integrable functions, $\mathcal{L}_2(\mathbb{R})$, is expressed as the limit of a sequence of close subspaces $\{V_j, j \in \mathbb{Z}\}$ where

$$..... \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset ..... \tag{1.9}$$

The nested spaces have an intersection that is trivial, that is,

$$\bigcap_j V_j = \{0\}, \tag{1.10}$$

and a union that is dense in $\mathcal{L}_2(\mathbb{R})$.

$$\overline{\bigcup_j V_j} = \mathcal{L}_2(\mathbb{R}) \tag{1.11}$$

(see Vidakovic, 1999).

Mallat(1989) introduced the notion of a multiresolution analysis, which is the fundamental concept necessary to construct and understand the wavelet paradigm. By his definition, a multiresolution analysis of $\mathcal{L}_2(\mathbb{R})$ consists of an increasing sequence of closed subspaces $V_j, j \in \mathbb{Z}$ such that

1. $\bigcap_j V_j = \{0\}$;
2. $\overline{\bigcup_j V_j} = \mathcal{L}_2(\mathbb{R})$;
3. there exists a scaling function $\phi \in V_0$ such that $\{\phi(x-k), k \in \mathbb{Z}\}$ is an orthonormal basis of V_0; that is, $V_0 = span\{\phi(x-k), k \in \mathbb{Z}\}$.

4. for all $k \in \mathbb{Z}$, $f(x) \in V_j \iff f(x-k) \in V_j$ and

5. $f(x) \in V_j \iff f(2x) \in V_{j+1}$, meaning that in passing from V_j to V_{j+1}, the resolution of the approximation is doubled.

Given any multiresolution analysis, it is possible to derive a function $\psi(x)$ such that the family $\{\psi_{j,k}(x) : j,k \in \mathbb{Z}\}$ is an orthonormal basis of $\mathcal{L}_2(\mathbb{R})$ (see Mallat, 1989).

To construct the primary wavelet, $\psi_{j,k}(x)$, we define for each $j \in \mathbb{Z}$ the difference space W_j to be the orthogonal complement of V_j such that

$$W_j \oplus V_j = V_{j+1}, \;\; W_j \perp V_j. \tag{1.12}$$

So, any function $f(x) \in V_{j+1}$ can be written as a linear combination or direct sum of functions in W_j and V_j. It can be shown that

$$V_j = V_0 \oplus W_j. \tag{1.13}$$

Iterating this infinitely many times, we find

$$\mathcal{L}_2(\mathbb{R}) = \bigcup_{j\in\mathbb{Z}}^{\infty} W_j = V_{j_0} \oplus \bigoplus_{j\geq j_0}^{\infty} W_j. \tag{1.14}$$

This implies that any $f \in \mathcal{L}_2(\mathbb{R})$ can be expressed as a series convergent in $\mathcal{L}_2(\mathbb{R})$:

$$f(x) = \sum_{k\in\mathbb{Z}} c_{j_0 k}\phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x). \tag{1.15}$$

Here c_{j_0k}, d_{jk} are coefficients and $\{\psi_{jk}\}$, $k \in \mathbb{Z}$ is a basis for W_j. The relation is called a multiresolution expansion of f. The space W_j is called resolution level of multiresolution analysis. In multiresolution analysis, there are many resolution levels which is the origin of its name.

1.2 Wavelet System Construction

The general procedure for constructing a wavelet system can be summarized in the following steps:

1. Choosing a scaling function ϕ such that $\{\phi_{0k}\}$ is an orthonormal system, and relation (1.10) is true.

2. Finding a primary wavelet function $\psi \in W_0$ such that $\{\psi_{0k}, k \in \mathbb{Z}\}$is an orthonormal basis in W_0. Then accordingly, $\{\psi_{jk}, k \in \mathbb{Z}\}$ is also an orthonormal basis in W_j.

3. Concluding that any $f \in \mathcal{L}_2(\mathbb{R})$ has the unique representation in terms of an $\mathcal{L}_2$-convergent series:

$$f(x) = \sum_{k\in\mathbb{Z}} c_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x) \tag{1.16}$$

where the wavelet coefficients are

$$c_{j_0k} = \int f(x)\phi_{j_0k}(x)dx, \text{ and } d_{jk} = \int f(x)\psi_{jk}(x)dx \tag{1.17}$$

Four constructions of the scaling function ϕ found in the literature (Strang, 1989 and Pinheiro and Vidacovic, 1997) are delineated here. The primary wavelet ψ can be computed by using (1.4) if $\phi(x)$ is known.

Construction 1. Iterate $\phi_j(x) = \sum h_k \phi_{j-1}(2x-k)$ with the box function. When $h_0 = 2$ the boxes get taller and thinner, approximating the delta function. For $h_0 = h_1 = 1$ the box is invariant: $\phi_1 = \phi_0$. For $\frac{1}{2}$, 1, $\frac{1}{2}$, the hat function appears. And $\frac{1}{8}$, $\frac{4}{8}$, $\frac{6}{8}$, $\frac{4}{8}$, $\frac{1}{8}$, yields the cubic B-spline. An example that will be important in our discussion has coefficients $\frac{1}{4}(1+\sqrt{3}), \frac{1}{4}(3+\sqrt{3}), \frac{1}{4}(3-\sqrt{3})$ and $\frac{1}{4}(1-\sqrt{3})$. This scaling function leads to orthogonal wavelets.

Construction 2. The second construction takes the Fourier transform of (1.3):

$$\begin{aligned} \hat{\phi}(\xi) &= \sum h_k \int \phi(2x-k) e^{i\xi x} dx \\ &= \frac{1}{2} \sum \left(h_k e^{ik\xi/2}\right) \int \phi(y) e^{iy\xi/2} dy \\ &= P\left(\frac{\xi}{2}\right) \hat{\phi}\left(\frac{\xi}{2}\right) \end{aligned} \tag{1.18}$$

The symbol $P(\xi) = \frac{1}{2} \sum h_k e^{ik\xi}$ is the crucial function in this theory. If $\xi = 0$ we find $P(0) = 1$. Repetition of (1.17) at $\frac{\xi}{2}, \frac{\xi}{4}, \ldots\ldots$ and noting that $\hat{\phi}(0) = \int \phi(x) dx = 1$ we

get an infinite product:

$$\hat{\phi}(\xi) = P\left(\frac{\xi}{2}\right).\hat{\phi}\left(\frac{\xi}{2}\right) \quad \hat{\phi}\left(\frac{\xi}{2}\right) = P\left(\frac{\xi}{2}\right)P\left(\frac{\xi}{4}\right) \quad \hat{\phi}\left(\frac{\xi}{4}\right) = \ldots\ldots = \prod_{j=1}^{\infty} P\left(\frac{\xi}{2^j}\right) \tag{1.19}$$

For $h_0 = 2$, $P \equiv 1$ and $\hat{\phi} \equiv 1$, the transform of the delta function. For $h_0 = h_1 = 1$, the products of the P's are geometric series:

$$P\left(\frac{\xi}{2}\right) = P\left(\frac{\xi}{4}\right) = \frac{1}{4}\left(1 + e^{i\xi/4}\right) = \frac{1 - e^{i\xi}}{4(1 - e^{i\xi/4})} \tag{1.20}$$

As $N \to \infty$ this approaches the infinite product $(1 - e^{i\xi})(-i\xi)$. This is $\int_0^1 e^{i\xi x} dx$, the transform of the box function. The hat function comes from squaring $P(\xi)$ which by (1.18) also squares $\hat{\phi}(\xi)$. The cubic B-spline comes from squaring again.

Construction 3. The construction of ϕ works directly with the recursion(1.3). Suppose ϕ is known at the integer $x = j$. The recursion (1.3) gives ϕ at the half-integers. Then it gives ϕ at the quarter integers, and ultimately at all dyadic points $x = k/2^j$. This is fast to program. The values of ϕ at the integers come from an eigenvector. With the four Daubechies coefficient $h_0 = \frac{1}{4}(1 + \sqrt{3}), h_1 = \frac{1}{4}(3 + \sqrt{3}), h_2 = \frac{1}{4}(3 - \sqrt{3}), h_3 = \frac{1}{4}(1 - \sqrt{3})$, set $x = 1$ and $x = 2$ in the dilation equation (1.3) and use the fact that $\phi = 0$ unless $0 < x < 3$, we get:

$$\phi(1) = \frac{1}{4}(3+\sqrt{3})\phi(1) + \frac{1}{4}(1+\sqrt{3})\phi(2) \tag{1.21}$$

$$\phi(2) = \frac{1}{4}(1-\sqrt{3})\phi(1) + \frac{1}{4}(3-\sqrt{3})\phi(2). \tag{1.22}$$

This is the eigenvalue problem $\phi = L\phi$, with matrix entries $L_{ij} = h_{2i-j}$. The eigenvalues are 1 and $\frac{1}{2}$, and the corresponding eigenvector for $\lambda = 1$ has components $\phi(1) = \frac{1}{2}(1+\sqrt{3}), \phi(2) = \frac{1}{2}(1-\sqrt{3})$, which are the heights on our graph of *daub2* in Figure 1.1. The other eigenvalue $\lambda = \frac{1}{2}$ means that the recursion can be differentiated:$\phi'(x) = \sum h_k 2\phi'(2x-k)$ leads similarly to $\phi'(1)$ and $\phi'(2)$. In some weak sense, $\phi = D_4$ has a dilative derivative. Here D_4 is the Daubechies wavelet with filter 8. For the hat function, the recursion matrix again has $\lambda = 1, \frac{1}{2}$. For the cubic spline the eigenvalues are $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$. When $\phi(1)$ and $\phi(2)$ is known, the dilation equation gives ϕ at half integers, such as

$$\phi\left(\frac{1}{2}\right) = \frac{1}{4}(1+\sqrt{3})\phi(1) = \frac{1}{4}(2+\sqrt{3}) \tag{1.23}$$

$$\phi\left(\frac{3}{2}\right) = \frac{1}{4}(3+\sqrt{3})\phi(2) + \frac{1}{4}(3-\sqrt{3})\phi(1) = 0. \tag{1.24}$$

Then the equation gives ϕ at quarter integers as combinations of ϕ at half integers.

Construction 4. The fourth construction is based on the Daubechies-Lagarias local pyramidal algorithm (Daubechies and Lagarias (1991, 1992)). The Daubechies-

Lagarias algorithm enables us to evaluate ϕ and ψ at a point with preassigned precision. The algorithm on wavelets from the Daubechies family will be illustrated; however, this algorithm works for all finite impulse response quadrature mirror filters.

Let ϕ be the scaling function of the Daubechies wavelet, D_N, with support $[0, 2N-1]$. Let $x \in (0,1)$ and define $dyad(x) = \{d_1, d_2, ..., d_n, ...\}$ as the set of $0-1$ digits in the dyadic representation of x. That is $x = \sum_{j=1}^{\infty} d_j 2^{-j}$. By $dyad(x,n)$, we denote the subset of the first n digits from $dyad(x)$, i.e., $dyad(x,n) = \{d_1, d_2, ..., d_n\}$. Let $\mathbf{h}= (h_0, h_1, ..., h_{2N-1})$ be the wavelet filter coefficients. Define two $(2N-1)\times(2N-1)$ matrices as:

$$T_0 = (h_{2i-j-1})_{1\leq i,j\leq 2N-1} \text{ and } T_1 = (h_{2i-j})_{1\leq i,j\leq 2N-1}. \tag{1.25}$$

Then the local pyramidal algorithm can be constructed based on Theorem 1.3.1(see Daubechies and Lagarias(1992) or Pinheiro and Vidacovic(1997)).

Theorem 1.3.1 $lim_{n\to\infty} = T_{d_1}.T_{d_2}.....T_{d_n} =$

$$\begin{bmatrix} \phi(x) & \phi(x) & \cdots & \phi(x) \\ \phi(x+1) & \phi(x+1) & \cdots & \phi(x+1) \\ \vdots & & & \\ \phi(x+2N-2) & \phi(x+2N-2) & \cdots & \phi(x+2N-2) \end{bmatrix} \tag{1.26}$$

The convergence of $\| T_{d_1}.T_{d_2}.....T_{d_n} - T_{d_1}.T_{d_2}.....T_{d_{n+m}} \|$ to zero, for fixed m, is exponential and constructive, i.e. effective decreasing bounds on the error can be established.

The following example is taken from Vidakovic (1999) to illustrate the topic:

Example 1.3.1 Consider the Daub2 scaling function(see Figure 1.1). The corresponding filter is $\mathbf{h}=\left(\frac{1+\sqrt{3}}{4}, \frac{3+\sqrt{3}}{4}, \frac{3-\sqrt{3}}{4}, \frac{1-\sqrt{3}}{4}\right)$. According to (1.24), the matrices T_0 and T_1 are given as:

$$T_0 = \begin{bmatrix} \frac{1+\sqrt{3}}{4} & 0 & 0 \\ \frac{3-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{4} & \frac{1+\sqrt{3}}{4} \\ 0 & \frac{1-\sqrt{2}}{4} & \frac{3-\sqrt{3}}{4} \end{bmatrix}$$

and

$$T_1 = \begin{bmatrix} \frac{3+\sqrt{3}}{4} & \frac{1+\sqrt{3}}{4} & 0 \\ \frac{1-\sqrt{3}}{4} & \frac{3-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{4} \\ 0 & 0 & \frac{1-\sqrt{3}}{4} \end{bmatrix}$$

Let us evaluate the scaling function at an arbitrary point, for instance , $x = 0.45$. Twenty decimals in the dyadic representation of 0.45 obtained through an s-plus code are dyad $(0.45, 20) = \{0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1\}$. In addition to the value at 0.45, we get the values at 1.45 and 2.45. the values $\phi(0.45), \phi(1.45)$

and $\phi(2.45)$ may be approximated as averages of the first, second and third row respectively in the following matrix:

$$\prod_{i \in dyad(0.45,20)} T_i = \begin{bmatrix} 0.86480582 & 0.86480459 & 0.86480336 \\ 0.08641418 & 0.08641568 & 0.08641719 \\ 0.04878000 & 0.04877973 & 0.04877945 \end{bmatrix}$$

The Daubechies-Lagarias algorithm gives only the values of the scaling function. The following theorem is useful in obtaining the values of the wavelet function.

Theorem 1.3.2 Let x be an arbitrary real number. And let the wavelet be given by its filter coefficients $\{h_0, h_1, ..., h_{2N-1}\}$. Define vector $\mathbf{u}$ with $2N-1$ components as

$$u(x) = \{(-1)^{1-[2x]} h_{i+1-[2x]}, i = 0, ..., 2N-2\} \tag{1.27}$$

If for some i, the index $i + 1 - [2x]$ is negative or larger than $2N - 1$, then the corresponding components of $\mathbf{u}$ is equal to 0. Here $[2x]$ represents the integer part of $2x$. Let the vector $\mathbf{v}$ be defined as

$$\mathbf{v}(x, n) = \frac{1}{2N-1} \mathbf{1}' \prod_{i \in dyad(\{2x\}, n)} T_i, \tag{1.28}$$

where $\mathbf{1}' = (1, 1, ..., 1)$ is the row-vector of ones. Then

$$\psi(x) = lim_{n \to \infty} \mathbf{u}(x)' \mathbf{v}(x, n), \tag{1.29}$$

and the limit is constructive.

Construction 4 is the easiest to implement considered in the computational context. Hence this construction has been used in this thesis to construct the Daubechies wavelet systems.

1.3 Some Important Wavelet Bases

In this section, some commonly used families of wavelets are described, namely Haar wavelet, multiwavelet and the Daubechies wavelet system.

Haar system

The Haar wavelet basis is the simplest example of a wavelet system on $\mathcal{L}^2(S)$. The scaling function is:

$$\phi(x) = I_{[0,1]}(x) = \begin{cases} 1, & if \ \ 0 \leq x < 1, \\ 0, & otherwise \end{cases} \tag{1.30}$$

The refining relations for the Haar wavelet basis are

$$\phi(x) = \phi(2x-1) + \phi(2x) \tag{1.31}$$

and

$$\psi(x) = \phi(2x) - \phi(2x-1) \tag{1.32}$$

Multiwavelet System

The multiwavelet system was constructed by Alpert (1992). In multiwavelet basis, instead of a single scaling function $\phi(x)$, there are several scaling functions $\phi_0, \phi_1, ..., \phi_{N-1}$ whose translates span the space V_0. Each scaling funciton is a dilated, translated and normalized Legendre polynomial in the interval $[0, 1)$:

$$\phi_i(x) = \begin{cases} \sqrt{2i+1}P_i(2x-1), & x \in [0,1) \\ 0, & otherwise. \end{cases} \tag{1.33}$$

where $P_i(i = 0, 1, ..., N-1)$ are the Legendre polynomials. The space $V_n, n \in \mathbb{Z}$ are dilates of V_0 and the difference spaces W_n are as defined previously. The primary wavelets denoted by ${}_N\omega_0, ..., {}_N\omega_{N-1}$ vanish outside the interval [0, 1) and are orthogonal to polynomials of maximum degree,

$$\int_S {}_N\omega_j(x)x^i dx = 0, \quad i = 0, 1, ..., N-1+j \tag{1.34}$$

When $N = 1$, the multiwavelet basis coincides with the Haar wavelet basis. For N=2 the scaling functions and primary wavelets are

$$\phi_0(x) = \begin{cases} 1, & if \ \ 0 \leq x < 1, \\ 0, & otherwise \end{cases} \tag{1.35}$$

$$\phi_1(x) = \begin{cases} \sqrt{3}(2x-1), & if \ \ 0 \le x < 1, \\ 0, & otherwise. \end{cases} \tag{1.36}$$

$$2\omega_0(x) = \begin{cases} \sqrt{3}(1-4x), & if \ \ 0 \le x < \frac{1}{2}, \\ \sqrt{3}(4x-3), & if \ \ \frac{1}{2} \le x < 1, \\ 0, & otherwise. \end{cases} \tag{1.37}$$

$$2\omega_1(x) = \begin{cases} 6x-1, & if \ \ 0 \le x < \frac{1}{2}, \\ 6x-5, & if \ \ \frac{1}{2} \le x < 1, \\ 0, & otherwise. \end{cases} \tag{1.38}$$

The refining relation for these multiwavelets ($N = 2$) are:

$$\phi_0(x) = \phi_0(2x) + \phi_0(2x-1) \tag{1.39}$$

$$\phi_1(x) = \frac{\sqrt{3}}{2}(\phi_0(2x-1) - \phi_0(2x)) + \frac{1}{2}(\phi_1(2x-1) + \phi_1(2x) \tag{1.40}$$

$$2\omega_0(x) = \phi_1(2x-1) - \phi_1(2x) \tag{1.41}$$

$$2\omega_1(x) = \frac{1}{2}(\phi_0(2x) - \phi_0(2x-1)) + \frac{\sqrt{3}}{2}(\phi_1(2x-1) + \phi_1(2x)) \tag{1.42}$$

Daubechies System

Daubechies(1992) was the first to construct compactly supported orthogonal wavelets

with a preassigned degree of smoothness. The scaling functions and primary wavelets of the Daubechies (1992) wavelet systems, commonly represented as ${}_N\phi(x)$ and ${}_N\psi(x)$ respectively, have no closed forms. They are constructed numerically for different values of the wavelet number N. Table 1.1 lists the filter coefficients ${}_nh_n$ for $N = 2$ through 10. Both ${}_N\phi$ and ${}_N\psi$ have support width $2N - 1$.

An important feature of wavelets is that the estimated functions inherit the smoothness properties of the wavelets employed in the estimation procedure. A Haar wavelet follow the general pattern of the function but show up as a step function. Multiwavelets show the cusps and jumps, at the points where the function changes its direction. Important feature of the Daubechies wavelets is their smoothness. Therefore the choice of an appropriate wavelet system is important depending on whether the experimenter expects the response to be a smooth function, contain discontinuities, or be a step function.

Table 1.1: The Filter Coefficients

	n	Nh_n
$N = 2$	0	0.4829629131445341
	1	0.8365163037378077
	2	0.2241438680420134
	3	-.1294095225512603
$N = 3$	0	0.3326705529500825
	1	0.8068915093110924
	2	0.4598775021184914
	3	-.1350110200102546
	4	-.0854412738820267
	5	0.0352262918857095
$N = 4$	0	0.2303778133088964
	1	0.8068915093110924
	2	0.6308807479398587
	3	-.0279837694168599
	4	-.1870348118190931
	5	0.0308613818355607
	6	0.0328830116668852
	7	-.0105974017850690
$N = 5$	0	0.1601023979741929
	1	0.6038292697971895
	2	0.7243085284377726
	3	0.1384281459013203
	4	-.2422948870663823
	5	0.0322448695846381
	6	0.0775714938400459
	7	-.0062414902127983
	8	-.0125807519990820
	9	0.0033357252854738
$N = 6$	0	0.1115407433501095
	1	0.4946238903984533
	2	0.7511339080210959
	3	0.3152503517091982
	4	-.2262646939654400
	5	0.1297668685672625
	6	0.0975016055873225
	7	0.0275228655303053
	8	-.0315820393174862
	9	0.0055538422011614
	10	0.0047772575109455
	11	-.0010773010853085
$N = 7$	0	0.0778520540850037
	1	0.3965393194818912
	2	0.7291320908461957
	3	0.4697822874051889
	4	-.1439060039285212
	5	-.2240361849938412
	6	0.0713092192668272
	7	0.0806126091510774
	8	-.0380299369350104
	9	-.0168745416306655
	10	0.0125509985560986
	11	0.0004295779729214
	12	-.0018016407040473
	13	0.0003537137999745

	n	Nh_n
$N = 8$	0	0.0544158422441072
	1	0.3128715909143166
	2	0.6756307362973195
	3	0.5853546836542159
	4	-.0158291052563823
	5	-.2840155429615824
	6	0.0004724845739124
	7	0.1287474266204893
	8	-.0173693010018090
	9	-.0440882539307871
	10	0.0139810279174001
	11	0.0087460940474065
	12	-.0048703529934520
	13	-.0003917403733770
	14	0.0006754494064506
	15	-.0001174767841248
$N = 9$	0	0.0380779473638778
	1	0.2438346746125858
	2	0.6048231236900955
	3	0.6572880780512736
	4	0.1331983858249883
	5	-.2932737832791663
	6	-.0968407832229492
	7	0.1485407493381256
	8	0.0307256814793385
	9	-.0676328290613279
	10	0.0002509471148340
	11	0.0223616621236798
	12	-.0047232047577518
	13	-.0042815036824635
	14	0.001847648830563
	15	0.0002303857635232
	16	-.0002519631889427
	17	0.0000393173203163
$N = 10$	0	0.0266700579005473
	1	0.1881768000776347
	2	0.5272011889315757
	3	0.6884590394534363
	4	0.2811723436605715
	5	-.2498464243271598
	6	-.1959462743772862
	7	0.1273693403357541
	8	0.0930573646035547
	9	-.0713941471663501
	10	-.0294575368218399
	11	0.0332126740593612
	12	0.0036065535669870
	13	-.0107331754833007
	14	0.0013953517470688
	15	0.0019924052951925
	16	-.0006858566979564
	17	-.0001164668551285
	18	0.0000935886703202
	19	-.0000132642028945

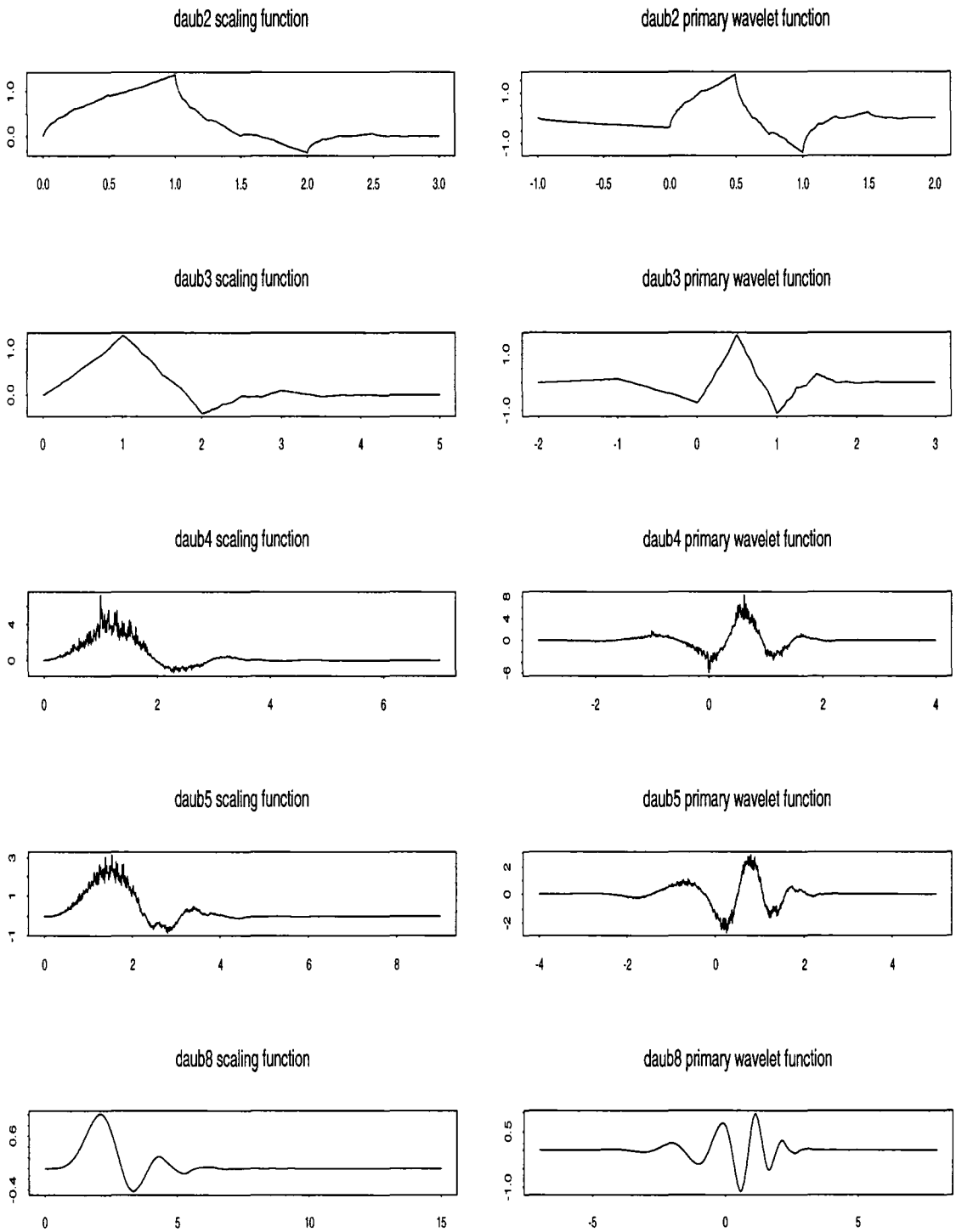

Figure 1.1: Scaling function and Primary wavelets of Daubechies wavelet for N = 2, 3, 4, 5 and 8

Chapter 2

Density Estimation

2.1 Notations and Model Setup

It is not always possible to follow every subject in an experiment in which subjects are followed over time until an event of interest, for example, death or other type of failure, occurs. So some lifetimes are known exactly and the remainder lifetimes are known to have occurred only within certain intervals which results in censored data. Censoring may occur as subjects may drop out of the study and be lost to follow-up, or be deliberately withdrawn, or the end of the data collection period may arrive before the event is observed to happen.

Let X_i, $i = 1, 2, ..., n$, be the lifetimes of n independent, identical units. We assume

that $X_1, X_2, ..., X_n$ are non-negative and independent and identically distributed, iid, with common continuous cumulative distribution function (CDF) F and continuous density f.

Also, associated with each X_i, let there be a random variable C_i, known as its censoring variable. It is common to assume that $C_1, C_2, ..., C_n$ are non-negative and iid with common continuous CDF G and continuous density g. The observed random variables are then $Z_i = min(X_i, C_i)$ and $\delta_i = I_{[X_i \leq C_i]}$. Here I_A denotes the indicator function of event A. So $\delta_i = 1$ indicates that the i-th subject's observed time is not censored.

However, the marginal distribution of X and C are not identifiable from observations (Z, δ) alone, unless specific assumptions are made on the dependence between X_i and C_i. The most used assumption of this kind is to let life times X_i and censoring times C_i be independent.

Let the density of those observations that are still to fail be $f^*(t)$, where,

$$f^*(t) = f(t)\{1 - G(t)\} \tag{2.1}$$

This is called the subdensity function and in this chapter, we describe the estimation

procedure of the subdensity function by the local histogram approach, the nearest neighbor approach and estimation of the density function by wavelet kernel approach. As we are interested in estimating the hazard rates as well, for censored data, subdensity estimates are on main focus so that we can use these estimates in Chapter 4.

2.2 Estimation Procedure

2.2.1 Local Histogram Approach

Let X be a discrete random variable with probability mass function $f(x)$. Consider the data $X_1, X_2, ..., X_n$. Then

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(x_i=x)}. \tag{2.2}$$

If X is a continuous random variable with probability density function $f(x)$, then

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} I_A(x_i), \quad A = \left[x - \frac{h}{2}, x + \frac{h}{2}\right], \tag{2.3}$$

where h is the width of the interval A, can be used as a rough estimate of $f(x)$.

In the censored case, for the paired observations $\{Z_i, \delta_i\}$, the histogram estimate of the subdensity $f^*(t)$ can be expressed as

$$\hat{f}^*(t) = \frac{1}{nh}\sum_{i=1}^{n} I_A(z_i)\delta_i, \quad A = \left[t - \frac{h}{2}, t + \frac{h}{2}\right]. \tag{2.4}$$

Antoniadis et.al.(1999) adopted this local histogram approach to obtain a crude estimate of the subdensity of the observed failures $f^*(t)$ by choosing a $\Delta > 0$ and binning the observed failures into K+1 bins of length Δ. By this method, estimates would only be computed over a finite interval $[0, \tau]$ where in practice $\tau = Z_{(n)}$. Let N be an integer that may depend on the sample size n and define a dyadic grid or evaluation points

$$t_k = \frac{k\tau}{2^N}, \quad k = 0, 1, 2, ..., \quad K = 2^N - 1$$

with the inter point distance on the grid $\Delta = 2^{-N}\tau$. Now, divide the time interval $[0, \tau]$ into $K + 1$ intervals of length Δ, centered on t_k with end points

$$\tau_0 = -\frac{\Delta}{2}; \quad \tau_k = t_k - \frac{\Delta}{2}, \quad k = 1, 2, ..., K; \quad \tau_{K+1} = \tau,$$

and denote the k-th interval by $J_k = [\tau_k, \tau_{k+1}], k = 0, 1, ..., K - 1$, and $J_K = [\tau_K, \tau]$. Using the observations, a new data set of $(K + 1)n$ records is created consisting of (Y_{ik}, t_k) where $Y_{ik} = I_{j_k}(Z_i)\delta_i, i = 1, 2, ..., n, k = 0, 1, ..., K$ is the indicator that an uncensored event for subject i falls within the time interval J_k. Finally, let U_k denote the proportion of failures observed in the interval J_k. i.e.

$$U_k = \frac{1}{n}\sum_{i=1}^{n} Y_{ik}, k = 0, 1, ..., K. \tag{2.5}$$

Then $\frac{U_k}{\Delta}$ are crude estimators of the subdensity values $f^*(t_k)$, defining a histogram type estimator at $K+1$, a power of 2, dyadic points. The binned data $\frac{U_k}{\Delta}$ is then smoothed by a discrete fast wavelet method via an appropriate linear wavelet smoother. One technique which we have applied in this thesis is called discrete wavelet transform. See Section 1.1 for a description of the transform. The idea underlying this approach is the fact that we can express any square integrable function on $[0,\tau]$ in the form

$$f(t) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0 k}\phi_{j_0 k}(t) + \sum_{j\geq j_0}^{\infty}\sum_{k=0}^{2^{j_0}-1} \beta_{jk}\psi_{jk}(t) \tag{2.6}$$

for collection of functions $\phi_{j_0,k}(t) = 2^{j_0/2}\phi(2^{j_0}t-k), j_0, k \in \mathbb{Z}$ and $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t-k), j, k \in \mathbb{Z}$ which form an orthogonal basis for $L^2([0,\tau])$.

Here, $\phi_{j_0,k}$ and $\psi_{j,k}$ are translated and dilated versions of a scaling function $\phi(x)$ and a primary wavelet $\psi(x)$ respectively. The $\phi_{j_0,k}$ allows an approximation of f at resolution j_0 whereas $\psi_{j,k}$'s represent the detail in f at resolutions finer than j_0.

Antoniadis et.al. (1999) chose the scaling function ϕ as a coiflet (Daubechies, 1992) of order L, with $L > m+1$ where m is the assumed order of differentiability of f.

The function f^* admits the following generalized Fourier expansion in L^2:

$$f^*(t) = \sum_{k=0}^{2^{j_0}-1} < f^*, \phi_{j_0 k} > \phi_{j_0 k}(t) + \sum_{j \geq j_0}^{\infty} \sum_{l=0}^{2^{j_0}-1} < f^*, \psi_{jl} > \psi_{jl}(t) \qquad (2.7)$$

with $< f, g >$ defined by $\int_0^\tau f(t)g(t)dt$. In application it is widely assumed that

$$< f^*, \phi_{N,k} > \approx 2^{-N/2} f^*(k/2^N),$$

but such an approximation is rarely justified. Antoniadis et.al. (1999) have shown that

$$< f^*, \phi_{N,k} > \approx 2^{-N/2} f^*(t_k), 0 \leq k \leq 2^N - 1,$$

with error

$$O(2^{-n/2} \times 2^{-Nm}).$$

Therefore a reasonable estimate of the projection, $\prod_N f^*$ of f^* onto the finest available scale N is

$$\tilde{f}_N^*(t) = 2^{-N/2} \sum_{k=0}^{K} \frac{U_k}{\Delta} \phi_{N,k}(t) \qquad (2.8)$$

To smooth the data with a better rate, a resolution $j(n) < N$ is chosen by using folded cross validation (See Nason, 1996). Then the wavelet coefficients of the binned data are computed at scale $j(n)$ and the resultant wavelet transform is taken as a smooth estimate of f^*. This method is sufficiently accurate and flexible to handle

peaks in the middle of the data. However, it does not work very well far out in the tails.

2.2.2 Nearest Neighborhood Approach

The estimator shown in equation 2.4 is the proportion of $z_1, ..., z_n$ in the interval $\left(x - \frac{h}{2}, x + \frac{h}{2}\right)$ divided by a fixed window width, h, which is the smoothing parameter. However, one would expect that the window width should be larger when trying to estimate the tails of a density than in its center, when fewer observations are expected to be available in the former situation. Moreover, if $f(x)$ is small and flat in the tails, it will not matter much that observations distant from x are employed. In contrast, when it is varying rapidly, as in the central part of the density, incorrect estimates are likely, unless observations close to x are used.

Attempts to estimate with a fixed window width are likely to lead to under-smoothing in some part of the range and over-smoothing in another. A procedure that responds to these problems is the nearest neighborhood estimator, first suggested by Fix and Hodges (1951), which is another naive estimator, but one where h is defined in terms of distances of the data points from x. Let $d_k(x)$ be the distance of x from its k-th nearest neighbor (k-NN) among $x_1, ..., x_n$. Then taking $h = 2d_k(x)$, we have,

$$\hat{f}(x) = \frac{1}{2nd_k(x)} \sum_{i=1}^{n} I_J(xi), \qquad J = [x - d_k(x), x + d_k(x)]\,. \tag{2.9}$$

To illustrate how $d_k(x)$ is computed, let $x_1 = 5, x_2 = 3, x_3 = 7$ and $x_4 = 13$. Also let $x = 2$. Then $|x_i - x| = |x_i - 2| = 3, 1, 5, 11$. After sorting the absolute differences, we get $1, 3, 5$ and 11. Then for $k = 2$, $d_k(x) = d_2(x) = 3$.

In general, we can write, for an appropriate kernel function, $K(.)$,

$$\hat{f}(x) = \frac{1}{2nd_k(x)} \sum_{i=1}^{n} K\left(\frac{x_i - x}{2d_k(x)}\right). \tag{2.10}$$

For censored data, this estimate can be expressed in the form

$$\hat{f}(x) = \frac{1}{2nd_k(x)} \sum_{i=1}^{n} I_J(x_i).\delta_i, \qquad J = [x - d_k(x), x + d_k(x)]\,, \tag{2.11}$$

or in general

$$\hat{f}(x) = \frac{1}{2nd_k(x)} \sum_{i=1}^{n} K\left(\frac{x_i - x}{2d_k(x)}\right).\delta_i \tag{2.12}$$

The degree of smoothing (h) is controlled by an integer, k; typically $k \equiv n^{\frac{1}{2}}$. In the

tails of the distribution, the distance $d_k(x)$, and hence h, will be larger than in the middle part of the distribution, if there are fewer observations in any given sample from the tail. This corrects a potential problem with the kernel estimator arising in the tails of the density, where, few observations will be encountered in the range, $x \pm \frac{h}{2}$, and therefore the estimate will tend to be undersmoothed. By effectively increasing h in the tails a smoother estimate is likely. As $n \to \infty$, and $k \to \infty$, $d_k(x)$ will tend to zero as more and more observations will be encountered that are close to x.

2.2.3 Wavelet Kernel Approach

For estimation of the density function $f(x)$, based on the censored data $\{Z_i, \delta_i\}$, $i = 1, 2, ..., n$, Xue (2004) also used the wavelet smoothing method.

Let ϕ be the scaling function of multiresolution analysis $(V_m)_{(m\in\mathbb{Z})}$ where $\mathbb{Z}$ is the set of all integers. We make the following assumptions:

- ϕ is bounded function with compact support and unit integral. i.e. there exist constants C and L such that

$$\sup_x \phi(x) \leq C,$$

with support $\phi(x) \subset [-L, L]$,

$$\int_{-\infty}^{\infty} \phi(x)dx = 1$$

- ϕ is of class $C^r(\mathbb{R})$ (Hölder space) and every derivative up to order r is rapidly decreasing

- The sequence $\{\phi(x-k), k \in \mathbb{Z}\}$ is an orthonormal family of $L^2(\mathbb{R})$ and V_0 be the subspace spanned.

- If we define, $\phi_{mk}(x) = 2^{m/2}\phi(2^m x - k), k \in \mathbb{Z}$, then $\{\phi_{0,k}, k \in \mathbb{Z}\}$ is an orthogonal basis of V_m (Watler, 1994).

For the scaling function ϕ, the Meyer wavelet kernel is defined as :

$K_m(x,u) = 2^m K(2^m x, 2^m u)$, $K(x,u) = \sum_{k\in\mathbb{Z}} \phi(x-k)\phi(u-k)$.

However, the kernel we used is

$K_m(x,u) = n q_m^T(x) q_m(u)$, where $q_m^T(x) = (\phi(x), \psi_{0,0}(x), \psi_{1,0}(x), ..., \psi_{m,2^m-1}(x))$.

This kernel can be obtained as follows. Let $f(x)$ be a square integrable mean response function and let $y = f(x) + \varepsilon$. Then, since $\{\phi(x), \psi_{j,k}(x)\}$ is a basis for the class of square integrable functions, we can write,

$$\begin{aligned} f(x) &= c\phi(x) + \sum_{j=0}^{m} \sum_{k=0}^{2^m-1} d_{jk}\psi_{jk}(x) + remainder \\ &= \sum_{i=1}^{2^{m+1}} \beta_i q_i(x) + remainder \\ &= q^T(x)\beta + remainder \end{aligned}$$

where

$$c = \int_S f(x)\phi(x)dx \tag{2.13}$$

$$\text{and} \quad d_{jk} = \int_S f(x)\psi_{jk}(x)dx. \tag{2.14}$$

Now, we let $S = \bigcup_{i=1}^{n} A_i$ where A_i's are disjoint. Here, A_i refers to the partitioning of the $[0, 1]$ intervals. Then,

$$c = \int_{\bigcup_n A_i} f(u)\phi(u)du = \sum_{i=1}^{n} \int_{A_i} f(u)\phi(u)du \tag{2.15}$$

Since $f(u)$ is unknown, we use the observed (or generated) data, Y_i to obtain an estimate of c as follows:

$$\hat{c} = \sum_{i=1}^{n} Y_i \int \phi(u)du \tag{2.16}$$

Similarly,

$$\hat{d}_{jk} = \sum_{i=1}^{n} Y_i \int \psi_{jk}(u)du \tag{2.17}$$

This results in

$$
\begin{aligned}
\hat{f}(x) &= \hat{c}\phi(x) + \sum_{j=0}^{m}\sum_{k=0}^{2^m-1} \hat{d}_{jk}\psi_{jk}(x) \\
&= \left(\sum_{i=1}^{n} Y_i \int_{A_i} \phi(u)du\right)\phi(x) + \sum_{j=0}^{m}\sum_{k=0}^{2^m-1}\left(\sum_{i=1}^{n} Y_i \int_{A_i} \psi_{jk}(u)du\right)\psi_{jk}(x) \\
&= \sum_{i=1}^{n} Y_i \int_{A_i} \phi(u)\phi(x)du + \sum_{i=1}^{n} Y_i \int_{A_i}\left(\sum_{j=0}^{m}\sum_{k=0}^{2^m-1}\psi_{jk}(u)\psi_{jk}(x)\right)du \\
&= \sum_{i=1}^{n} Y_i \left[\int_{A_i}\left(\phi(u)\phi(x) + \sum_{j=0}^{m}\sum_{k=0}^{2^m-1}\psi_{jk}(u)\psi_{jk}(x)\right)du\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i \int_{A_{i-1}}^{A_i} K_m(x,u)du
\end{aligned}
$$

which is the wavelet version of the Gasser Müller estimator, where

$$
K_m(x,u) = n\phi(u)\phi(x) + n\sum_{j=0}^{m}\sum_{k=0}^{2^m-1}\psi_{jk}(u)\psi_{jk}(x). \tag{2.18}
$$

The wavelet estimator of $f(x)$ is then defined as:

$$
f_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_m(x, Z_i)\frac{\delta_i}{1 - G(Z_i)} \tag{2.19}
$$

where $m = m_n$ is a positive integer dependent on n. When the CDF of censoring time G is unknown, the Kaplan-Meier estimator $G_n(x)$ of G can be used in equation (2.19). This estimator $G_n(x)$ is defined as:

$$G_n(x) = \begin{cases} 1 - \prod_{Z(j)\leq x} \left(\frac{n-j}{n-j+1}\right)^{I(\delta_{(j)}=0)}, & \text{if } x < Z_{(n)} \\ 1, & otherwise \end{cases} \tag{2.20}$$

Let $S_n(x)$ denote the Kaplan-Meier estimator of survival function $S(x) = 1 - F(x)$, where

$$S_n(x) = \begin{cases} \prod_{Z(j)\leq x} \left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}}, & \text{if } x < Z_{(n)} \\ 0, & otherwise \end{cases} \tag{2.21}$$

If we denote, $s_i = S_n(Z_{(i-1)}) - S_n(Z_{(i)})$ Then

$$\begin{aligned} s_i &= S_n(Z_{(i-1)}) - S_n(Z_{(i)}) \\ &= \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}} - \prod_{j=1}^{i} \left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}} \\ &= \left[\prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}}\right] \left[1 - \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}\right] \\ &= \left[\prod_{j=1}^{i-1} \left(\frac{n-j+1}{n-j}\right)^{-\delta_{(j)}}\right] \left[1 - \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}\right] \\ &= \left[\prod_{j=1}^{i-1} \left(\frac{n-j+1}{n-j}\right)^{-\delta_{(j)}}\right] \left[1 - \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}\right] \end{aligned} \tag{2.22}$$

Now,

$$\left[1-\left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}\right] = \begin{cases} \left[1-\left(\frac{n-i}{n-i+1}\right)\right], & \text{if} \quad \delta_i = 1 \\ 0 & \text{if} \quad \delta_i = 0 \end{cases}$$

$$= \delta_{(i)} \cdot \left[1-\left(\frac{n-i}{n-i+1}\right)\right]$$

$$= \delta_{(i)} \cdot \left(\frac{1}{n-i+1}\right)$$

$$= \delta_{(i)} \cdot \left(\frac{1}{n} \cdot \frac{n}{n-1} \cdot \cdots \cdot \frac{n-i+2}{n-i+1}\right)$$

$$= \frac{\delta_{(i)}}{n} \cdot \prod_{j=1}^{i-1}\left(\frac{n-j+1}{n-j}\right)$$

Substituting this value in equation (2.22), we get

$$s_i = \left[\prod_{j=1}^{i-1}\left(\frac{n-j+1}{n-j}\right)^{-\delta_{(j)}}\right]\left[\frac{\delta_{(i)}}{n} \cdot \prod_{j=1}^{i-1}\left(\frac{n-j+1}{n-j}\right)\right]$$

$$= \frac{\delta_{(i)}}{n} \cdot \prod_{j=1}^{i-1}\left(\frac{n-j+1}{n-j}\right)^{1-\delta_{(j)}}$$

$$= \frac{\delta_{(i)}}{n[1-G_n(Z_{(i)}-)]}$$

Hence, we have

$$\hat{f}_n(x) = \sum_{i=1}^{n} s_i K_m(x, Z_{(i)}) \tag{2.23}$$

The optimum value of m is usually determined through a simulation study.

Following the local histogram approach or nearest neighborhood approach, we may define the subdensity estimator as

$$\hat{f}_n^*(x) = \sum_{i=1}^{n} s_i K_m(x, Z_{(i)}).\delta_i. \tag{2.24}$$

A limitation of this Wavelet Kernel method, used by Xue (2004) is that the estimated $f(x)$ falls to zero very quickly as x goes beyond 1 even for a simple model. Another major limitation of this estimator is that it does not provide good estimate for densities with peaks. We applied this method to estimate gamma densities and found that this method is not very efficient in estimating such densities.

2.3 Average Mean Squared Error

The average mean squared errors, AMSE's, are calculated by averaging the mean squared errors

$$MSE(f^*(t)) = \frac{1}{K} \sum_{k} [\hat{f}_n^*(t_k) - f^*(t_k)]^2.$$

AMSE is used to compare the estimates obtained by different approaches and the results are shown in the next section.

2.4 Simulation Studies

To illustrate the methods, we conducted two simulation studies. First, we generated lifetimes X_i from a gamma distribution with parameter 5 and censoring times C_i from an exponential distribution with parameter 1/6. The choice of these parameters is to ensure that there is about 50 percent censoring. S-plus codes are used to perform the simulation study where the number of simulations was 200 with a sample of size 200.

We applied all the methods discussed in the previous section to estimate the subdensities. For the local histogram approach, we used the Daubechies wavelets with filter 16, for smoothing the crude estimates and $j(n)$ was chosen to be 1. For Xue's Wavelet Kernel approach, we used $m = 1$. Finally we constructed Figure 2.1 based on the estimates obtained from different methods.

From Figure 2.1, we observe that the subdensity by the local histogram approach (subdensity LH) works fairly well to estimate the true subdensity but it overestimates the true curve. The nearest neighbor approach (subdensity NN) results in some underestimation of the true curve. The wavelet kernel approach (subdensity WK) does not appear to be a good approximation for this subdensity.

We also compared the estimates based on their average mean squared errors, AMSE's.

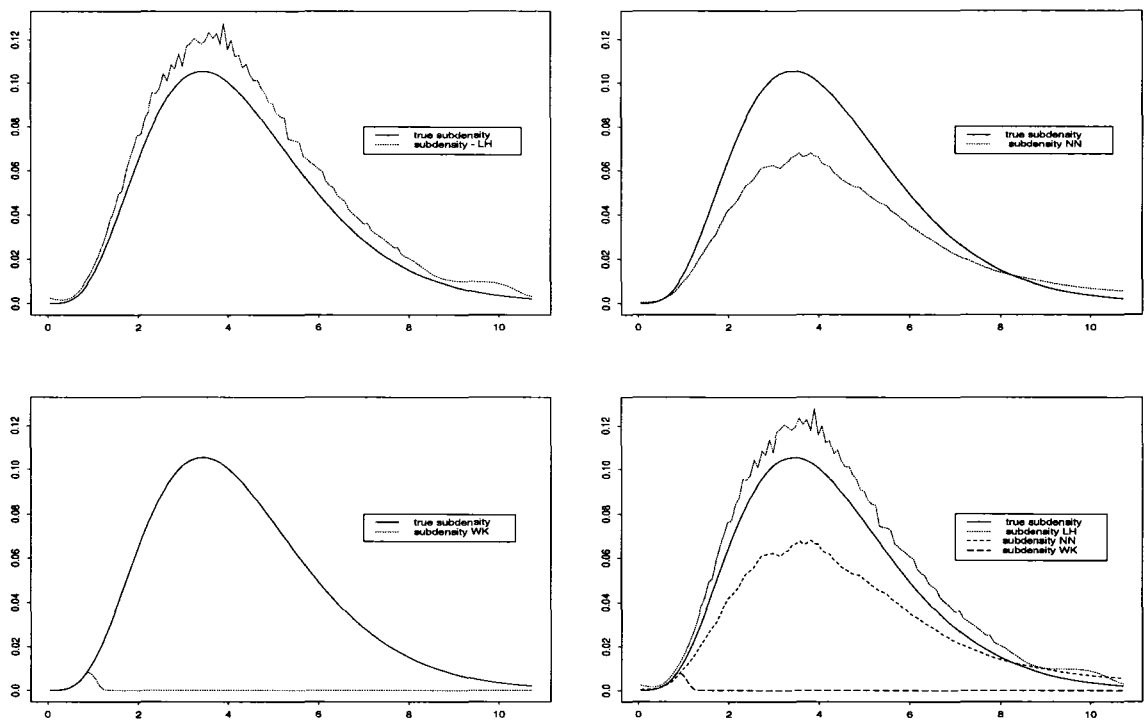

Figure 2.1: Subdensity estimates by different methods, $X_i \sim gamma(5,1)$, $C_i \sim exp(1/6)$.

The AMSE's are shown in Table 2.1. For the first example, we found the average mean squared error (AMSE) is the least for the estimates obtained by nearest neighbor approach which is a little smaller than the AMSE of the estimates by local histogram approach by Antoniadis et. al. (1999) The AMSE for Wavelet Kernel approach used by Xue (2004) is higher than that of Antoniadis et.al's estimates and nearest neighbor estimates.

However, Antoniadis et.al.(1999), in their paper reported the AMSE for the same situation to be .00025 which is a little lower than the one we reported for the same

Table 2.1: Table of AMSE of the subdensity estimates, $X_i \sim gamma(5,1)$ and $C_i \sim exp(1/6)$

Estimates	AMSE
$f_n^*(t)_{LH}$	0.0009
$f_n^*(t)_{NN}$	0.0006
$f_n^*(t)_{WK}$	0.0032

approach. And it should also be mentioned that the AMSE calculated in their paper was calculated restricting the sum over time to points with $t_k \leq 6$.

Secondly, we generated lifetimes X_i from an exponential distribution with parameter 1 and censoring times C_i from another exponential distribution with parameter 3/4. Again the parameters are chosen to ensure that there is about 40 percent censoring in the data. Sample size and the number of simulations were 200. All the methods described in Section 2.2 are again applied to estimate the subdensities. And we constructed Figure 2.2 based on the estimates obtained from different approaches.

From Figure 2.2 we observe that both smoothed local histogram approach and nearest neighbor approach underestimates the true curve to some extent and at the right tail, they are very close to the actual curve. Wavelet Kernel approach overestimates

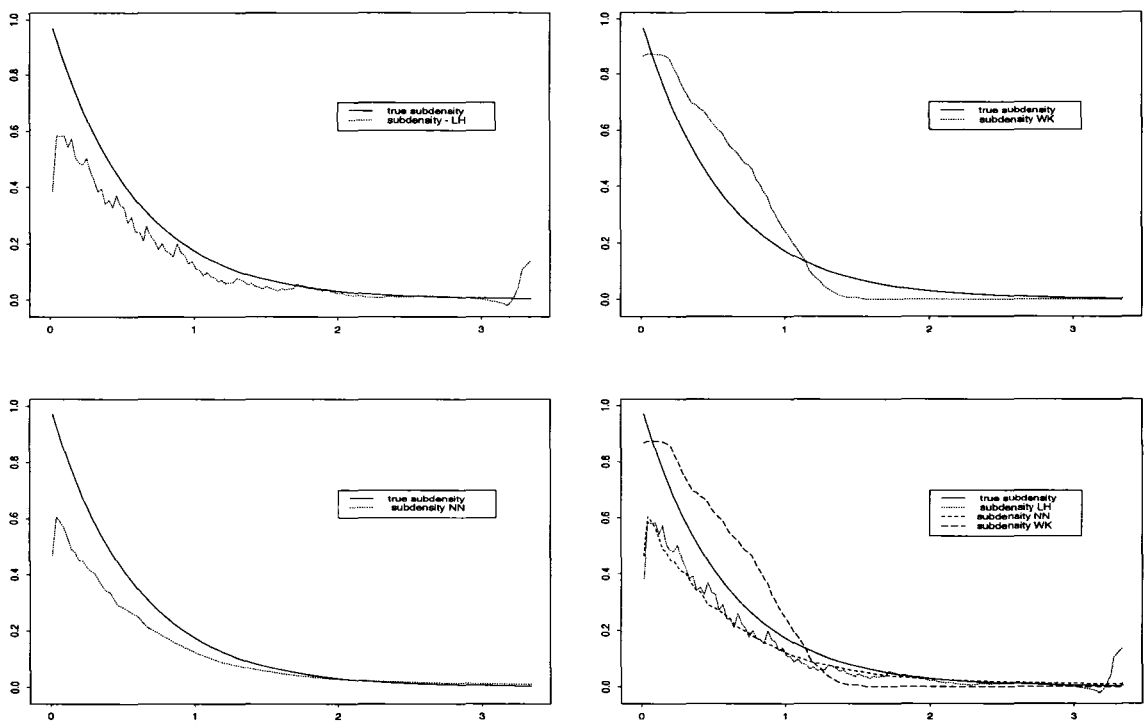

Figure 2.2: Subdensity Estimates by different methods, $X_i \sim exp(1)$, $C_i \sim exp(3/4)$.

the curve at the beginning, and underestimates the curve at the right tail.

The estimates are also compared based on their AMSE's as shown in Table 2.2. For the second example, we found that the average mean squared error (AMSE) is the least for the estimates obtained by nearest neighbor approach which is a little smaller than the AMSE of the estimates by smoothed local histogram approach suggested by Antoniadis et. al. The Wavelet Kernel approach has very close AMSE which is a little bigger than the two other estimates. So for the case where lifetimes are generated from an exponential distribution with parameter 1 and censoring times are

Table 2.2: Table of AMSE of the subdensity estimates, $X_i \sim exp(1)$ and $C_i \sim exp(0.75)$

Estimates	MSE
$f_n^*(t)_{LH}$	0.0174
$f_n^*(t)_{NN}$	0.0172
$f_n^*(t)_{WK}$	0.0179

also generated from another exponential distribution with parameter 3/4, all three approaches have very close average mean squared errors.

Chapter 3

Estimation of Distribution Function

3.1 Notations and Model Setup

In this Chapter, we discuss the estimation of the distribution function of $Z_i = min(X_i, C_i)$ where, X_i's are lifetimes and C_i's are censoring times as defined in Chapter 2. By definition, the distribution function of the random variable Z is

$$L(t) = P(Z \leq t) = 1 - P(Z > t).$$

Now,

$$\begin{aligned} P(Z > t) &= P[\min(X, C) > t] \\ &= P[(X > t) \text{ and } (C > t)] \\ &= P(X > t)P(C > t), \quad \text{from independence} \\ &= [1 - P(X \leq t)][1 - P(C \leq t)] \\ &= [1 - F(t)][1 - G(t)] \end{aligned}$$

where $F(t)$ and $G(t)$ are the distribution functions of lifetimes and censoring times respectively. Therefore,

$$\begin{aligned} L(t) &= P(Z \leq t) \\ &= 1 - P(Z \geq t) \\ &= 1 - [1 - F(t)][1 - G(t)], t \geq 0. \end{aligned} \tag{3.1}$$

So, the observed $Z_i's$ have a distribution function L satisfying

$$1 - L(t) = P(T \geq t) = [1 - F(t)][1 - G(t)], \quad t \geq 0 \tag{3.2}$$

The function $1 - L(t)$ is commonly referred to as the survival function for censored data.

Given the set of iid observations $Z_1, Z_2, ..., Z_n$, from the common distribution function L, the standard non-parametric estimator of L is the empirical distribution func-

tion L_n defined as

$$L_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{[Z_i \leq t]}. \tag{3.3}$$

In the absence of additional information about the shape of L, the empirical distribution function L_n is the optimal estimator for L in the asymptotically minimax sense (Dvoretsky et.al, 1956).

In Section 3.1, the procedure based on density function described by Antoniadis et. al. (1999) is discussed. Section 3.2 elaborates the estimation procedure by Series expansion. Kronmal and Tarter's (1968) approach followed by a Wavelet modification is discussed there. A Modified Kaplan-Meier estimate discussed by Diehl and Stute (1988) is described in Section 3.3. Finally in Section 3.4 we compare the estimates obtained by these approaches.

3.2 Estimation based on the Density Function

Considering that a continuous estimator of L is better adapted to fully account for the smoothness of L, Antoniadis et.al. (1999) defined an estimator of the distribution function L by

$$\hat{L}_n(t) = \int_0^t \hat{l}_n(x)dx, \quad t \in [0, \tau], \tag{3.4}$$

where $\hat{l}_n$ is a traditional histogram type estimator of the density l of L. Let $\phi(t) =$

$I_{[0,\tau]}(t)$ be the indicator function of $[0,\tau]$ and denote by $\phi_{j,k}(t)$ the translated and dilated functions, $\phi_{j,k} = 2^{j/2}\phi(2^j t - k)$.

Let $\tilde{j}(n)$ be a sequence of scales such that $\tilde{j}(n) \to \infty$ as $n \to \infty$. The resolution $\tilde{j}(n)$ was chosen by folded cross validation (Nason, 1996). Then,

$$\hat{l}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 2^{\tilde{j}(n)/2}\phi_{\tilde{j}(n),0}(t - Z_i). \tag{3.5}$$

For the Haar wavelets,

$$\phi(t) = I_{[0,1)}(t).$$

Thus, $\hat{l}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 2^{\tilde{j}(n)/2}\phi_{\tilde{j}(n),0}(t - Z_i)$ can be viewed as a Haar wavelet histogram type estimator of the density function of the random variable $Z_i = min(X_i, C_i)$. Now define

$$\Phi_{\tilde{j}(n),0}(t - Z_i) = \int_0^t 2^{\tilde{j}(n)/2}\phi_{\tilde{j}(n),0}(u - Z_i)du.$$

Then

$$\hat{L}_n(t) = \int_0^t \hat{l}_n(x)dx = \frac{1}{n}\sum_{i=1}^{n} \Phi_{\tilde{j}(n),0}(t - Z_i) \tag{3.6}$$

is an integrated Haar transform estimator of the distribution function of Z. This estimator can be viewed as a wavelet estimator of $L(t)$ and we denote it by $\hat{L}_n(t)_{DF}$.

3.3 Estimation by Series Expansion

Kronmal and Tarter (1968) suggested an estimation procedure of cumulatives by Fourier series expansion. We propose a non-parametric modification of Kronmal and Tarter's estimates of CDF by using its wavelet substitute. The most commonly used estimate of the population cumulative L is the sample cumulative or step function, L^*. Suppose, the set $\{Z_{(i)}\}$ represents the set of n order statistics corresponding to the censored random sample $\{Z_i\}$ and $a < Z_i < b$. The step function $L^*(t)$ can be defined as

$$L^*(t) = \frac{1}{n}\sum_{i=1}^{n}\left[I_{(Z_i,b)}(t) + \frac{1}{2}I_{[Z_i,Z_i]}(t)\right]. \tag{3.7}$$

For $L^*(t) = \sum A_k\varphi_k(t)$ and $A_k = \int \varphi_k(t)L(t)\omega(t)dt$, Kronmal and Tarter(1968) defined their estimated CDF as

$$\hat{L}_m(t) = \sum_{k=0}^{m}\hat{B}_k\varphi_k(t) \tag{3.8}$$

where the set $\varphi_k(t)$ consists of functions orthogonal with respect to a weight function

$\omega(t)$. Here

$$\begin{aligned}\hat{B}_k &= \int \varphi_k(u)L(u)\omega(u)du \\ &= \int \varphi_k(u)\left[\frac{1}{n}\sum_{i=1}^{n}\varepsilon(u-Z_i)\right]\omega(u)du \\ &= \frac{1}{n}\sum_{i=1}^{n}\int \varphi_k(u)\varepsilon(u-Z_i)\omega(u)du,\end{aligned}$$

where

$$\varepsilon(u) = \begin{cases} 0, & \text{if } u < 0 \\ 1, & \text{if} u \geq 0 \end{cases}$$

The limitation of this approach is that it depends on the characteristic functions of the particular density of interest and hence is not very flexible.

We propose Wavelet extension instead of Fourier expansion for $\hat{L}_m(x)$. This estimation method is more general and more flexible than the method discussed above as it does not depend on the density function of the observations. The Wavelet extension can be expressed as

$$\hat{L}_m(t) = \hat{c}\phi(t) + \sum_{j=0}^{m}\sum_{k=0}^{2^j-1}\hat{d}_{jk}\psi_{jk}(t) \tag{3.9}$$

where,

$$\begin{aligned}\hat{c} &= \int w(u)L(u)\phi_(u)du \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{Z_i}^{N-1} w(u)\phi_{j,k}(u)\varepsilon(u-Z_i)du, \end{aligned} \tag{3.10}$$

and

$$\hat{d}_{j,k} = \frac{1}{n}\sum_{i=1}^{n}\int_{Z_i}^{N} w(u)\psi_{j,k}(u)\varepsilon(u-Z_i)du, \tag{3.11}$$

where,

$$\psi_{jk}(u) = 2^{j/2}\psi(2^j u - k). \tag{3.12}$$

We note that, for any $x \in [a,b)$, $w = \frac{x-a}{b-a} \in [0,1]$. Furthermore,

$$\begin{aligned} L_W(w) &= P(W \leq w) \\ &= P\left(\frac{X-a}{b-a} \leq w\right) \\ &= P\left(X \leq a + w(b-a)\right) \\ &= L_X(x) \end{aligned} \tag{3.13}$$

Our approach is therefore to scale the values of the random variable Z into $[0,1]$ and then compute $L_Z(z)$, with $\omega(u) = \frac{1}{max(\hat{L}_m(z))}$.

3.4 Estimation Procedure Based on the Kaplan-Meier Method

The distribution function of the observed $Z_i's$ is L such that

$$1 - L(t) = P(T \geq t) = (1 - F(t))(1 - G(t)), \quad x \geq 0$$

where $1 - L(t)$ is the survival function for censored data. The familiar Kaplan Meier product limit estimator of the survival function $1-F(x) = P(X \geq x)$ can be expressed as

$$1 - \hat{F}_n(x) = \begin{cases} \prod_{j=1}^{n} \left(\frac{n-j}{n-j+1}\right)^{I_{[Z_j \leq x, \delta_j = 1]}} & x < \max\ Z_j, \quad 1 \leq j \leq n \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

The modified Kaplan Meier estimator of $1 - G(x)$ can therefore be written as

$$1 - \hat{G}_n(x) = \begin{cases} \prod_{j=1}^{n} \left(\frac{n-j+1}{n-j+2}\right)^{I_{[Z_j \leq x, \delta_j = 0]}} & x < \max\ Z_j, \quad 1 \leq j \leq n \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

An estimate of the survival function $1 - L(t)$ is then $1 - \hat{L}(t)$ where

$$\begin{aligned} \hat{L}(t) &= 1 - [1 - \hat{F}_n(t)][1 - \hat{G}_n(t)] \qquad (3.16) \\ &= \begin{cases} 1 - \prod_{j=1}^{n} \left(\frac{n-j}{n-j+1}\right)^{I_{[Z_j \le x, \delta_j = 1]}} \prod_{j=1}^{n} \left(\frac{n-j+1}{n-j+2}\right)^{I_{[Z_j \le x, \delta_j = 0]}} & x < \max\ Z_j, \quad 1 \le j \le n \\ 1 & \text{otherwise} \end{cases} \end{aligned}$$

We denote this estimator by $L_n(t)_{KM}$.

The average mean squared error AMSE's can be computed for all the methods by averaging the mean squared errors

$$MSE(L^*(t)) = \frac{1}{K} \sum_k [\hat{L}_n(t_k) - L^*(t_k)]^2.$$

AMSE is used to compare the estimates obtained by different approaches and the results are shown in the next section.

3.5 Simulation Studies

To illustrate the methods, we conducted two simulation studies as before. First, we used the lifetimes X_i generated from a gamma distribution with shape parameter equal to 5 and scale parameter equal to 1. The censoring times C_i were generated from an exponential distribution with mean 6 as described in Chapter 2. Censoring

rate was close to 50 percent. 200 samples each of size 200 were generated in the simulation study.

All the methods discussed in the previous section to estimate the cumulative distribution function were used and we constructed Figure 3.1 based on the estimates obtained from different approaches for the subdensity: where the lifetimes has a gamma distribution and the censoring time has an exponential distribution.

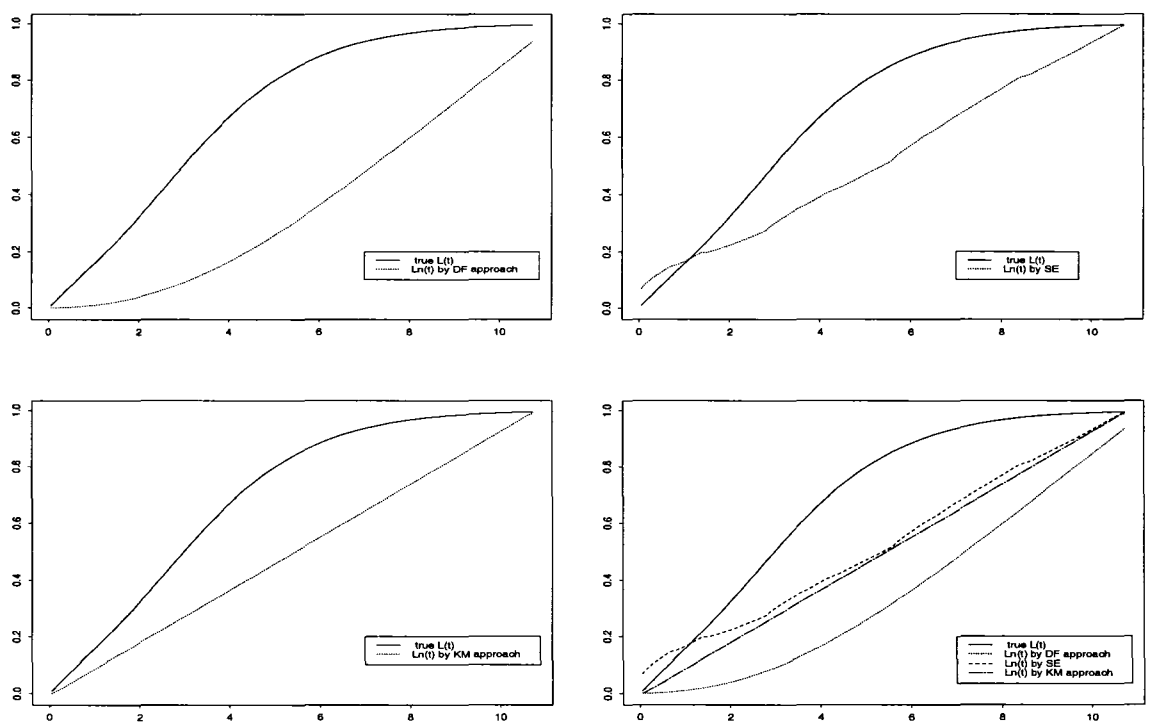

Figure 3.1: Estimates of CDF's by different methods: X_i $gamma(5,1), C_i$ $exp(\frac{1}{6})$.

For the subdensity, where the lifetimes are generated from gamma(5, 1) and censoring times are generated from exp(1/6), we observe from the 4 graphs in Figure 3.1 that all the approaches used for the estimation of the distribution function are resulting in underestimates of the actual CDF. But among all three methods considered, the series expansion approach with wavelet modification (Ln(t) SE) produces better estimates than the two other approaches, density function approach (Ln(t) DF) and the Kaplan Meier approach (Ln(t) KM). However, the estimates obtained by Kaplan Meier approach is quite close to the estimates obtained by Wavelet series expansion.

Table 3.1: Table of AMSE of the CDF estimates, $X_i \sim gamma(5, 1)$ and $C_i \sim exp(1/6)$

Estimates	AMSE
$L_n(t)_{DF}$	0.3499
$L_n(t)_{SE}$	0.0485
$L_n(t)_{KM}$	0.0550

The estimates are also compared based on their AMSE's. Table 3.1 is constructed for comparing the different approaches with respect to their average mean squared errors. For the first example, we found the average mean squared error (AMSE) is the least for the estimates obtained by Series expansion approach (Modified Tarter and

Kronmal (1968) approach). The AMSE of the estimates by modified Kaplan-Meier approach is very close to the first one. The AMSE of estimates by density function approach suggested by Antoniadis et.al(1999) is comparatively larger than the two other methods considered in this study.

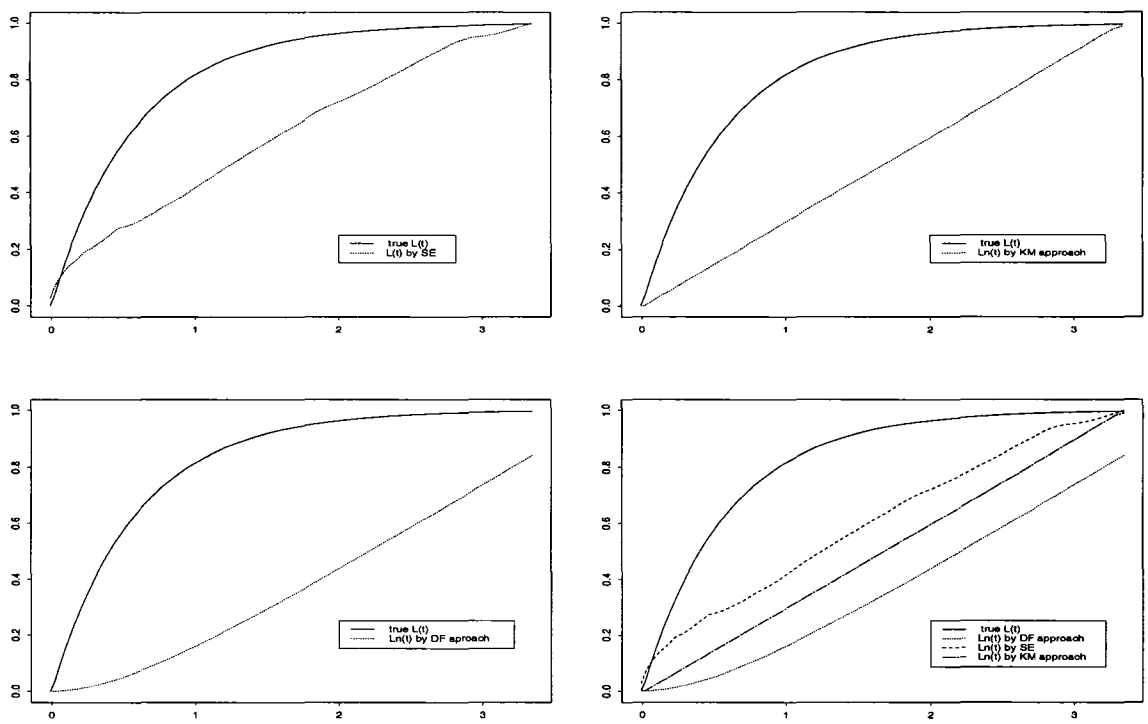

Figure 3.2: Estimates of CDF's by different methods when $X_i \sim exp(1)$, $C_i \sim exp(3/4)$.

Secondly, we generated lifetimes X_i from an exponential distribution with parameter 1 and censoring times C_i from exponential distribution with parameter 3/4 to ensure that there is about 40 percent censoring. All the methods are again applied to esti-

mate the subdensities and we constructed Figure 3.2 based on the estimates.

Table 3.2: Table of AMSE of the subdensity estimates, $X_i \sim exp(1)$ and $C_i \sim exp(0.75)$

Estimates	AMSE
$L_n(t)_{DF}$	0.2596
$L_n(t)_{SE}$	0.0682
$L_n(t)_{KM}$	0.1343

Figure 3.2 shows that for the CDF where the lifetimes are generated from exp(1) and censoring times are generated from exp(3/4), all the approaches used for estimating the distribution function are resulting in underestimates of the actual CDF. However, among all three methods considered, the series expansion approach with wavelet modification (Ln(t) SE) produces clearly better estimates than the two other approaches.

The estimates are also compared based on their AMSE's as given in Table 3.2. For the second example, we found that the average mean squared error (AMSE) is the least for the estimates obtained by wavelet series expansion.The AMSE for estimates obtained by Kaplan Meier approach are twice and the AMSE for estimates obtained

by density function approach are 4 times that of Wavelet approach.

Chapter 4

Hazard Estimation

4.1 Notations and Model Setup

In the analysis of lifetime data or time to event data, a primary interest is to assess the risk of an individual observing a particular event at certain times. This risk is what is called the hazard rate or hazard function. Let X_i, $i = 1, 2, ..., n$, be the lifetimes of n independent, identical units. $X_1, X_2, ..., X_n$ are non-negative and iid (independent and identically distributed) with common continuous cumulative distribution function (CDF) F and continuous density f. The risk of an individual at time t can be measured by the hazard rate or hazard function, defined by

$$\lambda(t) = lim_{\Delta t \to 0} \frac{P[t \leq T \leq t + \Delta t | T \geq t]}{\Delta t}. \tag{4.1}$$

Hazard rate inference is a widely used method to analyze the properties of durations between specific events, as it reflects the instantaneous probability that a duration will end in the next time instant. An increasing hazard rate indicates that the probability that a spell will be completed is increasing with the duration of the event, which is called positive duration dependence. Similarly, a decreasing hazard rate reflects negative duration dependence (Spierdijk, 2005).

If X is a continuous random variable, then,in the absence of censored individuals, the hazard function can be expressed as

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \quad F(t) < 1. \tag{4.2}$$

If, C_i is the corresponding censoring time of i-th individual with pdf g and CDF G, and $Z_i = min(X_i, C_i)$. Then Z_i has the distribution function L such that

$$\begin{aligned} L(t) &= P(Z_i \leq t) \\ &= 1 - [1 - F(t)][1 - G(t)], \quad \text{(which was given in equation 3.1).} \end{aligned} \tag{4.3}$$

Then

$$1 - L(t) = \{1 - F(t)\}\{1 - G(t)\}. \tag{4.4}$$

Then for censored data, if $G(t) < 1$, we have

$$\lambda(t) = \frac{f(t)(1 - G(t))}{(1 - F(t))(1 - G(t))}, \ F(t) < 1. \tag{4.5}$$

Substituting equations (4.4) and (4.5) into equation (4.6) gives

$$\lambda(t) = \frac{f^*(t)}{1 - L(t)}, \ L(t) < 1. \tag{4.6}$$

which is the hazard rate for censored data.

4.2 Estimation Procedure

Different approaches for estimating the hazard rate is available in the literature which includes parametric, semiparametric and nonparametric approaches. These methods include penalized likelihood methods (Antoniadis, 1989, Antoniadis and Gregoire, 1990), local likelihood methods(Loader, 1999), estimation by using orthogonal series (Kronmal and Tarter, 1968, Tanner and Wong, 1984), Spline models(Kooperberg and Stones, 1992), Kernel estimation (Ramlau-Hansen, 1983, Roussas, 1989, 1990) and Orthogonal Wavelet methods (Patil, 1997, Antoniadis et.al., 1999, Li, 2002). In this

chapter, we discuss the hazard estimates by Antoniadis et.al(1999), and compare it with some other possible estimates of hazard function.

Several estimation procedures of the subdensity function, $f^*(t)$, and the cumulative distribution function $L(t)$ for censored data were described in Chapter 2 and Chapter 3 respectively. Here, we use those estimates to compute the hazard function using equation 4.7.

4.3 Results and Discussion

We computed seven estimates of the hazard function:

- Hazard estimate obtained by using the ratio of subdensity estimate by local histogram approach and CDF estimate by Wavelet expansion. We refer to this approach as Model 1.

- Hazard estimate obtained by using the ratio of subdensity estimate by Wavelet Kernel approach and CDF estimate by Wavelet extension. This approach is referred to as Model 2.

- Hazard estimate obtained by using the ratio of subdensity estimate by local histogram approach CDF estimate by modified Kaplan Meier approach. We refer to this approach as Model 3.

- Hazard estimate obtained by using the ratio of subdensity estimate by Wavelet Kernel approach and CDF estimate by wavelet extension. We refer to this approach as Model 4.

- Hazard estimate obtained by using the ratio of Nearest Neighbor subdensity estimate and CDF estimate by wavelet extension. This model is referred to as Model 5.

- Hazard estimate obtained by using the ratio of Nearest Neighbor subdensity estimate and CDF estimate by modified Kaplan Meier approach. We refer to this approach as Model 6.

- Hazard estimate proposed by Antoniadis et.al (1999). We refer to this approach as Model 7.

These estimates are used to compute the hazard rate for both the lifetimes generated from gamma distribution and exponential distribution as discussed in the Simulation studies parts of Chapter 2 and Chapter 3. The hazard functions for the subdensity where lifetimes were generated from gamma density are shown in Figure 4.1.

For comparison, we computed average mean squared error (AMSE) for the hazard rates of all the models. The results are listed in Table 4.1. The AMSEs were estimated

Table 4.1: Table of AMSE of the hazard estimates, $X_i \sim gamma(5,1)$ and $C_i \sim exp(1/6)$

Estimates	AMSE
Model 1	3.5279
Model 2	6.5396
Model 3	3.6819
Model 4	6.5471
Model 5	4.4623
Model 6	4.6113
Model 7	4.3303

by averaging the mean squared errors

$$MSE(\lambda) = \frac{1}{K}\sum_{k}[\hat{\lambda}_n(t_k) - \lambda(t_k)]^2.$$

For the first example where lifetime is generated from a gamma distribution and censoring times are generated from an exponential distribution, we observe from the Figure 4.1 and the Table 4.1 that the hazard model 1 using subdensity estimate by smoothed local histogram as suggested by Antoniadis et.al.'s and CDF estimate by Wavelet series expansion gives the smallest AMSE. We obtained similar result for the model 3 using the same estimate for subdensity and CDF estimate obtained by Kaplan Meier approach. However, from figure 4.1, we observe that none of the

estimates are very well fitted to the actual hazard function. It would be worthy to mention that Antoniadis et.al.(1999) in their paper, reported the AMSE for hazard function estimation based on 200 repetitions of the simulations for sample size n = 200 to be 0.112. Also they reported that the hazard estimates were only computed at points were $L(t) > 0.5$ as the hazard estimates are very unstable and have little meaning when few subjects were left at risk.

Table 4.2: Table of AMSE of the hazard estimates, $X_i \sim exp(1)$ and $C_i \sim exp(0.75)$

Estimates	AMSE
Model 1	181.4185
Model 2	137.2793
Model 3	42.1125
Model 4	9.94
Model 5	36.6551
Model 6	39.8152
Model 7	42.09

While both lifetime and censoring time are generated from exponential distribution, from Figure 4.2, it can be viewed that the hazard model 1, hazard model 3 and hazard model 6 are close to the actual curve. For comparison, we computed the AMSE for the hazard rates of all the models and the results are listed in Table 4.2. From Table

4.2, we found, that the hazard model 1 using subdensity estimate by Wavelet kernel approach and CDF estimate by Kaplan Meier approach produces smallest AMSE. No simulation studies for calculating AMSE's for this situation is available in literature.

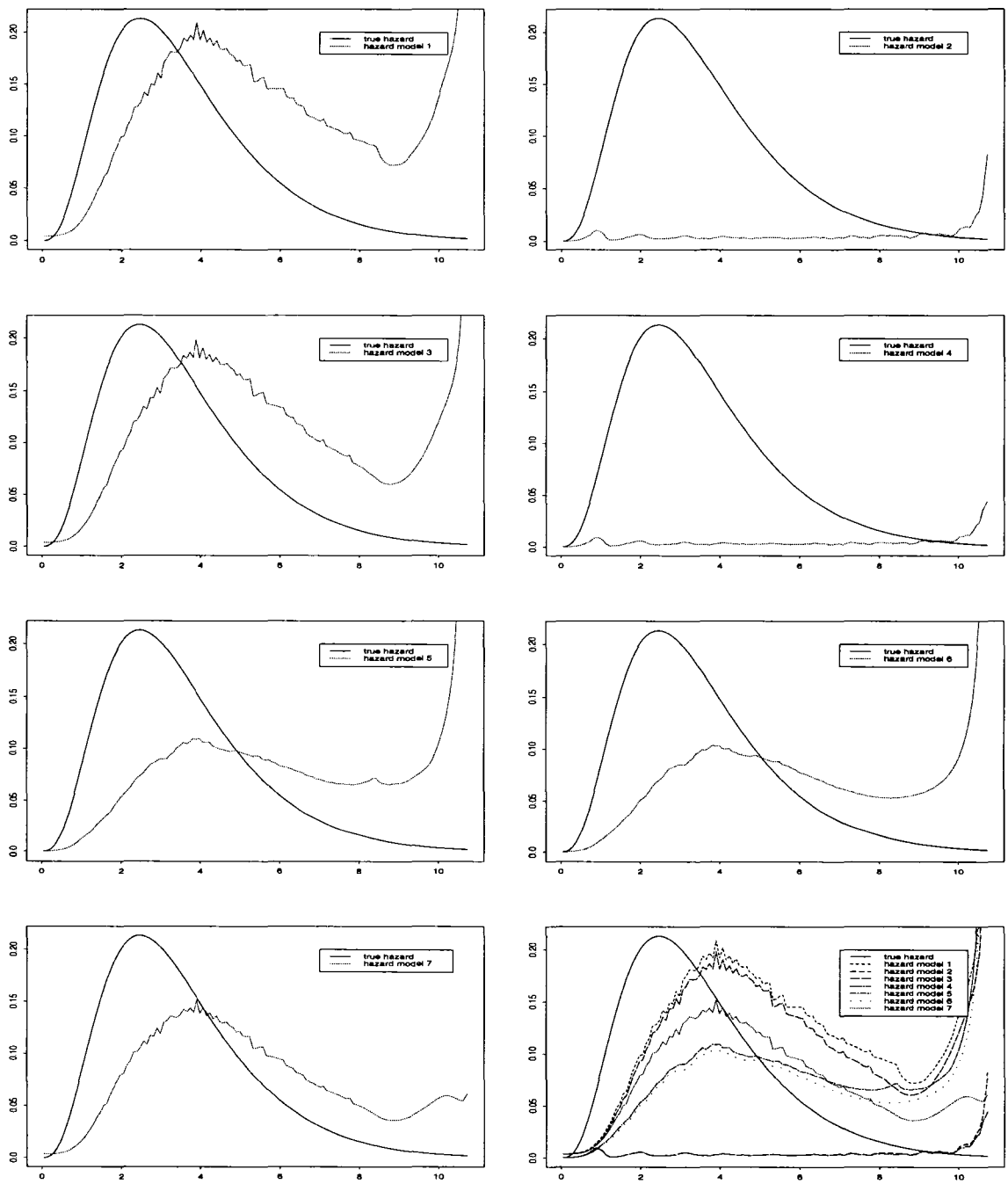

Figure 4.1: Estimated Hazard functions, $X_i \sim gamma(5,1)$ and $C_i \sim exp(1/6)$.

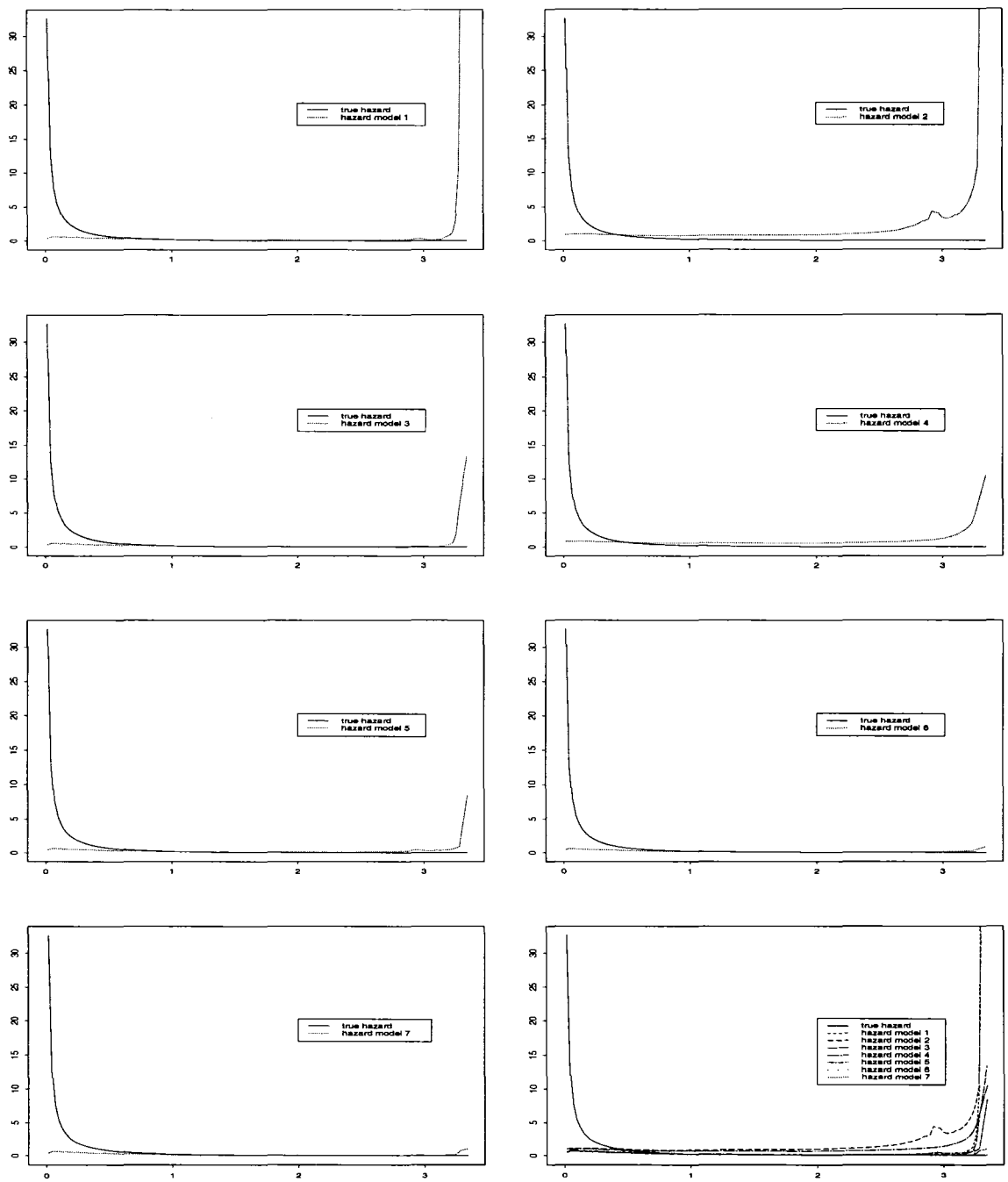

Figure 4.2: Estimated Hazard functions $X_i \sim exp(1)$, $C_i \sim exp(3/4)$.

Chapter 5

Concluding Remarks

In this study, we examined the subdensity, CDF and hazard rate estimation procedure for two cases. First, lifetimes are generated from gamma and censoring times are generated from exponential distribution and there is about fifty percent censoring in the data. Second, both lifetimes and censoring times are generated from exponential densities and about forty percent censoring occurs in the data.

For subdensity estimation, we applied local histogram approach, nearest neighbor approach and Wavelet Kernel approach and used wavelet smoothers for smoothing the crude estimate obtained by local histogram approach. For lifetimes generated from gamma distribution, we found that the first two approaches worked fairly well. Wavelet Kernel approach suggested by Xue (2002) does not provide good estimates

of the subdensity for this case. On the other hand, when lifetimes are generated from an exponential distribution, we found that all three approaches are competitive.

For CDF estimation, we used three approaches, density function approach suggested by Antoniadis et.al.(1999), Wavelet series expansion by modifying the Fourier series expansion suggested by Kronmal and Tarter(1968) and Kaplan Meier approach suggested by Diehl and Stute(1988). We found that Wavelet series expansion , i.e. modification of Kronmal and Tarter approach gives better estimate of the CDF than the two other methods adopted, for both the examples.

While estimating hazard rate, several possible combination of subdensity estimates and CDF estimates (as mentioned earlier) are tried for both the cases. We computed the 7 hazard models with different combinations of subdensity estimates and CDF estimates.

For the case where lifetime is generated from Gamma distribution and censoring times are generated from Exponential distribution, we found that the hazard model 1 using Antoniadis et.al.'s subdensity estimate and CDF estimate by Wavelet series expansion gives the smallest AMSE. We obtained similar result for the model 3 using Antoniadis et.al.'s subdensity estimate and CDF estimate by Kaplan Meier approach.

While both lifetime and censoring time are generated from Exponential distribution, we found, that the hazard model 1 using subdensity estimate by wavelet kernel approach and CDF estimate by Kaplan Meier approach produces smallest AMSE.

Bibliography

[1] Alpert, B. K. (1992). *Wavelets and Other Bases for Fast Numerical Linear Algebra in Wavelets: A Tutorial in Theory and Applications*, ed. C. K. Chui, Boston: Academic Press, 181-216.

[2] Antoniadis, A., Gregoire, G. and McKeague, I.(1990). Penalized likelihood estimation for rates with censored survival data. *Scand.J.Statist.*, **17**,43-63

[3] Antoniadis, A.(1989). A penalty method for nonparametic estimation of the intensity of a counting process. *Annals of Institute of Mathematical Statistiscs*, **41**, 781-807.

[4] Antoniadis, A., Gregoire, G. and McKeague, I.W. (1994). Wavelet methods for Curve Estimation. *Journal of the American Statistical Association*, **89**, 428, 1340-1353.

[5] Antoniadis, A., Gregoire, G. and Nason, G. (1999). Density and hazard rate estimation for right censored data by using wavelet methods. *Journal of the*

Royal Statistical Society, **B, 61**, 1, 63-84.

[6] Beran, R.(1981). Nonparametric regression with randomly censored survival data. *Technical Report.* Department of Statistics, University of California, Berkeley.

[7] Buckheit, J. B. and Donoho, D.L.(1995) Wavelet and Reproducible Research. *Lecture Notes on Statistics,* **103**

[8] Dabrowska, D. M.(1987). Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics*, **14**, 181-197.

[9] Daubechies, I. (1992). *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

[10] Daubechies, I. and Lagarias, J. (1991). Two-scale difference equations I. Existence and global regularity of solutions. *SIAM Journal of Mathematical Analysis.* **22(5)** 1388-1410.

[11] Daubechies, I. and Lagarias, J. (1992). Two-scale difference equations II. Local regularity, infinite products of matrices and fractals. *SIAM Journal of Mathematical Analysis.* **23** 1031-1079.

[12] David, L., Donoho, Iain M., et al., (1996). Density estimation by wavelet thresholding. *Annals of Statistics*, **24**,508-539.

[13] Diehl, S. and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. *Journal of Multivariate Analysis*, **25**, 299-310.

[14] Donoho, D. L. and Johnstone, I. M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425-455.

[15] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D.(1995) Wavelet shrinkage:asymptopia(with discussion)? *Journal of Royal Statistical Society,*B, , **57**, 301-369.

[16] Dvoretsky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and the classical multinomial estimator. *Annals of Mathematical Statistics*, **23**, 277-281.

[17] Fix, E. and Hodges, J. L. (1951). Discriminatory analysis, nonparametric estimation: consistency properties. *Report No. 4, Project No. 21-49-004*, USAF School of Aviation Medicine, Randolph Field, Texas.

[18] Földes, A., Rejto, L. and Winter, B. B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data. II: Estimation of density and failure rate. *Periodica Mathematica Hungarica* 12, 15-29.

[19] Gray, R. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of American Statistical Association* 87, 942-951.

[20] Hall, P. and Patil, P.(1996). Effect of threshold rules on performance of wavelet based curve estimators. *Statistical Sinica*, **6**, 331-345.

[21] Herrick, D. R. M., Nason G. P. and Silverman, B. W. (2001). Some new methods for wavelet density estimation. *Sankhya*, **A, 63**, 394-411.

[22] Huang, S. Y. (1999). Density estimation by wavelet based reproducing Kernels, *Statistical Sinica*, **9**, 137-151.

[23] Kalbfleisch, J.D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, New York, John Wiley.

[24] Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer-Verleg, New York, Inc.

[25] Kooperberg, C. and Stone, C. J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, **1**, 4, 137-151.

[26] Kronmal, R. and Tarter, M. (1968). The estimation of probability density and cumulatives by Fourier series methods. *Journal of American Statistical Association*, **63**, 925-952.

[27] Li, L. (2002). Hazard rate estimation for censored data by wavelet methods. *Communications in Statistics - Theory and Methods*, **31**, 6, 943-960.

[28] Lo, S. H., Mack, Y.P. and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan Meier estimator. *Probability Theory Related Fields*, **80**, 3, 461-473.

[29] Müller, P. and Vidakovic, B. (1998). Bayesian inference with wavelets: density estimation. *Journal of Computational and Graphical Statistics*, **7**, 4, 456-468.

[30] Mallat, S. (1989) Multiresolution approximations and wavelet orthonormal bases of L^2 **(R)**. *Transactions of the American Mathematical Society*, **315**,69-87.

[31] McNichols, D. T. and Padgett, W. J. (1985). Nonparametric methods for hazard rate estimation from right-censored samples. *Annals of Statistics*, **18,** 1172-1187.

[32] Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *Journal of Royal Statistical Society*, **B**, **58**, 463-479.

[33] Nason, G.P. and Silverman, B.W. (1994). The discrete wavelet transform in S. *Journal of Computational and Graphical Statistics*, **3**, 2, 163-191.

[34] O'Sullivan, F.(1988) Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing*, **9**,363-379.

[35] Patil, P. N. (1997). Nonparametric hazard rate estimation by orthogonal wavelet methods. *Journal of Statistical Planning and Inference*, **60**, 153-168.

[36] Pinherio, A. and Vidakovic, B (1997). Estimating the square root of a density via compactly supported wavelets. *Computational Statistics and Data Analysis* **25**, 399-415.

[37] Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Annals of Statistics*, 11 , 453-466.

[38] Sain, S. R.(2002). Multivariate locally adaptive density estimation. *Computational Statistics and Data Analysis*, **39**, 165-186.

[39] Semadeni, C., Davison, A. C. and Hinkley, D.V. (2004). Posterior prbability intervals in Bayesian wavelet estimation. *Biometrika*, **91**, 2, 497-505.

[40] Spierdijk, L. (2005). Nonparametric Conditional Hazard Rate Estimation.*http://wwwhome.math.utwente.nl/ spierdijkl/hazard.pdf*

[41] Strang, G. (1989). Wavelets and dilation equations: a brief introduction. *SIAM Review*, **31**, 4, 614-627.

[42] Sun, L. Q., Zhu, L.X., (1999). A Berry-Esseen type bound for kernel density estimators under random censorship. *Acta Mathematica Sinica*, **42**,627-636.

[43] Tanner, M. A. (1983). A note on the variable kernel estimator of the hazard function from randomly censored data. *Annals of Statistics*, 11, 994-998

[44] Tanner, M. A. and Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by kernel method. *Annals of Statistics*, **41**, 718-722.

[45] Tanner, M. A. and Wong, W. H. (1984). Data based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis. *Journal of the American Statistical Association*, **79**, 174-182.

[46] Truong, Y. K. and Patil P. N.(2001). Asymptotics for wavelet based estimates of piecewise smoth regression for stationary time series. *Annals of the Institute of Statistical Mathematics* , **53**, 1, 159-178.

[47] Vidakovic, B.(1999). *Statistical modeling by wavelets.* John Wiley & Sons, Inc., New York.

[48] Walter, G. G. (1994). Wavelet and other orthogonal systems with application. *CRC Press*, Boca Raton, FL.

[49] Wang,Q.H.,(1997). The smoothed Bootstrap approximation for the kernel estimator of probability density under random censorship. *Acta Mathematica Applicatae Sinica*, **20**,367-377.

[50] Watson, G. S. and Leadbetter, M.R. (1964). Hazard rate analysis. I. *Biometrika*, **51**, 175-184.

[51] Watson, G. S. and Leadbetter, M.R. (1964). Hazard rate analysis. II. *Sankhya*, Ser. A, **26**, 101-116.

[52] Wu, S. S. and Wells, M.T. (1999). Nonparametric estimation of hazard functions by wavelet methods. *Nonparametric Statistics*, **31**, 4, 614-627.

[53] Xue, L.G. (2004). Approximation rates of the error distribution of wavelet estimators of a density function under censorship. *Journal of Statistical Planning and Inference*, **118**, 167-183.

[54] Yandell, B.S. (1983). Nonparametric inference for rates with censored survival data. *Annals of Statistics*, 11, 1119-1135.