SOME CONTRIBUTIONS TO RENEWAL DENSITY ESTIMATION

Thesis submitted to the
University of Calicut

For the award of the Degree of
DOCTOR OF PHILOSOPHY

K.K. Hamsa

DEPARTMENT OF STATISTICS
FAROOK COLLEGE, KOZHIKODE
KERALA, INDIA-673632

2009

CERTIFICATE

This is to certify that the work reported in this thesis entitled **"Some Contributions to Renewal Density Estimation"**, that is being submitted by **Sri. K.K.Hamsa** for the award of Doctor of Philosophy, to the University of Calicut, is based on the bonafide research work carried out by him under my supervision and guidance in the Department of Statistics, Farook College. The results embodied in this thesis have not been included in any other thesis submitted previously for the award of any degree or diploma of any other university or institution.

Farook College
26-09-2009

Dr. P. Anilkumar
Head of the Department
Department of Statistics
Farook College

ACKNOWLEDGEMENT

I am immensely grateful to my research supervisor Dr. P.Anilkumar, Head, Department of Statistics, Farook College , Kozhikode. He has been a source of strength and inspiration to me in all the phases of this research. More words are insufficient to express my profound gratitude to him.

I am greatly indebted to my teacher and former guide Dr. N. Raju, Professor, Department of Statistics, Calicut University for his encouragements.

I would like to express my sincere thanks to Prof. A. Kuttialikutty, Principal of Farook College for providing necessary facilities to complete this work.

It is my duty to thank Managing Committee, Farook College, for providing me an opportunity to undergo this course.

The teacher fellowship received from University Grants Commission during 2005-07 through Faculty Improvement Program is gratefully acknowledged.

In addition I am thankful to my wife Ramla, daughter Shahana Farhath and son Shahan for their support throughout the period of my research.

Farook College
26-09-2009

K.K.HAMSA

Contents

List of Figures

Chapter 1

Introduction and Summary

1.1 Introduction

A fundemantal problem in statistics is to develop models based on a sample of observations so that further analysis can be done with statistical techniques using the model so developed. *Parametric modeling* has been the subject of investigations among various researchers. A disadvantage of parametric modeling is that it may not be robust in the sense that a slight contamination of the data by observations not following the particular parametric family might lead to erroneous conclusions. Further the data might be of such a type that there is no suitable parametric family that gives a good fit. These lead to the applications of non parametric modeling. Since the appearance of the first paper in the area of non parametric functional estimation by Rosenblatt(1956), several methods have been developed for non parametric estimation of density functions, distribution functions, regression functions, failure rates, etc. Some of the popular methods are the method of sieves, the method of histogram, kernel, the method of orthogonal series and the method of penalized likelihood. An extensive survey of various non parametric methods are given in Prakasa Rao(1983), Silverman(1986), Gibbons *et.al.*, (2003), Natalia M.(2007) and Wasserman(2007).

In this thesis we study some non parametric estimators of the functionals involved in stochastic processes. We consider the estimation of the renewal density function of a renewal process, estimation of the probability density function(PDF) and estimation of the intensity of a counting process in the multiplicative intensity model. The non-parametric estimator of the renewal density function is not available in the literature. In the estimation of renewal density function we have extended the notion of U-statistics. Martingale convergence theorem and projection techniques are used for showing the asymptotic properties of the estimator. For other problems we obtain estimators which are better alternatives to the estimators reported in the literature. In the estimation of the probability density function we have used a sequence of polynomials which have been used to prove the Stone- Weistrass theorem; Rudin (1976). The technique of likelihood cross validation is used to choose the appropriate value of the smoothing parameter of the proposed estimator. Orthogonal series are used to estimate the intensity of counting processes. More detailed discussion is differed to section 1.3.

In the next section we discuss some definitions and preliminaries associated with the stochastic processes and a brief review of the respective functionals. This section also gives a selective review of the techniques and theorems associated with the estimation methods developed in subsequent chapters.

1.2 Some Preliminaries and Review

1.2.1 Some Convergence Results

In this section we summarize the results needed in proving the asymptotic properties of certain estimates in the later chapters.

Definition 1.2.1 *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. The sequence of random variables $\{W_n\}$ is said to converge in probability to a constant c, if for every $\epsilon > 0$,*

$$\lim_{n \to \infty} P[|W_n - c| < \epsilon] = 1.$$

When $\{W_n\}$ is a sequence of estimators of a parameter γ, then W_n is termed a *consistent* estimator of γ if W_n converges to γ in probability for each value of γ.

Definition 1.2.2 *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. The sequence of random* *variables $\{W_n\}$ is said to converge in quadratic mean to a constant c, if*

$$\lim_{n\to\infty} E[(W_n - c)^2] = 0.$$

Definition 1.2.3 *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. The sequence of random variables $\{W_n\}$ is said to have a limiting distribution $F(w)$ (or to be asymptotically distributed according to $F(w)$), if*

$$\lim_{n\to\infty} P[(W_n \leq w] = F(w)$$

for all w values at which the d.f.F(w) is continuous.

Now we state some important theorems regarding various modes of convergence.

Theorem 1.2.1 *Convergence in quadratic mean implies convergence in probability.*

Theorem 1.2.2 *If $\{W_n\}$ converges in probability to c and if $k(t)$ is a function that is continuous at $t = c$, then $k(W_n)$ converges to $k(c)$.*

Theorem 1.2.3 (Slutsky's Theorem) *Let $\{W_n\}$ be a sequence of random variables with limiting distribution $F(w)$. Let $\{X_n\}$ denote a sequence of random variables that converges in probability to the constant c. Then the first and second members of each pair listed bellow have the same limiting distribution:*

(a) $(W_n + X_n)$ and $(W_n + c)$ *(b) $X_n W_n$ and cW_n* *(c) $\dfrac{W_n}{X_n}$ and $\dfrac{W_n}{c}$.*

Theorem 1.2.4 *If the sequence of random variables $\{V_n\}$ has an asymptotic distribution with d.f. $F(v)$ and if $\{W_n\}$ is a sequence ov random variables such that $\{W_n - V_n\}$ converges in probability to 0, then the limiting distribution of $\{W_n\}$ is also given by the d.f. $F(w)$.*

1.2.2 The Projection Principle

The notion of projection technique is popularized by Hajek (Serfling(1980)). Let $X_1, X_2,$

$\ldots, X_n$ denotes independent and identically distributed random variables, each with distribution function $F(.)$. Suppose we desire to show that a standardized version of a statistic $W = W(X_1, X_2, \ldots, X_n)$ has a limiting distribution, where $W(.)$ is such that it treats the n random variables $X_1, X_2, \ldots, X_n$ symmetrically. Since standardization will be necessary, we actually works with

$$W^* = W - E(W)$$

throughout this development. Hence $E(W^*) = 0$.

Consider a class of random variables, each members of which is a sum of independent and identically distributed random variables. Specifically let

$$\mathcal{V} = \left\{V/V = \sum_{i=1}^{n} k(X_i), \text{where } k(.) \text{ is some real valued function}\right\}.$$

The *projection* of W^* on to $\mathcal{V}$, this class of sums of independent and identically distributed random variables is given by

$$V^* = \sum_{i=1}^{n} k^*(X_i),$$

where

$$k^*(x) = E[W^*/X_i = x].$$

It can be seen that V^* is the member of $\mathcal{V}$ that is the closest in some sense to W^*; Randles, R.H and Wolfe, D.A.(1979). The projection V^* need not be a statistic, since it may depend on parameters or properties of the underlying distribution $F(.)$.

The following lemma shows that the projection principle produces a random variable in the class $\mathcal{V}$ that is as close as possible to the original statistic when closeness is measured by the expected squared difference.

Lemma 1.2.1 *Suppose $W^* = W^*(X_1, X_2, \ldots, X_n)$ treats the n i.i.d. random variables $X_1, X_2, \ldots, X_n$ symmetrically and $E[W^*] = 0$. Let V^* denote the projection of W^* onto $\mathcal{V}$. Specifically,*

$$V^* = \sum_{i=1}^{n} k^*(X_i),$$

where

$$k^*(x) = E[W^*/X_i = x].$$

Then, for any V in $\mathcal{V}$,

$$E[(W^* - V^*)^2] \leq E[(W^* - V)^2].$$

In chapter 2, the projection technique is used to establish the asymptotic normality of the proposed non parametric estimator of the renewal density function using some convergence rules and Slutski's theorem.

1.2.3 Martingales

Definition 1.2.4 *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. An increasing sequence of σ-fields $\{\mathcal{F}_n, n \geq 0\}$, $\mathcal{F}_n \subseteq \mathcal{F}$ is called a filtration.*

Definition 1.2.5 *Sequence $\{X_n\}$ is adapted to $\{\mathcal{F}_n, n \geq 0\}$ if X_n is $\mathcal{F}_n$ measurable.*

Definition 1.2.6 *A random sequence $X = \{X_1, X_2, \ldots,\}$ of $\mathcal{R}$-valued random variables with finite mean is a martingale with respect to a filtration $\{\mathcal{F}_n, n \geq 0\}$ to which it adapted if X_n is $\{\mathcal{F}_n\}$ measurable and*

$$E(X_{n+1}/\mathcal{F}_n) = X_n \quad a.s. \tag{1.2.1}$$

for all $n \geq 0$.

By relaxing the inequality (1.2.1) in to one sided inequalities, we can have two related definitions:

Definition 1.2.7 *A random sequence $X = \{X_0, X_2, \ldots,\}$ of $\mathcal{R}$-valued random variables with finite mean is a supermartingale with respect to a filtration $\{\mathcal{F}_n, n \geq$*

$0\}$ *to which it adapted if*

$$E(X_{n+1}/\mathcal{F}_n) \leq X_n \quad a.s. \tag{1.2.2}$$

for all $n \geq 0$.

Definition 1.2.8 *A random sequence* $X = \{X_0, X_2, \ldots,\}$ *of* $\mathcal{R}$*-valued random variables with finite mean is a submartingale with respect to a filtration* $\{\mathcal{F}_n, n \geq 0\}$ *to which it adapted if*

$$E(X_{n+1}/\mathcal{F}_n) \geq X_n \quad a.s. \tag{1.2.3}$$

for all $n \leq 0$.

The concept of reverse reverse martingale is defined using decreasing sequence of σ-fields called reverse filtration $\mathcal{G}_n$.

Definition 1.2.9 *A random sequence* $Z = \{Z_0, Z_2, \ldots,\}$ *is called a reverse martingale with respect to a reverse filtration* $\{\mathcal{G}_n, n \geq 1\}$ *if for all* n, Z_n *has finite mean, measurable with respect to* $\mathcal{G}_n$, *and satisfies*

$$E[Z_n/\mathcal{G}_{n+1}] = Z_{n+1} \quad a.s.$$

One important theorem of great consequences in martingale theory is *Doob decomposition Theorem*, stated below:

Theorem 1.2.5 *(Doob decomposition Theorem) Let* $\{X_n\}$ *be a submartingale with respect to a filtration* $(\mathcal{F}, n \geq 0)$. *There exists a unique random sequence* $\mathbf{Y} = (Y_0, Y_1, \ldots)$ *and* $\mathbf{V} = (V_0, V_1, \ldots)$ *such that*

1. *for* $n \geq 0, \quad X_n = Y_n + V_n$;

2. $\mathbf{Y}$ *is a martingale with respect to filtration* $(\mathcal{F}_n, n \geq 0)$;

3. $0 = V_0 \leq V_1 \leq V_2 \leq \ldots$;

4. *for all* $n > 0$, V_n *is measurable with respect to* $\mathcal{F}_{n-1}$.

In the estimation of renewal density in chapter 2, we will make use of the *reverse martingale convergence theorem* which we will now state.

Theorem 1.2.6 *(Reverse Martingale Convergence Theorem) Let* $\{Z_n : n \geq 0\}$ *be a reverse martingale with respect to reverse filtration* $(\mathcal{G}_n, n \geq 0)$. *Then there exists a random variable* Z_∞ *such that the sequence* $\{Z_n : n \geq 0\}$ *converges to* Z_∞ *a.s.. More over* Z_∞ *is* $\mathcal{G}_\infty$ *measurable where* $\mathcal{G}_\infty = \bigcap_{n=1}^{\infty} \mathcal{G}_n$.

1.2.4 Renewal Processes

Renewal theory began as the study of some particular probability problems connected with the failure and replacement of components, such as electric bulbs. Later it became clear that essentially the same problems arise also in connection with many other applications of probability theory and moreover that the fundamental mathematical theorems are of intrinsic interest in the theory of probability. Hence much recent work in the subject is not specifically connected with the replacement of components. Suppose that we have a population of *components*, each component charecterised by a non-negative random variable, X, called its *failure time.* It is best think of the components as physical objects such as electric bulbs, valves, etc., the failure time being the age of the component at which some clearly defined event called *failure* occur.

The random variable, X, is non-negative and there are in practice two main cases to consider.

1. The only possible values of X are $\{0, h, 2h, \ldots\}, h > 0$;
2. The random variable has an (absolutely) continuous distribution over the range $(0, \infty)$, its distribution being determined by a probability density function(PDF).

Case (1) is that of renewal theory in discrete time and have been discussed carefully by Feller(1957, Chapter 13). In chapter 2 of the thesis we shall consider the renewal theory of the case(2). Moreover the failure times $X_1, X_2, \ldots$ of different components will be assumed mutually independent.

Suppose that we have a population of components and that failure time, X_1, is a continuous random variable with PDF $f(x)$. Suppose that we start with a new component at time zero. This component fails at time X_1, say. Let it be replaced by a new component with failure time, X_2, say. Then the second failure will occur after time $X_1 + X_2$. Let this process be continued, a component being replaced immediately on failure by a new component. The failure time of the kth component used is X_k and the kth failure occurs at time S_k, where

$$S_k = X_1 + X_2 + \ldots + X_k \tag{1.2.4}$$

If $\{X_1, X_2, \ldots\}$ are independent and identically distributed random variables, with PDF.$f(x)$, we call the system an *ordinary renewal process.*

If in particular, the distribution is exponential with PDF. $\rho e^{-\rho x}$, we call the ordinary renewal process a *Poisson process* of rate ρ.

Following are definitions of two important functionals of our interest.

Definition 1.2.10 *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with absolute continuous distribution function F. Now define*

$$F^{(k)}(t) = P(S_k \leq t),$$

the $k-$fold convolution of F for $k \geq 1$. Define $N(t)$ as the biggest value of n for which $S_n \leq t$. The function

$$H(t) = E[N(t)]$$

is called renewal function.

Definition 1.2.11 *Consider for any time t, the function h(t), called renewal density, is defined by*

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0+} \frac{E(N(t, t+\Delta t))}{\Delta t} \\ &= H'(t) \end{aligned}$$

The renewal density specifies the mean number of renewals to be expected in a narrow interval near t. Its physical interpretation is that $h(t)\Delta t$ is asymptotically the chance of a renewal in the interval $(t, t+\Delta t)$. Mathematically, $h(t)$ is most easily calculated as the derivative of $H(t)$.

1.2.5 Probability Density Function

Probability density function(PDF) estimation is a fundamental step in statistics as it characterizes completely the behaviour of a random variable; see Silverman(1981),Silverman(1986) Prakash Rao(1983), Scott, D. W.(1992) and Natalia M.(2007). It is the construction of an estimate of the density function from the observed data. This provides a natural way to investigate the properties of a given data set, i.e. a realization of the random variable, and to carry out efficient data mining which consists of extraction of interesting informations or patterns from data in large databases. Density estimation plays an important role in finding out the shape of distributions of variables in the data/pattern analysis and information harvesting which are essential part of data mining. Histograms which are a type of density estimators, have been used for many years, both by statisticians in general and scientists in the social, physical and biological sciences as tools in drawing inferences from data. Recently a large class of non parametric methods have been developed for the estimation of distribution functions, density functions, etc., for data of several types. Density estimation is prominent part of non parametric estimation as it is applicable whether the data are periodic , vector valued, or not real. In these situations the distribution function may be meaningless or difficult to compute, and then one must estimate density. A brief survey of these methods is given in chapter 3.

Consider any random quantity X that has probability density function f. When we perform density estimation three alternatives can be considered. The first approach, known as *parametric* density estimation, assumes the data is drawn from a specific density model, for example the normal distribution with mean μ and variance σ^2.The density function f underlying data could be then estimated by finding estimates of μ and σ^2 from the data and substituting these estimates into the formula for the normal density. Unfortunately, an a-priori choice of the PDF model is in practice not suited since it might provide a false representation of the true PDF. An alternative is to build *non-parametric* PDF estimators in order to 'let the data speak for themselves'. Silverman(1981) explains that density estimates are of use in all three stages of statistical treatment of data, namely, exploratory, confirmatory and presentational. At the exploratory stage, density estimators give an indication of multi modality, skewness or dispersion, etc., of the data. For confirmatory purpose they can be used, for instance in non-parametric discriminant analysis (Prakash Rao,1983). For data presentation density estimators are the best information -presentation transformations given data; for instance histograms have been of wide use because they are easy to understand.

1.2.6 Counting Processes

A *counting process* over the half line $[0, \infty)$ may be thought of as a stochastic process recording, at any given time, the number of certain events having occurred before time t. Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a probability space with a filtration; that is $(\Omega, \mathcal{F}, P)$ is a usual probability space and $(\mathcal{F}_f)_{t \geq 0}$ is a family of sub $\sigma-$ algebras of F such as $\mathcal{F}_s \subset \mathcal{F}_t$ when $s \leq t$. A stochastic process $X = \{X(t), t \geq 0\}$ defined on $(\Omega, \mathcal{F})$ is adapted to $(\mathcal{F}_t)$ if $X(t)$ is $\mathcal{F}_t$- measurable.

Definition 1.2.12 *A counting process on a filtered probability space* $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ *is an adapted stochastic process* $N = \{N(t), t \geq 0\}$, *where each* $N(t)$ *takes values in* $\{0, 1, 2, \ldots \infty\}$ *with* $P[N(0) = 0] = 0$ *and such that almost all paths are non-*

decreasing and right continuous everywhere.

In general the increments of a counting process may be neither independent nor identically distributed. However, for large class of counting processes a conditional version of the intensity exists and it is called the stochastic intensity of the counting process. For details see Bremaud, P.(1981) and Karr, A.F.(1991). In this work the counting process will be denoted by $\{N(t) \geq 0\}$ and its stochastic intensity by $\Lambda(t)$. Sufficient conditions for the existence of $\Lambda(t)$ are discussed in Bremaud, P.(1981). Here we will consider only point processes for which the stochastic intensity is well defined.

Since N is increasing and hence a sub-martingale, it follows from the Doob-Meyer decomposition that $N = \Lambda + M$, where Λ is a predictable increasing process and M is a martingale. Since the stochastic intensity exists, we can assume that there exists non negative left continuous process $\Lambda(t)$, adapted to F_t, with right hand limits such that

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

Thus

$$N(t) = \int_0^t \lambda(s)ds + M(t)$$

or

$$M(t) = N(t) - \int_0^t \lambda(s)ds \tag{1.2.5}$$

where $M(t)$ is a square integrable martingale with variation process

$$< M, M > (t) = \int_0^t \lambda(s)ds.$$

We further assume that

$$\Lambda(t) = \alpha(t)Y(t), \tag{1.2.6}$$

where α is an unknown deterministic non-negative function called the intensity and $Y(t)$ is an observable stochastic process. Such models are usually called *multiplicative intensity models*; Aalen, O.(1978).

Multiplicative intensity model has many applications in the analysis of biological and medical data. By applying multiplicative intensity models and stochastic integrals, Aalen, O.(1976,1978) has shown how it is possible to develop non-parametric estimators for certain cumulative intensities.

1.3 Summary of Thesis

This section gives a brief summary of the thesis. In chapter 2 of the thesis we introduced two different non-parametric estimators for the *renewal density function*, $h(t)$ of a renewal process. The renewal density specifies the mean number of renewals to be expected in a narrow interval near t. Its physical interpretation is that $h(t)\Delta t$ is asymptotically the chance of a renewal in the interval $(t, t + \Delta t)$. Mathematically, $h(t)$ is most easily calculated as the derivative of the renewal function $H(t)$. Non-parametric estimation of the renewal density function is carried out using the method of histogram and then is generalized using a *kernel.* We suggest a non-parametric histogram estimator for the renewal density based on a random sample of size n. The almost sure consistency of the histogram estimator is established. The histogram estimator is then modified into a more refined estimator using kernel. Almost sure consistency as well as asymptotic normality of kernel estimator are discussed. A simulation study has been conducted to show how close are the estimators to the true renewal density functions in some standard renewal models.

In chapter 3, we introduced a novel non-parametric estimator of the *probability density function.* Probability density function estimation is a fundamental step in statistics as it characterizes completely the behavior of a random variable. It provides a natural way to investigate the properties of a given data set, i.e. a realization of the random variable, and to carry out efficient data mining. *Kernel density estimation* is the most popular technique used in statistics and data analysis to construct a smooth estimate of a density function from observed data. A limitation of kernel estimator is that it depends on the choice of the *smoothing kernel* and the *window width.* In various applications different smoothing kernels are recommended. One has to choose among *normal(Guasian)* , *box, triangle, Epanechnikov* and many others.

In the proposed technique use of a sequence of polynomials, namely

$$Q_d(x) = c_d(1 - x^2)^d, d = 1, 2, \cdots ,$$

is suggested for estimation. Here d is the *smoothing parameter.* This estimator is motivated from the proof of Stone-Weistrass theorem (Rudin W.(1976)), in which a continuous function on $[a, b]$ is approximated by the limit of a sequence of polynomials. We use the observed data to create such a polynomial that is close to the unknown density. Asymptotic properties like consistency and asymptotic normality of the proposed estimator are proved. An objective criterion for choosing the optimum value of the smoothing parameter d is proposed using *likelihood cross validation technique.* This approach is shown to be an excellent data-driven method of selecting the smoothing parameter of the proposed estimator. Unlike many other implementations, this method is immune to problems caused by multi-modal densities with widely separated modes. This is illustrated through simulation studies. Some data analyses are also done.

In chapter 4 we develop a non-parametric estimator of the intensity of a counting process. In some fields of science, data frequently consists of counts of the number of transitions between different statuses, such as the number of deaths or failures, the number of disablement or recoveries, or more generally the number of transitions between two states in a Markov chain. These counts may be subjected to various kinds of censoring. Even under very general censoring pattern, the number of such transitions observed may be described as a counting process. Multiplicative intensity model introduced by Aalen,(1978) has many applications in the analysis of biological and medical data. Specifically the multiplicative intensity model is the statistical model for counting process observes on a time interval for which the stochastic intensity admits the decomposition into a functional deterministic factor $\alpha(t)$ and a predictable stochastic process $Y(t)$. Aalen(1976,1978) obtained non-parametric estimators for certain cumulative intensities. Ramlau-Hansen (1983a) proposed a *kernel estimator* for the intensity α of a counting process by smoothing the martingale estimator of the cumulative intensity.

We discussed the method of constructing estimator of α using orthogonal series. Computation formulae for the estimator using the Fourier series is suggested. Apart from showing asymptotic unbiasedness, other properties of estimators such as consistency and asymptotic normality are also established under suitable conditions. Some discussions about the optimum choice of the smoothing parameter of the proposed estimator are included. An objective criterion for the choice of smoothing parameter is also suggested. Simulation studies are carried out for investigating the performance of the proposed estimator. Finally a data analysis is also included for dictating the practical use of the proposed estimator.

1.4 Extensions and Further research

In the renewal density estimation using kernel it remains to see how the kernel bandwidth p can be chosen in an optimum fashion. We had suggested only an ad hock procedure. Specifically we want to see whether a cross validation technique can be successfully employed.

Non parametric estimation of functions is an area of intrinsic interest in data mining and neural network; Brian D. Ripley(1996). In most of the applications kernel based are used in practise. We have suggested in chapter 3 a sequence of polynomials for density estimation is a potential tool that can be extended. We want to see how this method can be used in data mining and neural network as a powerful alternative to kernel methods. Its advantages (or limitations) in comparison with kernel based techniques are to be examined. Non parametric regression is yet another broad area of where such an estimation procedure can be attempted.

In chapter 4 we have suggested orthogonal series estimator for intensity function of a counting process. A procedure for the choice of optimum number of Fourier coefficients is also suggested. It of interest to see how some elementary cross validation technique can be employed here.

Chapter 2

Renewal Density Estimation

2.1 Introduction

Suppose that we have a population of components and that failure time, X, is a continuous random variable with PDF $f(x)$. Suppose that we start with a new component at time zero. This component fails at time X_1, say. Let it be replaced by a new component with failure time, X_2, say. Then the second failure will occur after time $X_1 + X_2$. Let this process be continued, a component being replaced immediately on failure by a new component. The failure time of the kth component used is X_k and the kth failure occurs at time S_k, where

$$S_k = X_1 + X_2 + \ldots + X_k \tag{2.1.1}$$

If $\{X_1, X_2, \ldots\}$ are independently identically distributed random variables, all with PDF $f(x)$, we call the system an *ordinary renewal process.*

If in particular, the distribution is exponential with PDF $\rho e^{-\rho x}$, we call the ordinary renewal process a *poisson process* of rate ρ.

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with absolute continuous distribution function F. Now define

$$F^{(k)}(t) = P(S_k \leq t),$$

the $k-$fold convolution of F for $k \geq 1$. Define $N(t)$ as the biggest value of n for which $S_n \leq t$. The function

$$H(t) = E[N(t)]$$

is called *renewal function.* In order to study $N(t)$, the number of renewals in the interval $(0, t)$, it is simplest to use the connection between $N(t)$ and the random variable S_k, the time up to the kth renewal. It is evident from the definition of the random variables $N(t)$ and S_k that

$$N(t) < k \text{ if and only if } S_k > t$$

Therefore

$$\begin{aligned} P(N(t) < k) &= P(S_k > t) \\ &= 1 - F^{(k)}(t) \end{aligned} \qquad (2.1.2)$$

Hence

$$P(N(t) = k) = F^{(k)}(t) - F^{(k+1)}(t) \tag{2.1.3}$$

The renewal function H can be, therefore, expressed in the following way:

$$\begin{aligned} H(t) &= \sum_{k=1}^{\infty} kP(N(t) = k) \\ &= \sum_{k=1}^{\infty} k(P(N(t) \le k) - P(N(t) \le k+1)) \\ &= \sum_{k=1}^{\infty} k(F^{(k)}(t) - F^{(k+1)}(t)) \\ &= \sum_{k=1}^{\infty} F^{(k)}(t), \text{ for } t > 0. \end{aligned} \tag{2.1.4}$$

Estimation of the renewal function is discussed by many authors. Most of the non-parametric estimators of $H(t)$ are based on a realization of the renewal process and on theorems , which yield simple approximations of $H(t)$ for asymptotically large values of t. For example , the result

$$\lim_{t\to\infty} H(t) - \frac{t}{\mu} = \frac{\sigma^2 + \mu - \mu^2}{2\mu^2} \tag{2.1.5}$$

suggests an estimator

$$\hat{H}(t) = \frac{t}{\hat{\mu}} + \frac{\hat{\sigma}^2 + \hat{\mu} - \hat{\mu}^2}{2\hat{\mu}^2}. \tag{2.1.6}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are estimators of μ and σ^2 based on the data recorded up to time t; Cox D. and Lewis P.(1966). Frees E.(1986) pointed out that estimators of the type

(2.1.6) do not perform well for small values of t. Frees, E.W. 1986) suggested a non-parametric estimator for $H(t)$ in the following way. Since $H(t)$, defined in(2.1.4) is the infinite sum of convolution of F, he defined a non-parametric estimator as

$$H_n(t) = \sum_{k=1}^{m} F_n^{(k)}(t) \tag{2.1.7}$$

where

$$F_n^{(k)}(t) = \frac{1}{{}^nC_k} \sum_c I(X_{i1} + X_{i2} + \ldots X_{ik} \leq t) \tag{2.1.8}$$

with $\sum_c$ denotes the sum over all ${}^nC_k$ distinct combination of $\{i_1, i_2, \ldots, i_k\}$, a subset of size k of $\{1, 2. \ldots, n\}$. Here $I(.)$ is the indicator function of a set and m is a function of n in the sense that $m \leq n$ and $m \to \infty$ as $n \to \infty$. The almost sure (a.s.)consistency of the estimator was proved by establishing the fact that $H_n(t)$ is a reverse martingale with respect to an appropriate sequence of σ-fields plus sum negligible terms.

2.2 Renewal Density Function

Consider for any time t, the renewal density function

$$h(t) = \lim_{\Delta t \to 0+} \frac{E(N(t, t + \Delta t))}{\Delta t}$$

which is the derivative $H'(t)$ of the renewal function. The renewal density specifies the mean number of renewals to be expected in a narrow interval near t. Since the random variables $\{X_k\}$ are continuous, with no concentration of probability at zero failure time, the probability of more than one failure in an interval of length Δt is $O(\Delta t)^2$. It follows that

$$h(t) = \lim_{\Delta t \to 0+} \frac{prob\{\text{one or more renewals in } (t, t+\Delta t)\}}{\Delta t} \tag{2.2.9}$$

An alternative interpretation of $h(t)$ is that if we have a very large number m of independent renewal processes in operation simultaneously, $mh(t)\Delta t$ is the number of renewals in the time interval $(t, t+\Delta t)$. Its physical interpretation is that $h(t)\Delta t$ is asymptotically the chance of a renewal in the interval $(t, t+\Delta t)$. Mathematically, $h(t)$ is most easily calculated as the derivative of $H(t)$. Therefore, for example,for an equilibrium renewal process (Cox D.R.(1961)), the renewal density function can be computed from the renewal function $H_0(t) = \dfrac{t}{\mu}$ as $h_0(t) = \dfrac{1}{\mu}$

From (2.1.4), the renewal density h admits the form

$$h(t) = \sum_{k=1}^{\infty} f^{(k)}(t), \text{ for } t > 0. \tag{2.2.10}$$

where

$$f^{(k)}(t) = \frac{d}{dt} F^{(k)}(t).$$

The renewal function and renewal density play an important role in warranty analysis but, unfortunately, there are only a few specified situations where analytical expressions can be obtained for $H(t)$ and $h(t)$. Estimation of the expected cost of a warranty for a stochastically failing unit is closely tied to estimation of the renewal function. The renewal function and renewal densities are basic tools also used in probabilistic models arising in other areas such as reliability theory, inventory theory, and continuous sampling plans. Blischke, W.R. and Murthy, D.R.P.(1996) survey asymptotic results, bounds, approximations numerical integrations and simulation procedures that have been used to give useful estimates of $H(t)$.

In this chapter non-parametric estimation of the renewal density function is carried out using the method of histogram and then generalized it using the kernel. In section 2 we suggest a non-parametric histogram estimator based on a random sample of size n. The almost sure consistency of the histogram estimator is established in section 3. In section 5 we introduced the estimator using kernel. Almost sure consistency as well as asymptotic normality of kernel estimator are discussed in sections. A simulation study has been conducted in the last section to show how close are the estimators to the true renewal density functions in some standard renewal models.

2.3 The Histogram Estimator

Histogram estimators were successfully employed in the non-parametric estimation of probability density function by Prakasa Rao, B.L.S.(1983) as well as in the estimation of the intensity of a counting process by Leskow J. and Razanzki R. (1983). Natalia M. Markovich(2007) proposed a histogram-type estimate of the renewal function and studied the convergence properties of the estimate. In this section we propose a histogram estimator for the renewal density function.

Assume that $t \in [0,1]$ and define a step function $\tilde{h}$, by

$$\tilde{h}(t) = \sum_{j=1}^{p} a_j I_{A_j}(t)$$

where

$$\begin{aligned} A_j &= \Big[\frac{j-1}{p}, \frac{j}{p}\Big), j = 1, 2, \ldots p-1 \\ A_p &= \Big[\frac{p-1}{p}, 1\Big] \text{ and} \\ a_j &= p \int I_{A_j} h(s) ds \end{aligned} \tag{2.3.11}$$

Selection of a_j is justified in the sense that

$$|h - \tilde{h}|^2 = \int (h(s) - \tilde{h}(s))^2 ds$$

is minimized.

To see this

$$\frac{\partial}{\partial a_k} \int |h - \tilde{h}|^2 = \frac{\partial}{\partial a_k} \int \Big(h(s) - \sum_{j=1}^{p} a_j I_{A_j}(s)\Big)^2 ds = 0$$

implies

$$\begin{aligned}\int_{A_j}\Big((h(s)-\sum_{j=1}^{p}a_jI_{A_j}(s)\Big)ds &= 0\\ \int_{A_j}h(s)ds-\int_{A_j}a_kds &= 0\\ \int I_{A_k}h(s)ds &= a_kl(A_k),\end{aligned}$$

where $l(A_k)$ the length of the interval A_k is $\frac{1}{p}$. Therefore

$$a_k = p\int I_{A_k}h(s)ds.$$

Now we can estimate a_j using

$$\begin{aligned}\hat{a}_j &= p\int I_{A_j}dH_n(s)\\ &= p\int I_{A_j}d\Big(\sum_{k=1}^{m}F_n^{(k)}(s)\Big)\end{aligned} \tag{2.3.12}$$

Thus an estimator of h can be defined by

$$h_{np}(t) = \sum_{j=1}^{p}\Big(\sum_{k=1}^{m}p\int I_{A_k}dF_n^{(k)}(s)\Big)I_{A_j}(t) \tag{2.3.13}$$

In the sequel we impose the following assumptions

1. The function h fulfill Holder condition with index set $\alpha, 0 < \alpha < 1$; that is there exists a constant c such that $|h(t) - h(s)| \leq c|t-s|^{\alpha}$.

2. The sequence p is chosen as an increasing function of m such that

$$\lim_{m} p\sum_{k>m}F^{(k)}(s) = 0$$

3. For each $k \geq 1$ define $m^{-1}(k) = \inf\{n : m(n) \geq k\}$, then

$$\sum_{k=1}^{\infty}(m^{-1}(k) - k)f^{(k)}(s_k) < \infty$$

whenever, s_k, $k = 1, 2, \ldots$ satisfies $|s_j - s_j| < \frac{1}{p}$.

Remark 2.3.1 *The assumption 1 is used to exclude drastically behaving functions from the parameter space. Assumption 2 is imposed in order to get the convergence of h_{np} towards h at a specific rate which is used while establishing the asymptotic properties. Notice that the assumption 3 is immediately satisfied when we choose $m = n$.*

2.4 Convergence of Histogram Estimator

In this section we shall establish, under some conditions, that the histogram estimator h_{np} is consistent.

Theorem 2.4.1 *Suppose that assumptions 1 and 2 are satisfied, then the estimator $h_{np}(t)$ converges in probability to $h(t)$. If in addition assumption 3 holds, then this convergence turns out to be almost sure convergence.*

To prove this theorem we we first consider the parameter $\tilde{h}$ which is close enough to the renewal density h. Then we shall show that the proposed estimator converges to the parameter $\tilde{h}$. Let us first consider the following lemma.

Lemma 2.4.1

$$\tilde{h}(t) \longrightarrow h(t) \ as \ p \longrightarrow \infty$$

Proof: For given $t \in [0,1]$, choose l such that $t \in A_l$, then

$$\begin{aligned} \tilde{h}(t) &= \sum_{j=1}^{p} a_j I_{A_j}(t) \\ &= p \int I_{A_l} h(s) ds \end{aligned} \tag{2.4.14}$$

and therefore

$$\begin{aligned} |h(t) - \tilde{h}(t)| &= \left| h(t) - p \int I_{A_l} h(s) ds \right| \\ &= \left| p \int I_{A_l} (h(t) - h(s)) ds \right| \\ &\leq p \int_{A_l} c|t-s|^{\alpha} ds \\ &= \frac{c}{p^{\alpha}} \longrightarrow 0 \text{ as } p \longrightarrow \infty \end{aligned} \tag{2.4.15}$$

Proof of the lemma is now complete.

Proof of the Theorem : Once again we choose l such that $t \in A_l$, then the histogram estimator can be written as

$$h_{np} = \sum_{k=1}^{m} p \int I_{A_l} dF_n^{(k)}(u). \tag{2.4.16}$$

Let $\{X_{1n}, X_{2n}, \ldots, X_{nn}\}$ be the order statistics associated with $\{X_1, X_2, \ldots, X_n\}$. We use $\chi_n = \sigma(X_{1n}, X_{2n}, \ldots, X_{nn}, X_{n+1}, X_{n+2}, \ldots), n \geq 1$ to denote the sequence of non-increasing sub σ-fields.

Define

$$RM_n = \int pI_{A_l} \sum_{k=1}^{n} dF_n^{(k)}(u) + \int pI_{A_l} d\Big(\sum_{k\geq n} I(S_k \leq u)\Big) - \int pI_{A_l} \sum_{k=1}^{\infty} dF^{(k)}(u)$$

We shall show that RM_n is a zero mean reverse martingale.

Note that

$$E(I(S_k \leq s)/\chi_n) = \begin{cases} F_n^{(k)}(s), & \text{if } k \leq n \\ I(S_k \leq s), & \text{if } k > n \end{cases} \tag{2.4.17}$$

and

$$E(F_n^{(k)}(u)/\chi_{n+1}) = F_{n+1}^{(k)}(u).$$

Therefore, by an application of dominated convergence theorem and Fubini's theorem, we have

$$\begin{aligned} E(RM_n/\chi_{n+1}) &= \int pI_{A_l} d\sum_{k=1}^{n} E(F_n^{(k)}(u)/\chi_{n+1}) + \int pI_{A_l} d\sum_{k>n} E(I(S_k \leq u)/\chi_{n+1}) \\ &\quad - \int pI_{A_l} d\sum_{k=1}^{\infty} F^{(k)}(u) \\ &= \int pI_{A_l} d\sum_{k=1}^{n} F_{n+1}^{(k)}(u) + \int pI_{A_l} dF_{n+1}^{(n+1)}(u) \\ &\quad + \int pI_{A_l} d\sum_{k>n+1} I(S_k \leq u) - \int pI_{A_l} d\sum_{k=1}^{\infty} F^{(k)}(u) \\ &= \int pI_{A_l} d\sum_{k=1}^{n+1} F_{n+1}^{(k)}(u) + \int pI_{A_l} d\sum_{k>n+1} I(S_k \leq u) - \int pI_{A_l} d\sum_{k=1}^{\infty} F^{(k)}(u) \\ &= RM_{n+1}. \end{aligned}$$

Also

$$\begin{aligned}
E(RM_n) &= \int pI_{A_l} d\sum_{k=1}^{n} E(F_n^{(k)}(u) + \int pI_{A_l} d\sum_{k>n} E(I(S_k \le u)) - \int pI_{A_l} d\sum_{k=1}^{\infty} F^{(k)}(u) \\
&= \int pI_{A_l} d\sum_{k=1}^{n} F^{(k)}(u) + \int pI_{A_l} d\sum_{k>n} F^{(k)}(u) - \int pI_{A_l} d\sum_{k=1}^{\infty} F^{(k)}(u) \\
&= \int pI_{A_l} d\sum_{k=1}^{\infty} F^{(k)}(u) - \int pI_{A_l} d\sum_{k=1}^{\infty} F^{(k)}(u) \\
&= 0,
\end{aligned}$$

since

$$E(I(S_k \le u) = F^{(k)}(u).$$

This shows that RM_n is a zero mean reverse martingale.

Define

$$Z_{1n} = \int I_{A_l} p \sum_{k=m+1}^{n} dF_n^{(k)}(u) \tag{2.4.18}$$

and

$$Z_{2n} = \int I_{A_l} p \sum_{k \ge n} I(S_k \le u). \tag{2.4.19}$$

Now the reverse martingale

$$\begin{aligned}
\int pI_{A_l} \sum_{k=1}^{n} dF_n^{(k)}(s) \quad &+ \quad \int pI_{A_l} (\sum_{k>n} I(S_k \le s)) \\
&= \int pI_{A_l} \sum_{k=1}^{m} dF_n^{(k)}(s) + Z_{1n} + Z_{2n},
\end{aligned}$$

has the expectation

$$\int pI_{A_l} \sum_{k=1}^{\infty} dF^{(k)}(s).$$

Therefore as a consequence of Doob's (reverse) martingale theorem we have

$$\begin{aligned}
\lim_n\Big(\int pI_{A_l}\sum_{k=1}^{m} dF_n^{(k)}(u)+Z_{1n}+Z_{2n}\Big) &= p\int I_{A_l}\sum_{k=1}^{\infty} dF^{(k)}(u)\\
&= p\int I_{A_l}dH(u)\\
&= p\int I_{A_l}h(u)du\\
&= \tilde{h}(t)\text{a.s.} \qquad (2.4.20)
\end{aligned}$$

Now allowing $p\to\infty$, then by virtue of Lemma (2.4.1), we have

$$\int I_{A_l}p\sum_{k=1}^{m} F_n^{(k)}(u)+Z_{1n}+Z_{2n}\to h(t).$$

Since Z_{2n} is monotone and bounded, $\lim Z_{2n}$ exists. By Fatou's lemma and assumption 2,

$$\begin{aligned}
E(\lim Z_{2n}) &\le \lim pE\Big[\int_{A_l} d\Big(\sum_{k>n} I(S_k\le u)\Big)\Big]\\
&= \lim p\int I_{A_l}\sum_{k>n} dF^{(k)}(u)\\
&= \lim p\sum_{k>n}\Big[F^{(k)}\Big(\frac{l}{p}\Big)-F^{(k)}\Big(\frac{l-1}{p}\Big)\Big]=0
\end{aligned}$$

Since $\lim Z_{2n}$ is non-negative and has non-positive expectation, it is zero a.s.

Also by the Markov inequality, for $\epsilon > 0$,

$$\begin{aligned} P(Z_{1n} \geq \epsilon) &\leq \epsilon^{-1} p \int I_{A_l} \sum_{k=m+1}^{n} dF^{(k)}(u) \\ &= \epsilon^{-1} p \sum_{k=m+1}^{n} \Big[F^{(k)}\Big(\frac{l}{p}\Big) - F^{(k)}\Big(\frac{l-1}{p}\Big)\Big] \\ &= \epsilon^{-1} \Big[p \sum_{k=m+1}^{n} F^{(k)}\Big(\frac{l}{p}\Big) - p \sum_{k=m+1}^{n} F^{(k)}\Big(\frac{l-1}{p}\Big)\Big] \\ &\to 0, \end{aligned}$$

by assumption 2. Thus $Z_{1n} \to 0$ in probability. Hence $h_{np}(t)$ converges to $h(t)$ in probability.

To prove almost sure convergence we need only show that

$$\limsup Z_{1n} = 0, a.s.$$

By Markov inequality, mean value theorem and assumption 3

$$\begin{aligned} \sum_{n=1}^{\infty} P(Z_{1n} \geq \epsilon) &\leq \epsilon^{-1} \sum_{n=1}^{\infty} p \int I_{A_l} \sum_{k=m+1}^{n} dF^{(k)}(u) \\ &= \epsilon^{-1} \sum_{n=1}^{\infty} (m^{-1}(k) - k) \Big[\frac{F^{(k)}\big(\frac{l}{p}\big) - F^{(k)}\big(\frac{l-1}{p}\big)}{\frac{1}{p}} \Big] \\ &= \epsilon^{-1} \sum_{n=1}^{\infty} (m^{-1}(k) - k) f^{(k)}(S_k) \\ &< \infty. \end{aligned}$$

This completes the proof.

Remark 2.4.1 *The choice of the design parameters m and p are dictated by practical considerations. Assumptions 1 and 2 give some theoretical guidelines* *for the choice. However the convolution $F^{(k)}(t)$ dies out quickly as k approaches infinity and typically m can be small compared to sample size. Similarly choice of p must be smaller than m.*

In the next section we shall introduce an estimator of renewal density using kernel.

2.5 The Kernel Estimator

Kernel estimators were successfully employed in the non parametric estimation of a probability density function; Silverman,B.W.(1986). Kernels are also used for the estimation of hazard rate (Ramlau-Hansen, H. (1983a)).

We define the kernel estimator h_{np} of the renewal density function h by

$$h_{np}(t) = \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dH_n(t) \qquad (2.5.21)$$

where $K(.)$ is the kernel which satisfy the condition

$$\int_{-\infty}^{\infty} K(s)ds = 1.$$

Usually, but not always, K will be a symmetric probability density function, like normal density, for instance. See Silverman, B.W.(1986) for a detailed discussion of various kernels. Here p is the smoothing parameter of the kernel and is called the *band width.*

Using the estimator $H_n(t) = \sum_{k=1}^{m} F_n^{(k)}(t)$, as given in the equation (2.1.7), the kernel estimator $h_{np}(t)$ can be expressed as follows

$$
\begin{aligned}
h_{np}(t) &= \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dH_n(s) \\
&= \frac{1}{p} \int_{-\infty}^{\infty} K\Big(\frac{t-s}{p}\Big) d\Big(\sum_{k=1}^{m} F_n^{(k)}(s)\Big) \\
&= \frac{1}{p} \sum_{k=1}^{m} \int_{-\infty}^{\infty} K\Big(\frac{t-s}{p}\Big) dF_n^{(k)}(s) \\
&= \frac{1}{p} \sum_{k=1}^{m} \frac{1}{{}^nC_k} \sum_{c} K\Big(\frac{t-(X_{i1}+X_{i2}+\ldots+X_{ik})}{p}\Big), \qquad (2.5.22)
\end{aligned}
$$

since

$$
F_n^{(k)}(t) = \frac{1}{{}^nC_k} \sum_{c} I(X_{i1} + X_{i2} + \ldots X_{ik} \leq t).
$$

The expression (2.5.22) is appropriate for the computation purposes.

2.6 Almost Sure Consistency

In this section we shall establish the convergence of the estimator $h_{np}(t)$ in probability to $h(t)$ under some conditions. By imposing an additional condition we strengthen the convergence to almost sure.

Consider the following conditions

A1: The function h is continuous.

A2: The sequence p is chosen as an increasing function of m such that

$$\int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d \sum_{k>m} F^{(k)}(s) \to 0 \text{ as } m \to \infty$$

A3: For each $k \geq 1$ define

$$m^{-1}(k) = \inf\{n : m(n) \geq k\},$$

then

$$\sum_{k=1}^{\infty} (m^{-1}(k) - k) f^{(k)}(s_k) < \infty,$$

whenever $s_k, k = 1, 2, \ldots$ satisfies $|s_i - s_j| < \dfrac{1}{p}$.

Theorem 2.6.1 *Suppose that assumptions A1 and A2 hold, then the estimator $h_{np}(t)$ converges in probability to $h(t)$. If in addition A3 holds then the convergence can be strengthened to almost sure convergence.*

To prove the theorem we first define a parameter $\tilde{h}$ which is close enough to the renewal density h. Then we shall show that the proposed estimator converges to the parameter $\tilde{h}$. Let us first consider the following lemma.

Lemma 2.6.1 *Let*

$$\tilde{h}(t) = \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) h(s) ds.$$

Then for sufficiently small value of the bandwidth p, $\tilde{h}(t)$ converges to $h(t)$. That is

$$\tilde{h}(t) \to h(t) \text{ as } p \to 0.$$

Proof of the Lemma 2.4.1: We have

$$\tilde{h}(t) = \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) h(s)ds.$$

Therefore

$$\begin{aligned} |h(t) - \tilde{h}(t)| &= \Big| h(t) - \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) h(s)ds \Big| \\ &= \Big| \int_{-\infty}^{\infty} h(t)K(u)du - \int_{-\infty}^{\infty} h(t-pu)K(u)du \Big| \\ &\leq \int_{-\infty}^{\infty} |h(t) - h(t-pu)| K(u)du \\ &\rightarrow 0 \text{ as } p \rightarrow 0. \end{aligned}$$

since

$$\int_{-\infty}^{\infty} K(u)du = 1$$

This proves the lemma.

Proof of the Theorem 2.4.1: Consider the estimator

$$h_{np}(t) = \sum_{k=1}^{m} \frac{1}{p} \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dF_n^{(k)}(s)$$

where

$$F_n^{(k)}(s) = \frac{1}{{}^nC_k} \sum_c I(X_{i1} + X_{i2} + \ldots + X_{ik} \leq s)$$

Let $\{X_{1n}, X_{2n}, \ldots, X_{nn}\}$ be the order statistics associated with $\{X_1, X_2, \ldots, X_n\}$. We use $\chi_n = \sigma(X_{1n}, X_{2n}, \ldots, X_{nn}, X_{n+1}, X_{n+2}, \ldots), n \geq 1$ to denote the sequence of non-increasing sub σ-field which are implicitly used in the following reverse martingales.

Consider

$$\begin{aligned} RM_n &= \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) \sum_{k=1}^{n} dF_n^{(k)}(s) + \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d \sum_{k>n} I(S_k \le s) \\ &\quad - \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) \sum_{k=1}^{\infty} dF^{(k)}(s). \end{aligned}$$

Now we shall show that RM_n is a zero mean reverse martingale.

Letting

$$g(s) = \frac{1}{p} K(\frac{t-s}{p})$$

we can express RM_n as

$$\begin{aligned} RM_n &= \int_{-\infty}^{\infty} g(s) \sum_{k=1}^{n} dF_n^{(k)}(s) + \int_{-\infty}^{\infty} g(s) d \sum_{k>n} I(S_k \le s) \\ &\quad - \int_{-\infty}^{\infty} g(s) d \sum_{k=1}^{\infty} F^{(k)}(s). \end{aligned}$$

Note that

$$E(I(S_k \le s)/\chi_n) = \begin{cases} F_n^{(k)}(s), & \text{if } k \le n \\ I(S_k \le s), & \text{if } k > n \end{cases} \tag{2.6.23}$$

Also

$$E(F_n^{(k)}(s)/\chi_{n+1}) = F_{n+1}^{(k)}(s).$$

Therefore

$$\begin{aligned}
E(RM_n/\chi_{n+1}) &= \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{n} E(F_n^{(k)}(s)/\chi_{n+1}) \\
&\quad + \int_{-\infty}^{\infty} g(s)d\sum_{k>n} E(I(S_k \le s)/\chi_{n+1}) - \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{\infty} F^{(k)}(s) \\
&= \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{n} F_{n+1}^{(k)}(s) + \int_{-\infty}^{\infty} g(s)dF_{n+1}^{(n+1)}(s) \\
&\quad + \int_{-\infty}^{\infty} g(s)d\sum_{k>n+1} I(S_k \le s) - \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{\infty} F^{(k)}(s) \\
&= \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{n+1} F_{n+1}^{(k)}(s) + \int_{-\infty}^{\infty} g(s)d\sum_{k>n+1} I(S_k \le s) \\
&\quad - \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{\infty} F^{(k)}(s) \\
&= RM_{n+1}.
\end{aligned}$$

Further

$$\begin{aligned}
E(RM_n) &= \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{n} E(F_n^{(k)}(s)) + \int_{-\infty}^{\infty} g(s)d\sum_{k>n} E(I(S_k \le s) \\
&\quad - \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{\infty} F^{(k)}(s) \\
&= \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{n} F^{(k)}(s) + \int_{-\infty}^{\infty} g(s)d\sum_{k>n} F^{(k)}(s) \\
&\quad - \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{\infty} F^{(k)}(s) \\
&= \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{\infty} F^{(k)}(s) - \int_{-\infty}^{\infty} g(s)d\sum_{k=1}^{\infty} F^{(k)}(s) \\
&= 0.
\end{aligned}$$

Hence RM_n is a reverse martingale.

Define

$$Z_{1n} \;=\; \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) \sum_{k=m+1}^{n} dF_n^{(k)}(s) \tag{2.6.24}$$

and

$$Z_{2n} \;=\; \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d\Big(\sum_{k>n} I(S_k \le s)\Big) \tag{2.6.25}$$

Therefore using (2.6.24) and (2.6.25) we have the reverse martingale:

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) \sum_{k=1}^{n} dF_n^{(k)}(s) \;&+\; \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d\Big(\sum_{k>n} I(S_k \le s)\Big) \\ &= \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) \sum_{k=1}^{m} dF_n^{(k)}(s) + Z_{1n} + Z_{2n},\end{aligned}$$

whose expectation is

$$\int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) \sum_{k=1}^{\infty} dF^{(k)}(s) = \tilde{h}(t)$$

Hence, Doob's reverse martingale theorem suggests that

$$\int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) \sum_{k=1}^{m} dF_n^{(k)}(s) + Z_{1n} + Z_{2n} \to \tilde{h}(t) \quad a.s.$$

Now by virtue of the Lemma (2.6.1) we have as $p \to 0$

$$\int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) \sum_{k=1}^{m} dF_n^{(k)}(s) + Z_{1n} + Z_{2n} \to h(t), \quad a.s. \tag{2.6.26}$$

As Z_{2n} is monotonic and bounded $\lim Z_{2n}$ exists. By Fatou's lemma and $A2$ we have

$$\begin{aligned} E(\lim Z_{2n}) &\leq \lim E \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d\{\sum_{k>n} I(S_k \leq s)\} \\ &= \lim \frac{1}{p} \int_{-\infty}^{\infty} K\Big(\frac{t-s}{p}\Big) d\{\sum_{k>n} F^{(k)}(s)\} \\ &= 0. \end{aligned}$$

Since $\lim Z_{2n}$ is non negative and has non positive expectation, it is zero a.s. That is

$$\lim Z_{2n} = 0, \quad a.s. \tag{2.6.27}$$

Again by Markov inequality and A2, for $\epsilon > 0$

$$\begin{aligned} P(Z_{1n} \geq \epsilon) &\leq \epsilon^{-1} \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d\{\sum_{k=m+1}^{n} F^{(k)}(s)\} \\ &\rightarrow 0. \end{aligned}$$

Therefore

$$\lim Z_{1n} = 0, \text{ in probability }. \tag{2.6.28}$$

Now from (2.6.26) and (2.6.27), we need only to show that $\limsup Z_{1n} = 0 \quad a.s.$

Again by the Markov inequality and a change of summation, for $\epsilon > 0$,

$$\begin{aligned} \sum_{n\geq 1} P(Z_{1n} > \epsilon) &\leq \epsilon^{-1} \sum_{n\geq 1} \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d\{\sum_{k=m+1}^{n} F^{(k)}(s)\} \\ &= \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d\{\sum_{k\geq 1} (m^{-1}(k) - k) F^{(k)}(s)\} \\ &< \infty \end{aligned}$$

by A3 and using the arguments in Frees, E.W.(1986), p. 1370. Therefore from Borel-Cantelli lemma,

$$Z_{1n} \to 0, \quad a.s. \tag{2.6.29}$$

Now the theorem follows from (2.6.26), (2.6.27) and (2.6.29).

2.7 Asymptotic Normality

In this section we shall study the asymptotic normality of the proposed estimator. This is done by applying the projection technique popularized by Hajek (Serfling(1980)). We have the estimator

$$\begin{aligned} h_{np}(t) &= \sum_{k=1}^{m} \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dF_n^{(k)}(s) \\ &= \sum_{k=1}^{m} f_{np}^{(k)}(t) \end{aligned} \tag{2.7.30}$$

where

$$f_{np}^{(k)}(t) = \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dF_n^{(k)}(s) \tag{2.7.31}$$

Define a truncated and smoothed version of $h(t)$ by

$$h_p^*(t) = \sum_{k=1}^{m} f_p^{(k)}(t) \tag{2.7.32}$$

where

$$f_p^{(k)}(t) = \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dF^{(k)}(s)$$

Now the projection of $h_{np}(t)$ on $h_p^*(t)$ is defined by

$$\hat{h}_{np}(t) = \sum_{j=1}^{n} E(h_{np}(t)/X_j) - (n-1)h_p^*(t) \tag{2.7.33}$$

Now we define an important covariance term required in this section by

$$\begin{aligned} \xi_{qr}(c) &= Cov[f_p^{(q-c)}(t-(X_1+X_2+\ldots+X_c)), \\ & \quad f_p^{(r-c)}(t-(X_1+X_2+\ldots+X_c))] \end{aligned} \tag{2.7.34}$$

where

$$f_p^{(k-c)}(t-(X_1+X_2+\ldots+X_c)) = \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-(X_1+X_2+\ldots+X_c)-s}{p}\Big) dF^{(k-c)}(s).$$

We are now able to prove the following theorem.

Theorem 2.7.1 *Assume that*

(i) $\sum_{k>m} f^k(t) \to 0$

(ii) $n \sum_{q,r=1}^{m} \frac{1}{{}^nC_q} \sum_{c=2}^{r} {}^rC_c{}^{n-r}C_{q-c}\xi_{qr}(c) \to 0$

(iii) $f^{(k)}(t)$ *satisfies the Holder's condition with some* $\alpha > 0$*; that is*

$$|f^{(k)}(t) - f^{(s)}(t)| \leq c|t-s|^{\alpha},$$

where c *is some constant and*

(iv) p *and* m *are chosen such that* $\sqrt{n}mp^{\alpha} \to 0$ *as* $n \to \infty$. *Then for each* $t > 0$

$$\sqrt{n}(h_{np}(t) - h(t)) \to N(0, \sigma_1^2)$$

where

$$\sigma_1^2 = \sum_{q,r=1}^{\infty} qr\xi_{qr}(1) < \infty.$$

To prove the theorem we consider some preparatory lemmas.

Lemma 2.7.1 $\sqrt{n}(\hat{h}_{np}(t) - h_p^*(t)) \to N(0, \sigma_1^2)$ *where*

$$\sigma_1^2 = \sum_{q,r=1}^{\infty} qr\xi_{qr}(1) < \infty.$$

Proof: Note that the projection $\hat{h}_{np}(t)$ is just sum of independently and identically distributed random variables and the usual theorems for double arrays of independent random variables are used to obtain a limiting distribution of $\hat{h}_{np}(t)$.

Using (2.5.33), we have

$$\hat{h}_{np}(t)t - h_p^*(t) = \sum_{j=1}^{n} E[\sum_{k=1}^{m} \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dF_n^{(k)}(s)/X_j] - nh_p^*(t)$$

Now, using the definition of $F_n^{(k)}$ and $f_{np}^{(k)}$ we have

$$E(f_{np}^{(k)}(t)/X_1) = \frac{k}{n} f_p^{(k-1)}(t - X_1) + (1 - \frac{k}{n}) f_p^{(k)}(t), \qquad (2.7.35)$$

since probability of including X_1 among k X_i's is $\frac{k}{n}$.

Now, since

$$f_p^{(k)}(t) = \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dF^{(k)}(s)$$

and using (2.6.30) and (2.6.33) we have

$$
\begin{aligned}
\hat{h}_{np}(t) - h_p^*(t) &= \sum_{j=1}^{n}\sum_{k=1}^{m}\Big\{\frac{k}{n}\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(k-1)}(s) \\
&\quad +(1-\frac{k}{n})\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF^{(k)}(s)\Big\} - \sum_{j=1}^{n}\sum_{k=1}^{m}\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF^{(k)}(s) \\
&= \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{m}k\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(k-1)}(s) - \\
&\quad -\frac{k}{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF^{(k)}(s)
\end{aligned}
$$

That is

$$
\begin{aligned}
\hat{h}_{np}(t) - h_p^*(t) &= \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{m}k\Big[\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(k-1)}(s) \\
&\quad -\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF^{(k)}(s)\Big] \\
&= \sum_{j=1}^{n}U_{nj}
\end{aligned}
$$

where

$$
\begin{aligned}
U_{nj} &= \frac{1}{n}\sum_{k=1}^{m}k\Big[\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(k-1)}(s) \\
&\quad -\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF^{(k)}(s)\Big] \qquad (2.7.36)
\end{aligned}
$$

Clearly

$$E(U_{nj}) = 0.$$

Also, consider

$$
\begin{aligned}
& Var(\hat{h}_{np}(t)) \\
= \; & Var(\hat{h}_{np}(t) - h_p^*(t)) \\
= \; & Var(\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{m} k\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(k-1)}(s)) \\
= \; & n^{-2}nVar\Big[\sum_{k=1}^{m} k\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(k-1)}(s)\Big] \\
= \; & \frac{1}{n}\sum_{q,r=1}^{m} qrCov\Big(\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(q-1)}(s), \int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(r-1)}(s)\Big)
\end{aligned}
$$

using (2.6.34) and the fact that $h_p^*(t)$ is non-random. That is

$$Var[\hat{h}_{np}(t)] = \frac{1}{n}\sum_{q,r=1}^{m} qr\xi_{qr}(1), \qquad (2.7.37)$$

since, putting $c = 1$ in (2.6.32) we have

$$\xi_{qr}(1) = Cov\Big(\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(q-1)}(s), \int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_j-s}{p}\Big)dF^{(r-1)}(s)\Big)$$

Or in other words

$$Var(\sum_{j=1}^{n} U_{nj}) = \frac{1}{n}\sum_{q,r=1}^{m} qr\xi_{qr}(1) \qquad (2.7.38)$$

Note that$(U_{nj}, j = 1, 2, \ldots n; n \geq 1)$ is a double array of random variable that are independent and identically distributed within raws. Thus the condition

$$\sum_{q,r=1}^{\infty} qr\xi_{qr}(1) < \infty$$

is enough to satisfy the Lindberg condition and

$$\sqrt{n}(\hat{h}_{np}(t) - h_p^*(t)) \to N(0, \sigma_1^2). \tag{2.7.39}$$

This proves the lemma.

Lemma 2.7.2 *Assume that*

(i) $\sigma^2 < \infty$ *and (ii)* $n \sum_{q,r=1}^{m} \frac{1}{{}^nC_q} \sum_{c=2}^{r} {}^rC_c {}^{n-r}C_{q-c} \xi_{qr}(c) \to 0.$

Then

$$nE(h_{np}(t) - \hat{h}_{np}(t))^2 \to 0, \qquad as \qquad n \to 0.$$

Proof: Since

$$F_n^{(k)}(t) = \frac{1}{{}^nC_k} \sum_c I(X_{i1} + X_{i2} + \ldots X_{ik} \le t),$$

we have

$$\begin{aligned}
& Cov(f_{np}^{(q)}(t), f_{np}^{(r)}(t)) \\
= \; & Cov\Big(\int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dF_n^{(q)}(s), \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) dF_n^{(r)}(s)\Big) \\
= \; & Cov\Big(\int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d[\frac{1}{{}^nC_q} \sum_{\alpha} I(X_{a_1} + X_{a_2} + \ldots + X_{a_q} \le t)], \\
& \int_{-\infty}^{\infty} \frac{1}{p} K\Big(\frac{t-s}{p}\Big) d[\frac{1}{{}^nC_r} \sum_{\beta} I(X_{b_1} + X_{b_2} + \ldots + X_{b_r} \le t)]\Big)
\end{aligned}$$

where $\{a_1, a_2, \ldots, a_q\}$ and $\{b_1, b_2, \ldots, b_r\}$ be two subsets of $\{1, 2, \ldots, n\}$ with q and r elements. Let $c \le min(q, r)$ elements are common in these subsets. Now to compute

$$Cov(f_{np}^{(q)}(t), f_{np}^{(r)}(t)) = E[f_{np}^{(q)}(t) f_{np}^{(r)}(t)] - E[f_{np}^{(q)}(t)]E[f_{np}^{(r)}(t)],$$

we consider, for a fixed c,

$$\begin{aligned}
& E[f_{np}^{(q)}(t)f_{np}^{(r)}(t)] \\
= \;& E[\int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF_n^{(q)}(s) \int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF_n^{(r)}(s)] \\
= \;& E\{E[\int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF_n^{(q)}(s) \int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF_n^{(r)}(s)/X_1, X_2, \ldots, X_c]\} \\
= \;& E\{E\int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-(X_1+X_2+\ldots+X_c)-s}{p}\Big)dF_n^{(q)}(s-(X_1+X_2+\ldots+X_c)) \\
& \cdot \int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-(X_1+X_2+\ldots+X_c)-s}{p}\Big)dF_n^{(r)}(s-(X_1+X_2+\ldots+X_c))\} \\
= \;& E\{\int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-(X_1+X_2+\ldots+X_c)-s}{p}\Big)dF^{(q-c)}(s) \\
& \cdot \int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-(X_1+X_2+\ldots+X_c)-s}{p}\Big)dF^{(r-c)}(s)\} \\
= \;& E\{f_p^{(q-c)}(t-(X_{i_1}+X_{i_2}+\ldots+X_{i_q})) \\
& \cdot f_p^{(r-c)}(t-(X_{i_1}+X_{i_2}+\ldots+X_{i_r}))\} \qquad (2.7.40)
\end{aligned}$$

Now

$$\begin{aligned}
E[f_p^{(q-1)}(t-X_1)] &= E\int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-s-X_1}{p}\Big)dF^{(q-1)}(s) \\
&= E[\int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-s-X_1}{p}\Big)dE(I(X_{i_1}+X_{i_2}+\ldots+X_{i_{q-1}} \leq s))] \\
&= \int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-u}{p}\Big)dE(I(X_{i_1}+X_{i_2}+\ldots+X_{i_q} \leq u-X_1) \\
&= \int_{-\infty}^{\infty} \frac{1}{p}K\Big(\frac{t-u}{p}\Big)dF^{(q)}(u) \\
&= f_p^{(q)}(t)
\end{aligned}$$

Using similar arguments we get

$$E\Big(f_p^{(q)}(t)f_p^{(r)}(t)\Big)$$

$$= E[f_p^{(q-c)}(t-(X_{i_1}+\ldots+X_{i_q}))f_p^{(r-c)}(t-(X_{i_1}+\ldots+X_{i_c}))] \qquad (2.7.41)$$

Therefore using (2.7.40) and (2.7.41) we have, for given c,

$$\begin{aligned}
& Cov(f_{np}^{(q)}(t), f_{np}^{(r)}(t)) \\
= \; & E(f_{np}^{(q)}(t)f_{np}^{(r)}(t)) - E(f_{np}^{(q)}(t))E(f_{np}^{(r)}(t)) \\
= \; & E\{f_p^{(q-c)}(t-(X_{i_1}+X_{i_2}+\ldots+X_{i_q}))f_p^{(r-c)}(t-(X_{i_1}+X_{i_2}+\ldots+X_{i_r}))\} \\
& -f_{np}^{(q)}(t)f_{np}^{(r)}(t) \\
= \; & \xi_{qr(c)}. \qquad (2.7.42)
\end{aligned}$$

Again

$$\begin{aligned}
& Cov\Big(h_{np}(t), \hat{h}_{np}(t)\Big) \\
= \; & Cov(h_{np}(t), \hat{h}_{np}(t) - h_p^*(t)) \\
= \; & Cov\Big(\sum_{k=1}^{m} f_{np}^{(k)}(t), \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{m} k\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-s-X_j}{p}\Big)dF^{(k-1)}(s)\Big) \\
= \; & Cov\Big(\sum_{k=1}^{m}\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF_n^{(k)}(s), \sum_{k=1}^{m} k\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_1-s}{p}\Big)dF^{(k-1)}(s)\Big) \\
= \; & \sum_{q,r=1}^{m} rCov\Big(\int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-s}{p}\Big)dF_n^{(q)}(s), \int_{-\infty}^{\infty}\frac{1}{p}K\Big(\frac{t-X_1-s}{p}\Big)dF^{(r-1)}(s)\Big) \\
= \; & \sum_{q,r=1}^{m} rCov\Big(f_{np}^{(q)}(t), f_p^{(r-1)}(t-X_1)\Big) \qquad (2.7.43)
\end{aligned}$$

Now

$$
\begin{aligned}
& Cov(f_{np}^{(q)}(t), f_p^{(r-1)}(t-X_1)) \\
= \; & E[f_{np}^{(q)}(t) f_p^{(r-1)}(t-X_1)] - E[f_{np}^{(q)}(t)]E[f_p^{(r-1)}(t-X_1)] \\
= \; & E[E(f_{np}^{(q)}(t) f_p^{(r-1)}(t-X_1)/X_1)] - E[f_{np}^{(q)}(t)]E[f_p^{(r-1)}(t-X_1)] \\
= \; & E[f_p^{(r-1)}(t-X_1)\{\frac{q}{n} f_p^{(q-1)}(t-X_1) + (1-\frac{q}{n}) f_p^{(q)}(t)\}] \\
& -E[f_{np}^{(q)}(t)]E[f_p^{(r-1)}(t-X_1)]. \\
= \; & \frac{q}{n}\{f_p^{(r-1)}(t-X_1) f_p^{(q-1)}(t-X_1) - f_p^{(q)}(t) f_p^{(r-1)}(t-X_1)\} \\
& + f_p^{(q)}(t)E[f_p^{(r-1)}(t-X_1)] - f_p^{(q)}(t) f_p^{(r-1)}(t-X_1), \qquad (2.7.44)
\end{aligned}
$$

using the fact that

$$f_p^{(q)}(t) = E[f_{np}^{(q-1)}(t-X_1)].$$

Therefore

$$
\begin{aligned}
Cov(f_{np}^{(q)}(t), f_p^{(r-1)}(t-X_1)) &= \frac{q}{n} Cov(f_p^{(r-1)}(t-X_1), f_p^{(q-1)}(t-X_1)) \\
&= \frac{q}{n}\xi_{qr}(1), \text{from definition.} \qquad (2.7.45)
\end{aligned}
$$

Using (2.7.43) in (2.7.45) we get

$$Cov(h_{np}(t), \hat{h}_{np}(t)) = \sum_{q,r=1}^{m} \frac{rq}{n}\xi_{qr}(1) \qquad (2.7.46)$$

Now to calculate $Var(h_{np}(t))$ we examine the covariance between $f_{np}^{(q)}(t)$ and $f_{np}^{(r)}(t)$. For fixed $c = \min(r, q)$, we have

$$Cov(f_{np}^{(q)}(t), f_{np}^{(r)}(t)) = \xi_{qr}(c).$$

Therefore, for $r \leq q$,

$$\begin{aligned} Cov(f_{np}^{(q)}(t), f_{np}^{(r)}(t)) &= \frac{1}{{}^nC_r}\frac{1}{{}^nC_q}\sum_{c=1}^{r} {}^nC_r {}^rC_c {}^{n-r}C_{q-c}\xi_{qr}(c), \\ &= \frac{1}{{}^nC_q}\sum_{c=1}^{r} {}^rC_c {}^{n-r}C_{q-c}\xi_{qr}(c). \end{aligned}$$

since the number of distinct choices for two subsets of size r and q, respectively having exactly c elements in common is ${}^nC_r\ {}^rC_c\ {}^{n-r}C_{q-c}$.

Therefore

$$\begin{aligned} Var(h_{np}(t)) &= Var\sum_{k=1}^{m} f_{np}^{(k)}(t) \\ &= \sum_{q,r=1}^{m} Cov(f_{np}^{(q)}(t), f_{np}^{(r)}(t)) \\ &= \sum_{q,r=1}^{m} \frac{1}{{}^nC_q}\sum_{c=1}^{r} {}^rC_c {}^{n-r}C_{q-c}\xi_{qr}(c). \end{aligned} \quad (2.7.47)$$

Now consider

$$\begin{aligned} nE(h_{np}(t) - \hat{h}_{np}(t))^2 &= nVar(h_{np}(t) - \hat{h}_{np}(t)) \\ &= n[Var(h_{np}(t) + Var(\hat{h}_{np}(t)) - 2Cov(h_{np}(t), \hat{h}_{np}(t))] \\ &= n\sum_{q,r=1}^{m}[\frac{1}{{}^nC_q}\sum_{c=1}^{r} {}^rC_c {}^{n-r}C_{q-c}\xi_{qr}(c) + \frac{qr}{n}\xi_{qr}(1) - 2\frac{qr}{n}\xi_{qr}(1)] \\ &= \sum_{q,r=1}^{m}[n\frac{1}{{}^nC_q}\sum_{c=1}^{r} {}^rC_c {}^{n-r}C_{q-c}\xi_{qr}(c) - qr\xi_{qr}(1)], \end{aligned}$$

using (2.7.37),(2.7.46) and (2.7.47).

Now to show that $nE[h_{np}(t) - \hat{h}_{np}(t)]^2 \to 0$, it is sufficient to use the assumption:

$$n\sum_{q,r=1}^{m} \frac{1}{{}^nC_q}\sum_{c=2}^{r} {}^rC_c {}^{n-r}C_{q-c}\xi_{qr}(c) \to 0$$

and prove that

$$\sum_{q,r=1}^{m} [n\frac{1}{{}^{n}C_q}{}^{r}C_1{}^{n-r}C_{q-1} - qr]\xi_{qr}(1) \to 0$$

To prove this, consider

$$\begin{aligned} \left|n\frac{1}{{}^{n}C_q}{}^{r}C_1{}^{n-r}C_{q-1} - qr\right| &= \left|\frac{n(n-q)!}{n!}q!r\ {}^{m-r}C_{q-1} - qr\right| \\ &= \left|qr\left[\frac{{}^{n-r}C_{q-1}}{{}^{n-1}C_{q-1}} - 1\right]\right| \\ &\leq qr\left[\frac{{}^{n-r}C_{q-1}}{{}^{n-1}C_{q-1}} + 1\right] \\ &\leq 2qr \end{aligned}$$

since ${}^{n-r}C_{q-1} \leq {}^{n-1}C_{q-1}$ is always true.

Therefore

$$\left|\sum_{q,r=1}^{m} [n\frac{1}{{}^{n}C_q}{}^{r}C_1{}^{n-r}C_{q-1} - qr]\xi_{qr}(1)\right| \leq 2qr|\xi_{qr}(1)| < \infty,$$

since $\sigma_1^2 < \infty$.

Thus using dominated convergence theorem, we have

$$\sum_{q,r=1}^{m} [n\frac{1}{{}^{n}C_q}{}^{r}C_1{}^{n-r}C_{q-1} - qr]\xi_{qr}(1) \to 0.$$

Hence

$$nE[h_{np}(t) - \hat{h}_{np}(t)]^2 \to 0, \text{ as } n \to 0.$$

The lemma is now proved.

Lemma 2.7.3 *Suppose $f^{(k)}(t)$ satisfies the Holder's condition with some $\alpha > 0$. That is*

$$|f^{(k)}(t) - f^{(s)}(t)| \leq c|t-s|^{\alpha},$$

where c is some constant and p is chosen such that

$$\sqrt{n}mp^{\alpha} \to 0 \qquad as \qquad n \to \infty, then$$

$$\sqrt{n}\Big(h_p^*(t) - \sum_{k=1}^{m} f^{(k)}(t)\Big) \to 0, \qquad as \qquad n \to 0.$$

Proof: We have

$$\sqrt{n}\Big(h_p^*(t) - \sum_{k=1}^{m} f^{(k)}(t)\Big) = \sqrt{n}\Big(\sum_{k=1}^{m} \frac{1}{p}\int_{-\infty}^{\infty} K(\frac{t-s}{p}) f^{(k)}(s)ds - \sum_{k=1}^{m} f^{(k)}(t)\Big).$$

Therefore

$$\begin{aligned}
&\Big|\sqrt{n}\Big(\sum_{k=1}^{m} \frac{1}{p}\int_{-\infty}^{\infty} K(\frac{t-s}{p}) f^{(k)}(s)ds - \sum_{k=1}^{m} f^{(k)}(t)\Big)\Big| \\
\leq &\ \sqrt{n}\sum_{k=1}^{m}\Big|\frac{1}{p}\int_{-\infty}^{\infty} K(\frac{t-s}{p}) f^{(k)}(s)ds - f^{(k)}(t)\Big| \\
= &\ \sqrt{n}\sum_{k=1}^{m}\Big|\int_{-\infty}^{\infty} K(u)\Big(f^{(k)}(t-up) - f^{(k)}(t)\Big)\Big|du,
\end{aligned}$$

by applying the change of variable using the transformation $u = \dfrac{t-s}{p}$. Now, using the Holder's condition we have

$$\begin{aligned}
&\Big|\sqrt{n}\Big(\sum_{k=1}^{m} \frac{1}{p}\int_{-\infty}^{\infty} K(\frac{t-s}{p}) f^{(k)}(s)ds - \sum_{k=1}^{m} f^{(k)}(t)\Big) \\
\leq &\ \sqrt{n}\sum_{k=1}^{m}\int_{-\infty}^{\infty} K(u)\Big|f^{(k)}(t-up) - f^{(k)}(t)\Big|du \\
\leq &\ \sqrt{n}\sum_{k=1}^{m} cp^{\alpha}\int_{-\infty}^{\infty} K(u)du.
\end{aligned}$$

Applying the property

$$\int_{\infty}^{\infty} K(u)du = 1$$

for any kernel K we have

$$\begin{aligned}\sqrt{n}\Big(h_p^*(t) - \sum_{k=1}^{m} f^{(k)}(t)\Big) &\leq c\sqrt{n}mp^{\alpha} \\ &\to 0 \qquad as \qquad n \to \infty.\end{aligned}$$

Proof of the Theorem: Consider

$$\begin{aligned}
& \sqrt{n}(h_{np}(t) - h(t)) \\
= \; & \sqrt{n}\Big(h_{np}(t) - \hat{h}_{np}(t) + \hat{h}_{np}(t) - h_p^*(t) + h_p^*(t) \\
& - \sum_{k=1}^{m} f^{(k)}(t) + \sum_{k=1}^{m} f^{(k)}(t) - h(t)\Big) \\
= \; & \sqrt{n}\Big(h_{np}(t) - \hat{h}_{np}(t)\Big) + \sqrt{n}\Big(\hat{h}_{np}(t) - h_p^*(t)\Big) + \sqrt{n}\Big(h_p^*(t) - \sum_{k=1}^{m} f^{(k)}(t)\Big) \\
& + \sqrt{n}\Big(\sum_{k=1}^{m} f^{(k)}(t) - h(t)\Big).
\end{aligned} \tag{2.7.48}$$

Now

$$\begin{aligned}
\sqrt{n}\Big(\sum_{k=1}^{m} f^{(k)}(t) - h(t)\Big) &= \sqrt{n}\Big(\sum_{k=1}^{m} f^{(k)}(t) - \sum_{k=1}^{\infty} f^{(k)}(t)\Big) \\
&= \sqrt{n}\Big(\sum_{k=1}^{m} f^{(k)}(t) - \sum_{k=1}^{m} f^{(k)}(t) + \sum_{k>m} f^{(k)}(t)\Big) \\
&\to 0,
\end{aligned} \tag{2.7.49}$$

by assumption (i) of the theorem. Now from lemma(2.7.1), (2.7.2), (2.7.3) and equations (2.7.48) and (2.7.49), the theorem follows.

2.8 Simulation Study

Here a simulation study is carried out to see how the kernel estimate for renewal density can be computed in practice. We have considered two life time models: exponential and gamma distributions. It is well known that the renewal density of the exponential model

$$f(x, \lambda) = \frac{1}{\lambda} exp(-\frac{x}{\lambda}), \quad x > 0$$

is λ. The simulation study is carried out with $\lambda = 2$ for different sample sizes and band width p. The graph of the exact renewal density $h(t) = \lambda$ and the proposed estimators are displayed in the same graph for making comparisons. The *Epanechnikov kernel*

$$K(t) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}t^2) & \text{if } -\sqrt{5} \leq t \leq \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$$

is used throughout to compute the estimator. The computation of $\sum_c$ creates practical difficulty as we have to consider all the ${}^nC_k$ combinations of

$$K\Big(\frac{t - (X_{i_1} + X_{i_2} + \ldots + X_{i_k})}{p}\Big), k = 1, 2, \ldots, m.$$

To overcome this practical difficulty we have generated subsamples with replacement of sufficient sizes, say, between 2000 and 5000 and

$$\sum K\Big(\frac{t-(X_{i_1}+X_{i_2}+\ldots+X_{i_k})}{p}\Big)$$

is computed for those samples. Finally instead of dividing the expression by ${}^{n}C_k$ we use the division factor as the number of subsample selected. We found that even with this approximation the estimator performs satisfactory.

As a second illustration we used the gamma distribution with the density

$$f(x)=\frac{\beta^{\alpha}}{\Gamma\alpha}xe^{-\frac{x}{\beta}}, \quad x>0$$

with $\alpha=2$ and $\beta=1$. The renewal density of this model is given by

$$h(t)=e^{-t}\sum_{k=1}^{\infty}\frac{t^{2k-1}}{(2k-1)!}.$$

The plots of true renewal density and the proposed estimators in various samples are plotted.

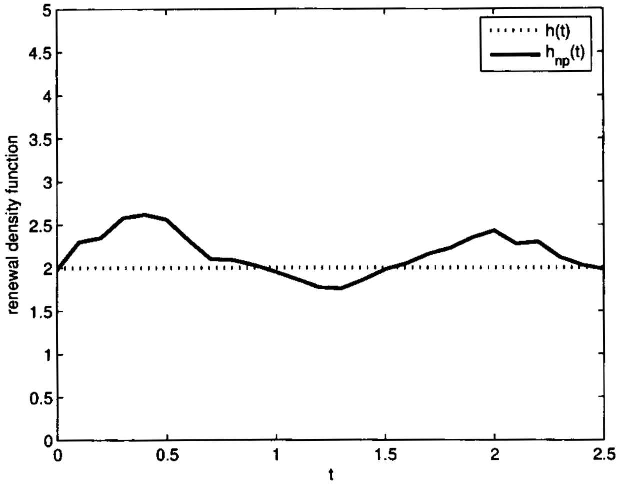

Figure 2.1: *Simulation results from exponential model with n=100, m=4 and p=1/3*

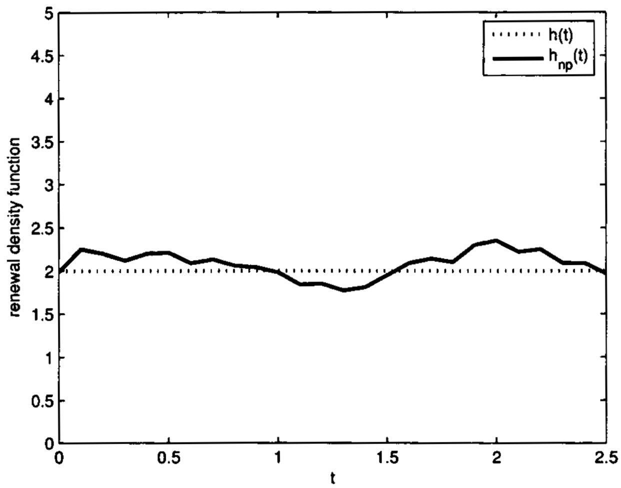

Figure 2.2: *Simulation results from exponential model with n=100, m=4 and p=1/4*

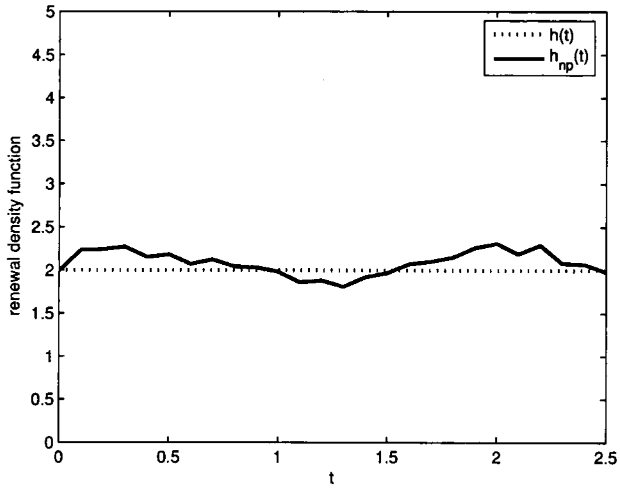

Figure 2.3: *Simulation results from exponential model with n=100, m=6 and p=1/4*

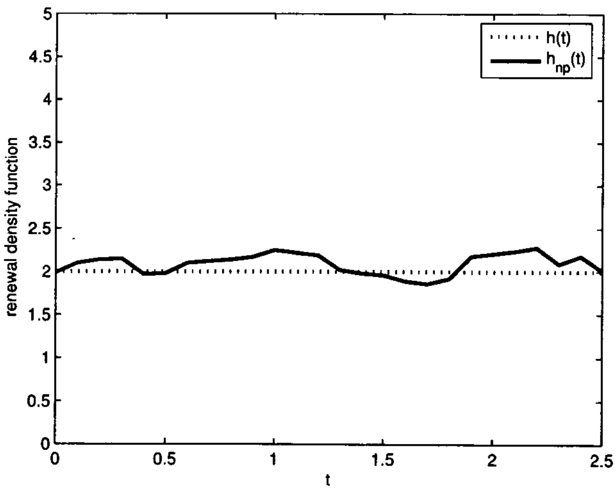

Figure 2.4: *Simulation results from exponential model with n=200, m=4 and p=1/3*

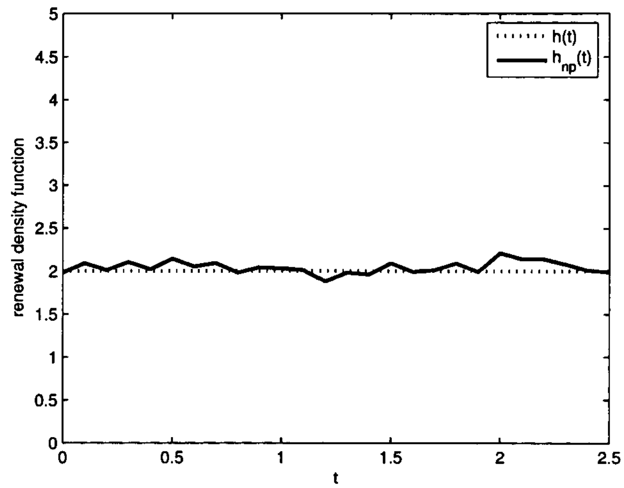

Figure 2.5: *Simulation results from exponential model with n=200, m=4 and p=1/4*

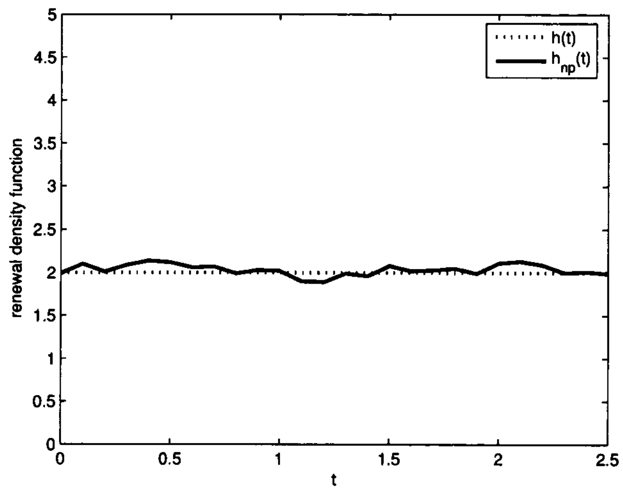

Figure 2.6: *Simulation results from exponential model with n=200, m=6 and p=1/4*

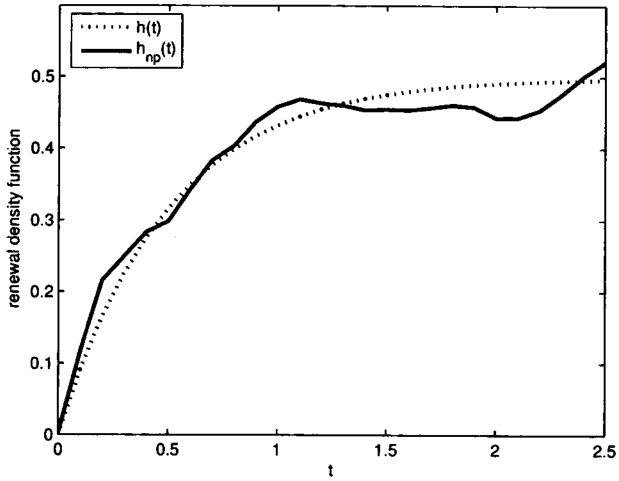

Figure 2.7: *Simulation results from gamma model with n=100, m=5 and p=1/3*

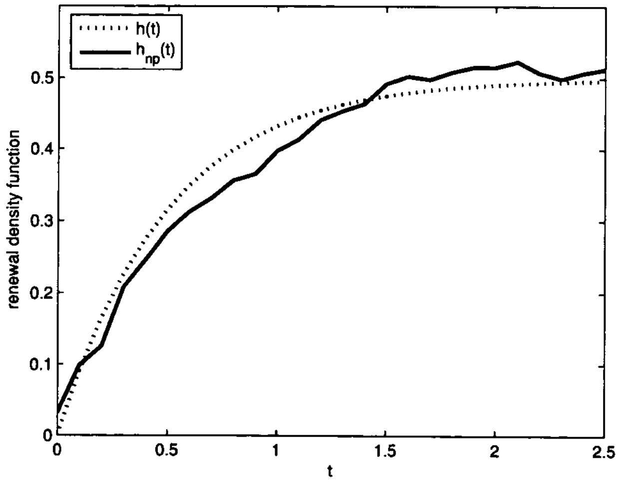

Figure 2.8: *Simulation results from gamma model with n=100, m=5 and p=1/4*

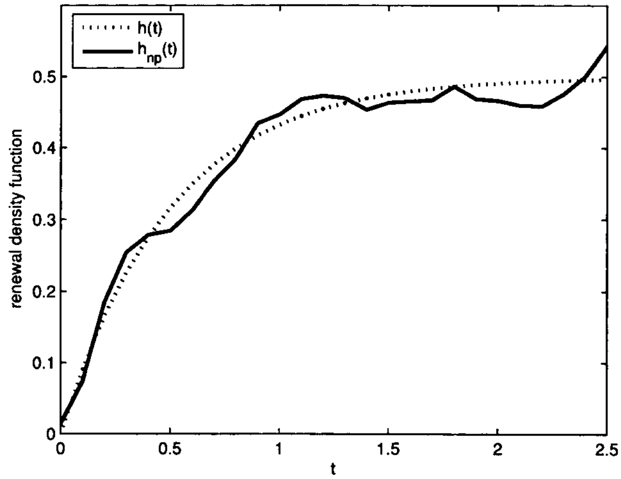

Figure 2.9: *Simulation results from gamma model with n=200, m=4 and p=1/3*

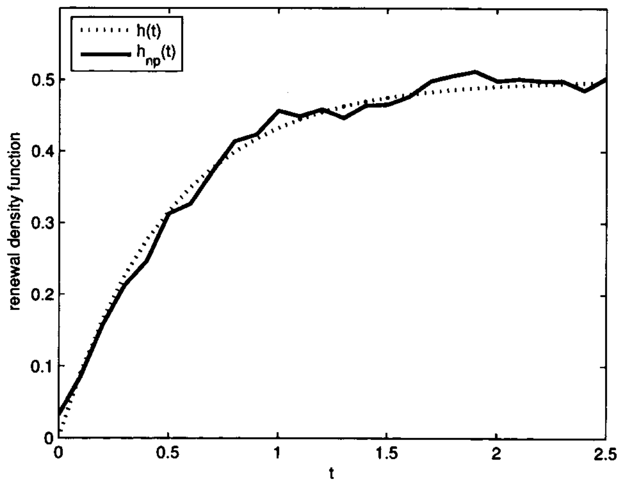

Figure 2.10: *Simulation results from gamma model with n=200, m=6 and p=1/4*

2.9 Concluding Remarks

The simulation study is carried out for different sample sizes, band width p and truncation parameter m. The plots of the exact renewal density and the proposed estimators for independent set of trials to each model are displayed in figures 2.1 to 2.10. The plots reveal that the proposed estimator behaves satisfactorily. The optimum choice of the kernel, m and p for a given sample is yet to be settled. In the probability density estimation it is established through studies that the choice of the kernel is not critical; Silverman, B.W.(1986). This fact remains intact in the estimation of renewal density. However the choice of the parameters m and p are to be made with care. Some guiding principles are already made earlier. It remains to see how a cross validation type technique can be applied here.

Chapter 3

Density Estimation using Polynomial

3.1 Introduction

Numerical data are found in many applications of data analysis. Most numerical data come from measurements and experiments, thus resulting from the sampling of a random variable, the mathematical concept that characterizes the numerical results of experiments. To analyse data, one may choose to handle directly the results of the experiments. A random variable is completely characterized by its probability density functions (PDF), i.e. a function that represents the probability of an event to occur when the random variable is equal to (or contained in an interval around) a specific value. Estimating PDFs based on a sample is thus of primary importance in many contexts. However, and despite the vast literature on the topic, even in the univariate case there is no consensus about which method to use, nor about the pros and cons of these methods.

Probability density function(PDF) estimation is a fundamental step in statistics as it characterizes completely the behaviour of a random variable. It is the construction of an estimate of the density function from the observed data. This provides a natural way to investigate the properties of a given data set, i.e. a realization of the random variable, and to carry out efficient data mining. Consider any random quantity X that has probability density function f. When we perform density estimation mainly two alternatives can be considered. The first approach, known as *parametric* density estimation, assumes the data is drawn from a specific density model, for example the normal distribution with mean μ and variance σ^2. The density function f underlying data could then be estimated by finding estimates of μ and σ^2 from the data and substituting these estimates into the formula for the normal density. Unfortunately, an a-priori choice of the PDF model in practice is not suited since it might provide a false representation of the true PDF. An alternative is to build *non-parametric* PDF estimators in order to let the data speak for themselves. Silverman(1981) explains that density estimates are of use in all three stages of statistical treatment of data, namely, exploratory, confirmatory and presentational. At the exploratory stage, density estimators give an indication of multi modality, skewness or dispersion, etc., of the data. For confirmatory purpose they can be used, for instance in non-parametric discriminant analysis (Prakash Rao,1983). For data presentation density estimators are the best information -presentation transformations given data.

In this chapter we introduce a polynomial estimator of the probability density function using a polynomial used in proving the Stone-Weistrass theorem; Rudin, W.(1976). In fact the proposed estimator belongs to the general class of estimators given in section 3.2. The appropriate degree of the polynomial, d will be fixed based on the sample using an objective criterion.

The chapter is organized as follows. In section 2 we summarize the main methods available for univariate density estimation. The polynomial estimator is defined in section 3. Asymptotic properties like consistency and asymptotic normality are proved in section 4. In section 5 an objective criterion for choosing the optimum value of the smoothing parameter d is proposed. Simulation studies and data analyses are done in sections 6 and section 7.

3.2 Survey of existing Methods

In this section we review current methods available for probability density estimation. Many of the important applications of density estimation are to multivariate data, but since all the multivariate methods are generalizations of univariate methods, it is worth getting a feel for the univariate case.

It is convenient to define some standard notations. Except where otherwise stated, it will be assumed that we are given a sample of n real observations $X_1, X_2, \ldots, X_n$ whose underlying density f is to be estimated. The symbol $\hat{f}$ will be used to denote whatever density estimate is currently being considered.

The oldest and most widely used density estimate is histogram. Let $X_i, 1 \leq i \leq n$, be a random sample of size n from a population with density function f. Let $k_n(a, b)$ denote the number of observations in the sample that falls in $[a, b]$.

Observe that $\int_a^b f(x)dx$ can be estimated by the corresponding relative frequency, namely,

$$\frac{1}{n} k_n(a, b).$$

Under reasonable conditions, for example, if the interval $[a, b]$ is small, $a \leq x \leq b$, and f is continuous, then $f(x)$ is close to

$$\frac{1}{b-a} \int_a^b f(y)dy,$$

and hence, one can estimate $f(x)$ by

$$\hat{f}_n(x) = \frac{k_n(a, b)}{n(b-a)}.$$

This method of histogram has been used in many fields of applications.

Prakash Rao B.L.S. (1983) presented similar but more refined method. He defined the partitions $\{x_k^{(n)}, k = 0, \pm 1, \pm 2, \cdots,\}$ of real line R for every $n \geq 1$

such that

$$\cdots < x_k^{(n)} < x_{k+1}^{(n)} < \cdots$$

These partitions may be fixed in advance or may depend on the sample. Let

$$x_{k+1}^{(n)} - x_k^{(n)} = h_n, k = 0, \pm 1, \pm 2, \cdots ,$$

where $h_n \downarrow 0$ as $n \to \infty$. He defined the histogram by

$$\hat{f}_n(x) = \frac{k_n(x_k^{(n)}, x_{k+1}^{(n)})}{nh_n} \quad \text{as} \quad x_k^{(n)} \le x < x_{k+1}^{(n)},$$

and studied the properties of the estimate. In particular, the asymptotic unbiasedness and consistency of the histogram estimate are established.

The *naive estimator* is a simple alternative to the histogram estimator. From the definition of a probability density, if the random variable X has density f, then

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x - h < X < x + h)$$

For any given h, we can estimate $P(x - h < X < x + h)$ by the proportion of the sample observations falling in the interval $(x - h, x + h)$. A natural estimator of the density is given by choosing a small number h_n and setting

$$\hat{f}(x) = \frac{1}{2nh_n} [\text{Number of} \quad X_1, X_2, \cdots , X_n \quad \text{falling in} \quad (x - h_n, x + h_n)]$$

It is easy to see that the naive estimator can be expressed as

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} w(\frac{x - X_i}{h_n})$$

with the weight function

$$w(x) = \frac{1}{2}, \quad \text{if} \quad |x| < 1 \quad \text{and} \quad 0, \quad \text{otherwise.}$$

The naive estimator is not wholly satisfactory from the point of view of using density estimates for presentation. $\hat{f}$ is not a continuous function, but has jumps at the points $X_i \pm h_n$ and has zero derivative everywhere else. This gives estimates a somewhat ragged character which is not only eisthetically undesirable, but more significantly, could provide a misleading impression.

Kernel estimator is used to partially overcome this difficulty and partially for other technical reasons. In this method we replace the weight function w by a *kernel function* K which satisfy the condition

$$\int_{-\infty}^{\infty} K(x)dx = 1 \tag{3.2.1}$$

Usually, but not always, K will be a symmetric probability density function. By analogy with the definition of naive estimator, the kernel estimator with kernel K is defined by

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{x - X_i}{h_n}) \tag{3.2.2}$$

where h_n is the *window width* (or *band width* or *smoothing parameter*). It follows from (3.2.2) that the kernel estimator can be considered as a sum of *bumps* placed at the observations. There is an enormous literature on the asymptotic properties on the kernel estimate . The weak uniform consistency are established for kernel estimates (Prakash Rao.L.S.(1983)). A different and somewhat weaker kind of consistency is discussed by Deveroy and Gyorfi (1985). Making no assumptions on the unknown density f, they show that the conditions (i) $h_n \to 0$ and (ii) $nh_n \to \infty$ as $n \to \infty$ are necessary and sufficient conditions for the convergence:

$$\int |\hat{f}(x) - f(x)|dx \to 0$$

with probability one as $n \to \infty$.

Another method is to define the k^{th} *nearest neighbour density* which we denote by

$$\hat{f}(t) = \frac{k-1}{2nd_k(t)}$$

where

$$d_1(t) \le d_2(t) \le \cdots \le d_n(t)$$

to be distances, arranged in ascending order, from t to the points of the sample. While the naive estimator is based on the number of observations falling in a box of fixed width centered at the point of interest, the nearest neighbour estimate is inversely proportional to the size of the box needed to contain a given number of observations. In the tails of the distribution, the distance $d_k(t)$ will be larger than in the main part of the distribution, and so the problem of under smoothing in the tails should be reduced. Like the naive estimator, the nearest neighbor estimate is not a smooth curve. $\hat{f}$ will be positive and continuous everywhere, but will have discontinuous derivative at all the same points as d_k. This estimate is unlikely to be appropriate if an estimate of the entire density is required.

It is possible to generalize the nearest neighbour estimate in the form:

$$\hat{f}(t) = \frac{1}{nd_k(t)} \sum_{j=1}^{n} K\left(\frac{t - X_j}{d_k(t)}\right)$$

where $K(x)$ be a kernel function integrating to one. Clearly $\hat{f}(t)$ is the kernel estimate evaluated at t with window width $d_k(t)$. The overall amount of smoothing is governed by the choice of the integer k, but the window width used at any particular point depends on the density of observations near that point.

Another related method to the nearest neighbour method is *the variable kernel method.* In this the estimate is constructed similar to the classical kernel estimate, but the scale parameter of the bumps placed on the data points is allowed to vary from one data point to the other.

Another class of estimators of the probability density is that of *orthogonal series estimators.* Suppose we are trying to estimate a density f on the unit interval $[0, 1]$. The idea of orthogonal series method is then to estimate f by estimating the coefficients of its Fourier series expansion. The PDF f can be represented by the Fourier series

$$\sum_{\nu=0}^{\infty} f_\nu \phi_\nu,$$

where for each $\nu \geq 0$

$$f_\nu = \int_0^1 f(x)\phi_\nu(x)dx.$$

Suppose X is a random variable with density f. Then

$$f_\nu = E[\phi_\nu(X)]$$

and hence a natural and unbiased estimator of f_ν based on a sample $X_1, X_2, \cdots, X_n$ from f is

$$\hat{f}_\nu = \frac{1}{n}\sum_{i=1}^{n} \phi_\nu(X_i).$$

Unfortunately, the sum $\sum_{\nu=0}^{\infty} \hat{f}_\nu \phi_\nu$ will not be a good estimate of f, but will converge to a sum of delta functions at the observations; see Silverman B.W.(1986). In order to obtain a useful estimator of the density f, we truncate the expansion $\sum_{\nu=0}^{\infty} \hat{f}_\nu \phi_\nu$ at some point. Choose an integer k and define the density estimate $\hat{f}$ by

$$\hat{f}(x) = \sum_{\nu=1}^{k} \hat{f}_\nu \phi_\nu(x)$$

The choice of the cutoff point k determines the amount of smoothing.

Finally, a general class of density estimators has been proposed as follows. Suppose $w(x,y)$ is a function of two arguments, which in most cases will satisfy the conditions

$$\int_{-\infty}^{\infty} w(x,y)dy = 1 \quad \text{and} \quad w(x,y) \geq 0 \quad \text{for all } x \text{ and } y.$$

An estimate of the density underlying the data may be obtained by putting

$$\hat{f}(t) = \frac{1}{n}\sum_{i=1}^{n} w(X_i, t).$$

$\hat{f}(t)$ is a probability function and the smoothness properties of $\hat{f}$ will be inherited from the conditions. This class is helpful in two ways. Firstly it is a unifying concept which makes it possible to obtain theoretical results applicable to a whole range of apparently distinct estimators . On the other hand it is possible to define useful estimators which do not fall into any of the classes discussed earlier.

In this chapter we introduce a polynomial estimator of the probability density function using a polynomial used in proving the Stone-Weistrass theorem; Rudin, W.(1976). In fact the proposed estimator belongs to the general class of estimators discussed above. The appropriate degree of the polynomial, d will be fixed based on the sample using an objective criterion. The chapter is organized as follows.

The polynomial estimator is defined in section 3. Asymptotic properties like consistency and normality are proved in section 4. In section 5 an objective criterion for choosing the optimum value of the smoothing parameter d is proposed. Simulation studies and data analyses are done in sections 6 and section 7.

3.3 Polynomial Density Estimator

It will be assumed that we have a sample $X_1, X_2, \cdots, X_n$ of independent and identically distributed observations from a continuous univariate distribution with probability density function f, which we are trying to estimate and have the support $[a, b]$. Without loss of generality we assume that $a = 0$ and $b = 1$.

We propose a non-parametric estimator for the probability density function f using a sequence of real polynomials.

Definition 3.3.1 *Let*

$$Q_d(x) = c_d(1 - x^2)^d, d = 1, 2, \cdots$$

where c_d is chosen so that

$$\int_{-1}^{1} Q_d(x)dx = 1, d = 1, 2, \cdots \tag{3.3.3}$$

We define a non-parametric estimator f_n of f by

$$\begin{aligned} f_n(x) &= \frac{1}{n}\sum_{i=1}^{n} Q_d(X_i - x) \\ &= \frac{1}{n}\sum_{i=1}^{n} c_d[1 - (X_i - x)^2]^d \end{aligned}$$

This estimator is a special case of the general class of density estimators given in section 2. The degree of the polynomial, d, is called the *smoothing parameter* of the polynomial estimator. c_d's are constants depends on the degree of the polynomial. Now we shall look into some of the asymptotic properties of the proposed estimator.

The values of the constants c_d for various d can be evaluated in the following way. We have

$$
\begin{aligned}
1 &= \int_{-1}^{1} c_d(1-x^2)^d dx \\
&= 2\int_{0}^{1} c_d(1-x^2)^d dx
\end{aligned}
$$

Substituting $x = \sin\theta$ we get

$$
\begin{aligned}
\frac{1}{c_d} &= 2\int_{0}^{\frac{\pi}{2}} \cos^{2d}\theta \cos\theta d\theta \\
&= 2\int_{0}^{\frac{\pi}{2}} \cos^{2d+1}\theta d\theta \\
&= 2\frac{(2d)}{(2d+1)}\frac{(2d-2)}{(2d-1)}\cdots\frac{2}{3}
\end{aligned}
$$

That is

$$
\begin{aligned}
c_d &= \prod_{i=1}^{d}\frac{2i+1}{2i} \\
&= \frac{1}{2}\frac{(2d+1)!}{2^{2d}(d!)^2} \\
&\approx \sqrt{\frac{d}{\pi}} \quad , \qquad \text{using sterling's approximation.} \qquad (3.3.4)
\end{aligned}
$$

3.4 Asymptotic Properties

The usual asymptotic frame work in which theorems about density estimation are provided is to assume that the unknown density f is fixed and satisfy given regularity conditions. The density estimates considered are constructed from the first n observations in an independent identically distributed sequence $X_1, X_2, \cdots$ drawn from f. It is assumed that the smoothing parameter d depends in some way on the sample size. Limiting results are then obtained on the behavior of the estimate as n tends to infinity. We shall write d_n for the smoothing parameter d in order to make the dependence on n explicit.

3.4.1 Consistency of Polynomial Density Estimator

In this section we shall establish the consistency of the polynomial estimator of the density function. The following theorem establishes asymptotic unbiasedness of the polynomial estimator f_n under certain regularity conditions.

Theorem 3.4.1 *Let $f(x)$ be probability density function with support $[a, b]$. Then we have*

$$Ef_n(x) \to f(x)$$

as $n \to \infty$ and $d_n \to \infty$.

Proof: Assume that $f(0) = f(1) = 0$. Further we define $f(x)$ to be zero outside $[0,1]$. Then f is uniformly continuous on the whole line.Now

$$\begin{aligned} E[f_n(x)] &= E[\frac{1}{n}\sum_{i=1}^{n} Q_{d_n}(X_i - x)] \\ &= E[Q_{d_n}(X_1 - x)] \\ &= \int_0^1 Q_{d_n}(t-x)f(t)dt \end{aligned} \tag{3.4.5}$$

We need some information about the order of magnitude of c_d. We have

$$\begin{aligned} \int_{-1}^{1} Q_{d_n}(x)dx &= \int_{-1}^{1} c_{d_n}(1-x^2)^{d_n}dx \\ &= \int_0^1 c_{d_n}(1-x^2)^{d_n}dx \\ &\geq 2c_{d_n}\int_0^{\frac{1}{\sqrt{d_n}}} (1-x^2)^{d_n}dx \end{aligned}$$

Now consider the function

$$g(x) = (1-x^2)^{d_n} - 1 + d_n x^2$$

Clearly $g(x) = 0$ at $x = 0$. Also $g'(x) = 2xd_n(1-(1-x^2)^{d_n-1})$ and $g'(x) > 0$ in $(0,1)$; hence $g(x)$ is increasing in$[0,1]$. Therefor we have the inequality,

$$(1-x^2)^{d_n} - 1 + d_n x^2 > 0 \quad \text{in} \quad [0,1], \quad \text{or}$$

$$(1-x^2)^{d_n} > 1 - d_n x^2 \quad \text{in} \quad [0,1]$$

Therefore

$$\begin{aligned} \int_{-1}^{1} Q_{d_n}(x)dx &\geq 2c_{d_n}\int_0^{\frac{1}{\sqrt{d_n}}} (1-d_n x^2)dx. \\ &= \frac{4c_{d_n}}{3\sqrt{d_n}} \end{aligned}$$

since $\int_0^{\frac{1}{\sqrt{d_n}}}(1-d_nx^2)dx = \frac{2}{3\sqrt{d_n}}$.

It follows from (3.3.3) that $\frac{4}{3}\frac{c_{d_n}}{\sqrt{d_n}} \leq 1$ implying that

$$c_{d_n} < \sqrt{d_n} \tag{3.4.6}$$

This expression provides an upper bound for c_{d_n}.

Now, for any $\delta > 0$, $\quad Q_{d_n}(x) \leq \sqrt{d_n}(1-\delta^2)^{d_n}$ in $\delta \leq |x| \leq 1$, so that

$$Q_{d_n}(x) \rightarrow 0 \quad \text{uniformly in} \delta \leq |x| \leq 1. \tag{3.4.7}$$

Let, for $0 \leq x \leq 1$,

$$\begin{aligned} P_{d_n}(x) &= \int_{-1}^{1} f(x+t)Q_{d_n}(t)dt \qquad (3.4.8) \\ &= \int_{-1}^{-x} f(x+t)Q_{d_n}(t)dt + \int_{-x}^{1-x} f(x+t)Q_{d_n}(t)dt \\ &\quad + \int_{1-x}^{1} f(x+t)Q_{d_n}(t)dt \end{aligned}$$

For $-1 \leq t \leq -x$, $\quad f(x+t) = 0$, by assumption and similarly for $\quad 1-x \leq t < 1$, $\quad f(x+t) = 0$. Therefore, using (3.4.5)

$$\begin{aligned} P_{d_n}(x) &= \int_{-x}^{1-x} f(x+t)Q_{d_n}(t)dt \\ &= \int_0^1 f(t)Q_{d_n}(t-x)dt, \quad \text{using a simple change of variable.} \\ &= Ef_n(x), \qquad (3.4.9) \end{aligned}$$

Clearly

$$P_{d_n}(x) = \int_0^1 f(t)Q_{d_n}(t-x)dt, \quad \text{is a polynomial sequence.}$$

Again, given $\epsilon > 0$, we choose $\delta > 0$ such that $|y-x| < \delta$ implies $|f(y)-f(x)| < \frac{\epsilon}{2}$. Let $M = \sup |f(x)|$ using (3.4.8) and (3.4.9) and the fact that

$$\int_{-1}^{1} Q_d(x)dx = 1, \quad d = 1, 2, \cdots$$

for $0 \leq x \leq 1$,

$$\begin{aligned}
|Ef_n(x) - f(x)| &= |\int_{-1}^{1} f(x+t)Q_{d_n}(t)dt - \int_{-1}^{1} f(x)Q_{d_n}(t)dt| \\
&\leq \int_{-1}^{1} |f(x+t) - f(x)|Q_{d_n}(t)dt \\
&= \int_{-1}^{-\delta} |f(x+t) - f(x)|Q_{d_n}(t)dt + \int_{-\delta}^{\delta} |f(x+t) - f(x)|Q_{d_n}(t)dt \\
&\quad + \int_{\delta}^{1} |f(x+t) - f(x)|Q_{d_n}(t)dt \\
&\leq 2M \int_{-1}^{-\delta} Q_{d_n}(t)dt + \frac{\epsilon}{2} \int_{\delta}^{\delta} Q_{d_n}(t)dt + 2M \int_{\delta}^{1} Q_{d_n}(t)dt \\
&\leq 4M\sqrt{d_n}(1-\delta^2)^{d_n} + \frac{\epsilon}{2} \\
&< \epsilon, \quad \text{for large enough values of } d_n.
\end{aligned}$$

This proves the theorem.

Theorem 3.4.2 *Assume that* $\frac{d_n}{n} \to 0$ *as* $n \to \infty$. *Then*

$$Var(f_n(x)) \to 0 \text{ as } n \to \infty.$$

Proof: We have

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} Q_{d_n}(X_i - x).$$

Then

$$
\begin{aligned}
Var(f_n(x)) &= \frac{1}{n^2} n Var[Q_{d_n}(X_i - x)] \\
&= \frac{1}{n} c_{d_n}^2 Var(1-(X_i-x)^2)^{d_n}
\end{aligned}
$$

Since $V = Var(1-(X_i-x)^2)^{d_n} < \infty$, we have

$$
\begin{aligned}
Var(f_n(x)) &= \frac{c_{d_n}^2}{n} V \\
&< \frac{d_n}{n} V, \quad \text{since} \quad c_{d_n} < \sqrt{d_n}
\end{aligned}
$$

Now, from assumption it follows that

$$
Var(f_n(x)) \to 0 \quad as \quad n \to \infty.
$$

This proves the theorem.

Remark 3.4.1 *Theorems 3.4.1 and 3.4.2 together implies that the proposed estimator is uniformly and weakly consistent.*

We shall now prove a theorem which provide a sufficient condition for the uniform consistency with probability one for the proposed estimator of probability density function.

Theorem 3.4.3 *Suppose the series* $\sum_{n=1}^{\infty} exp(-\gamma n d_n^{-2})$ *converges for every* $\gamma > 0$. *Then*

$$
\sup_x |f_n(x) - f(x)| \to 0
$$

with probability one as $n \to \infty$ *if the density* f *is uniformly continuous.*

Proof: We have the polynomial estimator

$$\begin{aligned} f_n(x) &= \frac{1}{n}\sum_{i=1}^{n} Q_{d_n}(X_i - x) \\ &= \int_0^1 Q_{d_n}(y-x)dF_n(y), \end{aligned}$$

where

$$Q_{d_n}(x) = c_{d_n}(1-x^2)^{d_n}, d_n = 1, 2, \cdots$$

Also

$$E[f_n(x)] = \int_0^1 Q_{d_n}(y-x)f(y)dy.$$

Now let

$$\begin{aligned} \bar{V}_n &= \sup_x |f_n(x) - Ef_n(x)| \\ &= \sup_x \left| \int_0^1 c_{d_n}(1-(y-x)^2)^{d_n} dF_n(y) - \int_0^1 c_{d_n}(1-(y-x)^2)^{d_n} dF(y) \right| \\ &\leq c_{d_n} \sup_x \int_0^1 |(1-(y-x)^2)^{d_n}| \quad |d(F_n(y) - F(y))| \\ &= c_{d_n} \sup_x \int_0^1 |F_n(y) - F(y)| \quad |d(1-(y-x)^2)^{d_n}| \\ &\leq \sup_x |F_n(y) - F(y)| \mu c_d \end{aligned} \qquad (3.4.10)$$

where

$$\mu = \int_0^1 |d((1-(y-x)^2)^{d_n})|,$$

the variation of $(1-(y-x)^2)^{d_n}$. Let

$$D_n = \sup_x |F_n(x) - F(x)|.$$

It follows that there exist positive constants c and $\alpha, 0 < \alpha \leq 2$, such that

$$P(D_n > \lambda n^{-\frac{1}{2}}) \leq c\ exp(-\alpha\lambda^2) \tag{3.4.11}$$

for every $\lambda > 0$ and any continuous distribution function F; Dvoretzky *et al.*(1956).

From (3.4.10) and (3.4.11), it follows that

$$\begin{aligned} P(\sup_x |f_n(x) - Ef_n(x)| > \epsilon) &= P(\sup_x |F_n(x) - F(x)|\mu c_{d_n} > \epsilon) \\ &= P(\sup_x |F_n(x) - F(x)| > \epsilon\mu^{-1}c_{d_n}^{-1}) \\ &= P(D_n > \epsilon\mu^{-1}c_{d_n}^{-1}) \\ &\leq C_1 exp(-\alpha\epsilon^2\mu^{-2}c_{d_n}^{-2}n) \\ &= C_1 exp(-\beta n c_{d_n}^{-2}) \end{aligned}$$

where $\beta = \alpha\epsilon^2\mu^{-2} > 0$.

Since the series $\sum_{n=1}^{\infty} exp(-\gamma n c_{d_n}^{-2})) < \infty$ for every $\gamma > 0$,

$$\sum_{n=1}^{\infty} P(\sup_x |f_n(x) - Ef_n(x)| > \epsilon) \leq C_1 \sum_{n=1}^{\infty} exp(-\beta n c_{d_n}^{-2})$$

implies

$$P(\sup_x |f_n(x) - Ef_n(x)| > \epsilon) = 0,$$

using Borel-Cantelli lemma.

Or in other words

$$\sup_x |f_n(x) - Ef_n(x)| \to 0 \quad a.s. \tag{3.4.12}$$

Again using theorem 3.4.1, we have

$$E(f_n(x)) \to f(x) \quad \text{as} \quad n \to \infty$$

That is

$$\sup_x |Ef_n(x) - f(x)| \to 0 \quad a.s. \quad n \to 0. \tag{3.4.13}$$

Using (3.4.12) and (3.4.13), it follows that $f_n(x)$ is uniformly consistent with probability one and the theorem is complete.

3.4.2 Limit Distribution of Polynomial Density Estimator

In this section we shall show that the polynomial estimator of the density function is asymptotically normal. To show the asymptotic normality we verify the Lyaponove's sufficient condition and then apply the C_r-inequality; Loeve M. (1963), p. 155 . The following theorem shows that the proposed polynomial estimator is asymptotically normal under a very general regularity condition. We have the following theorem.

Theorem 3.4.4 *The random variable* $\dfrac{f_n(x) - E(f_n(x))}{\left[Var(f_n(x))\right]^{\frac{1}{2}}}$ *is asymptotically normal if*

$$\frac{d_n^{\frac{1}{2}}}{n} \to 0$$

.

Proof: The estimator

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} Q_{d_n}(X_i - x)$$

can be written as

$$\begin{aligned} f_n(x) &= \frac{1}{n}\sum_{i=1}^{n} c_{d_n}[1-(X_i-x)^2]^{d_n} \\ &= \frac{1}{n}\sum_{i-1}^{n} Z_{ni} \end{aligned}$$

where

$$Z_{ni} = c_{d_n}[1-(X_i-x)^2]^{d_n} \tag{3.4.14}$$

are independent identically distributed random variables. The Lyaponove's sufficient condition for

$$\frac{f_n(x) - Ef_n(x)}{[Var(f_n(x))]^{\frac{1}{2}}} \longrightarrow N(0,1)$$

is that

$$\frac{E|Z_{n1} - E(Z_{n1})|^3}{n^{\frac{1}{2}}[Var(Zn1)]^{\frac{3}{2}}} \longrightarrow 0 \quad \text{as} \quad n \to \infty. \tag{3.4.15}$$

Now

$$\begin{aligned} E(Z_{n1}) &= \int_0^1 c_{d_n}[1-(y-x)^2]^{d_n} f(y)dy \\ &= \int_{-1}^1 c_{d_n}(1-u^2)^{d_n} f(u+x)du \\ &\approx f(x)c_{d_n}\int_{-1}^1 (1-u^2)^{d_n}du \\ &= f(x), \end{aligned}$$

using the arguments used in the proof of theorem 3.4.1 and since

$$\int_{-1}^{1} c_{d_n}(1-x^2)^{d_n}dx = 1$$

Similarly,

$$\begin{aligned} E(Z_{n1}^2) &= \int_0^1 (c_{d_n})^2[1-(y-x)^2]^{2d_n} f(y)dy \\ &= (c_{d_n})^2 \int_{-1}^{1} (1-u^2)^{2d_n} f(y)dy \\ &\approx \frac{(c_{d_n})^2}{c_{2d_n}} f(x). \end{aligned} \tag{3.4.16}$$

Also

$$\begin{aligned} E|Z_{n1}^3| &= E(Z_{n1}^3) \\ &\approx \frac{(c_{d_n})^3}{c_{3d_n}} f(x). \end{aligned} \tag{3.4.17}$$

Now applying C_r-inequality(Loeve(1963), page 155), we get

$$\begin{aligned} \frac{E|Z_{n1}-E(Z_{n1})|^3}{n^{\frac{1}{2}}\Big[Var(Z_{n1})\Big]^{\frac{3}{2}}} &\leq \frac{2^2E|Z_{n1}|^3 + 2^2|E(Z_{n1})|^3}{n^{\frac{1}{2}}\Big[Var(Z_{n1})\Big]^{\frac{3}{2}}} \\ &\approx \frac{4\Big[\frac{(c_{d_n})^3}{c_{3d_n}} f(x) + (f(x))^3\Big]}{n^{\frac{1}{2}}\Big[\frac{(c_{d_n})^2}{c_{2d_n}} f(x) - (f(x))^2\Big]} \\ &= \frac{4\Big[\frac{(\frac{d_n}{\pi})^{\frac{3}{2}}}{\sqrt{\frac{3d_n}{\pi}}} f(x) + (f(x))^3\Big]}{n^{\frac{1}{2}}\Big[\frac{(\frac{d_n}{\pi})}{\sqrt{\frac{2d_n}{\pi}}} f(x) - (f(x))^2\Big]} \end{aligned}$$

using the approximation

$$c_d \approx \sqrt{\frac{d}{\pi}}.$$

That is

$$\frac{E|Z_{n1} - E(Z_{n1})|^3}{n^{\frac{1}{2}}\Big[Var(Z_{n1})\Big]^{\frac{3}{2}}} \approx \frac{d_n}{n^{\frac{1}{2}} d_n^{\frac{3}{4}}}$$
$$= \frac{d_n^{\frac{1}{4}}}{n^{\frac{1}{2}}} = \left(\frac{\sqrt{d_n}}{n}\right)^{\frac{1}{2}} \to 0$$

since $\frac{d_n^{\frac{1}{2}}}{n} \to 0$.

Remark 3.4.2 *When we consider the distributions whose support is not the unit interval* $[0, 1]$, *we shall use transformation of the original data into the unit interval. Suppose* $X_1, X_2, \ldots, X_n$ *be the n observations on a random variable* X *whose probability density function is* $f(x)$, $a \leq x \leq b$. *We transform this data using* $Y_i = \frac{X_i - a}{b - a}$. *Then the estimator of PDF* $f(x)$ *will be*

$$f_n(x) = \frac{1}{b-a} g_n\left(\frac{x-a}{b-a}\right)$$

where $g_n(.)$ *is the estimated probability density function of the transformed variable.*

3.5 Choice of Smoothing Parameter

The accuracy of the kernel method of estimation depends on the bandwidth parameter and on the number of observed data points. If the bandwidth is too high with respect to the number of observed data points, the resulting estimate will be over-smooth, if the bandwidth is too low, the estimate will be too noisy; see Silverman B.W.(1986) . In estimating the density function obtaining a smoothened estimate is desirable. In polynomial estimation the degree of the proposed polynomial, d, is the smoothing parameter. We shall establish that the choice of d is not much sensitive to the proposed estimation in large samples as in the case of kernel estimation.

When we consider estimation at a single point, the natural measure of discrepancy between the estimator from the true density is the *mean square error*, MSE, defined by

$$\begin{aligned} MSE(f_n(x)) &= E\{f_n(x) - f(x)\}^2 \\ &= \{Ef_n(x) - f(x)\}^2 + varf(x) \qquad (3.5.18) \end{aligned}$$

which is the sum of the bias and the variance at at x. It can be seen that, as in many branches of statistics, there is a trade-off between the bias and variance terms in (3.5.18). The bias can be reduced at the expense of increasing the variance, and vice versa, by adjusting the amount of smoothing. Various methods for choosing the smoothing parameter are discussed using kernel estimators in Silverman(1986). The appropriate choice of smoothing parameter will be influenced by the purpose for which the density estimate is to be used. If the purpose of density estimation is to explore the data in order to suggest possible models and hypothesis then it is sufficient to choose the smoothing parameter subjectively. When using density estimation for presenting conclusions, some objective smoothing is required. However, many applications require an automatic choice of the smoothing parameter and such a choice can be used as a starting point for subsequent subjective adjustment. Investigators reporting and comparing their results will want to make reference to a standard method. When density estimation becomes a routine on large set of data or a part of large procedure an automatic method of choosing the smoothing parameter is essential.

A natural method for choosing the smoothing parameter is to plot out several curves and to choose the estimate that is most in accordance with ones prior ideas about the density. The process of examining several plots of the data, all smoothed by different amounts will give more insight into the data than simply considering a single automatically produced curve.

Another method is to apply the technique of *least-square cross validation* for finding the optimum value of d, the degree of polynomial, in the sense of minimizing the integrated square error. It is a completely automatic method for choosing the smoothing parameter; Silverman B.W.(1986). Given any estimator $\hat{f}$ of a density f, the integrated square error can be written as

$$\int(\hat{f}-f)^2 = \int \hat{f}^2 - 2\int \hat{f}f + \int f^2.$$

The ideal choice of degree d will correspond to the choice of degree will correspond to the choice which minimize the quantity R defined by

$$R(\hat{f}) = \int \hat{f}^2 - 2\int \hat{f}f.$$

The basic principle of least-square cross validation is to construct an estimate of $R(\hat{f})$ from the data themselves and then to minimize this estimator over d to give the appropriate choice of d. The term $\int \hat{f}^2$ can be found from the estimate $\hat{f}$. Define $\hat{f}_{-i}$ to be the estimate constructed from all the data points except X_i as

$$\hat{f}_{-i}(x) = (n-1)^{-1}\sum_{j\neq i} Q_d(X_j - x).$$

Now define

$$M_0(d) = \int \hat{f}^2 - 2n^{-1}\sum_i \hat{f}_{-i}(X_i). \tag{3.5.19}$$

Clearly the score M_0 depends only on the data. The idea of cross validation is to minimize the score M_0 over d.

An expression of score M_0 for computational purpose can be obtained as follows:

$$\begin{aligned}
\int \hat{f}^2 &= \int \hat{f}^2(x)dx \\
&= \int_0^1 n^{-1}\sum_{i=1}^{n} Q_d(X_i - x)n^{-1}\sum_{i=1}^{n} Q_d(X_j - x)dx \\
&= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} c_d^2 \int_0^1 [1-(X_i - x)^2]^d[1-(X_j - x)^2]^d dx \\
&= n^{-2}c_d^2\sum_{i=1}^{n}\sum_{j=1}^{n} \int_0^1 [1-(X_i - x)^2]^d[1-(X_j - x)^2]^d dx. \qquad (3.5.20)
\end{aligned}$$

Now

$$n^{-1}\sum_{i=1}^{n}\hat{f}_{-i}(X_i) = n^{-1}\sum_{i=1}^{n}(n-1)^{-1}\sum_{i\neq i}Q_d(X_j - X_i)$$

$$= n^{-1}(n-1)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}Q_d(X_j - X_i) - n^{-1}(n-1)^{-1}\sum_{i=1}^{n}Q_d(0)$$

$$= (n-1)^{-1}c_d\{n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}[1-(X_i - X_j)^2]^d - 1\}. \qquad (3.5.21)$$

Using (3.5.19) and (3.5.20) in (3.5.21) we get

$$M_0(d) = n^{-2}c_d^2\sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^1[1-(X_i - x)^2]^d[1-(X_j - x)^2]^d dx$$
$$-(n-1)^{-1}c_d\{n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}[1-(X_i - X_j)^2]^d - 1\}. \qquad (3.5.22)$$

The integral in (3.5.22) can be evaluated using numerical integration. $M_0(d)$ can be computed for various choices of d we chose that value of d for which $M_0(d)$ is minimum. This gives the optimum choice of the smoothing parameter d.

A much easier method for choosing the smoothing parameter d in polynomial estimation is obtained by using the *technique of Likelihood cross- validation.* The method of Likelihood cross-validation is a natural development of the idea of using likelihood to judge the adequacy of fit of a statistical model. As applied to the density estimation the rationale behind using this technique is the following. Suppose that, in additional to the original data set, an *independent observation* Y

from f were available. Then the log likelihood of f as the density underlying the observation Y would be $\log f(Y)$. Regarding f_n as a parametric family of densities depending on d, but with the data $X_1, X_2 \ldots, X_n$ fixed, we have $log f_n(Y)$, a function of d, and is the log likelihood of the smoothing parameter d. Now, since an independent observation Y is not available, we omit one of the given observations , say X_i from the original data set and use X_i instead of Y. This would give log likelihood $\log \hat{f}_{-i}(X_i)$, where

$$\begin{aligned} \hat{f}_{-i}(X_i) &= \frac{1}{n-1} \sum_{j(\neq i)}^{n} Q_{d_n}(X_j - X_i) \\ &= \frac{1}{n-1} \sum_{j(\neq i)}^{n} c_{d_n}[1 - (X_j - X_i)^2]^{d_n} \end{aligned}$$

Since the choice of X_i is independent and arbitrary, the log likelihood is averaged over each choice of the omitted X_i, we define a *score function,*

$$CV(d) = \frac{1}{n} \sum_{i=1}^{n} log \hat{f}_{-i}(X_i)$$

where $\hat{f}_{-i}$ is the density estimate constructed from all the data points except X_i, given by

$$\hat{f}_{-i}(X_i) = \frac{1}{n-1} \sum_{j(\neq i)=1}^{n} c_{d_n}[1 - (X_j - X_i)^2]^{d_n}.$$

Therefore

$$CV(d) = \frac{1}{n} \sum_{i=1}^{n} log \Big[\frac{1}{n-1} \sum_{j(\neq i)=1}^{n} c_{d_n}[1 - (X_j - X_i)^2]^{d_n} \Big].$$

The likelihood cross-validation choice of d is then the value of d which maximizes the score function $CV(d)$ for the given data $(X_1, X_2, \dots, X_n)$.

3.6 Simulation Study

In this section we investigate the performance of the proposed polynomial density estimator for various standard probability distributions. We plot the estimated density function for the optimum choice of d using likelihood cross validation as suggested in section 5, along with the true density function. The proposed non-parametric polynomial estimators for the density function $f(x)$ based on the sample $(X_1, X_2, \dots, X_n)$ are computed by

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} c_d[1 - (X_i - x)^2]^d,$$

for $x \in [0, 1]$. The constants c_d's can be computed for various values of the smoothing parameter d using the formula

$$c_d = \frac{1}{2}\frac{(2d+1)!}{2^{2d}(d!)^2}$$

as given in section 3. The table bellow gives some values of c_d for different choices of d.

d	c_d	d	c_d	d	c_d
1	0.7500	20	2.5701	55	4.2126
2	0.9375	25	2.8630	60	4.3974
3	1.0938	30	3.1286	65	4.5748
4	1.2305	35	3.3734	70	4.7456
5	1.3535	40	3.6016	75	4.9504
10	1.8501	45	3.8161	80	5.0699
15	2.2192	50	4.0193	84	5.1939

Table 4.1 : c_d values for various choices of smoothing parameter d.

We first consider Beta distribution $B(\alpha, \beta)$ with parameters $\alpha = 3$ and $\beta = 3$ with probability density function

$$f_B(x) = \frac{1}{B(3,3)} x^2(1-x)^2, 0 < x < 1.$$

and the Normal distribution $N(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma^2 = 1$ with probability density function

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty.$$

Note that the beta distribution is of support $[0, 1]$. We have generated random samples from $B(3, 3)$ and the plots of polynomial density estimators with different choices of the smoothing parameter, namely, $d = 1, 2, 3, 4, \ldots$ are obtained. The

d	c_d	d	c_d	d	c_d
1	0.7500	20	2.5701	55	4.2126
2	0.9375	25	2.8630	60	4.3974
3	1.0938	30	3.1286	65	4.5748
4	1.2305	35	3.3734	70	4.7456
5	1.3535	40	3.6016	75	4.9504
10	1.8501	45	3.8161	80	5.0699
15	2.2192	50	4.0193	84	5.1939

NB 5987
519.23 HAM/S

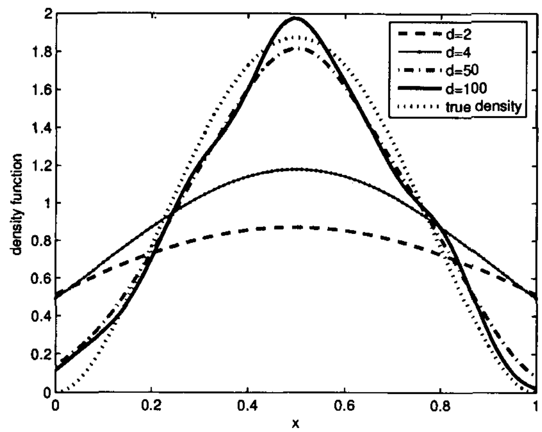

Figure 3.1: *Performance of the polynomial estimator with increasing d*

figure 3.1 shows the way in which the estimator performs as d increases when the sample size n=20. The dotted curve plots the true density. It can be easily seen that as d becomes larger the estimated density plot becomes closer to the true density plot. After a certain value of the smoothing parameter the curve becomes under smooth and depart from the shape of the true density. However the optimum choice of the parameter d can be found using the optimization technique discussed earlier in section 5.

The plots of polynomial density estimators for random samples from various standard symmetrical distributions with different choices of the smoothing parameter, namely, $d = 1, 2, 3, 4, \ldots$ revealed that as d becomes larger the estimated density plot becomes closer to the true density plot, upto a certain limit. Simulations of random samples from symmetric distributions like $B(3,3)$, $N(0,1)$ and from the mixture of two normal distributions $N(0,1)$ and $N(5,1)$ for small samples also revealed some useful methods of choosing the smoothing parameter d. It is found that the $CV(d)$ curve for symmetric data increases rapidly at the beginning and then seems to attain a maximum value and then decreases at a lower rate with increasing vales of d. We take that value of d corresponding to this unique maximum as the optimum choice for the smoothing parameter. It can be observed that this method is less sensitive to the optimum choice of the parameter d for large samples. This is clear from the figures (3.2) to (3.5).

Throughout in the simulation study we take the value of the optimum value of the smoothing parameter d as obtained by the plot of $CV(d)$ against various d.

It may be further observed that the polynomial estimator coincides with the well known kernel estimator in the case of Beta distribution $B(3,3)$, for the optimum choice of the smoothing parameter. The polynomial estimator and kernel estimator are plotted for 4 trials of the simulation experiment, each with a sample size, n=20, from $B(3,3)$ and are given in figure 3.6 with the optimum choice of $d = 30$. From the plots we realize that the polynomial estimator is very close to the true density of the underlined distribution and the kernel estimator. Polynomial

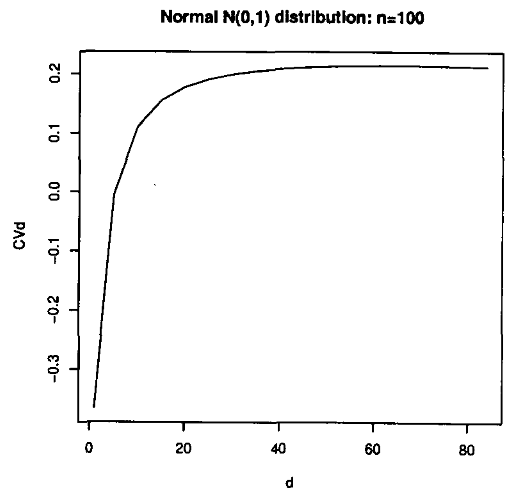

Figure 3.2: $CV(d)$ *plot of a sample of size* $n = 100$ *from* $N(0,1)$ *distribution:* $d = 65$

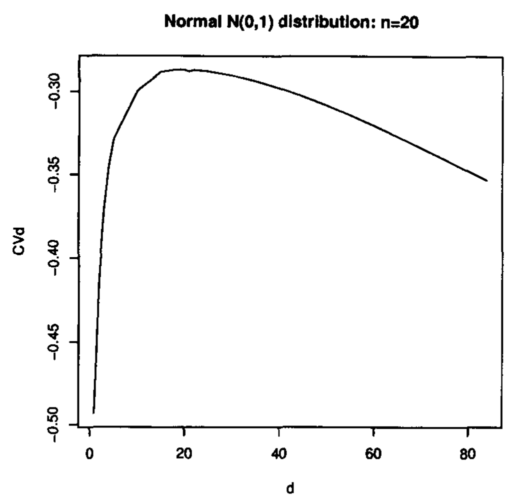

Figure 3.3: $CV(d)$ *plot of a sample of size* $n = 20$ *from* $N(0,1)$ *distribution:* $d = 20$

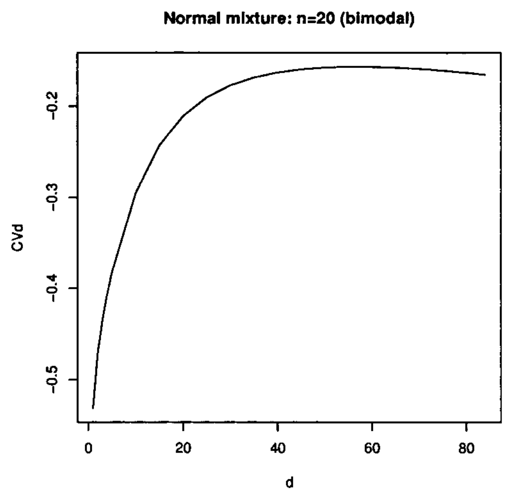

Figure 3.4: *$CV(d)$ plot of a sample of size $n = 20$ from the mixture of $N(0,1)$ and $N(5,1)$: $d = 57$*

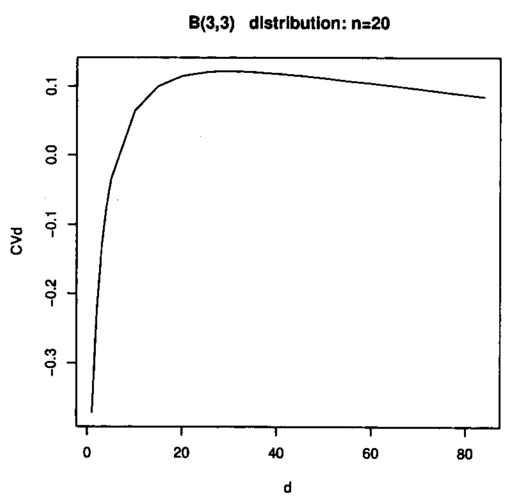

Figure 3.5: *$CV(d)$ plot of a sample of size $n = 20$ from $B(3,3)$ distribution: $d = 30$*

estimator works well even for distributions with support $[-\infty, \infty]$. To show this we consider the standard normal distribution and draw 4 random samples. The performance of the estimator in 4 trials are shown in figure 3.7. Comparison with the corresponding kernel estimator is also made in e ach trial. We use the sample size $n = 100$ and $d = 65$ as revealed from the $CV(d)$ plot in figure3.2.

We have investigated the performance of the estimator in multi- modal distributions, which are of practical importance in exploratory purpose. To do this we have generated samples of size $n = 20$ from a mixture of two normal distributions $N(0, 1)$ and $N(5, 1)$. It is found that the performance of the estimator is satisfactory even in the case of multi modal distributions. The estimated density plots along with the their kernel density estimators are displayed in figure 3.8.

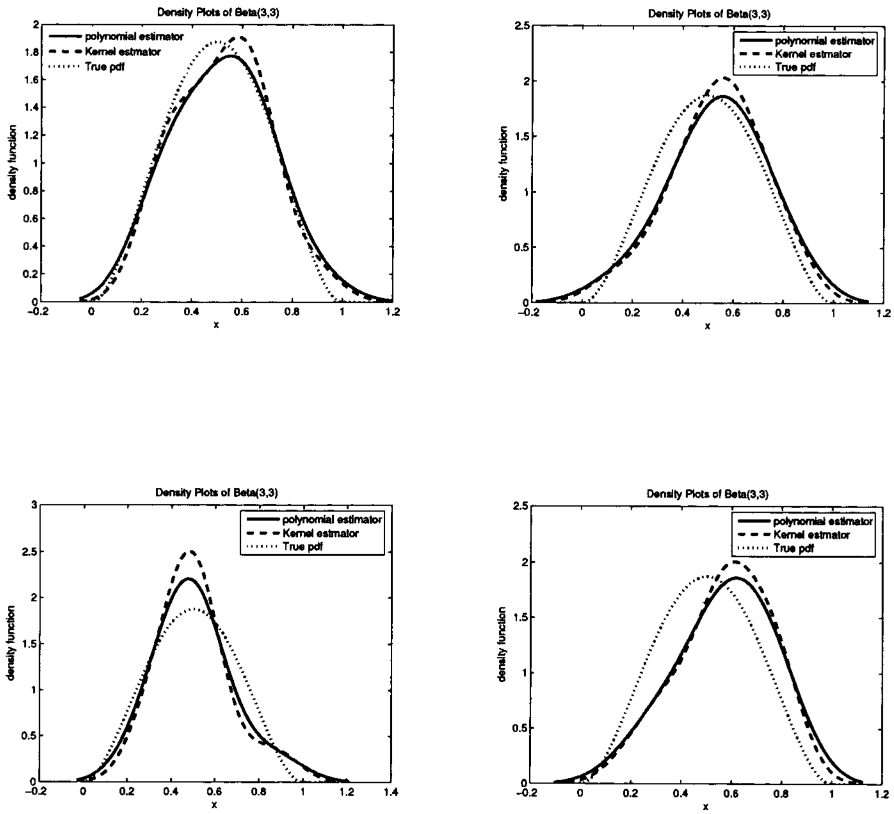

Figure 3.6: *Polynomial density estimator plots of 4 trials of simulation experiment of random sampling of size* $n = 20$ *from* $B(3,3)$*in comparisons with kernel density and true density:* $d = 30$.

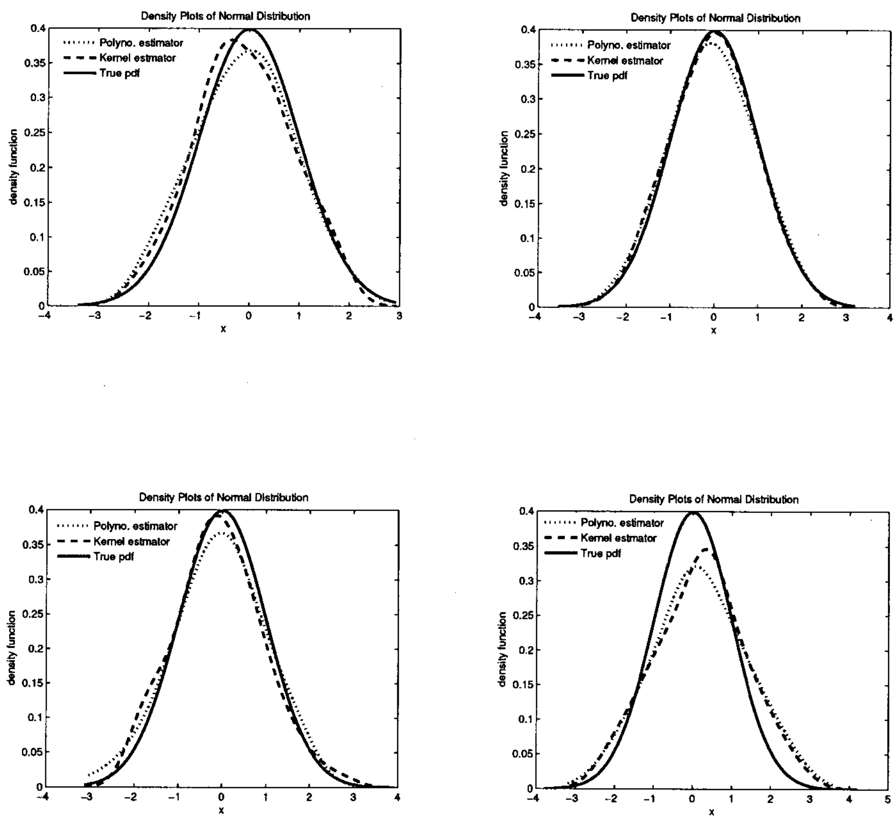

Figure 3.7: *Polynomial density estimator plots of 4 trials of simulation experiment of random sampling of size* $n = 100$ *from* $N(0,1)$ *in comparisons with kernel density and true density:* $d = 65$.

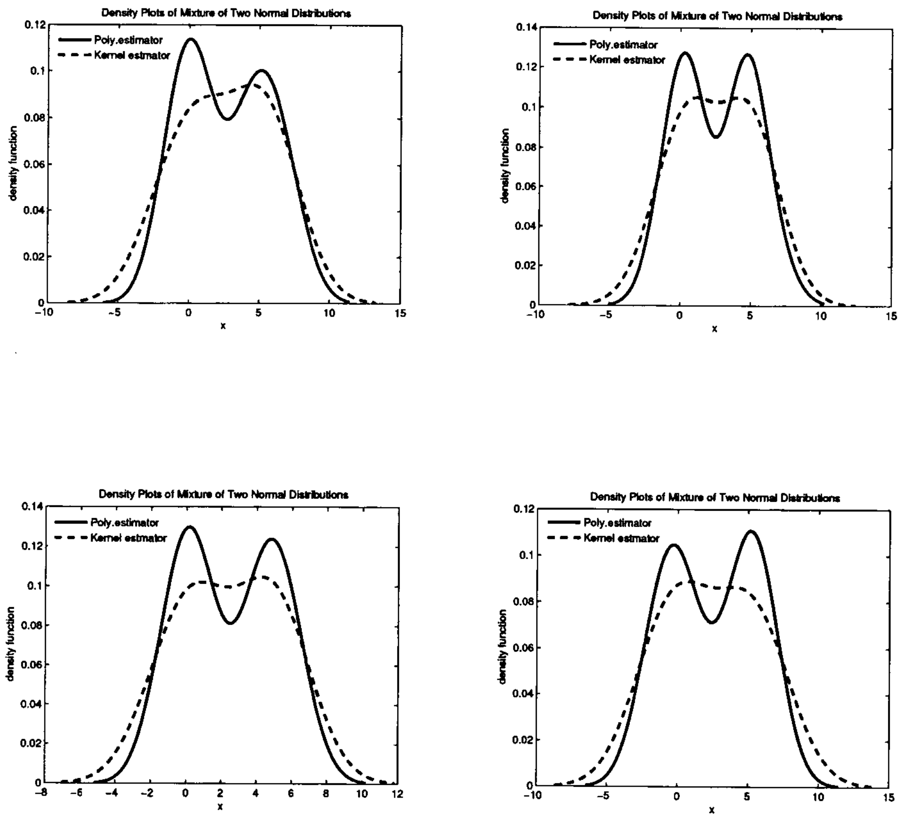

Figure 3.8: *Polynomial density estimator plots of 4 trials of simulation experiment of random sampling of size* $n = 20$ *from a mixture of two normal distributions* $N(0,1)$ *and* $N(5,1)$ *in comparisons with kernel density :* $d = 57$.

3.7 Data Analysis

In this section we first estimate the probability distribution of the data associated with the Winter Snowfall Data in Buffalo, NewYork. (Source: Scott D. W.(1992)) in 63 consecutive years from 1910 to 1972. Snowfall totals are given in inches. As an initial tool of exploration of data the histogram (Fig.3.9(left)) revealed that the snowfall data is approximately symmetrical. The density plot of the proposed polynomial estimator is drawn in Fig.3.10(left), with the appropriate value of the smoothing parameter $d = 56$ as clear from the $CV(d)$ plot displayed in Fig.3.9(right), justifies these findings.

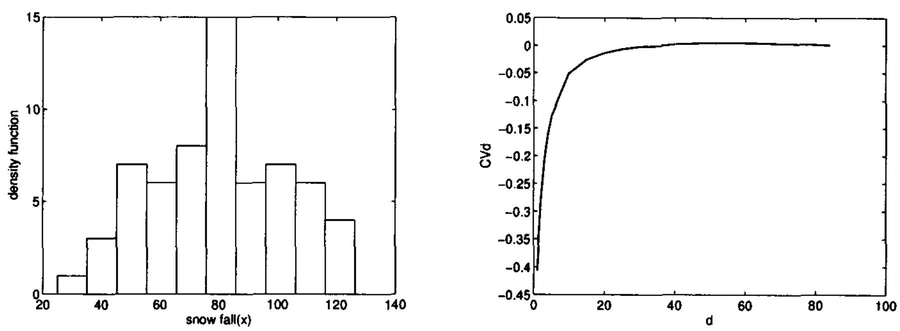

Figure 3.9: *Plots of Histogram (left) and $CV(d)$ curve (right) for the Winter Snowfall Data in Buffalo, New York: Optimum $d = 56$.*

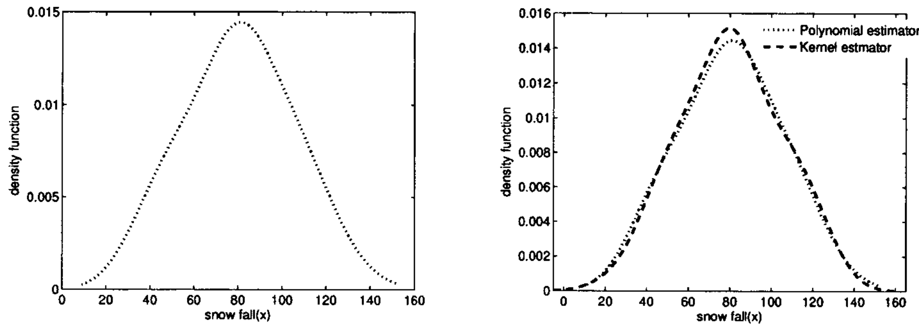

Figure 3.10: *Plots of Polynomial Estimator (left) and Comparison of Kernel Estimator and Polynomial Estimator (right) for the Winter Snowfall Data in Buffalo, New York.*

It is also verified that the kernel estimator and the proposed polynomial estimator are almost same for the snow fall data as revealed in the Fig.3.9(right). Here the kernel density is plotted with Gaussian kernels and with an automatic choice of band width.

Secondly, we estimate the probability distribution using the data on duration in minutes of eruptions of the Old Faithful geyser in Yellow Stone National Park, Wyoming, USA.; see Azzalini,A. and Bowman,A.W.(1990). 20 observations on eruptions are used to estimate the probability density function. The histogram of the data reveals that the underlined distribution is bimodal;Fig.3.11-right. The appropriate band width is chosen as $d = 48$ using the $CV(d)$ plot Fig.3.11-left.

The polynomial density estimator and comparison of proposed estimator with kernel density are also plotted in Fig.3.10.

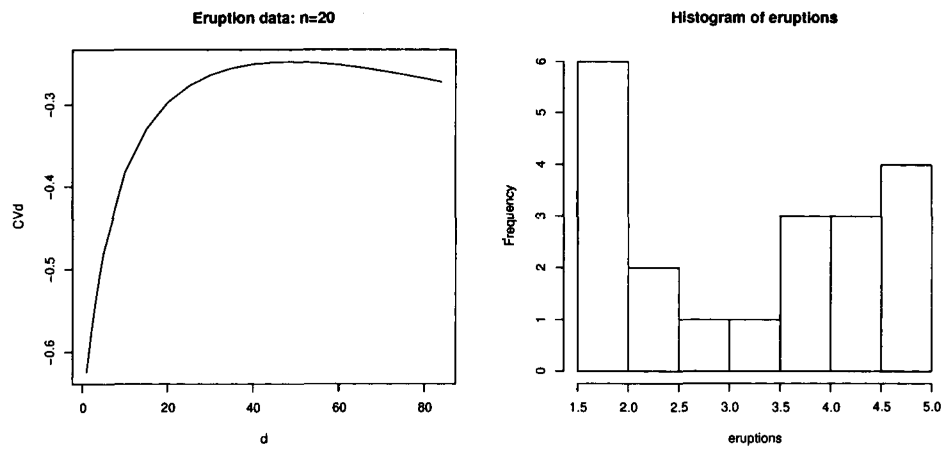

Figure 3.11: *Plots of $CV(d)$ curve(left), Histogram (right) for the Old Faithful Geyser Data: Optimum $d = 48$.*

Remark 3.7.1 *A limitation of kernel estimator is that it depends on the choice of the smoothing kernel and is sensitive to the window width. In various applications different smoothing kernels are recommended. One can choose among normal, box, triangle, epanechnikov and many others. Through out in the simulation studies and data analyses we have used normal kernel. In the proposed technique of estimation use of a sequence of polynomials*

$$Q_d(x) = c_d(1 - x^2)^d, d = 1, 2, \cdots$$

as given in section 3, is suggested. Further the simulation studies shown that the

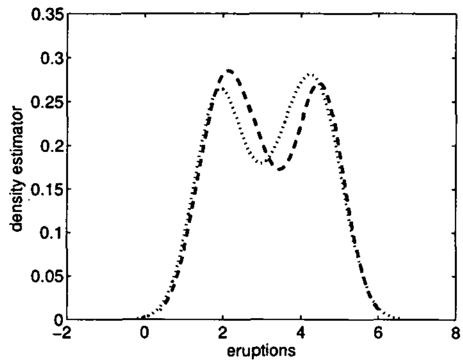

Figure 3.12: *Comparison of Kernel Estimator and Polynomial Estimator of probability density function for the Old Faithful Geyser Data: Optimum* $d = 48$.

optimum choice of smoothing parameter d, using the maximum CV(d) score is not sensitive to large samples. This approach is an excellent data-driven method of selecting the smoothing parameter of the proposed estimator. Unlike many other implementations this is immune to problems caused by multimodal densities with widely separated modes.

3.8 Concluding remarks

In most of the applications kernel based are used in practise. The proposed sequence of polynomials for density estimation is a potential tool that can be extended. We want to see how this method can be used in data mining and neural network as a powerful alternative to kernel methods. Its advantages (or limitations) in comparison with kernel based techniques are to be examined. Non parametric regression is yet another broad area of where such an estimation procedure can be attempted.. This method can be used to obtain a smooth estimator for various other models. For example to the estimation of $f^{(p)}$, the p^{th} derivative of a density f, which is useful in estimating Fisher-information when non parametric form of f is known (Prakasa Rao.B.L.S.(1983) p. 237). Another possible applications of this technique are to estimate distribution function similar to the estimation by the method of kernels (Prakasa Rao.B.L.S.(1983) p. 397) and to estimate the time dependent covariate effects in a Cox-type regression model discussed by Zukcker and Karr(1999) and Murphy and Sen(1991).

SOME CONTRIBUTIONS TO RENEWAL DENSITY ESTIMATION

Thesis submitted to the
University of Calicut

For the award of the Degree of
DOCTOR OF PHILOSOPHY

K.K. Hamsa

DEPARTMENT OF STATISTICS
FAROOK COLLEGE, KOZHIKODE
KERALA, INDIA-673632
2009

Chapter 4

Orthogonal Series Estimation of the Intensity of a Counting Process

4.1 Introduction

In some fields of science, data frequently consists of counts of the number of transitions between different statuses, such as the number of deaths or failures, the number of disablement or recoveries, or more generally the number of transitions between two states in a Markov chain. These counts may be subjected to various kinds of censoring. Even under very general censoring pattern, the number of such transitions observed may be described as a counting process.

Since N is increasing and hence a sub-martingale, it follows from the Doob-Meyer decomposition that $N = \Lambda + M$, where Λ is a predictable increasing process and M is a martingale. Since the stochastic intensity exists, we can assume that there exists non negative left continuous process $\Lambda(t)$, adapted to F_t, with right hand limits such that

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

Thus

$$N(t) = \int_0^t \lambda(s)ds + M(t)$$

or

$$M(t) = N(t) - \int_0^t \lambda(s)ds \tag{4.1.1}$$

where $M(t)$ is a square integrable martingale with variation process

$$< M, M > (t) = \int_0^t \lambda(s)ds.$$

We further assume that

$$\Lambda(t) = \alpha(t)Y(t), \tag{4.1.2}$$

where α is an unknown deterministic non-negative function called the intensity and $Y(t)$ is an observable stochastic process. Such models are usually called *multiplicative intensity models*; Aalen, O.(1978).

Multiplicative intensity model has many applications in the analysis of biological and medical data. By applying multiplicative intensity models and stochastic integrals, Aalen, O.(1976,1978) has shown how it is possible to develop non-parametric estimators for certain cumulative intensities.

Ramlau-Hansen (1983a) proposed a *kernel estimator* for the intensity α of a counting process $\{(N_n(t))\}$ by smoothing the martingale estimator of the cumulative intensity. The estimator is given by

$$\hat{\alpha}_n(t) = \frac{1}{b}\int_0^1 K\Big(\frac{t-s}{b}\Big)d\hat{\beta}(s),$$

where

$$\hat{\beta}(s) = \int_0^s \frac{J_n(u)}{Y_n(u)}dN_n(u) \quad \text{with} \quad J_n(s) = I\{Y_n(s) > 0\}.$$

Apart from usual consistency, a stronger consistency viz. *mean square consistency*, of the kernel estimator is established in the sense that

$$E[\sup_{t\in[z_0,z_1]} |\hat{\alpha}(t) - \alpha(t)|^2] \to 0$$

when $n \to \infty$, $b \to 0$ and $nb^2 \to 0$ and for fixed interval $[z_0, z_1]$ with $0 < z_0 < z_1 < 1$. Further, asymptotic normality is also established, under some conditions, and in the sense that

$$(nb_n)^{\frac{1}{2}}\{\hat{\alpha}_n(t) - \alpha_n^*(t)\},$$

where

$$\alpha_n^*(t) = \frac{1}{b}\int_0^\infty K\Big(\frac{t-s}{b}\Big)d\beta^*(s)$$

and

$$\beta^*(t) = \int_0^t \alpha(s)J(s)ds,$$

converges in distribution towards a normal distribution with mean 0 and variance

$$\frac{\alpha(t)}{\tau(t)}\int_{-1}^{1} K^2(u)du$$

when $n \to \infty$, $b_n \to 0$ and $nb_n \to \infty$, assuming that

$$\frac{nJ_n}{Y_n} \to \frac{1}{\tau}.$$

Here b_n is the *band width* of the kernel estimator.

Another development was initiated by Karr, A.F. (1987) in this direction. He observed that the method of maximum likelihood cannot be used directly to estimate the intensity and suggested a sieve using the regularity criteria

$$S(m_n) = \{\alpha \geq 0 : (\frac{1}{m_n} \leq \alpha \leq m_n, |\alpha'| < m_n\alpha)\},$$

where m_n is a parameter chosen based on the sample and α' is the derivative of α. Later, Leskoww and Rozanski(1989) derived a histogram sieve estimator for α and studied their asymptotic properties. The asymptotic variance of each of these estimators is closely related with that of Ramlau-Hansen's estimator.

Anilkumar and Naik-Nimbalkar (1995) used a kernel function to smooth the likelihood and obtained an estimate of the intensity. The technique thus developed is then applied to estimate the hazard rate in survival analysis and the transition intensities in Markov chain models under very general type of censoring. The technique is also extended to estimate the time dependent infection rate in simple as well as general epidemic models. Since the direct maximization of the log-likelihood function

$$L_n = \sum_{i=1}^{n} \Big[\int_0^1 (1 - \alpha(u)) Y_i(u) du + \int_0^1 log(\alpha(u)) dN_i(u) \Big]$$

is meaningless as it is unbounded, in order to estimate $\alpha(s)$, they considered a portion of the likelihood which gives more emphasis to the behavior of the process around s. This is accomplished by computing the kernel log-likelihood approximation

$$L_n^*(K, h, s) = (1-\alpha(s))\frac{1}{h}\int_0^1 K(\frac{s-u}{h})Y_n(u)du + \log(\alpha(s))\frac{1}{h}\int_0^1 K(\frac{s-u}{h})dN_n(u),$$

where K is a non-negative smooth symmetric function having support $[-1, 1]$ and satisfying

$$\int_{-1}^{1} K(u)du = 1$$

and h is the bandwidth parameter. The Maximum modified kernel likelihood estimator (MMKL estimator) is thus obtained as

$$\hat{\alpha}_h(s) = \frac{\int_0^1 K(\frac{s-u}{h})dN_n(u)}{\int_0^1 K(\frac{s-u}{h})Y_n(u)du}$$

In this chapter we have derived an orthogonal series estimator for the intensity function $\alpha(t)$. This chapter is orgonised as follows. In section 2 we defined a sequence of orthogonal series estimators of the intensity function α. The asymptotic unbiasedness of the estimator is also established under some regularity conditions. A computation formula is provided for the proposed estimator using Fourier series. The consistency and asymptotic normality are established in section 3. In section 4 we proposed a powerful criterion for selecting the appropriate value of the smoothing parameter d of the estimator using Nelsen-Aalen estimator of the cumulative hazard rate β. Simulation studies are conducted to show the closeness of the proposed estimator to the true intensity.

4.2 Orthogonal Series Estimation

Assume that the intensity function $\alpha \in L_2[0,1]$. Let $\{\phi_k, k = 1, 2, \ldots\}$ be a complete orthonormal basis of $\alpha \in L_2[0,1]$. Therefore there exists a unique square summable sequence $\{c_k\}$, so that

$$\alpha(t) = \sum_{k=0}^{\infty} c_k \phi(t)$$

where

$$c_k = \int_0^1 \alpha(t)\phi_k(t)dt.$$

Also

$$\begin{aligned}\|\alpha\|^2 &= \int_0^\infty \alpha^2(t)\phi(t)dt \\ &= \sum_{k=1}^{\infty} c_k^2.\end{aligned}$$

The current non-parametric theory concentrates on the cumulative intensity function

$$\beta(t) = \int_0^t \alpha(s)ds.$$

From (4.1.1) and(4.1.2) we have the representation

$$dN(t) = \alpha(t)Y(t)dt + dM(t) \tag{4.2.3}$$

Here $dM(t)$ may be considered as a *random noise*; Karg Anderson et.al.(1993, p.178). Corresponding to a sequence of $\{N_n\}$ of one-dimensional counting process, each with an intensity process of the form:

$$\Lambda_n(t) = \alpha(t)Y_n(t),$$

one may construct a corresponding sequence of *orthogonal series estimators*, which we denote by $\hat{\alpha}_n(t)$ and is

$$\hat{\alpha}_n(t) = \sum_{k=0}^{d_n} \hat{c}_k\phi_k(t),$$

d_n being constants depend on the number n and is to be chosen in a suitable manner and $\hat{c}_k$'s are defined by

$$\hat{c}_k = \int_0^1 \phi_k(s)d\hat{\beta}_n(s) \tag{4.2.4}$$

where $\hat{\beta}(t)$ is a martingale estimator of the cumulative intensity function

$$\beta(t) = \int_0^t \alpha(s)ds$$

From (4.2.3) $\hat{\beta}_n$ can be defined by

$$\hat{\beta}_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)} \tag{4.2.5}$$

From (4.2.4) and (4.2.5) we have

$$\hat{c}_k = \int_0^1 \phi_k(t)\frac{dN_n(t)}{Y_n(t)} \tag{4.2.6}$$

Theorem 4.2.1 *$E(\hat{\alpha}_n(t)) \to \alpha(t)$ as $n \to \infty$ if $P(Y_n(s) > 0) = 1$ and $d_n \to \infty$ as $n \to \infty$, and $d_n \to \infty$, $\frac{d_n}{n} \to 0$.*

Proof: We have from (4.2.6)

$$\begin{aligned}
E(\hat{c}_k) &= E\int_0^1 \phi_k(t)\frac{dN_n(t)}{Y_n(t)} \\
&= E\int_0^1 \phi_k(t)\frac{1}{Y_n(t)}[\alpha(t)Y_n(t)dt + dM_n(t)], \text{using (4.2.3)} \\
&= E\int_0^1 \phi_k(t)\alpha(t)dt \\
&= c_k.
\end{aligned}$$

$$\begin{aligned}
\text{Now} \quad E(\hat{\alpha}_n(t) &= E\sum_{k=1}^{d_n} \hat{c}_k\phi_k(t) \\
&= \sum_{k=1}^{d_n} c_k\phi_k(t)
\end{aligned}$$

Clearly $\hat{\alpha}_n(t)$ is a biased estimate for α with biase,

$$B(\hat{\alpha}_n(t)) = \sum_{k=d_n+1}^{\infty} c_k\phi_k(t) \tag{4.2.7}$$

tends to zero as $d_n \to \infty$.

This proves the theorem.

Remark 4.2.1 *A formula for computation of the estimator using the Fourier series (sine series) can be provided as follows:*

$$\hat{\alpha}_n(t) = \sum_{k=0}^{d} \hat{c}_k \phi_k(t)$$

Assuming that the process is observed over $[0, L]$*,* L *being finite, we have*

$$\hat{\alpha}_n(t) = \frac{2}{L} \sum_{k=1}^{d} \hat{c}_k \sin(\frac{k\pi t}{L})$$

where

$$\begin{aligned}
\hat{c}_k &= \int_0^L \phi_k(t) \frac{dN_n(t)}{Y_n(t)} \\
&= \sum_{i=1}^{n} \frac{\phi_k(T_i)}{Y_i(T_i)} \\
&= \frac{2}{L} \sum_{i=1}^{n} \frac{\sin(\frac{k\pi T_i}{L})}{Y_i}
\end{aligned}$$

A similar formula using the cosine series is

$$\begin{aligned}
\hat{\alpha}_n(t) &= \sum_{k=0}^{d} \hat{c}_k \phi_k(t) \\
&= \frac{\hat{c}_0}{2} + \frac{2}{L} \sum_{k=1}^{d} \hat{c}_k \cos(\frac{k\pi t}{L})
\end{aligned}$$

where

$$\begin{aligned}\hat{c}_0 &= \int_0^L \phi_0(t)\frac{dN_n(t)}{Y_n(t)} \\ &= \frac{2}{L}\sum_{i=1}^{n}\frac{1}{Y_i}\end{aligned}$$

and

$$\begin{aligned}\hat{c}_k &= \int_0^L \phi_k(t)\frac{dN_n(t)}{Y_n(t)} \\ &= \sum_{i=1}^{n}\frac{\phi_k(T_i)}{Y_i(T_i)} \\ &= \frac{2}{L}\sum_{i=1}^{n}\frac{\cos(\frac{k\pi T_i}{L})}{Y_i}\end{aligned}$$

for $k = 1, 2, \ldots$.

Therefore

$$\hat{\alpha}_n(t) = \frac{1}{L}\sum_{i=1}^{n}\frac{1}{Y_i} + \frac{2}{L}\sum_{k=1}^{d}\hat{c}_k\cos(\frac{k\pi t}{L})$$

4.3 Asymptotic Properties

4.3.1 Consistency

We shall now study the asymptotic properties of the proposed estimator. The consistency of the estimator $\hat{\alpha}_n(t)$ is established in the following theorem.

Theorem 4.3.1 *If*

(i) $nE\Big(\frac{1}{Y_n(t)}\Big) \to \frac{1}{\tau}$, *uniformly in the neihbourhood of* t *and*

(ii) α and τ are continuous at the point t. Then

$$E\|\hat{\alpha}_n - \alpha\|^2 \to 0 \quad as \quad n \to 0.$$

Proof: We have

$$\begin{aligned} E\|\hat{\alpha}_n - \alpha\|^2 &= E\|\hat{\alpha}_n - \alpha^* + \alpha^* - \alpha\|^2 \\ &= E\|\hat{\alpha}_n - \alpha^*\|^2 + E\|\alpha^* - \alpha\|^2 \end{aligned} \tag{4.3.8}$$

where

$$\alpha^*(t) = \sum_{k=1}^{d_n} c_k \phi_k(t),$$

so that

$$\hat{\alpha}_n(t) - \alpha^*(t) = \sum_{k=1}^{d_n} (\hat{c}_k - c_k)\phi_k(t).$$

Therefore

$$\|\hat{\alpha}_n - \alpha^*\|^2 = \sum_{k=1}^{d_n} (\hat{c}_k - c_k)^2$$

using the property of orthogonal basis.

Hence,

$$\begin{aligned} E\|\hat{\alpha}_n - \alpha^*\|^2 &= E\sum_{k=1}^{d_n} (\hat{c}_k - c_k)^2 \\ &= \sum_{k=1}^{d_n} E(\hat{c}_k - c_k)^2. \end{aligned}$$

We have

$$c_k = \int_0^1 \alpha(t)\phi_k(t)dt$$

and

$$\hat{c}_k = \int_0^1 \phi_k(t) \frac{dN_n(t)}{Y_n(t)}$$

so that

$$\hat{c}_k - c_k = \int_0^1 \phi_k(t) \frac{dM_n(t)}{Y_n(t)}$$

Hence

$$\begin{aligned} E(\hat{c}_k - c_k)^2 &= \int_0^1 \frac{\phi_k^2(t)}{Y_n^2(t)} d < M_n(t) > \\ &= E \int_0^1 \frac{\phi_k^2(t)}{Y_n^2(t)} Y_n(t) \alpha(t) dt, \end{aligned}$$

since, $< M_n(t) >= \int_0^1 \Lambda_n(t) ds$ and $\Lambda_n(t) = \alpha(t) Y_n(t)$. Therefore

$$\begin{aligned} E(\hat{c}_k - c_k)^2 &= E\Big[\int_0^1 \frac{\phi_k^2(t)}{Y_n(t)} \alpha(t) dt\Big] \\ &= \frac{1}{n} \int_0^1 \phi_k^2(t) n E\Big[\frac{1}{Y_n(t)}\Big] \alpha(t) dt \end{aligned}$$

By assumption (i) $E[\frac{1}{Y_n(t)}] \to \frac{1}{\tau}$, uniformly in the neibourhood of t and also that α and τ are continuous at the point t we have

$$E(\hat{c}_k - c_k)^2 \to \frac{1}{n} \int_0^1 \phi_k^2(t) \frac{\alpha(t)}{\tau(t)} dt.$$

Hence

$$\begin{aligned} E\|\hat{\alpha} - \alpha^*\|^2 &\to \frac{1}{n} \sum_{k=1}^{d_n} \int_0^1 \phi_k^2(t) \frac{\alpha(t)}{\tau(t)} dt \\ &\to 0, \end{aligned} \tag{4.3.9}$$

assuming $\tau > 0$ as $\dfrac{d_n}{n} \to 0$, $\dfrac{\alpha(t)}{\tau(t)}$ is bounded and since

$$\int_0^1 \phi_k^2(t)dt = 1.$$

Further, since

$$\begin{aligned}\alpha^*(t) - \alpha(t) &= \sum_{k=1}^{d_n} c_k\phi_k(t) - \sum_{k=1}^{\infty} c_k\phi_k(t) \\ &= -\sum_{k=d_n+1}^{\infty} c_k\phi_k(t)\end{aligned}$$

we have

$$\begin{aligned}E\|\alpha^* - \alpha\|^2 &= \sum_{d_n+1}^{\infty} c_k^2 \\ &\to 0 \quad \text{as} \quad d_n \to \infty. \qquad (4.3.10)\end{aligned}$$

Using(4.3.9) and (4.3.10) in (4.3.8) we can easily see that

$$E\|\hat{\alpha}_n - \alpha\|^2 \to 0 \quad as \quad n \to 0 \quad and \quad d_n \to \infty.$$

This proves the result.

4.3.2 Asymptotic Normality

Ramlau-Hansen (1983) proved that the kernel estimator of the intensity is asymptotically normal. This was established by making use of the result about the asymptotic distribution of a martingale triangular array, proved by Lipster and Shiryayev(1980) and Shiryayev(1981).

Consider a sequence of counting process N_n on $[0,1]$ with a corresponding sequence of martingales given by

$$M_n(t) = N_n(t) - \int_0^1 \Lambda_n(s)ds$$

where $\{\Lambda_n\}$ is the sequence of intensity process. Let H_n be a sequence of predictable processes where

$$E\int_0^1 H_n^2(s)\Lambda_n(s)ds < \infty$$

and introduce

$$\tilde{M}_n(t) = \int_0^t H_n(s)dM_n(s)$$

Then, we have the following proposition, proved by Ramlau-Hansen(1983).

Proposition 4.3.1 *Suppose that*

(i) For every $\epsilon > 0 : \int_0^1 H_n^2(s)I(|H_n(s)| > \epsilon)\Lambda_n(s)ds \to 0$ *in probability*

and

(ii) $\int_0^1 H_n^2(s)\Lambda_n(s)ds \to 1$ *in probability when* $n \to \infty$.

Then $\tilde{M}_n(1) \to N(0,1)$ *in distribution, where* $N(0,1)$ *is the standard normal distribution.*

Here instead of establishing the asymptotic normality of the estimator

$$\hat{\alpha}_n(t) = \sum_{k=1}^{d_n} \hat{c}_k\phi_k(t),$$

our discussion will first confined to showing the asymptotic normality of the vector $(\hat{c}_1.\ldots,\hat{c}_d)$ for a fixed d. This entails the asymptotic distribution of $\sum_{k=1}^{d} c_k\phi_k(t)$ for fixed d.

To do this we will consider a linear combination of the vector $(c_1, c_2, \ldots, c_d)$ for an arbitrary fixed value of d. Let $l_1c_1+l_2c_2+\ldots l_dc_d$ be a linear combination of $(c_1, c_2, \ldots, c_d)$. In the following theorem we shall show that the the linear combination $l_1c_1+l_2c_2+\ldots l_dc_d$ follows univariate normal distribution which ensures that the vector of coefficients $(c_1, c_2, \ldots, c_d)$ follows multivariate normal distribution.

Theorem 4.3.2 *Assume that*

(i) $\frac{n}{Y_n} \to \frac{1}{\tau}$ *uniformly in a neihbourhood of* t *as* $n \to \infty$

and

(ii) the functions α *and* τ *are continuous at the point* t.

Then, for any fixed, but arbitrary, d

$$\sqrt{n}(\sum_{k=1}^{d} l_k\hat{c}_k - \sum_{k=1}^{d} l_kc_k) \to N(0,\xi_d)$$

where

$$\xi_d = \int_0^1 \frac{\alpha(s)}{\tau(s)} \sum_{p=1}^{d}\sum_{q=1}^{d} l_pl_q\phi_p(s)\phi_q(s)ds.$$

Proof: We have

$$\hat{c}_k - c_k = \int_0^1 \phi_k(s)\frac{dM_n(s)}{Y_n(s)} \tag{4.3.11}$$

Therefore

$$\begin{aligned}
\sqrt{n}(\sum_{k=1}^{d} l_k\hat{c}_k - \sum_{k=1}^{d} l_k c_k) &= \sqrt{n}\sum_{k=1}^{d} l_k(\hat{c}_k - c_k) \\
&= \sqrt{n}\sum_{k=1}^{d} l_k \int_0^1 \phi_k(s)\frac{dM_n(s)}{Y_n(s)} \\
&= \int_0^1 \Big[\frac{\sqrt{n}\sum_{k=1}^{d} l_k\phi_k(s)}{Y_n(s)}\Big] dM_n(s) \\
&= \int_0^1 H_n(s)dM_n(s)
\end{aligned}$$

where

$$H_n(s) = \frac{\sqrt{n}\sum_{k=1}^{d} l_k\phi_k(s)}{Y_n(s)}.$$

Introduce

$$\tilde{M}_n(t) = \int_0^t H_n(s)dM_n(s).$$

We shall now apply the proposition (4.3.1) by verifying conditions (i) and (ii).

Consider

$$\begin{aligned}
I\Big(|H_n(s)| > \epsilon\Big) &= I\Big(\Big|\frac{\sqrt{n}\sum_k^d l_k\phi_k(s)}{Y_n(s)}\Big| > \epsilon\Big) \\
&= I\Big(\Big|\frac{n\sum_k^d l_k\phi_k(s)}{Y_n(s)}\Big| > \sqrt{n}\epsilon\Big) \\
&\to I\Big(\Big|\frac{1}{\tau(s)}\sum_k^d l_k\phi_k(s)\Big| > \sqrt{n}\epsilon\Big) \\
&\leq I\Big(\frac{1}{\tau(s)}M > \frac{\sqrt{n}\epsilon}{\sum_k^d l_k}\Big),
\end{aligned}$$

M bieng an upper bound of $\phi_k(t)$, which shows that

$$I\Big(|H_n(s)| > \epsilon\Big) \to 0 \text{ as } n \to 0$$

uniformly on $[0, 1]$.

Thus

$$\int_0^1 H_n^2(s) I(|H_n(s)| > \epsilon)\alpha(s) Y_n(s) ds \to 0$$

in probability and the condition (i) of the proposition (4.3.1) is verified.

Again since

$$\begin{aligned}
\int_0^1 H_n^2(s)\Lambda_n(s)ds &= \int_0^1 H_n^2(s)\alpha(s)Y_n(s)ds \\
&= \int_0^1 \frac{n}{Y_n^2(s)}\Big(\sum_{k=1}^d l_k\phi_k(s)\Big)^2 \alpha(s)Y_n(s)ds \\
&\to \int_0^1 \frac{\alpha(s)}{\tau(s)} \sum_{p=1}^d \sum_{q=1}^d l_p l_q \phi_p(s)\phi_q(s)ds
\end{aligned}$$

in probability, the condition (ii) is verified.

Applying the proposition (4.3.1), it follows immediately that

$$\hat{M}_n(1) = \sqrt{n}\Big(\sum_{k=1}^d l_k\hat{c}_k - \sum_{k=1}^d l_k c_k\Big) \to N(0, \xi_d)$$

where

$$\xi_d = \int_0^1 \frac{\alpha(s)}{\tau(s)} \sum_{p=1}^d \sum_{q=1}^d l_p l_q \phi_p(s)\phi_q(s)ds.$$

This proves the theorem.

This theorem implies that the vector of coefficients in the proposed estimator, $(c_1, c_2, \ldots, c_d)$ follows multivariate normal distribution. The result is stated in the following corollary.

corollary 4.3.1 *Let* $\hat{\mathbf{c}}_\mathbf{d}$=$(\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_d)$. *Assume that (i)* $\frac{n}{Y_n} \to \frac{1}{\tau}$ *uniformly in the neigbourhood of* t *as* $n \to \infty$, *and (ii) the functions* α *and* τ *are continuous at the point* t. *Then*

$$\hat{\mathbf{c}}_\mathbf{d} \to N_d(\mathbf{c}_\mathbf{d}, \mathbf{\Sigma}_\mathbf{d})$$

where

$$\begin{aligned} \mathbf{c}_\mathbf{d} &= (c_1, c_2, \ldots, c_d) \textit{ the mean vector and }, \\ \mathbf{\Sigma}_\mathbf{d} &= [\xi_{pq}], \textit{ the variance covariance matrix} \end{aligned}$$

with the covariances given by

$$\xi_{pq} = \int_0^1 \frac{\alpha(s)}{\tau(s)} l_p l_q \phi_p(s) \phi_q(s) ds,$$

$p, q = 1, 2, \ldots, d$

Now for any fixed t the proposed estimator $\hat{\alpha}_n(t)$ will follow normal distribution. The result is given in the following corollary.

corollary 4.3.2 *Assume that (i)* $\frac{n}{Y_n} \to \frac{1}{\tau}$ *uniformly in the neigbourhood of* t *as* $n \to \infty$, *and (ii) the functions* α *and* τ *are continuous at the point* t. *Then*

$$\hat{\alpha}_n(t) \to N(\alpha^*(t), \sigma_d^2(t))$$

where

$$\alpha^*(t) = \sum_{k=1}^{d} c_k \phi_k(t)$$

$$\sigma_d^2(t) = \int_0^1 \frac{\alpha(s)}{\tau(s)} \sum_{p=1}^{d} \sum_{q=1}^{d} \phi_p(t)\phi_q(t)\phi_p(s)\phi_q(s)ds$$

Remark 4.3.1 *To be more realistic we should tackle the situation where the parameter d depends on the sample size n and $d \to \infty$ as $n \to \infty$. But the situation looks intractable as we will not get an expression for the limiting variance.*

Example 4.3.1 *Consider independent identically distributed failure times $X_1, X_2, \ldots, X_n$ with values in $[0, \infty)$ and a hazard rate α and a distribution function F. Let $T_1, T_2, \ldots, T_n$ be the corresponding censoring times with distribution function H. Assume that the failure time is independent of the censoring time. The number of failures in time $[0, t]$, that is,*

$$N_n(t) = \sum_{i=1}^{n} I[X_i \le t, X_i \le T_i]$$

is a counting process with stochastic intensity $\Lambda_n(t) = \alpha(t)Y_n(t)$ where

$$Y_n(t) = \sum_{i=1}^{n} I[X_i \ge t, T_i \ge t]$$

denotes the number of individuals under observations just before time t. The orthogonal series estimator of α is given by

$$\hat{\alpha}(t) = \sum_{k=1}^{d} \hat{c}_k \phi_k(t), \text{ where } \hat{c}_k = \int_0^L \phi_k(t) \frac{dN_n(t)}{Y_n(t)} = \sum_{i=1}^{n} \frac{\phi_k(T_i)}{Y_i(T_i)} D_i$$

where D_i is the indicator of the death of the i^{th} individual. To verify the consistency of the of the above estimator it is enough to see whether $E\Big(\frac{n}{Y_n(s)}\Big)$ is convergent. Since $Y_n(s)$ is binomially distributed with parameter n and $[(1 - F(s)(1 - H(s-))]$, it follows that (Aalen,1976, Lemma4.2)

$$\begin{aligned} E[\frac{n}{Y_n(s)}] &\rightarrow \Big[(1 - F(s)(1 - H(s-))\Big]^{-1} \\ &= \Big(1 - H(s-)\Big)^{-1} exp\Big[\int_0^s \alpha(u)du\Big] \end{aligned}$$

uniformly in [0, 1]. Thus if we assume that $\alpha(s)$ is continuous, the proposed estimator will be consistent.

Example 4.3.2 *Now we extend example 4.3.1. to a multiple decrement model, which is a continuous time Markov chain with one transient state* 0 *and m absorbing states* 1 *to m. The model is often used to analyse different causes of decrement in demography and actuarial science. We denote the transient probability and intensity from state 0 to 1 by α_i and $P_{0i}(s,t)$ respectively. Assume $P_{00}(0,1) = exp\{-\int_0^t \alpha(s)ds\} > 0$, where $\alpha = \sum \alpha_i$.*

Consider n independent Markov chains of that kind and assume that each process starts from state 0 at time 0. If we denotes the sample paths of an individual process by $S_j(t)$, it follows that

$$N_n^i(t) = \sum_{j=i}^{n} I(S_j(t) = i), \quad i = 1, 2, \ldots, m,$$

is the number of transitions to state i during $[0, t]$. Then each N_n^i is a counting process with corresponding intensity process

$$\Lambda_n^i(t) = \alpha_i(t) Y_n(t), \quad i = 1, 2, \ldots, m,$$

where $Y_n(t) = n - N_n(t^-)$, and $N_n(t) = \sum_{j=1}^{n} N_n^i(t)$. The orthogonal series estimator of $\alpha_i(s)$ is

$$\hat{\alpha}_i(t) = \sum_{k=1}^{d} \hat{c}_k \phi_k(t), \ \textit{where}\ \hat{c}_k = \int_0^L \phi_k(t) \frac{dN_n^i(t)}{Y_n(t)}.$$

4.4 Choice of Smoothing Parameter

In the proposed estimator the choice of d, the smoothing parameter is crucial as this decides the shape of the plotted curve of the estimator. We have the Nelsen-Aalen estimator for the cumulative hazard rate,

$$\beta(t) = \int_0^t \alpha(s) ds,$$

given by

$$\hat{\beta}(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)}.$$

Using the proposed estimator of the intensity function $\alpha(t)$, given by

$$\hat{\alpha}(t) = \sum_{k=1}^{d} \hat{c}_k \phi_k(t),$$

a reasonable choice of the parameter d can be obtained by minimizing

$$\begin{aligned} S(d) &= \sup_t \left| \int_0^t \hat{\alpha}(s)ds - \hat{\beta}(t) \right| \\ &= \sup_t \left| \sum_{k=1}^{d} \hat{c}_k \int_0^t \phi_k(s)ds - \hat{\beta}(t) \right|. \end{aligned} \quad (4.4.12)$$

This criterion is used in the simulation studies with Fourier series.

4.5 Simulation Study

In this section we study the performance of the proposed non-parametric estimator of the intensity where the failure times follows a Weibull distribution with shape parameter a and scale parameter b and has density given by

$$f(x) = (a/b)(x/b)^{(a-1)} exp(-(x/b)^a)$$

for $x \geq 0$. The cumulative distribution function is

$$F(x) = 1 - exp(-(x/b)^a) \text{ on } x \geq 0,$$

and the intensity function of the corresponding counting process is given by

$$\alpha(t) = \frac{a}{b}\left(\frac{t}{b}\right)^{a-1}.$$

Weibull distribution is particularly useful in investigating the behavior of the proposed estimator because for various choices of the shape parameter a and scale parameter b it gives definite shapes for the intensity function α. This enable to compare the estimators using simulation samples with the true intensity functions.

Simulation experiments are conducted for for $a = 1, 2,$ and 3 fixing $b = 1$. We plot the estimated intensity functions for deferent choices of d along with the true intensity functions. The proposed non-parametric orthogonal series estimators for the intensity function $\alpha(t)$ based on the sample $(X_1, X_2, \ldots, X_n)$ are computed by Fourier sine series

$$\begin{aligned} \hat{\alpha}_n(t) &= \sum_{k=0}^{d} \hat{c}_k \phi_k(t) \\ &= \frac{2}{L} \sum_{k=1}^{d} \hat{c}_k \sin(\frac{k\pi t}{L}) \end{aligned}$$

where

$$\begin{aligned} \hat{c}_k &= \int_0^L \phi_k(t) \frac{dN_n(t)}{Y_n(t)} \\ &= \sum_{i=1}^{n} \frac{\phi_k(T_i)}{Y_i(T_i)} \\ &= \frac{2}{L} \sum_{i=1}^{n} \frac{\sin(\frac{k\pi T_i}{L})}{Y_i}. \end{aligned}$$

Since

$$\begin{aligned} \int_0^t \sin \frac{k\pi s}{L} ds &= -\Big(\frac{1}{k\pi/L}\Big) \cos \frac{k\pi s}{L}\Big]_0^t \\ &= \frac{L}{k\pi}\Big(1 - \cos \frac{k\pi t}{L}\Big) \end{aligned}$$

and

$$\hat{\beta}(t) = \sum_{T_i<t} \frac{1}{Y_n(T_i)} \tag{4.5.13}$$

the choice of the parameter d is obtained by minimizing

$$S(d) = \sup_t \left| \sum_{k=1}^{d} \hat{c}_k \frac{L}{k\pi}\left(1 - \cos\frac{k\pi t}{L}\right) - \sum_{T_i \le t} \frac{1}{Y_n(T_i)} \right|, \tag{4.5.14}$$

using the criterion given equation(4.5.12).

The figure 4.1 shows the way in which the orthogonal series estimator performs as d differs; we used a random sample of size $n = 500$ from Weibull distribution with shape parameter $a = 2$ and scale parameter $b = 1$. The dotted curve plots the true intensity function, $\alpha(t) = 2t$. It can be easily seen that as d becomes larger the estimated intensity plot becomes closer to the true intensity plot. After a certain value of the smoothing parameter the curve becomes under smooth and depart from the shape of the true density.

However the optimum choice of the parameter d can be found using the optimization technique discussed earlier in equation (4.5.14). Table 4.1 shows that the suprema for various values differ and minimum of them attains at $d = 3$.

d	1	2	3	4	5
$S(d)$	111.570	10.882	9.639	10.499	13.227
d	6	7	8	9	10
$S(d)$	12.366	11.699	12.087	12.087	12.193

Table 4.1 : $S(d)$ values for various choices of smoothing parameter d; Weibull with $a = 2$ and $b = 1$.

Using the appropriate value of the smoothing parameter, namely $d = 3$ we have the optimal orthogonal series estimator of the intensity function α as plotted in the figure 4.2.

Secondly we used a random sample of size $n = 500$ from Weibull distribution with shape parameter $a = 3$ and scale parameter $b = 1$. whose true intensity function is, $\alpha(t) = 3t^2$. Plots of the intensity function are given different values of the smoothing parameter d in figure 4.3. The optimum choice of the parameter d can be found using the optimization technique discussed earlier. Table 4.2 shows that the suprema for various values differ and minimum of them attains at $d = 3$.

d	1	2	3	4	5
$S(d)$	493.121	5.238	2.765	5.852	3.851
d	6	7	8	9	10
$S(d)$	6.187	4.541	5.336	4.269	4.953

Table 4.2 : $S(d)$ values for various choices of smoothing parameter d.:Weibull with $a = 3$ and $b = 1$.

Using the appropriate value of the smoothing parameter, namely $d = 2$ we have the optimal orthogonal series estimator of the intensity function α is given in Figure 4.4.

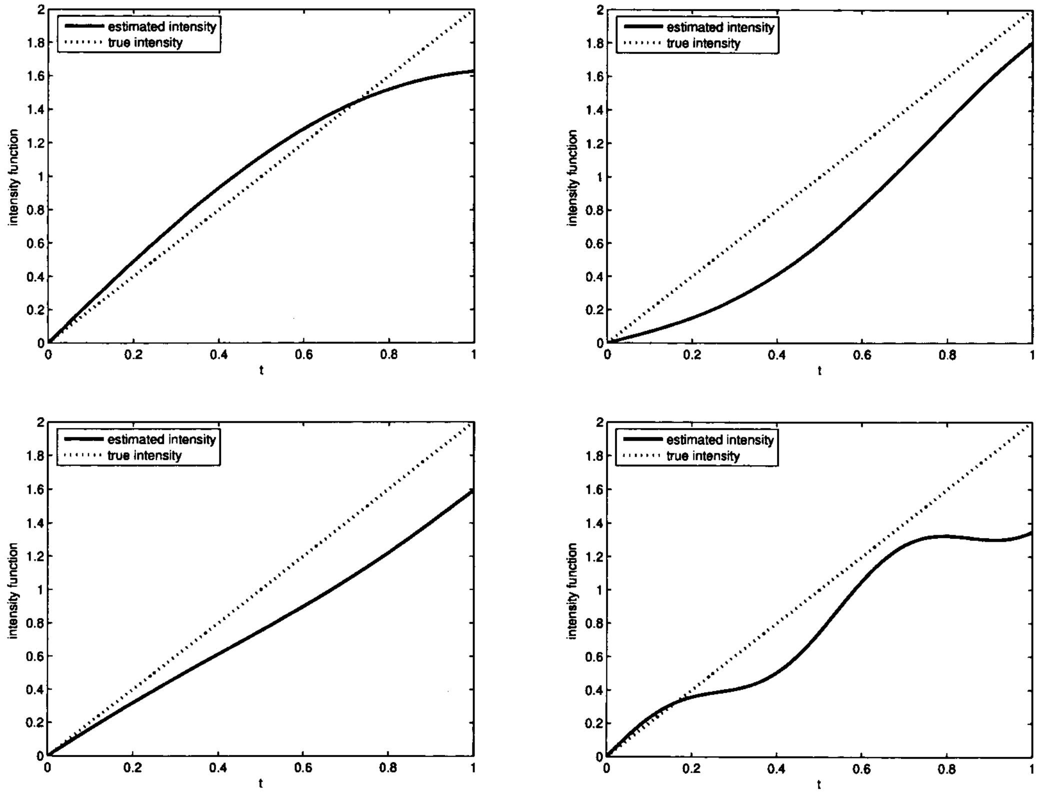

Figure 4.1: *Orthogonal Series Estimators of the Intensity using random sampling of size* $n = 500$ *from Weibull distribution with shape parameter* $a = 2$ *and scale parameter* $b = 1$ *in comparisons with true intensity :* $d = 2, 3, 4, 10$ *respectively, from left to right.*

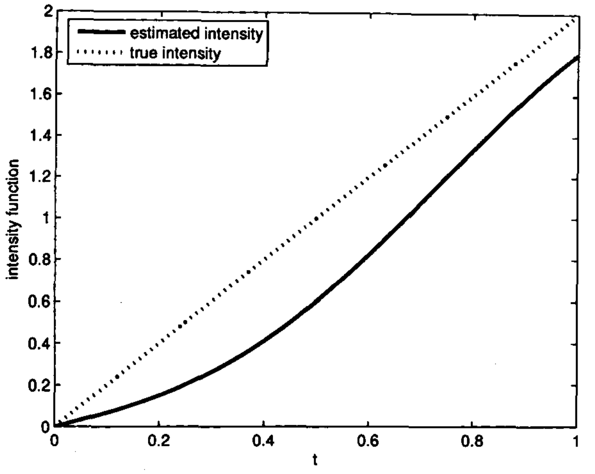

Figure 4.2: *Orthogonal Series Estimators of the Intensity using random sampling of size* $n = 500$ *from Weibull distribution with shape parameter* $a = 3$ *and scale parameter* $b = 1$ *in comparisons with true intensity :* $d = 3$

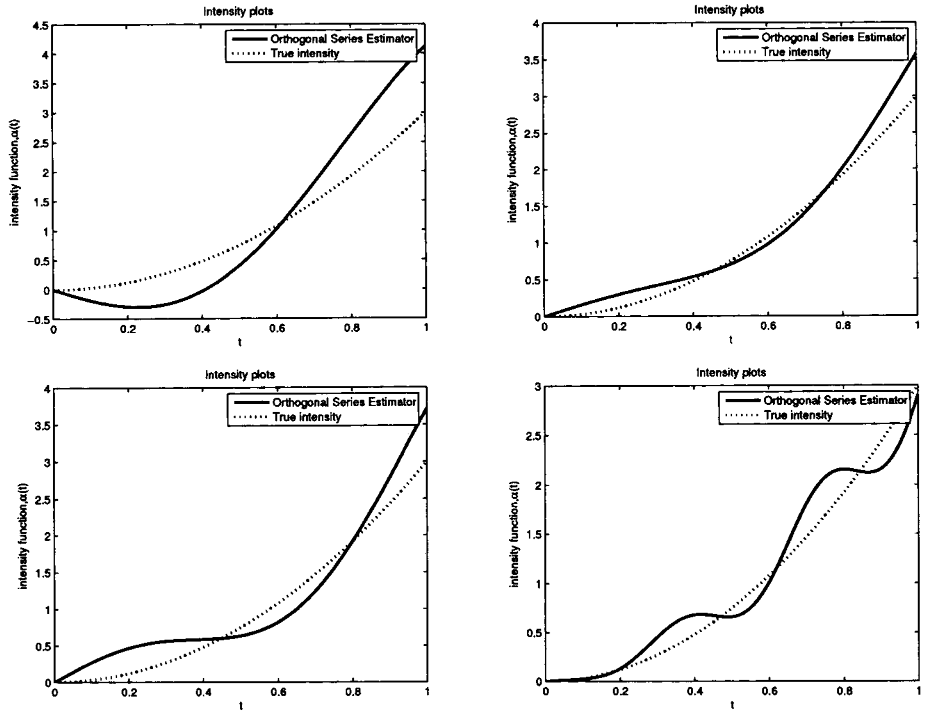

Figure 4.3: *Orthogonal Series Estimators of the Intensity using random sampling of size $n = 500$ from Weibull distribution with shape parameter $a = 3$ and scale parameter $b = 1$ in comparisons with true intensity : $d = 2, 3, 4, 10$ respectively, from left to right.*

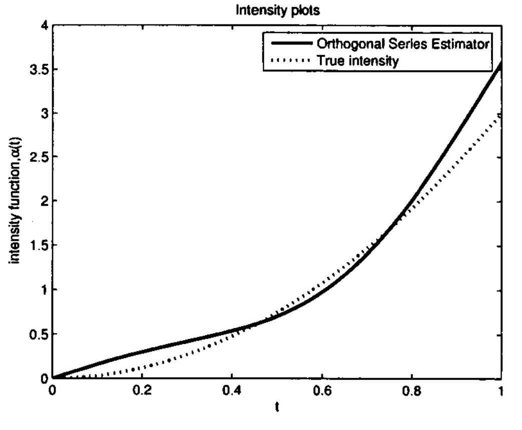

Figure 4.4: *Orthogonal Series Estimators of the Intensity using random sampling of size* $n = 500$ *from Weibull distribution with shape parameter* $a = 3$ *and scale parameter* $b = 1$ *in comparisons with true intensity* : $d = 3$

4.6 Data Analysis

In this section we estimate the hazard rate of the survival data associated with 205 patients in Denmark with malignant melanoma. (Source: P. K. Andersen, O. Borgan, R. D. Gill and N. Keiding (1993)). The survival time is given in days, possibly censored. The survival data is attached with three statuses namely, 1: died from melanoma, 2: alive, and 3: dead from other causes. In our analysis we take the data with statuses 2 and 3 as censored. We assume that the patiemts behave independently of each other. Then the number $N(t)$ of deaths up to time t is a counting process

$$\Lambda(t) = \alpha(t)Y(t),$$

where α is the hazard rate(force of mortality) and $Y(t)$ is the number of patients alive just before time t. The Nelson-Aalen estimator

$$\hat{\beta}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s)$$

d	1	2	3	4	5
$S(d)$	307.65	28.711	15.273	12.670	8.613
d	6	7	8	9	10
$S(d)$	4.9146	3.1056	1.6613	1.8594	2.0898

Table 4.3 : $S(d)$ values : Malignent Melanoma Data.

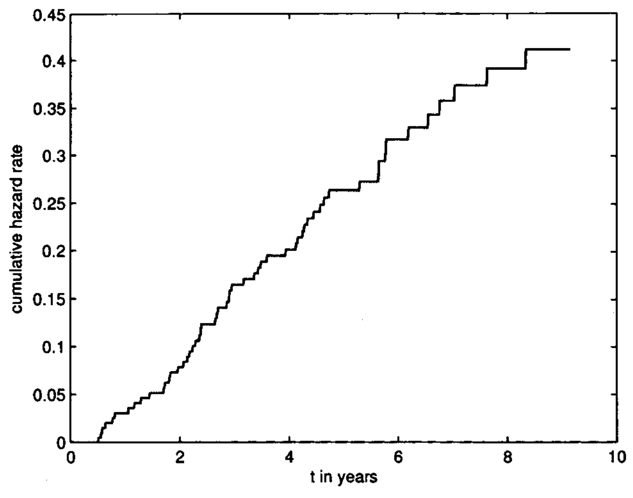

Figure 4.5: *Nelson-Aalen estimators of the cumulative hazard rate using the survival data associated with 205 patients in Denmark with Malignant Melanoma*

of the cumulative mortality rate $\beta(t)$ has been plotted in figure 4.5. It is difficult to get a precise estimate of α from figure 4.5, but seems that the mortality is steady at a higher level in the beginning up to 5 years after which it stabilizes at a lover level. These features are much more clear from the plot of the proposed orthogonal series estimator, drawn in figure.4.6. The appropriate value of the smoothing parameter d for the proposed estimator is chosen as **8** which is clear from the values of min$S(d)$ displayed in table 4.3.

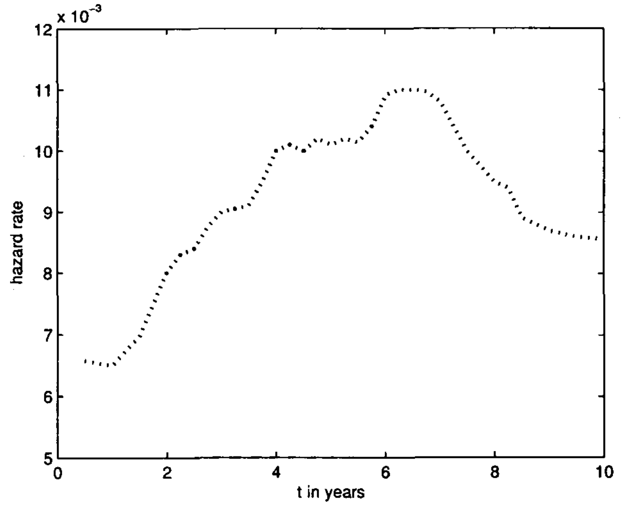

Figure 4.6: *Orthogonal Series Estimators of the hazard rate using the survival data associated with 205 patients in Denmark with Malignant Melanoma* : $d = 8$

4.7 concluding Remarks

In this chapter we have suggested orthogonal series estimator for intensity function of a counting process. A procedure for the choice of optimum number of Fourier coefficients is also suggested. It of interest to see how some elementary cross validation technique can be employed here for a better choice of the number of Fourier coefficients.

REFERENCES

1. Aalen, O. (1976). Nonparametric inference in connection with multiple dicrement models. *Scand J. Statist.*, **3**, 15-27.

2. Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701-726.

3. Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag.

4. Anilkumar, P. and Naik-Nimbalkar, U.V.(1995). Maximum Modified Kernel Likelihood Estimation of the Intensity for a Counting Process. *Analysis of Censored data, IMS Lecture Notes-Monographs Series* **27**, 1-9.

5. Azzalini,A. and Bowman, A.W.(1990). A look at some data on the Old Faithful geyser. *Applied Statistics* **39**,357-365.

6. Blischke, D.R and Murthy, D.R.P.,(Eds)(1996). *Product Warranty Handbook*, Marcel Dekker, New York.

7. Bremaud, P. (1981). *Point Process and Queues : Martingale Dynamics*, Springer Verlag, New York.

8. Brian, D.Ripley (1996). *Pattern Recognition and Neural Networks*, Cambridge: Cambridge University Press

9. Cox, D.R.(1961). *Renewal Theory*, Chapman and Halt, London.

10. Cox, D and Lewis, P.(1966).*The Statistical Analysis of Series of Events*, Methuen, London.

11. Dvoretzky, A., Kiefer,J., and Wolfowitz,J.(1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642-669.

12. Feller, W.(1971) *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Weiley, New York.

13. Frees, E.W.(1986a). Warranty analysis and renewal function. *Naval Res.logist Quart* **33**, 361-372.

14. Frees, E.W.(1986b). Non-parametric renewal function estimation. *Annals of statistics* Vol. **14**, No.**4**, 1366-1378.

15. Gibbons, Jean Dickinson and Chakraborti, S.(2003) *Nonparametric Statistical Inference*, 4th ed. CRC, New York.

16. Karr, A.F. (1987). Maximum likelihood estimstion in the multiplicative intensity model via sieves., *Ann. Statist.,* **15(2)**, 437-490.

17. Karr, A.F. (1991). *Point Processes and their Statistical Inference.*, Marcel Dekker, Inc., New York.

18. Leskow, J. and Raazanzki, R.(1989). Histogram Maximum Likelihood Estimator in the Multiplicative Intensity Model, *Stochastic Models and their applications*, **31**, 151-159.

19. Loeve, M.(1963). Probability Theory . 2^{nd} ed. Van Nostrand-Reinhold, Princetone, New Jersey.

20. Natalia Markovich(2007), *Nonparametric analysis of univariate heavy-tailed data: research and practice* John Wiley and sons Ltd., London.

21. Murphy, S.A. and Sen, P.K.(1991). Time-dependent coefficients in a Cox-type regression model *Stoch.Processos. and appl.,* **39**, 153-180.

22. Prakasa Rao, B.L.S. (1983). *Non-parametric functional estimation.* Academic Press, New York.

23. Randles, R.H. and Wolfe, D.A (1979). *Introduction to the theory of non-parametric statistics.* John wiley and sons, New York.

24. Ramlau-Hansen, H. (1983a). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.,* **11(2)**,453-466.

25. Rudin,W.(1976). *Principles of Mathematical Analysis,* 3^{rd} edn. McGraw Hill, New York.

26. Scott, D. W.(1992). Multivariate Density Estimation:Theory, Practice, and Visualization, Wiley,New York.

27. Sefling, R.J.(1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

28. Silverman, B.W.(1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc.* **B** , No.1, 97-99.

29. Silverman, B.W.(1986), *Density estimation for statistics and Data Analysis.* Chapman and Hall, New york.

30. Smith, W. (1958). Renewal theory and its ramifications (with discussions). *J. Roy. Statist. Soc. Ser.* **B** **20** 243-302.

31. Wasserman, L.(2007).*All of Nonparametric Statistics*, Springer, New York.

32. Zuckar, D.M. and Karr, A.F. (1990). Nonparmetric survival analysis with time-dependent covariate effects-A penalized likelihood approach, *Ann. Statist.*, **18**, 327-353.

519.23 HAM/S
NB 5987
MOHAMMED KOYA LIBRARY
UNIVERSITY OF CALICUT