DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

GEOMETRIC MODELLING IN REGRESSION AND DISTRIBUTION ESTIMATION

by
Mahmut Sami ERDOĞAN

March, 2019
İZMİR

GEOMETRIC MODELLING IN REGRESSION AND DISTRIBUTION ESTIMATION

A Thesis Submitted to the
Graduate School of Natural And Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics

by
Mahmut Sami ERDOĞAN

March, 2019
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**GEOMETRIC MODELLING IN REGRESSION AND DISTRIBUTION ESTIMATION**" completed by **MAHMUT SAMİ ERDOĞAN** under supervision of **PROF. DR. ÖZLEM EGE ORUÇ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

....

Prof. Dr. Özlem EGE ORUÇ

Supervisor

....

Assoc. Prof. Dr. Neslihan DEMİREL

Thesis Committee Member

....

Assoc. Prof. Dr. Çetin DİŞİBÜYÜK

Thesis Committee Member

....

Assoc. Prof. Dr. Sevcan DEMİR ATALAY

Examining Committee Member

....

Assoc. Prof. Dr. Femin YALÇIN

Examining Committee Member

Prof. Dr. Kadriye ERTEKİN
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

Completion of this doctoral dissertation was possible with the support of several people. First of all, I would like to express my deepest gratitude to my supervisor Prof. Dr. Özlem EGE ORUÇ for this continuous support, selfless guidance, generous advice and helpful comments. Especially, I wish to thank my supervisor, also for her valuable insighths, patience and encouragement during the hard times throughout my research.

Besides my advisor, I would like to thank Assoc. Prof. Dr. Neslihan DEMİREL and Assoc. Prof. Dr. Çetin DİŞİBÜYÜK and Prof. Dr. Halil ORUÇ, for their contributions and valuable help. I'm also grateful for the insights and efforts put forth.

Special thanks are to my precious brother and sister, for their kind patience to all my complaints and their support, whenever I needed.

This thesis is dedicated to my parents Ayşe and Bülent ERDOĞAN and my lovely fiancee Merve ŞENDAL because of their confidence, encouragement and support in my life. Words cannot express my thanks to my lovely family.

Mahmut Sami ERDOĞAN

GEOMETRIC MODELLING IN REGRESSION AND DISTRIBUTION ESTIMATION

ABSTRACT

The contribution of this thesis to inferential statistics by using geometric modeling methods is comprised of three parts. First, we propose a new nonparametric regression model that uses the rational Bézier curves. The main advantages of the proposed regression model are the flexibility over parametric models and having no restriction on the assumptions in regression function estimation. Moreover, rational Bézier curves provide many conveniences such as more control over the shape of a curve and projective invariance. We provide numerical examples to demonstrate superiority of the proposed model and verify it. Secondly, we introduce a new distribution function estimation method using the rational Bernstein polynomials (RBP). The performance of the proposed new method is compared with Bernstein polynomials and empirical distribution function methods using simulation. The new method guarantees monotone nondecreasing function by applying linear constraints on the coefficients of the rational Bernstein basis functions and smooth the empirical distribution function. Furthermore, as a special case, it reduces to Bernstein polynomial estimator method. Some theoretical properties of the new estimator are investigated. Simulation study shows that the proposed estimator is preferable to the Bernstein polynomials and empirical distribution function estimator methods. Thirdly, we introduce a unimodal density estimator based on the Schoenberg's spline operator and thus generalize that of Bernstein polynomials and the beta density. The advantage of this method is the local property. That is, refining the knots while keeping the degree fixed of B-splines yields better estimates. We also give a numerical example to verify our results.

Keywords: Geometric modelling, nonparametric regression, distribution function estimation, rational Bernstein polynomials, B-splines, unimodal estimator

REGRESYON VE DAĞILIM TAHMİNİNDE GEOMETRİK MODELLEME

ÖZ

Bu tezin geometrik modelleme yöntemleriyle çıkarımsal istatistiğe katkısı üç bölümden oluşmaktadır. İlk olarak, rasyonel Bézier eğrilerini kullanan yeni bir parametrik olmayan regresyon modeli öneriyoruz. Önerilen regresyon modelinin temel avantajları, parametrik modellerin üzerindeki esneklik ve regresyon fonksiyonu tahminindeki varsayımlar üzerinde herhangi bir kısıtlamaya sahip olmamasıdır. Ayrıca, rasyonel Bézier eğrileri, eğrinin şekli üzerinde daha fazla kontrol imkanı ve projektif değişmezlik gibi birçok kolaylık sağlamaktadır. Önerilen modelin üstünlüğünü göstermek ve doğrulamak için sayısal örnekler verilmiştir. İkinci olarak, rasyonel Bernstein polinomları (RBP) kullanarak yeni bir dağılım fonksiyonu kestirim yöntemi tanıtıyoruz. Önerilen yeni yöntemin performansı, Bernstein polinomları ve ampirik dağılım fonksiyon yöntemleriyle simülasyon kullanarak karşılaştırılmıştır. Yeni yöntem, rasyonel Bernstein temel fonksiyonlarının katsayıları üzerinde lineer kısıtlamalar uygulayarak ve ampirik dağılım fonksiyonunu düzleştirerek monoton azalmayan fonksiyonu garanti etmektedir. Ayrıca, özel bir durum olarak, Bernstein polinomu kestirim yöntemine indirgenmektedir. Yeni tahmin edicinin bazı teorik özellikleri araştırılmıştır. Simülasyon çalışması, önerilen tahmin edicinin Bernstein polinomları ve ampirik dağılım fonksiyonu kestirim yöntemlerine tercih edildiğini göstermektedir. Üçüncü olarak, Schoenberg'in spline operatörüne dayanan tek modlu yoğunluk kestiricisi tanıtılmıştır. Böylece Bernstein polinomları ve beta yoğunluğunu genelleştiriyoruz. Bu yöntemin avantajı yerel özelliğidir. Yani, B-spline'ların derecesini sabit tutarken, düğümleri rafine etmek daha iyi tahminler verir. Ayrıca sonuçlarımızı doğrulamak için sayısal bir örnek verilmiştir.

Anahtar kelimeler: Geometrik modelleme, parametrik olmayan regresyon, dağılım fonksiyonu tahminleme, rasyonel Bernstein polinomları, B-spline'lar, tekmodlu kestirici

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

Nowadays, geometric modeling is used in many areas such as computer aided design (CAD), image processing, shape reconstruction, solid modeling, geographical information systems, robotics, production and product design, and molecular biology. Bézier and B-spline curves are the best known and applied modeling tools that represent the geometric modeling curve that is interested in creating and representing geometric shapes. In this thesis, geometric modeling is applied to three different areas of statistical inference. The first of these is nonparametric regression models. Nonparametric regression models are widely used in the literature. The main advantages of these models are to have great flexibility compared to parametric models and not to require restrictive assumptions in the estimation of regression function. In recent years, many approaches to nonparametric models have been proposed. These are the kernel method (Silverman (1986), Wand & Jones (1995), Fan & Gijbels (1996), Loader (1999, 2006), Wasserman (2006)), smoothing splines (Wahba (1990), Green & Silverman (1993), Eubank (1999)), series-based smoothers (Tarter & Lock (1993), Ogden (2012)), regression splines (Hastie & Tibshirani (1990), Friedman (1991); Hansen et al. (2002)), penalized splines (Eilers & Marx (1996); Ruppert et al. (2009)) and Bézier curve smoothing.

The Bézier Curve is a very important tool for the mathematical representations of curves and surfaces, and is especially used in computational graphs, in the production of fonts and in geometric modeling studies. However, the use of the Bézier curve in statistical science began in the late twentieth century. The Bézier curve was first used to estimate the density function by Kim (1996). Then, it was also used for estimation in the measurement error model by Kim et al. (2000) and for smoothing of the bivariate Kaplan-Meier estimator by Bae et al. (2005).

In this thesis, firstly a new nonparametric regression model is proposed using the rational Bézier curve, which is a general case of Bézier curve and it is showed that the proposed model is superior to the polynomial model with the simulation studies. The model mimicking the shape of data very well is obtained since each control point is represented by a scalar weight in rational Bézier curve. Least square, which is the most common method to obtain regression coefficients, is used in the selection of control points.

One of the important issues of statistical inference is the estimation of the cumulative distribution function of the data set. As another objective of this thesis, the new distribution function estimator with geometric modeling is proposed, providing more control and flexibility on the shape of the curve. In statistics, nonparametric smoothing methods are used to estimate the population's cumulative distribution function and density function. In the literature, among these methods, kernel density estimation and orthogonal least squares (Cencov (1962)) are used. Recently, Bernstein polynomials and Bézier curves are among the subjects that attract the attention of researchers in the distribution function and density estimation (Leblanc (2010)). Vitale (1975) was first proposed using Bernstein polynomials to produce smooth density estimates. Kim et al. (1999) showed that estimators using Bézier curve smoothing in density estimation and regression function estimation provide similar results to kernel type smoothing as well as has the same asymptotic properties as classical kernel estimators. Babu et al. (2002) investigated the asymptotic properties of using Bernstein polynomials to approximate bounded and continuous density functions. Kakizawa (2004) showed that Bernstein polynomials could be used as a nonparametric prior for continuous densities. Thongjaem et al. (2013) showed superiority from Kernel method of Bernstein polynomials in distribution function estimation and density estimation. Turnbull & Ghosh (2014) presented a method of unimodal density estimation using Bernstein polynomials. Xue & Wang (2010) proposed to estimate the cdf by applying the constrained B-spline. Their proposed estimator is not only smooth but also monotone nondecreasing. Also, B-spline smoothing has been accomplishedly applied in estimating various nonparametric and semiparametric models (Stone (1985), He & Shi (1998), Huang (1998), Huang (2003), Qu & Li (2006), Zhou & He (2008), Wang (2009), Xue (2009), Zhou (2009)).

Bernstein polynomials have been generalized in many different ways - see for example Bernšteın (1912), Barry et al. (1992), Budakçı & Oruç (2012) - one of which is rational Bernstein polynomials. Rational Bernstein polynomials reduce to Bernstein polynomial when all weights are equal. Rational Bernstein polynomials provide several advantages, e.g. more control over the shape of a curve and exact representation of conics. This study introduces a new nonparametric estimation method for the distribution function using rational Bernstein polynomials. In conducting this study, first we obtain essential parts of rational Bernstein polynomials by finding control points $F_n(k/m)$ and weights w_k. Optimal weights are calculated based on empirical distribution function. The weights are estimated by minimizing the sum of squares according to ECDF. Then, we calculate averaged squared errors (ASE) by simulation studies for the proposed method, the Bernstein polynomials and the empirical distribution function.

In this thesis, we finally introduce a unimodal density estimator based on the Schoenberg spline operator to generalize both the Bernstein polynomials given by Turnbull & Ghosh (2014) and the beta density. We show that better results are achieved by taking advantage of the remarkable properties of B-splines.

This thesis is organized as follows; the general structures of Bernstein polynomials, Bézier curves, rational Bézier curves, and B-splines are discussed in Chapter Two. In Chapter Three, the topics of regression, distribution function and density estimation are examined and information is given about the available methods in the literature. The proposed methods are then introduced in Chapter Four. The performances of these proposed methods are compared with those of available methods by simulation in Chapter Five. Chapter Six presents the conclusion.

CHAPTER TWO
FUNDAMENTAL CONCEPTS OF BÉZIER CURVES AND B-SPLINES

Geometric modelling is a field that examines the methods and tools used to create geometric shapes, to provide mathematical identification of shapes and to manipulate them. The tools of the geometric modelling have been correlated with technology using computer aided design (CAD) in particular in recent years. There are many approaches and techniques for geometric modeling in the literature and the choice of these techniques usually depends on the applicability and the difficulty of the problem. Nowadays, geometric modelling is widely used in many areas of statistical science such as regression analysis, density and distribution function estimation. In this chapter, basic concepts such as Bézier curves, rational Bézier curves, B-splines, used in geometric modelling are introduced.

2.1 Bézier Curves

One of the most frequently benefited curve representations in computer graphics and geometric modeling is the Bézier Curve. Bézier curves and surfaces were originally developed independently of each other by different mathematical approaches by Pierre Etienne Bézier, a French engineer between, 1958 and 1960 and a French mathematician Paul de Faget de Casteljau.

Bézier Curve is a curve type that has a special mathematical representation using the Bernstein polynomial as the basis function. Bernstein polynomial forms the foundation of the Bézier curve. A Bézier curve of degree m has $m + 1$ control points. Control polygon is obtained by combining the control points with line segments. Bézier curve is interrelated with control points that will define the curve itself. Only the first and last of the control points are on the curve. Other control points contribute to defining the degree of the curve and thus the shape.

2.1.1 Bézier Curves in Low Degrees

2.1.1.1 Linear Bézier Curves

Linear Bézier curves are a straight line between given two control points because they do not contain the slope factor. Accordingly, linear Bézier curves are shown below to be parametrized by control points b_0 and b_1.

$$B(x) = (1-x)b_0 + xb_1,\ x \in [0,1] \tag{2.1}$$

Equation (2.1) is the equation of the linear Bézier curve which starts at $x = 0$ and ends at $x = 1$. Thus, the start and the end points of the curve are equal to b_0 and b_1, respectively.

2.1.1.2 Quadratic Bézier Curves

The formula of a quadratic Bézier curves is given through three control points b_0, b_1, b_2 by

$$B(x) = (1-x)^2b_0 + 2(1-x)xb_1 + x^2b_2,\ x \in [0,1]. \tag{2.2}$$

Also, the values at $x = 0$ and $x = 1$ of quadratic Bézier curves are b_0 and b_2, respectively. An example of a quadratic Bézier curves is shown in Figure 2.1.

2.1.1.3 Cubic Bézier Curves

Cubic Bézier curves are defined as follows when the control points b_0, b_1, b_2 and b_3 are given:

$$B(x) = (1-x)^3b_0 + 3(1-x)^2xb_1 + 3(1-x)x^2b_2 + x^3b_3,\ x \in [0,1]. \tag{2.3}$$

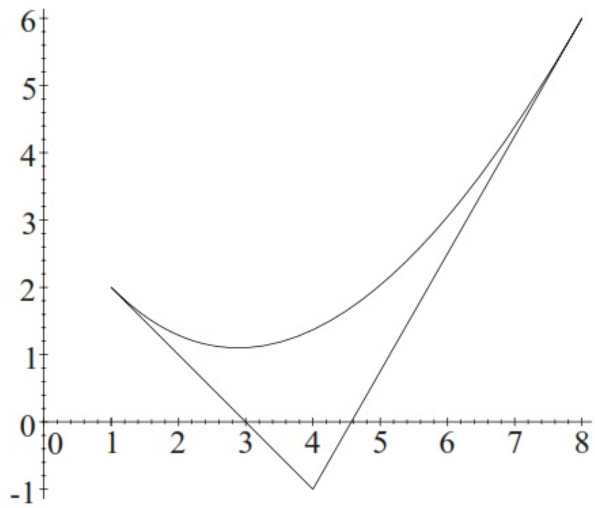

Figure 2.1 An example of quadratic Bézier curves

Despite their simple structures, cubic Bézier curves are the most common used Bézier curve type in application. Cubic Bézier curves generate an amplitude of shapes than quadratic Bézier curves, as they can model loops as in Figure 2.2(a), cusps (sharp corners) as shown in Figure 2.2(b), and inflections (Marsh (2006)).

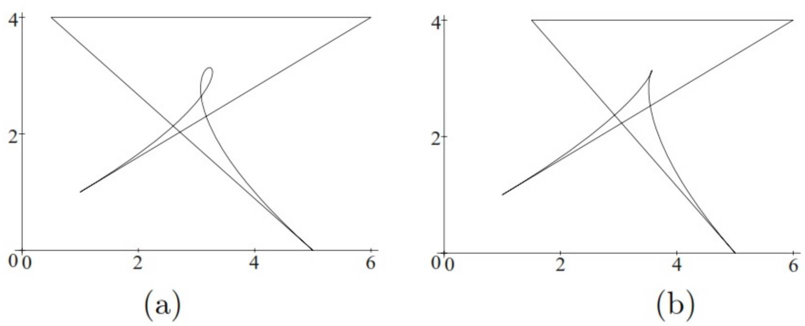

Figure 2.2 (a) Cubic Bézier curve containing a loop, (b) cubic Bézier curve containing a cusp

2.1.2 The General Bézier Curves

A Bézier curve of degree m is defined by a sequence of $m+1$ control points b_0, b_1,...,$b_m \in \mathbb{R}^2$ or $\mathbb{R}^3$

$$B(x) = \sum_{k=0}^{m} b_k B_{k,m}(x) \tag{2.4}$$

where

$$B_{k,m}(x) = \begin{cases} \frac{m!}{(m-k)!k!}(1-x)^{m-k}x^k, & \text{if } 0 \le k \le m \\ 0, & \text{otherwise.} \end{cases} \quad (2.5)$$

Note that the $B_{k,m}(x)$ are known as the Bernstein basis functions or Bernstein polynomials in a Bézier curve of degree m. These polynomials are quite easy to write down: the coefficients $\binom{m}{k}$ can be obtained from Pascal's triangle; the exponents on the x term increase by one as k increases; and the exponents on the $(1 - x)$ term decrease by one as k increases. The graphs of the Bernstein polynomials of degrees two and three are shown in Figure 2.3. In addition, Bézier curves can be defined at

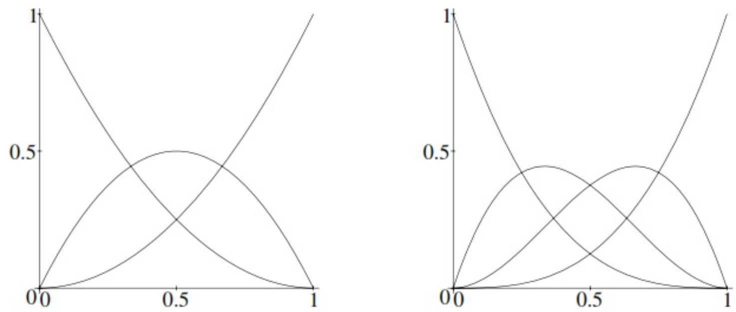

Figure 2.3 (a) Bernstein polynomials of degree two, (b) Bernstein polynomials of degree three

arbitrary interval of parameters. A Bézier curve of degree m with control points b_0, b_1,...,b_m is defined for any range of parameters as $[x_{\text{min}}, x_{\text{max}}]$,

$$B(x) = \sum_{k=0}^{m} b_k B_{k,m} \left(\frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \right). \quad (2.6)$$

As $B_{k,m}(x)$ is the Bernstein polynomial, the Bézier curve defined in arbitrary range is termed as the reparametrization of the ordinary Bézier curve.

2.1.2.1 The Properties of Bernstein Polynomials

The Bernstein basis functions have four different properties and by using these properties, important characteristics of the Bézier curve are arisen.

Non-negativity: The Bernstein basis functions of degree m are nonnegative on the standard parameter interval $[0, 1]$,

$$B_{k,m}(x) \geq 0,\ x \in [0, 1].$$

Partition of unity: The sum of the Bernstein polynomials of degree m is equal to one,

$$\sum_{k=0}^{m} B_{k,m}(x) = 1,\ x \in [0, 1].$$

Symmetry: The Bernstein polynomials are symmetric in the sense that

$$B_{m-k,m}(x) = B_{k,m}(1 - x), \text{for}\ \ k = 0, ..., m\ \ x \in [0, 1].$$

Recursion: The Bernstein polynomials of degree m can be defined recursively.

$$B_{k,m}(x) = (1 - x)B_{k,m-1}(x) + xB_{k-1,m-1}(x), \text{for}\ \ k = 0, ..., m$$

where $B_{-1,m-1}(x) = 0$ and $B_{m,m-1}(x) = 0$.

The partition of unity and positivity properties reveal two important properties of the Bézier curves known as invariance under transformations and the convex hull property. These properties are given in the next subsection.

2.1.2.2 The Properties of Bézier Curves

Important properties of Bézier curves of degree m with control points b_0, b_1,..., b_m are summarized below.

Endpoint interpolation property: As a very useful property of a Bézier curve, the initial and final control points are the endpoints of the Bézier curve of degree m. That is

$$B(0) = b_0 \text{ and } b(1) = b_m.$$

Endpoint tangent property: The first derivation of the Bézier curve can be described as follows,

$$B'(0) = m(b_1 - b_0) \text{ and } B'(1) = m(b_m - b_{m-1}).$$

Convex hull property (CHP): The Bézier curve is always included within the convex hull of the control points.

$$\text{For all } x \in [0,1], B(x) \in \text{CH}\{b_0, ..., b_m\}$$

Invariance under affine transformations: The Bézier curve is invariant under affine transformations such as rotation, reflection or scaling.

$$\mathrm{T}\left(\sum_{k=0}^{m} b_k B_{k,m}(x)\right) = \sum_{k=0}^{m} \mathrm{T}\ (b_k) B_{k,m}(x)$$

Variation diminishing property (VDP): For a planar Bézier curve $B(x)$, the VDP designates that the number of intersections of a given line with $B(x)$ is less than or equal to the number of intersections of a line with the control polygon (Marsh (2006)).

2.1.3 The de Casteljau Algorithm

Although the algorithm described in this section is simple, it is the most basic algorithm for curve and surface design. This algorithm combines geometry and algebra in an interesting way. The De Casteljau algorithm proposes a method for computing a point on a Bézier curve. The following theorem expresses the De Casteljau algorithm.

Theorem 2.1 *Let a Bézier curve of degree m be given by control points b_0, b_1,...,b_m and let $x \in [0,1]$ be any parameter value. Then $B(x) = b_0^m$, where $b_k^0 = b_k$, and*

$$b_k^j = b_k^{j-1}(1-x) + b_{k+1}^{j-1}x \tag{2.7}$$

for $j = 1, ..., m$ and $k = 0, ..., m-j$ (Marsh (2006)).

A point on a cubic Bézier curve is formed as follows with the help of control points for the specified value of x.

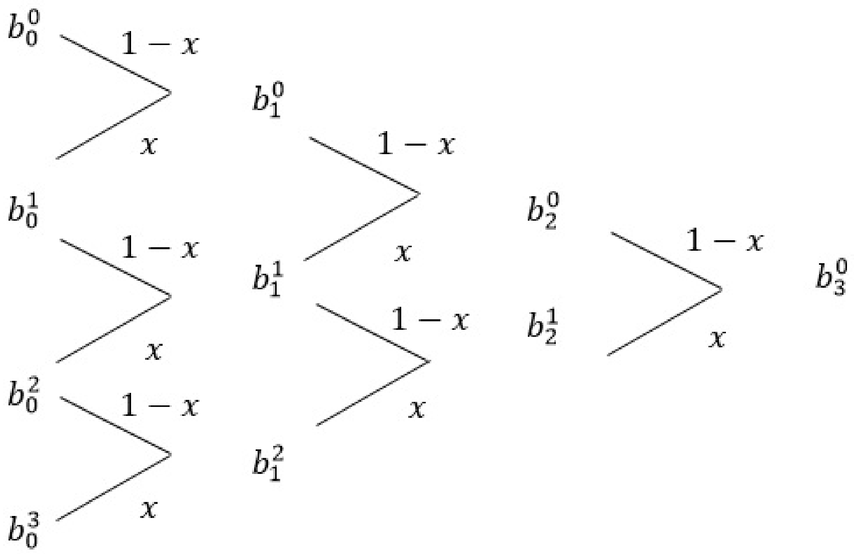

Figure 2.4 The de Casteljau Algorithm for Bézier curve

2.1.4 Derivatives of Bézier Curves

In order to use the Bézier curves and Bernstein polynomials to match the data in the field of statistics, it is necessary to calculate the first and second derivatives. In the following theorems, the first and second derivatives of the Bézier curve are given.

Theorem 2.2 *The first and second derivatives for a Bézier curve of degree m are*

$$B'(x) = \sum_{k=0}^{m-1} b_k^{(1)} B_{k,m-1}(x) \tag{2.8}$$

and

$$B''(x) = \sum_{k=0}^{m-2} b_k^{(2)} B_{k,m-2}(x) \tag{2.9}$$

where $b_k^{(1)} = m(b_{k+1} - b_k)$ *and* $b_k^{(2)} = m(m-1)(b_{k+2} - 2b_{k+1} + b_k)$.

Theorem 2.3 *The first and second derivatives for Bernstein polynomial of degree m are*

$$B'_{k,m}(x) = \frac{k - mx}{x(1-x)} B_{k,m}(x) \tag{2.10}$$

and

$$B''_{k,m}(x) = \left(\frac{k(k-1) - 2k(m-1)x + m(m-1)x^2}{x^2(1-x)^2} \right) B_{k,m}(x). \tag{2.11}$$

2.1.5 Rational Bézier Curves

Rational Bézier curves are obtained by assigning a scalar weight to each control point to obtain better approximations with Bézier curves. These curves provide a closer approach to the random shape by adding adjustable weights to the control points. Rational Bézier curves include the Bézier curve on the numerator and denominator, and the sum of Bernstein polynomials. A rational Bézier curve of degree m with control points b_0, b_1,..., b_m and associated positive scalar weights w_0,..., w_k can be described by:

$$B(x) = \frac{\sum_{k=0}^{m} w_k b_k B_{k,m}(x)}{\sum_{k=0}^{m} w_k B_{k,m}(x)}, x \in [0, 1]. \tag{2.12}$$

Not all the weights are zero (Marsh (2006)). When the weights of the all control points are the same, rational Bézier curve is reduced to Bézier curve. The rational Bézier curve can be described in basis form as follows,

$$B(x) = \sum_{k=0}^{m} b_k R_{k,m}(x) \tag{2.13}$$

where

$$R_{k,m}(x) = \begin{cases} \frac{w_k B_{k,m}(x)}{\sum_{j=0}^{m} w_j B_{j,m}(x)}, & \text{if } w_k \neq 0 \\ \frac{B_{k,m}(x)}{\sum_{j=0}^{m} w_j B_{j,m}(x)}, & \text{if } w_k = 0. \end{cases} \tag{2.14}$$

Rational Bézier curves are beneficial for many reasons. Rational functions may be

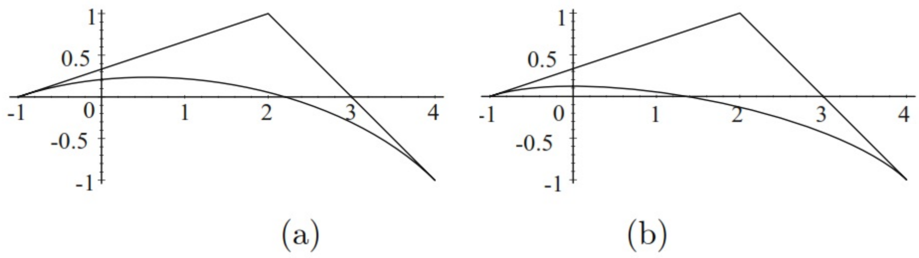

Figure 2.5 Rational Bézier curve of degree 2 with the same control points, and (a) weights 1, 1, 2, and (b) weights 1, 0.6, 2

used for hyperbolas and ellipses when polynomial functions are needed to define the parabolas. Also, higher ratings may be needed in the Bézier curve most of the time. It is possible to achieve the desired shape by keeping the degree of the curve constant by using the rational Bézier curve. Figure 2.5 above shows how to change the curve by changing the weights of the control points.

2.1.5.1 The Properties of Rational Bézier Curves

Rational Bézier curves resemble the properties of Bézier curves and these properties are summarized below.

Endpoint interpolation property: The rational Bézier curve of degree m passes through the first and the last control point.

$$B(0) = b_0 \text{ and } B(1) = b_m$$

Endpoint tangent property: First derivative at points $x = 0$ and $x = 1$ of the rational Bézier curve are given as follows,

$$B'(0) = m\frac{w_1}{w_0}(b_1 - b_0) \text{ and } B'(1) = m\frac{w_{m-1}}{w_m}(b_m - b_{m-1}).$$

Convex hull property (CHP): Assuming that each $w_k > 0$, the rational Bézier curve is always contained within the convex hull of the control points.

$$\text{For all } x \in [0, 1], B(x) \in \text{CH}\{b_0, ..., b_m\}$$

Invariance under affine transformations: If an affine transformation T is applied, the rational Bézier curve is invariant.

$$\text{T}\left(\frac{\sum_{k=0}^{m} w_k b_k B_{k,m}(x)}{\sum_{k=0}^{m} w_k B_{k,m}(x)}\right) = \frac{\sum_{k=0}^{m} w_k \text{T}(b_k) B_{k,m}(x)}{\sum_{k=0}^{m} w_k B_{k,m}(x)}$$

Variation diminishing property (VDP): Assuming that each $w_k > 0$, rational Bézier curve satisfies the variation diminishing property.

2.1.6 Derivatives of Rational Bézier Curves

The following iterative formula is utilized to obtain the derivative of rational Bézier curve. For $P(t) = u(t)/v(t)$;

$$P'(t) = \frac{v(t)u'(t) - v'(t)u(t)}{v(t)^2} = \frac{u'(t) - v'(t)P(t)}{v(t)}. \tag{2.15}$$

The Leibnitz rule (Spivak (1967)) used to obtain the derivatives of a product of two functions yields that derivative $u(t) = v(t)P(t)$ is

$$\begin{aligned} u^{(r)}(t) &= \sum_{k=0}^{r} \binom{r}{k} v^k(t) P^{(r-k)}(t) \\ &= v(t)P^{(r)}(t) + \sum_{k=1}^{r} \binom{r}{k} v^k(t) P^{(r-k)}(t). \end{aligned} \tag{2.16}$$

Therefore,

$$P^{(r)}(t) = \frac{u^{(r)}(t) - \sum\limits_{k=1}^{r} \binom{r}{k} v^k(t) P^{(r-k)}(t)}{v(t)}. \tag{2.17}$$

Thus the rth derivative of $P(t)$ can be obtained in terms of the first $r - 1$ derivatives of $P(t)$, and the first r derivatives of $u(t)$ and $v(t)$. The formula of rational Bézier curve of degree m can be divided into two part as $u(t)$ and $v(t)$ when defined as $u(t) = \sum_{k=0}^{m} w_k b_k B_{k,m}(x)$ and $v(t) = \sum_{k=0}^{m} w_k B_{k,m}(x)$. The $u(t)$ and $s(t)$ are obtained by Bézier curve, where $w_k b_k$ are considered to be the control points of $u(t)$ and the weights w_k are considered in Marsh (2006). From this point on,the first and second derivatives of the rational Bézier curve are given by

$$B'(x) = \frac{(\sum_{k=0}^{m} w_k b_k B_{k,m}(x))' - (\sum_{k=0}^{m} w_k B_{k,m}(x))' \, B(x)}{\sum_{k=0}^{m} w_k B_{k,m}(x)} \tag{2.18}$$

and

$$\begin{aligned} B''(x) &= \frac{(\sum_{k=0}^{m} w_k b_k B_{k,m}(x))'' - 2(\sum_{k=0}^{m} w_k B_{k,m}(x))' B'(x)}{\sum_{k=0}^{m} w_k B_{k,m}(x)} \\ &\quad - \frac{(\sum_{k=0}^{m} w_k B_{k,m}(x))'' B(x)}{\sum_{k=0}^{m} w_k B_{k,m}(x)}. \end{aligned} \tag{2.19}$$

Starting from here, we can verify the endpoint tangent property,

$$B'(0) = \frac{m(w_1 b_1 - w_0 b_0) - m(w_1 - w_0) b_0}{w_0} = m \frac{w_1}{w_0}(b_1 - b_0)$$

and in the same way,

$$B'(1) = m \frac{w_{m-1}}{w_m}(b_m - b_{m-1}).$$

2.2 B-splines

In this section, B-splines as the basis function is examined. Although the polynomial approach provides great benefit in functional approach methods, the higher the number of points given in the determination of an unknown function, the higher the polynomial degree. This can lead to erroneous results because it would cause large oscillations in functions. An alternative way is to approach with lower degree piecewise polynomials to a set of points given to represent the searched function and to provide ease of operation. The process here is to approach the searched function using low-degree polynomials on each piece by breaking into small ranges. Piecewise polynomials, which have such properties, are called spline functions.

Spline functions are useful in hand and computer calculations are utilized in many areas such as computer aided design, computer aided manufacturing, computer graphics representations, numerical analysis, programming. Both derivatives and integrals of spline functions are again spline functions. Low-degree spline functions are very flexible and do not oscillate as in polynomials. B-spline functions are also spline functions. However, according to polynomial degree and disintegration of solution domain, B-spline functions are functions with minimal support. Each spline function at a certain degree can be represented by a linear combination of B-spline functions at the same degree. For this reason, B-spline functions form a basis for spline functions. At the same time, a B-spline function is a generalization of Bézier curves. The B-spline term was first used by Schoenberg (1946) as an abbreviation for the basis spline term.

Any $[a, b]$ interval is divided by $n+1$ points as $x_0 < x_1 < \ldots < x_n$. Given any a nondecreasing (extended) knot sequence $x_0, \ldots, x_n$ with a sequence of $m+1$ control points b_0, b_1,...,b_m, the equation of the B-spline curve of degree d on an interval $[a, b]$ defined recursively by the de Boor algorithm (Goldman (2002)) is given by

$$B(x) = \sum_{k=0}^{m} b_k N_{k,d}(x) \tag{2.20}$$

where

$$N_{k,0}(x) = \begin{cases} 1, & \text{if } x \in [x_k, x_{k+1}) \\ 0, & \text{otherwise.} \end{cases} \tag{2.21}$$

$$N_{k,d}(x) = \frac{x - x_k}{x_{k+d} - x_k} N_{k,d-1}(x) + \frac{x_{k+d+1} - x}{x_{k+d+1} - x_{k+1}} N_{k+1,d-1}(x), \tag{2.22}$$

for $k = 0, \ldots, m$ and d. Here, $N_{k,d}(x)$ represents B-spline basis functions of degree d that extremely dependents on the knot sequence. If there are a sufficient number of repeated knot values in the knot vector, then a division of the form $N_{k,d-1}(x)/(x_{k+d} -$

$x_k) = 0/0$ (for some k) may be found during the execution of the iteration. It is assumed that $0/0 = 0$ (Marsh (2006)). The graph of the basis functions of degree two are shown with knots $x_0 = 0, x_1 = 0, x_2 = 0, x_3 = 1/4, x_4 = 1/2, x_5 = 3/4, x_6 = 1, x_7 = 1, x_8 = 1$ in the Figure 2.6 below. Also, despite a Bézier curve of degree of d has $m+1$

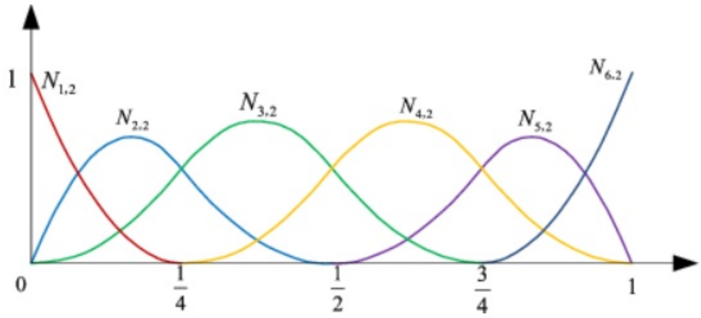

Figure 2.6 Basis functions of degree 2 for the knot vector $x_0 = 0, x_1 = 0, x_2 = 0, x_3 = 1/4, x_4 = 1/2, x_5 = 3/4, x_6 = 1, x_7 = 1, x_8 = 1$

control points, the number of control points increases with the number of knots in B-splines. Thus, while defining complex shapes, more satisfying results can be achieved without increasing the degree of the curve.

2.2.1 The Properties of B-spline Basis Functions

The basis function of B-spline $N_{k,d}(x)$ has important properties similar to the basis functions of Bézier. We have summarized these properties by using the reference Marsh (2006) as follows.

Non-negativity: For all k, d and x, $N_{k,d}(x)$ has positivity property,

$$N_{k,d}(x) > 0, \text{for } x \in (x_k, x_{k+d+1}).$$

Local support: İf x is outside the interval (x_k, x_{k+d+1}), the basis function $N_{k,d}(x)$ is equal to zero,

$$N_{k,d}(x) = 0, \text{for } x \notin (x_k, x_{k+d+1}).$$

Piecewise polynomial: The B-spline basis function of degree d is a piecewise polynomial function of degree d.

Partition of unity: The sum of the B-spline basis functions of degree d is equal to one,

$$\sum_{k=r-d}^{r} N_{k,d}(x) = 1, \text{for } x \in [x_r, x_{r+1}).$$

2.2.2 The Properties of B-spline Curve

A B-spline curve of degree d on the knot sequence $x_0, ..., x_n$ has many useful properties. Some of the most important features of B-spline curves is shown below.

Local control: Each segment is determined by $m + 1$ control points. If $x \in [x_r, x_{r+1})$ $(d \leq r \leq n - d - 1)$, then

$$B(x) = \sum_{k=r-d}^{r} b_k N_{k,d}(x) = 1, \text{for } x \in [x_r, x_{r+1}).$$

Thus to evaluate $B(x)$ it is sufficient to evaluate $N_{r-d,d}(x), ..., N_{r,d}(x)$.

Convex hull property (CHP): If $x \in [x_r, x_{r+1})$ $(d \leq r \leq n - d - 1)$, then

$$B(x) \in \text{CH}\{b_{r-d}, ..., b_r\}.$$

Invariance under affine transformations: Let T be an affine transformation. Then,

$$\text{T}\left(\sum_{k=0}^{m} b_k N_{k,d}(x)\right) = \sum_{k=0}^{m} \text{T}\ (b_k) N_{k,d}(x).$$

Continuity: If p_k is the multiplicity of the breakpoint $x = u_k$, then $B(x)$ is C^{n-p_k} (or greater) at $x = u_k$, and C^{∞} elsewhere (Marsh (2006)).

2.2.3 Derivatives of B-splines

Although the B-spline curves are more complex than the Bézier curves, the derivatives of the two curve types are very similar. The derivative of the B-spline curve is very important in engineering applications, and especially in the calculation of distribution function and density function estimation in statistical science. For this reason, our goal is to achieve the derivative of a B-spline curve of degree d. The following theorem gives the derivatives of the B-spline curve and B-spline basis function, respectively (for the proof, see Marsh (2006)).

Theorem 2.4 *The first and second derivatives for a B-spline curve of degree* d *are*

$$B'(x) = \sum_{k=0}^{m-1} b_k^{(1)} N_{k,d-1}^{(1)}(x) \tag{2.23}$$

and

$$B''(x) = \sum_{k=0}^{m-2} b_k^{(2)} N_{k,d-2}^{(2)}(x) \tag{2.24}$$

where $b_k^{(0)} = b_k$.

Theorem 2.5 *The first derivatives for B-spline basis function of degree* d *on the knot vector* $x_0, ..., x_n$ *are*

$$N'_{k,d}(x) = \frac{d}{x_{k+d} - x_k} N_{k,d-1}(x) + \frac{d}{x_{k+d+1} - x_{k+1}} N_{k+1,d-1}(x). \tag{2.25}$$

CHAPTER THREE
BRIEF INFORMATION ABOUT POLYNOMIAL REGRESSION, DISTRIBUTION FUNCTION AND UNIMODAL DENSITY ESTIMATION METHODS

In this chapter, brief information about polynomial regression method, distribution function and unimodal density function estimation methods are introduced.

3.1 Polynomial Regression

Regression analysis involves identifying the relationship between a dependent variable and one or more independent variables. It is one of the most important statistical tools used extensively in almost all sciences. Multiple regression refers to regression applications in which there are more than one independent variables. Multiple regression includes a technique called polynomial regression. In polynomial regression, we regress a dependent variable on powers of the independent variables. In this section, the polynomial regression method, which is used to model nonlinear relationships, will be briefly explained.

The objective of the polynomial regression is to model nonlinear relationships between a dependent variable and an independent variable. Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear in the sense that the regression function $E(y|x)$ is linear in the unknown parameters $\beta_0, \beta_1, ..., \beta_m$ that are estimated from the data. For this reason, polynomial regression is considered a special case of multiple linear regression.

The m^{th} order polynomial model in one variable is given by

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + ... + \beta_m x^m + \epsilon \tag{3.1}$$

where y is dependent variable, x is independent variable and also m indicates the degree of the regression model, ϵ is an error term. Also, ϵ shows the difference between the actual value of the dependent variable and the predicted value of the model in each observation pair and has a normal distribution with mean 0 and σ^2. The basic multiple regression model of a dependent variable Y on a set of m independent variables $x_1, x_2, ..., x_m$ can be expressed as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + ... + \beta_m x_i^m + \epsilon_i \tag{3.2}$$

where y_i is the value of the dependent variable Y for the i^{th} case, x_i is the value of the independent variable for i^{th} observation, β_0 is the intercept of the regression surface, each β_j, $j = 1, 2, ..., m$ is the slope of the regression surface with respect to variable X and ϵ_i is the random error component for the i^{th} case.

In matrix notation, we can rewrite model (3.2) as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ 1 & x_3 & x_3^2 & \cdots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{3.3}$$

This can be rewritten as follows,

$$\vec{y} = X\vec{\beta} + \vec{\epsilon}. \tag{3.4}$$

We estimate the regression parameters by the method of least squares. This is an extension of the procedure used in simple linear regression. First, we calculate the sum of the squared errors and, second, find a set of estimators that minimize the sum. The least squared estimator for β is given in equation (3.5).

$$\hat{\vec{\beta}} = (X^T X)^{-1} X^T \vec{y} \tag{3.5}$$

Vector $\hat{\vec{\beta}}$ is an unbiased estimator of β. The fitted values for the mean of $\hat{Y}$, are computed by

$$\hat{\vec{y}} = X\hat{\vec{\beta}} = X(X^TX)^{-1}X^T\vec{y} \tag{3.6}$$

Polynomial regression models are among the most frequently used empirical models for fitting functions. However, these models have some limitations. Sometimes it is exhibited in the data that a lower order polynomial does not provide a good fit. A possible solution in such situation is to increase the order of the polynomial but it may not always work. The higher order polynomial may not improve the fit significantly. Such situations can be analyzed through residuals, e.g., the residual sum of square may not stabilize or the residual plots fail to explain the unexplained structure. One possible reason for such happening is that the response function has different behaviors in different ranges of independent variables. This type of problems can be overcome by fitting an appropriate function in different ranges of explanatory variable. Due to the deficiencies mentioned above, polynomial basis functions are used with new basis functions such as splines, radial basis functions, and wavelets. These families of basis functions are useful for many data types and provide a better fit.

There are other issues with polynomial regression. For example, it is inherently non-local, i.e., changing the value of Y at one point in the training set can affect the fit of the polynomial for data points that are very far away. Hence, to avoid the use of high degree polynomial on the whole data set, we can substitute it with many different small degree polynomial functions.

3.2 Distribution Function Estimation

The basis of the most statistical analyzes is based on the use of probability density and cumulative distribution functions. If the density functions and distribution functions of the underlying population are known, these functions can be used for statistical analysis. Generally, these functions are not known, so it is necessary to estimate them. The estimation methods of these functions are comprised of two parts, parametric and nonparametric. There are strong assumptions in parametric estimations. When the parametric assumption yields invalid results in modeling the function of the underlying population, the nonparametric estimation is a powerful alternative. There are no rigid assumptions in nonparametric estimation and this situation provides a great convenience and flexibility to researchers.

Several estimators have been proposed to estimate the density or distribution function nonparametrically. These include the histograms, the empirical distribution function, the kernel density estimator by Rosenblatt (1956) and Parzen (1962), the nearest neighborhood estimator by Fix & Hodges Jr (1951), the series estimator by Cencov (1962), the penalized likelihood estimator by Good & Gaskins (1971, 1980), and more recently the Bernstein polynomial estimators by Babu et al. (2002), Leblanc (2012) and Dutta (2016), among which the empirical distribution function estimator is the oldest one. It is also the most commonly used estimator. Therefore, we will firstly focus on empirical distribution function estimation.

3.2.1 Empirical Distribution Function

The most basic method used to estimate the distribution function underlying a random variable is empirical distribution function. It is the distribution function associated with the empirical measure of a sample. This cumulative distribution function is a step function that jumps up by $1/n$ at each of the n data points. Its value at any specified value of the measured variable is the fraction of observations of the measured variable that are less than or equal to the specified value. The empirical distribution function is an estimate of the cumulative distribution function that generated the points in the sample. The empirical distribution function is obtained as follows.

Let $X_1, X_2, ..., X_n$ be independent and identically distributed (i.i.d.) real valued random variables with the cumulative distribution function $F(X) = P(X \leq x)$. Empirical cumulative distribution function (ECDF) or also called as empirical distribution function is defined as follows,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \tag{3.7}$$

where I represents the indicator function which is one if $(X_i \leq x)$ and zero else.

Note that ECDF is a nondecreasing and right-continuous, also takes values on $[0, 1]$ because it is jumping with probabilities $1/n$ in the points $X_1, X_2, ..., X_n$. The following are some important properties of the ECDF.

Note that, for a fixed point $x \in \mathbb{R}$, the quantity $F_n(x)$ has a binomial distribution with parameters n and success probability $F(x)$. Therefore,

$$E(F_n(x)) = F(x) \tag{3.8}$$

and

$$Var(F_n(x)) = MSE(F_n(x)) = n^{-1}F(x)(1 - F(x)) = n^{-1}\sigma^2(x) \tag{3.9}$$

where $\sigma^2(x) = F(x)(1 - F(x))$.

The Glivenko - Cantelli theorem implies a stronger convergence result, namely that strong convergence holds uniformly in x:

Theorem 3.1 *Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with the cumulative distribution function $F(X) = P(X \leq x)$ and let $F_n(x)$ denotes empirical distribution function. In this case, $F_n(x)$ converges uniformly to true function. Then,*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \overset{a.s.}{\to} 0 \tag{3.10}$$

as $n \to \infty$ where a.s. represents almost sure convergence.

The ECDF is known to have optimum properties as an estimator of the underlying distribution function. However, it may not be appropriate for estimating continuous distributions because of its jump discontinuities. When it is known that $F(x)$ is continuous, the estimation of $F(x)$ is naturally made by using smooth functions instead of using the discontinuous empirical distribution function,. We can do this by using one of the famous Bernstein polynomial approximation, in which $f(x)$ is supported on the unit interval.

3.2.2 Bernstein Polynomials Method

The problem of distribution function (df) estimation arises naturally in many contexts. The empirical and the kernel df estimators are well known. There is another df estimator based on a Bernstein polynomials of degree m. This section is concerned with df estimator based on sequences of polynomials named after their creator S.N. Bernstein (see Bernšteın (1912)). Given a function $f(x)$ on $[a, b]$, we define the Bernstein polynomial,

$$B_m(x,f) = \sum_{k=1}^{m} f(a + \frac{k-1}{m-1}(b-a)) \binom{m-1}{k-1} \left(\frac{x-a}{b-a}\right)^{k-1} \left(\frac{b-x}{b-a}\right)^{m-k} \tag{3.11}$$

for $a \leq x \leq b$. S.N. Bernstein developed Bernstein polynomials as a probabilistic proof of the Weierstrass Approximation Theorem. He showed that any continuous function, on a closed interval can be uniformly approximated using them by, $\| B_m(.,f) - f(.) \|_\infty \equiv \sup_{a \leq x \leq b} |B_m(x,f) - f(x)| \to 0$ as $m \to \infty$ (Lorentz (2012)).

It is well known that the Bernstein polynomial is a useful tool for interpolating functions defined on a closed interval and can therefore be used to approximate a density function on such an interval. Furthermore, Bernstein polynomials are an example of polynomial approximation with a simple interpretation in terms of probability and possess good shape preservation properties. The smooth estimates of a density function and distribution function with a finite known support can be made based on Bernstein - Weierstrass Approximation Theorem. Bernstein polynomials are an attractive approach to density and distribution function estimation as they are the simplest example of a polynomial approximation with a probabilistic interpretation.

As a nonparametric estimation method, Bernstein polynomial method proposed by Babu et al. (2002) is to estimate the distribution function $F(x)$. This method estimates the $F(x)$ distribution function using the smoothing technique by help of the Bernstein polynomial. The Bernstein estimator described by Babu et al. (2002) is given by

$$\hat{F}_{m,n}(x) = \sum_{k=0}^{m} F_n(k/m) B_{k,m}(x) \tag{3.12}$$

where $B_{k,m}(x) = \binom{m}{k} x^k (1-x)^{m-k}$. Then, $B_m(x)$ is defined as the Bernstein polynomial of degree m of the function of F by Leblanc (2012).

$$B_m(x) = \sum_{k=0}^{m} F(k/m) B_{k,m}(x) \tag{3.13}$$

It is easy to see that $E[\hat{F}_{m,n,w}(x)] = B_m(x)$ as a result of examining equations (3.12) and (3.13) for all $x \in [0,1]$ and $n \geq 1$. Also, the following equations can be written due to the limit property of the cdf for $m \geq 1$ in Bernstein estimator.

$$\hat{F}_{m,n}(0) = F(0) = B_m(0) = 0 \text{ and } \hat{F}_{m,n}(1) = F(1) = B_m(1) = 1 \tag{3.14}$$

In addition, it is possible to achieve the following equation based on the first derivative of Bernstein polynomials.

$$\hat{f}_{m,n}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\hat{F}_{m,n}(x) = m \sum_{k=0}^{m-1} [F_n((k+1)/m) - F_n(k/m)] B_{k,m}(x) \tag{3.15}$$

3.2.2.1 Some Theoretical Properties of Bernstein Estimator

In this part of the study, we present some theoretical properties of the Bernstein estimator such as the variance, bias and convergence rate. First, we focus on the following theorem presented by Leblanc (2012).

Theorem 3.2 *Let F be a continuous and differentiable function on the interval $[0, 1]$. Under these assumptions, $B_m(x)$ is defined as follows*

$$B_m(x) = F(x) + m^{-1}b(x) + o(m^{-1}) \tag{3.16}$$

where $b(x) = x(1 - x)f'(x)/2$.

The some basic properties of the Bernstein estimator are given using Theorem 3.2.

Theorem 3.3 *Let F be a continuous and differentiable function on the interval $[0, 1]$. Under these assumptions, we have*

$$Bias[\hat{F}_{m,n}(x)] = E[\hat{F}_{m,n}(x)] - F(x) = m^{-1}b(x) + o(m^{-1}) \tag{3.17}$$

and

$$Var[\hat{F}_{m,n}(x)] = n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V(x) + o_x(m^{-1/2}n^{-1}) \tag{3.18}$$

where $\sigma^2(x) = F(x)(1 - F(x))$ and $V(x) = f(x)[2x(1 - x)/\pi]^{1/2}$.

Also, it is easy to calculate the MSE of the Bernstein estimator after the variance and bias values are obtained.

$$\begin{aligned} MSE[\hat{F}_{m,n}(x)] &= n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V(x) + m^{-2}b^2(x) \\ &\quad + o(m^{-2}) + o_x(m^{-1/2}n^{-1}) \end{aligned} \tag{3.19}$$

The following theorem proved by Babu et al. (2002) shows the convergence rate of the Bernstein estimator.

Theorem 3.4 *Let F be a continuous probability distribution function on the interval $[0, 1]$. If $m, n \to \infty$, then*

$$||\hat{F}_{m,n}(x) - F(x)|| \to 0. \tag{3.20}$$

It is possible to find proof of all these theorems in Babu et al. (2002) and Leblanc (2012)'s works.

3.3 Unimodal Density Estimation

In common statistical inference, we are often willing to make the assumption that the underlying density f is unimodal. Indeed, unimodality comes more natural than symmetry; the real data usually show some skewness, but unimodality is more common. For this reason, statistical inference is typically based on an assumed family of unimodal parametric models.

Some of the available approaches to estimating unimodal densities are derived from monotonic densities. The mode is located on the right or left side of the density's support in monotone densities, which is a special case of unimodal densities. The nonparametric maximum likelihood (MLE) estimator was suggested on monotone densities for the first time by Grenander (1956). This MLE estimator is designed as the derivative of the least concave majorant of the ECDF on nonincreasing densities and the derivative of the greatest convex minorant of the ECDF on nondecreasing densities. This estimator was suggested by Grenander.

Recent unimodal density estimators are concentrated on producing smooth estimates by implementing the unimodality constraint. For this reason, Bernstein polynomials are preferred unimodal density estimations. Turnbull & Ghosh (2014) proposed a new unimodal density estimator using Bernstein polynomials. Next section is discussed considering the studies of Turnbull & Ghosh (2014) .

3.3.1 Unimodal Density Estimation with Bernstein Polynomials

Let $X_1, X_2, ..., X_n$ be i.i.d. real valued random variables with the continuous probability density function $f(.)$ which is known to be unimodal. In other words, let us assume that there exists an $x^* \in \mathbb{R}$. Namely, the probability function $f(x)$ is nondecreasing and nonincreasing density on the interval $(-\infty, x^*)$ and (x^*, ∞) respectively. Since the aim is to construct an estimate of $f(x)$, $\hat{f}(x)$ must provide the following constraints.

a) $\hat{f}(x) \geq 0$, for all $x \in \mathbb{R}$,

b) $\int \hat{f}(x)dx = 1$,

c) $\hat{f}(x)$ is unimodal, i.e., there exists an $x^* \in \mathbb{R}$, $\hat{f}(x)$ is nondecreasing and nonincreasing density on the interval $(-\infty, x^*)$ and (x^*, ∞), respectively.

We start by introducing a Bernstein polynomial of order $m-1$ to estimate $f(x)$ on a closed interval $[0, 1]$,

$$B_m(x, f) = \sum_{k=1}^{m} f(\frac{k-1}{m-1}) \binom{m-1}{k-1} x^{k-1}(1-x)^{m-k}. \tag{3.21}$$

The Bernstein polynomial $B_m(x, f)$ described in equation (3.21) is rescaled by Turnbull & Ghosh (2014) because it is not a proper density function. The rescaled version of $B_m(x, f)$ is defined as follows:

$$f_m(x) = \sum_{k=1}^{m} f(\frac{k-1}{m-1}) m \binom{m-1}{k-1} x^{k-1}(1-x)^{m-k} / \sum_{k=1}^{m} f(\frac{k-1}{m-1}). \tag{3.22}$$

Equation (3.22) as a proper density function can also be expressed as a mixture of Beta densities,

$$f_m(x, w) = \sum_{k=1}^{m} w_k f_b(x; k, m-k+1) \tag{3.23}$$

where $w = (w_1, w_2, ..., w_m)^T$ is a vector of weights of size m, and f_b is the Beta density function with shape parameters k and $m-k+1$. Notice that $f_m(x, w)$ will obtain all the desired properties of $\hat{f}(.)$ if the vector of weights, w, satisfies the following constraints:

1. $w_k \geq 0$, for $k = 1, ..., m$,
2. $\sum_{k=1}^{m} w_k = 1$,

3. $w_1 \leq w_2 \leq \cdots \leq w_{k^*} \geq w_{k^*+1} \geq \cdots \geq w_m$, for some $k^* \in \{1, 2, ..., m\}$, i.e. the w_k's are nondecreasing for $k \leq k^*$ and nonincreasing for $k \geq k^*$ (Turnbull & Ghosh (2014)).

It can be said that the 1, 2 and 3 constraints provide a, b and c properties respectively.

CHAPTER FOUR
NEW METHODS ON REGRESSION, DISTRIBUTION FUNCTION AND DENSITY ESTIMATION

In this chapter, we introduce a new method for regression, distribution function and unimodal density estimation, based on the geometric modelling.

4.1 Regression Modeling Based on Rational Bézier Curves

The most basic and simplest way to perform parametric regression analysis is probably the least squares method. Polynomial models are preferred in situations where the investigator knows that curvilinear effects are present in the regression function. These models are also used as functions that approximate nonlinear relations in very complex structures among variables. In this section, a new polynomial regression model is proposed that uses rational Bézier curves to add adjustable weights to provide closer approximation to the desired shape. Rational Bézier curves provide several advantages such as providing more control over a curve shape.

Polynomial regression equation is similar to rational Bézier curves regression equation. The difference between two equations is parameters. Instead of parameters $\beta_0, \beta_1, ..., \beta_m$ used in m^{th} degree polynomial regression, new parameters are defined as weighted control points $b_0, b_1, ..., b_m$. So the regression equations are as follows,

$$y_i = \frac{\sum_{k=0}^{m} w_k b_k B_{k,m}(x_i)}{\sum_{k=0}^{m} w_k B_{k,m}(x_i)} + \epsilon, x \in [0,1], i = 1, ..., n. \tag{4.1}$$

The goal is to determine all of the coefficients $b_0, b_1, ..., b_m$ of the curve that best fits the data meaning that the sum squares of residuals is minimum. The linear Bézier curves case is dealt with for simplicity. According to the minimization techniques, the following two equations need to be solved in order to obtain minimum values:

$$
\begin{aligned}
0 &= \frac{\partial \left(y_i - \frac{w_0 b_0 (1-x_i) + w_1 b_1 x_i}{w_0(1-x_i) + w_1 x_i} \right)^2}{\partial b_0} \\
&= 2 \sum_{i=1}^{n} \left(y_i - \frac{w_0 b_0 (1-x_i) + w_1 b_1 x_i}{w_0(1-x_i) + w_1 x_i} \right) \left(- \frac{w_0 (1-x_i)}{w_0(1-x_i) + w_1 x_i} \right), \quad (4.2)
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{\partial \left(y_i - \frac{w_0 b_0 (1-x_i) + w_1 b_1 (1-x_i)}{w_0(1-x_i) + w_1 (1-x_i)} \right)^2}{\partial b_1} \\
&= 2 \sum_{i=1}^{n} \left(y_i - \frac{w_0 b_0 (1-x_i) + w_1 b_1 x_i}{w_0(1-x_i) + w_1 x_i} \right) \left(- \frac{w_1 x_i}{w_0(1-x_i) + w_1 x_i} \right). \quad (4.3)
\end{aligned}
$$

Equations (4.2) and (4.3) can be rearranged in equation (4.4) and (4.5).

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{y_i w_0 (1-x_i)}{w_0(1-x_i) + w_1 x_i} &= \sum_{i=1}^{n} \frac{w_0^2 b_0 (1-x_i)^2}{(w_0(1-x_i) + w_1 x_i)^2} \\
&\quad + \sum_{i=1}^{n} \frac{w_0 w_1 b_1 (1-x_i) x_i}{(w_0(1-x_i) + w_1 x_i)^2}, \quad (4.4)
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{y_i w_1 x_i}{w_0(1-x_i) + w_1 x_i} &= \sum_{i=1}^{n} \frac{w_0 w_1 b_0 (1-x_i) x_i}{(w_0(1-x_i) + w_1 x_i)^2} \\
&\quad + \sum_{i=1}^{n} \frac{w_1^2 b_1 x_i^2}{(w_0(1-x_i) + w_1 x_i)^2}. \quad (4.5)
\end{aligned}
$$

Let

$$
X = \begin{bmatrix} \frac{w_0(1-x_1)}{w_0(1-x_1)+w_1 x_1} & \frac{w_1 x_1}{w_0(1-x_1)+w_1 x_1} \\ \frac{w_0(1-x_2)}{w_0(1-x_2)+w_1 x_2} & \frac{w_1 x_2}{w_0(1-x_2)+w_1 x_2} \\ \vdots & \vdots \\ \frac{w_0(1-x_n)}{w_0(1-x_n)+w_1 x_n} & \frac{w_1 x_n}{w_0(1-x_n)+w_1 x_n} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ and } b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}. \quad (4.6)
$$

Two equations in (4.4) and (4.5) can be written in the matrix form $X^T X b = X^T y$. In other words, (4.4) and (4.5) are the system of normal equations associated with regression equation in (4.1). The least squares estimator of control points is founded in the matrix form by $b = (X^T X)^{-1} X^T y$, where

$$X^T X = \begin{bmatrix} \sum_{i=1}^n \frac{w_0^2(1-x_i)^2}{(w_0(1-x_i)+w_1x_i)^2} & \sum_{i=1}^n \frac{w_0w_1(1-x_i)x_i}{(w_0(1-x_i)+w_1x_i)^2} \\ \sum_{i=1}^n \frac{w_0w_1(1-x_i)x_i}{(w_0(1-x_i)+w_1x_i)^2} & \sum_{i=1}^n \frac{w_1^2x_i^2}{(w_0(1-x_i)+w_1x_i)^2} \end{bmatrix} \tag{4.7}$$

and

$$(X^T X)^{-1} = \frac{\begin{bmatrix} \sum_{i=1}^n \frac{w_1^2x_i^2}{(w_0(1-x_i)+w_1x_i)^2} & -\sum_{i=1}^n \frac{w_0w_1(1-x_i)x_i}{(w_0(1-x_i)+w_1x_i)^2} \\ -\sum_{i=1}^n \frac{w_0w_1(1-x_i)x_i}{(w_0(1-x_i)+w_1x_i)^2} & \sum_{i=1}^n \frac{w_0^2(1-x_i)^2}{(w_0(1-x_i)+w_1x_i)^2} \end{bmatrix}}{\sum_{i=1}^n \frac{w_0^2(1-x_i)^2}{(w_0(1-x_i)+w_1x_i)^2} \sum_{i=1}^n \frac{w_1^2x_i^2}{(w_0(1-x_i)+w_1x_i)^2} - (\sum_{i=1}^n \frac{(1-x_i)x_i}{(w_0(1-x_i)+w_1x_i)^2})^2}. \tag{4.8}$$

Let $v_i = w_0(1-x_i) + w_1x_i$. Then,

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^n \frac{x_i^2}{v_i^2} \sum_{i=1}^n \frac{(1-x_i)y_i}{v_i} - \sum_{i=1}^n \frac{(1-x_i)x_i}{v_i^2} \sum_{i=1}^n \frac{x_iy_i}{v_i}}{w_0 \sum_{i=1}^n \frac{w_0^2(1-x_i)^2}{v_i^2} \sum_{i=1}^n \frac{w_1^2x_i^2}{v_i^2} - (\sum_{i=1}^n \frac{(1-x_i)x_i}{v_i^2})^2} \\ \frac{-\sum_{i=1}^n \frac{(1-x_i)x_i}{v_i^2} \sum_{i=1}^n \frac{(1-x_i)y_i}{v_i} - \sum_{i=1}^n \frac{(1-x_i)^2}{v_i^2} \sum_{i=1}^n \frac{x_iy_i}{v_i}}{w_1 \sum_{i=1}^n \frac{w_0^2(1-x_i)^2}{v_i^2} \sum_{i=1}^n \frac{w_1^2x_i^2}{v_i^2} - (\sum_{i=1}^n \frac{(1-x_i)x_i}{v_i^2})^2} \end{bmatrix}. \tag{4.9}$$

4.1.1 Properties of the New Regression Coefficients Estimator (Control Points)

The new regression coefficients estimator satisfy the two properties, unbiased and minimum variance. Two coefficients $(\hat{b}_0, \hat{b}_1)$ are calculated by $a_i = \frac{x_i}{w_0(1-x_i)+w_1x_i}$ and $k_i = \frac{1-x_i}{w_0(1-x_i)+w_1x_i}$. The estimators of control points can be written as follows:

$$\hat{b}_0 = \frac{1}{w_0} \left(\frac{\sum_{i=1}^n a_i^2 \sum_{i=1}^n k_iy_i - \sum_{i=1}^n a_ik_i \sum_{i=1}^n a_iy_i}{\sum_{i=1}^n a_i^2 \sum_{i=1}^n k_i^2 - \sum_{i=1}^n a_ik_i \sum_{i=1}^n a_ik_i} \right) \tag{4.10}$$

and

$$\hat{b}_1 = \frac{1}{w_1} \left(\frac{-\sum_{i=1}^n a_ik_i \sum_{i=1}^n k_iy_i - \sum_{i=1}^n k_i^2 \sum_{i=1}^n a_iy_i}{\sum_{i=1}^n a_i^2 \sum_{i=1}^n k_i^2 - \sum_{i=1}^n a_ik_i \sum_{i=1}^n a_ik_i} \right). \tag{4.11}$$

Expected values of these coefficients are

$$E(\hat{b}_0) = \frac{1}{w_0} \left(\frac{\sum_{i=1}^n a_i^2 E(\sum_{i=1}^n k_iy_i) - \sum_{i=1}^n a_ik_iE(\sum_{i=1}^n a_iy_i)}{\sum_{i=1}^n a_i^2 \sum_{i=1}^n k_i^2 - \sum_{i=1}^n a_ik_i \sum_{i=1}^n a_ik_i} \right) \tag{4.12}$$

and

$$E(\hat{b}_1) = \frac{1}{w_1}\left(\frac{-\sum_{i=1}^{n} a_i k_i E(\sum_{i=1}^{n} k_i y_i) - \sum_{i=1}^{n} k_i^2 E(\sum_{i=1}^{n} a_i y_i)}{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} k_i^2 - \sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} a_i k_i}\right). \tag{4.13}$$

$E(\sum_{i=1}^{n} k_i y_i)$ and $E(\sum_{i=1}^{n} a_i y_i)$ can be easily found. For equations $E(\sum_{i=1}^{n} k_i y_i) = \sum_{i=1}^{n} k_i E(y_i)$ and $E(\sum_{i=1}^{n} a_i y_i) = \sum_{i=1}^{n} a_i E(y_i)$,

$$\sum_{i=1}^{n} k_i \left(\frac{w_0 b_0 (1-x_i) + w_1 b_1 x_i}{w_0(1-x_i) + w_1 x_i}\right) = w_0 b_0 \sum_{i=1}^{n} k_i^2 + w_1 b_1 \sum_{i=1}^{n} a_i k_i \tag{4.14}$$

$$\sum_{i=1}^{n} a_i \left(\frac{w_0 b_0 (1-x_i) + w_1 b_1 x_i}{w_0(1-x_i) + w_1 x_i}\right) = w_0 b_0 \sum_{i=1}^{n} a_i k_i + w_1 b_1 \sum_{i=1}^{n} a_i^2. \tag{4.15}$$

Then, the following result is achieved when the values obtained in equations (4.14) and (4.15) are used in equations (4.12) and (4.13).

$$E(\hat{b}_0) = b_0 \text{ and } E(\hat{b}_1) = b_1 \tag{4.16}$$

Thus, $\hat{b}_0$ and $\hat{b}_1$ are unbiased estimators of b_0 and b_1 control points. The variances of $\hat{b}_0$ and $\hat{b}_1$ are

$$Var(\hat{b}_0) = Var\left(\frac{1}{w_0}\left(\frac{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} k_i y_i - \sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} a_i y_i}{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} k_i^2 - \sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} a_i k_i}\right)\right) \tag{4.17}$$

and

$$Var(\hat{b}_1) = Var\left(\frac{1}{w_1}\left(\frac{-\sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} k_i y_i - \sum_{i=1}^{n} k_i^2 \sum_{i=1}^{n} a_i y_i}{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} k_i^2 - \sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} a_i k_i}\right)\right). \tag{4.18}$$

Using the property of variance, we find

$$\begin{aligned} Var(\hat{b}_0) \quad = \quad & \left(\frac{(\sum_{i=1}^{n} a_i^2)^2 \sum_{i=1}^{n} k_i^2 \sigma^2 + (\sum_{i=1}^{n} a_i k_i)^2 \sum_{i=1}^{n} a_i^2 \sigma^2}{w_0^2 (\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} k_i^2 - \sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} a_i k_i)^2}\right. \\ & \left. - \frac{2\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} a_i k_i (\sum_{i=1}^{n} k_i \sum_{i=1}^{n} a_i)\sigma^2}{w_0^2 (\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} k_i^2 - \sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} a_i k_i)^2}\right) \end{aligned} \tag{4.19}$$

and

$$
\begin{aligned}
Var(\hat{b}_1) &= \Bigg(\frac{(\sum_{i=1}^{n} a_i k_i)^2 \sum_{i=1}^{n} k_i^2 \sigma^2 + (\sum_{i=1}^{n} k_i^2)^2 \sum_{i=1}^{n} a_i^2 \sigma^2}{w_1^2(\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} k_i^2 - \sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} a_i k_i)^2} \\
&\quad - \frac{2\sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} k_i^2 (\sum_{i=1}^{n} k_i \sum_{i=1}^{n} a_i)\sigma^2}{w_1^2(\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} k_i^2 - \sum_{i=1}^{n} a_i k_i \sum_{i=1}^{n} a_i k_i)^2}\Bigg). \qquad (4.20)
\end{aligned}
$$

We now check the minimum variance property for the parameters of the proposed regression model. Let $\tilde{b} = Cy$ be another linear estimator of the control points b and $C = (X^T X)^{-1} X^T + K$, where K is a mxn matrix. The aim is to show that the variance of our estimator is smaller. We obtain the expectation of $\tilde{b}$ as

$$
\begin{aligned}
E(Cy) &= E\left(((X^T X)^{-1} X^T + K)(Xb + \epsilon)\right) \\
&= ((X^T X)^{-1} X^T + K)Xb + ((X^T X)^{-1} X^T + K)E(\epsilon) \\
&= (X^T X)^{-1} X^T X b + KXb \\
&= (I + KX)b. \qquad (4.21)
\end{aligned}
$$

Hence, $\tilde{b}$ is unbiased if and only if $KX = 0$, the zero matrix. Furthermore, the variance of $\tilde{b}$ is

$$
\begin{aligned}
Var(\tilde{b}) &= Var(Cy) = CVar(y)C^T = \sigma^2 CC^T \\
&= \sigma^2((X^T X)^{-1} X^T + K)(X(X^T X)^{-1} + K^T) \\
&= \sigma^2 (X^T X)^{-1} + \sigma^2 (X^T X)^{-1}(KX)^T + \sigma^2 KX(X^T X)^{-1} + \sigma^2 KK^T \\
&= \sigma^2 (X^T X)^{-1} + \sigma^2 KK^T. \qquad (4.22)
\end{aligned}
$$

Since KK^T is a positive semi definite matrix, $Var(\tilde{b})$ is greater than $Var(\hat{b})$.

4.2 Distribution Function Estimation Method Using Rational Bernstein Polynomials

Bernstein polynomials are an attractive topic in density and distribution function estimation as they are the simplest examples of polynomial approximation with probabilistic interpretation. Bernstein polynomials have been generalized in many different ways, one of them is the rational Bernstein polynomials. Rational Bernstein polynomials reduce to Bernstein polynomial when all weights are equal. The rational Bernstein polynomial in geometric modelling (Computer Aided Geometric Design) provides many advantages for further control over the shape of a curve and the complete representation of the cones. This section presents a new nonparametric estimation method for the distribution function using rational Bernstein polynomials. The proposed method has significant advantages over the Bernstein polynomial estimation method.

Nonparametric distribution function estimation benefiting from linear combinations of Bernstein polynomials which has some disadvantages. Bernstein polynomials in Babu et al. (2002) depend only on the control points $F_n(k/m)$ whereas our rational Bernstein polynomials depend on the control points $F_n(k/m)$ and the associated weight w_k. In both cases, the control points are fixed and therefore the shape of Bernstein polynomial curves (Bézier curves) cannot be manipulated. However, in our method, we may use weights to change the shape of the rational Bernstein polynomial curve. As in the Bernstein polynomials, rational Bernstein polynomials converge uniformly for continuous probability distribution. The proof is given in Theorem 4.3. However, for a fixed degree, although the error in the Bernstein polynomial approximation is fixed, the error in the approximation using rational Bernstein polynomials depends on the weights. Thus, to get a better approximation with rational Bernstein polynomials, suitable weights can be chosen. For calculation of such optimal weights are given in the subsection 4.2.2. Also, a Bernstein polynomial cannot be perfect when it encounters a data set, which is very complicated and challenging. It is possible to make predictions much more appropriate and accurate, since the weights of the w_k are calculated from the empirical distribution in rational Bernstein polynomials. Considering all of these, we present a new method of distribution function estimation using rational Bernstein polynomials based on Bernstein - Weierstrass Approximation Theorem. This new method benefits from the mathematical advantages of rational function and combination of rational polynomial and probabilistic approach.

The initial goal of this section is to obtain a smooth estimate of the cumulative distribution function (cdf), $F(X) = P(X \leq x)$ where X is a random variable and we know that $F(.)$ is continuous. Beginning with a simple case, we consider random variables that are known to take values in the interval $[0, 1]$. The rational Bernstein estimator of order $m > 0$ of the distribution F is defined as

$$\hat{F}_{m,n,w}(x) = \sum_{k=0}^{m} F_n(k/m) R_{k,m}(x), \tag{4.23}$$

where

$$R_{k,m}(x) = \frac{w_k \binom{m}{k} x^k (1-x)^{m-k}}{\sum\limits_{j=0}^{m} w_j \binom{m}{j} x^j (1-x)^{m-j}}.$$

Similarly, the rational Bernstein polynomial of degree m for the function F, say $R_m(x)$, is defined as

$$R_m(x) = \sum_{k=0}^{m} F(k/m) R_{k,m}(x). \tag{4.24}$$

It is easy to see that $E[\hat{F}_{m,n,w}(x)] = R_m(x)$ by examining equations (4.23) and (4.24) for all $x \in [0, 1]$ and $n \geq 1$. Also, the following equations can be written deduced from the limit property of the cdf for $m \geq 1$ in the proposed estimator:

$$\hat{F}_{m,n,w}(0) = F(0) = R_m(0) = 0 \text{ and } \hat{F}_{m,n,w}(1) = F(1) = R_m(1) = 1. \tag{4.25}$$

Now, we can also check the first derivative of the proposed estimator. It follows from the first derivative of rational Bernstein polynomials (see Floater (1992)), that

$$\hat{F}'_{m,n,w} = \sum_{k=0}^{m-1} \lambda_k(x)(F_n((k+1)/m) - F_n(k/m)),$$

where

$$\begin{aligned}\lambda_k(x) &= \frac{m}{w_{0,m}^2(x)} \sum_{j=0}^{k} \sum_{l=k+1}^{m} \Big[B_{j,m-1}(x) B_{l-1,m-1}(x) \\ &\quad - B_{j-1,m-1}(x) B_{l,m-1}(x) \Big] w_j w_l \\ &= \frac{1}{(1-x) x w_{0,m}^2(x)} \sum_{j=0}^{k} \sum_{l=k+1}^{m} (l-j) B_{j,m}(x) B_{l,m}(x) w_j w_l,\end{aligned}$$

$w_{0,m}^2(x) = (\sum_{k=0}^{m} w_k B_{k,m}(x))^2$ and $B_{j,m}(x) = \binom{m}{j} x^j (1-x)^{m-j}$. Thus, we conclude that $\hat{F}_{m,n,w}(x)$ is a cdf, provided that the conditions $w_k > 0$ for $k = 0, 1, \ldots, m$ and $F_n((k+1)/m) - F_n(k/m) \geq 0$ for $k = 0, 1, \ldots, m-1$. As a special case, if all weights are equal, then $\hat{F}_{m,n,w}(x)$ reduces to the Bernstein polynomial, and thus the proposed cdf estimator reduces to the Bernstein polynomials cdf estimator given in Babu et al. (2002).

Our estimator method can also be used for any interval with a suitable transformation. If the support is any closed interval $[a, b]$, then the transformation $Y = (X - a)/(b - a)$ transforms $[a, b]$ to $[0, 1]$. Similarly the transformations $Y = X/(1 + X)$ and $Y = 1/2 + (1/\pi) tan^{-1} X$ transform the supports $[0, \infty)$ and $(-\infty, \infty)$ respectively to the interval [0,1], (see Babu et al. (2002)).

4.2.1 Theoretical Properties

In this section, we present the theoretical properties of the proposed estimator. First, we focus on basic properties such as the variance and bias of our estimator based on rational Bernstein polynomials. Then, we establish a convergence rate by examining the asymptotic properties of the rational Bernstein estimator $\hat{F}_{m,n,w}(x)$. $\mathcal{O}$ and o notations are used to characterize the boundary on approximation error in equation (4.26).

Theorem 4.1 *Let F be a continuous and differentiable function on the interval* $[0, 1]$.

Then, we have

$$R_m(x) = F(x) + \mathcal{O}(m^{-1/4}) + o(m). \tag{4.26}$$

The proof of Theorem 4.1 is given below.

Proof of Theorem 4.1: Let $R_m(x) = \sum_{k=0}^{m} F(k/m) R_{k,m}(x)$, where $R_{k,m}(x) = \frac{w_k \binom{m}{k} x^k (1-x)^{m-k}}{\sum_{j=0}^{m} w_j \binom{m}{j} x^j (1-x)^{m-j}}$. From the Taylor's theorem, we may write

$$F(k/m) = F(x_0) + F'(c_k)(k/m - x_0), \tag{4.27}$$

for some c_k is between k/m and x_0, i.e. $c_k \in (0,1)$. Multiplying equation (4.27) by $R_{k,m}(x_0)$ and sum from $k = 0$ to $k = m$ yield

$$\begin{aligned} R_m(x_0) &= \sum_{k=0}^{m} F(x_0) R_{k,m}(x_0) + \sum_{k=0}^{m} F'(c_k)(k/m - x_0) R_{k,m}(x_0) \\ &= F(x_0) \sum_{k=0}^{m} R_{k,m}(x) + \sum_{k=0}^{m} F'(c_k)(k/m - x_0) R_{k,m}(x_0) \\ &= F(x_0) + \sum_{k=0}^{m} F'(c_k)(k/m - x_0) R_{k,m}(x_0). \end{aligned}$$

Since $w_k > 0$ for all k, we may write

$$\begin{aligned} |R_m(x_0) - F(x_0)| &= |\sum_{k=0}^{m} F'(c_k)(k/m - x_0) R_{k,m}(x_0)| \\ &\leq \sum_{k=0}^{m} |F'(c_k)(k/m - x_0)| R_{k,m}(x_0) \\ &\leq M \sum_{k=0}^{m} |(k/m - x_0)| R_{k,m}(x_0) \end{aligned}$$

where $M = \sup_{0 \leq x \leq 1} |F'(x)|$. On the other hand, since $R_{k,m}(x) \leq W B_{k,m}(x)$, $k = 0, 1, ..., m$ where $W = \frac{\max_{0 \leq k \leq m} \{w_k\}}{\min_{0 \leq k \leq m} \{w_k\}}$ and $B_{j,m}(x) = \binom{m}{j} x^j (1-x)^{m-j}$, we may write

$$|R_m(x_0) - F(x_0)| \leq MW \sum_{k=0}^{m} |(k/m - x_0)| B_{k,m}(x_0).$$

Since x_0 is an arbitrary, we obtain

$$\begin{aligned}
|R_m(x) - F(x)| &\leq MW \sum_{k=0}^{m} |k/m - x| B_{k,m}(x) \\
&\leq MW \Big\{ \sum_{|\frac{k}{m}-x|<m^{-\frac{1}{4}}} |k/m - x| B_{k,m}(x) \\
&\quad + \sum_{|\frac{k}{m}-x|\geq m^{-\frac{1}{4}}} |k/m - x| B_{k,m}(x) \Big\} \\
&\leq MW \Big\{ m^{-\frac{1}{4}} \sum_{|\frac{k}{m}-x|<m^{-\frac{1}{4}}} B_{k,m}(x) \\
&\quad + \sum_{|\frac{k}{m}-x|\geq m^{-\frac{1}{4}}} |k/m - x| B_{k,m}(x) \Big\} \\
&\leq MW \Big\{ m^{-\frac{1}{4}} \sum_{k=0}^{m} B_{k,m}(x) + \sum_{|\frac{k}{m}-x|\geq m^{-\frac{1}{4}}} |k/m - x| B_{k,m}(x) \Big\} \\
&\leq MWm^{-\frac{1}{4}} + MW \sum_{|\frac{k}{m}-x|\geq m^{-\frac{1}{4}}} |k/m - x| B_{k,m}(x).
\end{aligned}$$

For $0 \leq \frac{k}{m} \leq 1$ and $x \in [0,1]$, we have $-1 \leq \frac{k}{m} - x \leq 1$. That is $|\frac{k}{m} - x| \leq 1$ for all k. Thus,

$$|R_m(x) - F(x)| \leq MWm^{-\frac{1}{4}} + MW \sum_{|\frac{k}{m}-x|\geq m^{-\frac{1}{4}}} B_{k,m}(x).$$

From Lemma 6.35 in Davis (1975), there is a constant K independent of m, such that for all $x \in [0,1]$,

$$\sum_{|\frac{k}{m}-x|\geq m^{-\frac{1}{4}}} B_{k,m}(x) \leq \frac{K}{m^{\frac{3}{2}}}.$$

Thus,

$$|R_m(x) - F(x)| \leq MWm^{-\frac{1}{4}} + MWKm^{-\frac{3}{2}}.$$

Therefore, we have

$$R_m(x) = F(x) + \mathcal{O}(m^{-1/4}) + o(m).$$

Theorem 4.2 *Let F be a continuous and differentiable function on the interval* $[0,1]$.

Then,

$$Bias[\hat{F}_{m,n,w}(x)] = E[\hat{F}_{m,n,w}(x)] - F(x) = \mathcal{O}(m^{-1/4}) \tag{4.28}$$

and

$$Var[\hat{F}_{m,n,w}(x)] = n^{-1}\sigma^2(x) - W^2 m^{-1/2} n^{-1} V(x) + \mathcal{O}(m^{-1/4}n^{-1}) \tag{4.29}$$

where $\sigma^2(x) = F(x)(1 - F(x))$ *and* $V(x) = f(x)[2x(1-x)/\pi]^{1/2}$.

Before giving the proof of this theorem, let us mention an important observation. One can calculate the MSE of our estimator using the Theorem 4.2 as

$$\begin{aligned} MSE[\hat{F}_{m,n,w}(x)] &= n^{-1}\sigma^2(x) - W^2 m^{-1/2} n^{-1} V(x) + \mathcal{O}(m^{-1/4}n^{-1}) \\ &\quad +(\mathcal{O}(m^{-1/4}))^2. \end{aligned} \tag{4.30}$$

Since, the MSE of the empirical cdf is known to be (see Leblanc (2012))

$$MSE[F_n(x)] = Var[F_n(x)] = n^{-1}\sigma^2(x),$$

we can conclude that our proposed estimator $\hat{F}_{m,n,w}(x)$ asymptotically reaches superior performance than the empirical cdf in the open interval $(0, 1)$ for x according to MSE criteria.

Proof of Theorem 4.2: Let $R_m(x) = \sum_{k=0}^{m} F(k/m) R_{k,m}(x)$. We know from Theorem 4.1 that, $|R_m(x) - F(x)| \leq MWm^{-\frac{1}{4}} + MWKm^{-\frac{3}{2}}$, where K is a constant independent of m, $M = \sup_{0\leq x\leq 1} |F'(x)|$ and $W = \frac{\max_{0\leq k\leq m}\{w_k\}}{\min_{0\leq k\leq m}\{w_k\}}$. Thus we have $R_m(x) = F(x) + \mathcal{O}(m^{-1/4})$. We also know that

$$E[\hat{F}_{m,n,w}(x)] = R_m(x)$$

for all $x \in [0,1]$. Now, we can define a new equation to calculate the variance and bias of our estimator $\hat{F}_{m,n,w}(x)$ any $x \in [0,1]$,

$$\Delta_i(x) = I(X_i \leq x) - F(x),$$

where I is the indicator function. Therefore the random variables $\Delta_i(x)$ are independent and identically distributed with zero mean. Then,

$$\begin{aligned}
\hat{F}_{m,n,w}(x) - R_m(x) &= \sum_{k=0}^{m} F_n(k/m)R_{k,m}(x) - \sum_{k=0}^{m} F(k/m)R_{k,m}(x) \\
&= \sum_{k=0}^{m} (F_n(k/m) - F(k/m))R_{k,m}(x) \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_{i,m}
\end{aligned}$$

where

$$Y_{i,m} = \sum_{k=0}^{m} \Delta_i(k/m)R_{k,m}(x).$$

Now, we can obtain the variance of the proposed estimator using the random variables $Y_{1,m}, Y_{2,m}, ..., Y_{n,m}$ which are i.i.d. with mean zero.

$$\begin{aligned}
Var[\hat{F}_{m,n,w}(x)] &= E[\hat{F}_{m,n,w}(x) - E(\hat{F}_{m,n,w}(x))]^2 \\
&= E[\hat{F}_{m,n,w}(x) - R_m(x)]^2 \\
&= E[\frac{1}{n}\sum_{i=1}^{n} Y_{i,m}]^2 = \frac{1}{n}E[Y_{1,m}^2].
\end{aligned}$$

Finally, it is possible to get $E[Y_{1,m}^2]$ using the following equation.

$$E[\Delta_1(x)\Delta_1(y)] = min(F(x), F(y)) - F(x)F(y)$$

for any $x, y \in [0,1]$. Subsequently,

$$\begin{aligned} E[Y_{1,m}^2] &= \sum_{k=0}^{m}\sum_{l=0}^{m} E[\Delta_1(k/m)\Delta_1(l/m)]R_{k,m}(x)R_{l,m}(x) \\ &= \sum_{k=0}^{m} F(\frac{k}{m})R_{k,m}^2(x) + 2\sum_{0\le k<l\le m} F(\frac{k}{m})R_{k,m}(x)R_{l,m}(x) - R_m^2(x). \end{aligned} \tag{4.31}$$

Since $F(k/m) = F(x) + \mathcal{O}(|k/m - x|)$, we have

$$\sum_{k=0}^{m} F(k/m)R_{k,m}^2(x) = F(x)\sum_{k=0}^{m} R_{k,m}^2(x) + \mathcal{O}(I_m(x)) \tag{4.32}$$

where $I_m(x) = \sum_{k=0}^{m} |k/m - x|R_{k,m}^2(x)$. Similarly, since

$$F(k/m) = F(x) + (k/m - x)f(x) + \mathcal{O}((k/m - x)^2)$$

where $f(x) = F'(x)$, and

$$1 = \sum_{k=0}^{m}\sum_{l=0}^{m} R_{k,m}(x)R_{l,m}(x) = 2\sum_{0\le k<l\le m} R_{k,m}(x)R_{l,m}(x) + \sum_{k=0}^{m} R_{k,m}^2(x).$$

We may write

$$\begin{aligned} \sum_{0\le k<l\le m} F(k/m)R_{k,m}(x)R_{l,m}(x) &= F(x)\sum_{0\le k<l\le m} R_{k,m}(x)R_{l,m}(x) \\ &\quad + f(x)\sum_{0\le k<l\le m} (\frac{k}{m} - x)R_{k,m}(x)R_{l,m}(x) \\ &\quad + \mathcal{O}(\sum_{0\le k<l\le m} (\frac{k}{m} - x)^2 R_{k,m}(x)R_{l,m}(x)) \\ &= \frac{1}{2}F(x)[1 - \sum_{k=0}^{m} R_{k,m}^2(x)] \\ &\quad + f(x)\sum_{0\le k<l\le m} (\frac{k}{m} - x)R_{k,m}(x)R_{l,m}(x) \\ &\quad + \mathcal{O}(\sum_{0\le k<l\le m} (\frac{k}{m} - x)^2 R_{k,m}(x)R_{l,m}(x)). \end{aligned} \tag{4.33}$$

Substituting equations (4.32) and (4.33) into equation (4.31), we get

$$\begin{aligned} E[Y_{1,m}^2] &= F(x) + 2f(x) \sum_{0\le k<l\le m} (k/m - x) R_{k,m}(x) R_{l,m}(x) \\ &\quad + \mathcal{O}(I_m(x)) + \mathcal{O}\Big(\sum_{0\le k<l\le m} (k/m-x)^2 R_{k,m}(x) R_{l,m}(x)\Big) - R_m^2(x). \end{aligned} \tag{4.34}$$

Since $I_m(x) = \sum_{k=0}^{m} |k/m - x| R_{k,m}^2(x) \le W^2 \sum_{k=0}^{m} |k/m - x| B_{k,m}^2(x)$, by using equality (20) in Leblanc (2012), we obtain $I_m(x) \le W^2 \left[\frac{\sum_{k=0}^{m} B_{k,m}^2(x)}{4m}\right]^{1/2}$. Hence using Lemma 2 in Leblanc (2012) gives

$$I_m(x) = \mathcal{O}(m^{-3/4}). \tag{4.35}$$

On the other hand, we may write

$$\begin{aligned} \sum_{0\le k<l\le m} \Big(\frac{k}{m} - x\Big)^2 R_{k,m}(x) R_{l,m}(x) &\le W^2 \sum_{0\le k<l\le m} \Big(\frac{k}{m} - x\Big)^2 B_{k,m}(x) B_{l,m}(x) \\ &\le W^2 (4m)^{-1}. \end{aligned} \tag{4.36}$$

Similarly, we have

$$\begin{aligned} \sum_{0\le k<l\le m} \Big(\frac{k}{m} - x\Big) R_{k,m}(x) R_{l,m}(x) &\le W^2 \sum_{0\le k<l\le m} \Big(\frac{k}{m} - x\Big) B_{k,m}(x) B_{l,m}(x) \\ &\le W^2 (m)^{-\frac{1}{2}} \Big\{-\Big[\frac{x(1-x)}{2\pi}\Big]^{\frac{1}{2}} + \mathcal{O}(1)\Big\}. \end{aligned} \tag{4.37}$$

Finally, from equation (4.26), we may write

$$R_m^2(x) = F^2(x) + \mathcal{O}(m^{-1/4}). \tag{4.38}$$

Now substituting (4.38) into (4.34) and using equations (4.35), (4.36) and (4.37) gives

$$\begin{aligned} E[Y_{1,m}^2] &= F(x)(1 - F(x)) - 2W^2 f(x) m^{-1/2} [x(1-x)/2\pi]^{1/2} + \mathcal{O}(m^{-1/4}) \\ &= \sigma^2(x) - W^2 f(x) m^{-1/2} [2x(1-x)/\pi]^{1/2} + \mathcal{O}(m^{-1/4}). \end{aligned} \tag{4.39}$$

Next theorem shows the uniform convergence of $\hat{F}_{m,n,w}(x)$.

Theorem 4.3 *Let F be a continuous probability distribution function on interval [0,1]. If $w_k > 0$ for all k then for any $\epsilon > 0$ there exists m_0 such that*

$$|\hat{F}_{m,n,w}(x) - F(x)| < \epsilon \tag{4.40}$$

for all $m > m_0$.

Proof of Theorem 4.3:

$$\begin{aligned}
|F(x) - \hat{F}_{m,n,w}(x)| &= \left| F(x) - \frac{\sum_{k=0}^{m} F_n(k/m) w_k \binom{m}{k} x^k (1-x)^{m-k}}{\sum_{j=0}^{m} w_j \binom{m}{j} x^j (1-x)^{m-j}} \right| \\
&= \left| \frac{\sum_{k=0}^{m} (F(x) - F_n(k/m)) w_k \binom{m}{k} x^k (1-x)^{m-k}}{\sum_{j=0}^{m} w_j \binom{m}{j} x^j (1-x)^{m-j}} \right|
\end{aligned}$$

Then,

$$|F(x) - \hat{F}_{m,n,w}(x)| \leq W \left| \sum_{k=0}^{m} (F(x) - F_n(k/m)) \binom{m}{k} x^k (1-x)^{m-k} \right|$$

where $W = \frac{\max\limits_{0 \leq k \leq m} \{w_k\}}{\min\limits_{0 \leq k \leq m} \{w_k\}}$. Since $F \in C[0,1]$ then there exists $\xi_k \in [0,1]$ such that

$$F(\xi_k) = F_n(k/m) \text{ for all } k.$$

Then,

$$W \left| \sum_{k=0}^{m} (F(x) - F_n(\frac{k}{m})) \binom{m}{k} x^k (1-x)^{m-k} \right| = W \left| \sum_{k=0}^{m} (F(x) - F(\xi_k)) B_{k,m}(x) \right|.$$

For fixed $x \in [0,1]$ choose $\delta > 0$. Let S_δ denote the set of all values of k satisfying $|\xi_k - x| \geq \delta$ and $|\frac{k}{m} - x| \geq \delta$.

$$\begin{aligned}
W\left|\sum_{k=0}^{m}(F(x)-F(\xi_k))B_{k,m}(x)\right| &= W\Bigg|\sum_{k\in S_\delta}(F(x)-F(\xi_k))B_{k,m}(x) \\
&\quad + \sum_{k\notin S_\delta}(F(x)-F(\xi_k))B_{k,m}(x)\Bigg| \\
&\leq W\Bigg(\sum_{k\in S_\delta}|F(x)-F(\xi_k)|B_{k,m}(x) \\
&\quad + \sum_{k\notin S_\delta}|F(x)-F(\xi_k)|B_{k,m}(x)\Bigg).
\end{aligned}$$

Since $F \in C[0,1]$, it is bounded and $|F(x)| \leq 1$ and

$$|F(x) - F(\xi_k)| \leq 1$$

for all k and all $x \in [0,1]$. Then, we have

$$\sum_{k\in S_\delta}|F(x)-F(\xi_k)|B_{k,m}(x) \leq \sum_{k\in S_\delta} B_{k,m}(x).$$

On the other hand, since $|\frac{k}{m} - x| \geq \delta$, we have

$$\frac{1}{\delta^2}\left|\frac{k}{m} - x\right|^2 \geq 1.$$

So,

$$\begin{aligned}
\sum_{k\in S_\delta}|F(x)-F(\xi_k)|B_{k,m}(x) &\leq \frac{1}{\delta^2}\sum_{k\in S_\delta}\left|\frac{k}{m}-x\right|^2 B_{k,m}(x) \\
&\leq \frac{1}{\delta^2}\sum_{k=0}^{m}\left(\frac{k}{m}-x\right)^2 B_{k,m}(x) \\
&\leq \frac{1}{\delta^2}\frac{x(1-x)}{m}. \text{ (see Phillips (2003))}
\end{aligned}$$

Thus, it follows from the inequality $0 \leq x(1-x) \leq \frac{1}{4}$ that

$$\sum_{k \in S_\delta} |F(x) - F(\xi_k)| B_{k,m}(x) \leq \frac{1}{4m\delta^2}.$$

Since $F(x)$ is continuous on $[0,1]$, for any $\epsilon > 0$, there is a number $\delta > 0$, depending on ϵ such that

$$|x - \xi_k| < \delta \Rightarrow |F(x) - F(\xi_k)| < \frac{\epsilon}{2}$$

for all x and $\xi_k \in [0,1]$. Thus,

$$\begin{aligned}
\sum_{k \notin S_\delta} |F(x) - F(\xi_k)| B_{k,m}(x) &\leq \frac{\epsilon}{2} \sum_{k \notin S_\delta} B_{k,m}(x) \\
&\leq \frac{\epsilon}{2} \sum_{k=0}^{m} B_{k,m}(x) \\
&\leq \frac{\epsilon}{2}.
\end{aligned}$$

That is, we have

$$|F(x) - \hat{F}_{m,n,w}(x)| \leq W \left(\frac{1}{4m\delta^2} + \frac{\epsilon}{2} \right)$$

Now, choosing $m_0 > \frac{W}{2\epsilon\delta^2(2-W)}$, we obtain

$$|F(x) - \hat{F}_{m,n,w}(x)| < \epsilon. \tag{4.41}$$

4.2.2 Determination of Weights

One can approximate the unknown distribution function F by a linear combination of rational Bernstein basis, where the unknown coefficients w_k can be estimated by minimizing the following sum of square based on ECDF.

$$\hat{\boldsymbol{w}} = \text{argmin} \sum_{i=1}^{n} \left[F_n(X_i) - \hat{F}_{m,n,w}(x) \right]^2 \tag{4.42}$$

As a result, the rational Bernstein polynomial estimator of the distribution function is given as

$$\hat{F}_{m,n,w}(x) = \sum_{k=0}^{m} F_n(k/m) \frac{\hat{w}_k \binom{m}{k} x^k (1-x)^{m-k}}{\sum\limits_{j=0}^{m} \hat{w}_j \binom{m}{j} x^j (1-x)^{m-j}}. \tag{4.43}$$

Also while w_k are being estimated, the specified constraints must be provided to ensure monotone nondecreasing and the cdf conditions ($w_k > 0$ for $k = 0, 1, \ldots, m$ and $F_n((k+1)/m) - F_n(k/m) \geq 0$ for $k = 0, 1, \ldots, m-1$). Thus, we can produce a smooth and monotone distribution estimator.

The solution of this optimization problem can be easily obtained by using quadratic programming techniques. The coefficients w_k are obtained by Wolfram Mathematica 10.1 programming in our simulation studies. The details of the simulation are presented in Chapter Five.

4.3 Unimodal Estimator Using B-Splines

Since the fundamental concepts of statistical inference are based on unimodal distribution families, it is very important to estimate these distributions. Geometric modeling tools such as the Bernstein polynomials are used extensively in the estimation of unimodal distributions. One of the most known of these methods is the work of Turnbull & Ghosh (2014). The work (Turnbull & Ghosh (2014)) introduced a method of unimodal density estimation via Bernstein polynomials and beta density. We generalize their approach using the Schoenberg operator that replaces polynomials by B-splines. An important advantage of the latter is their local control property and refinement. In practice, for data or continuous approximation purposes we refine knot sequences at each stage until a desired outcome is attained, while the degree of B-splines is fixed. Cubic B-splines and uniform (equally spaced) knot sequences are common choices.

To establish a probability density function based on the B-splines, we require the following important fact about their integration (see for example Goldman (2002)):

$$\int_{-\infty}^{+\infty} N_{k,d}(x)dx = \frac{1}{d+1}(x_{k+d+1} - x_k).$$

Schoenberg (1959) introduced the following piecewise polynomial a variation diminishing operator as a generalization of the Bernstein operator,

$$S_k(f;x) = \sum_{k=-d}^{n-1} f(\overline{x}_{k,d}) N_{k,d}(x), \tag{4.44}$$

with positive integers n, d and a nondecreasing (extended) knot sequence

$$T_{n,d} = (x_{-d} = \cdots = x_0 = 0 < x_1 < \cdots < x_n = \cdots = x_{n+d} = 1), \tag{4.45}$$

and $\overline{x}_{k,d} = (x_{k+1} + \cdots + x_{k+d})/d$.

Now for each nonnegative integer d, a positive integer n, and a given nondecreasing extended knot sequence similar to (4.45) and a function f on $[a, b]$, we define a new probability density function by

$$g_{n,d}(f, x) = \sum_{k=-d}^{n-1} \frac{d+1}{x_{k+d+1} - x_k} f(\overline{x}_{k,d}) N_{k,d}(x) \,/\, \left(\sum_{k=-d}^{n-1} f(\overline{x}_{k,d}) \right). \tag{4.46}$$

It is easy to verify that

$$\int_{-\infty}^{+\infty} g_{n,d}(f, x)dx = \int_{a}^{b} g_{n,d}(f, x)dx = 1.$$

This allows us to construct a family of unimodal density estimates

$$h_{n,d}(x, w) = \sum_{k=-d}^{n-1} w_k \frac{d+1}{x_{k+d+1} - x_k} N_{k,d}(x), \tag{4.47}$$

when weights satisfy the following properties:

1. $w_k \geq 0$ for each k and $\sum\limits_{k=-d}^{n-1} w_k$=1,

2. w is unimodal, i.e. there exists an integer m such that

   $w_{-d} \leq w_{-d+1} \leq \cdots \leq w_m \geq w_{m+1} \geq \cdots \geq w_{n-1}$.

The equations (4.46) and (4.47) generalize the method given by Turnbull & Ghosh (2014) based on the Bernstein polynomials.

Special case: When $n = 1$, $N_{k,d}(x)$ with the knot sequence $T_{n,d} = (0, \ldots, 0, 1, \ldots, 1)$ in which both 0 and 1 repeat $d + 1$ times yields Bernstein basis polynomials of degree d on $[0, 1]$,

$$\binom{d}{k} x^k (1-x)^{d-k}, \quad k = 0, 1, \ldots, d$$

or the Binomial distribution.

It follows that the density in (4.46) reduces to

$$g_{n,d}(f, x) = \sum_{k=0}^{d} (d+1) f(k/d) \binom{d}{k} x^k (1-x)^{d-k} \Big/ \left(\sum_{k=0}^{d} f(k/d) \right),$$

the beta density function with shape parameters d and k. Therefore the functions on the right of (4.47), $(d + 1)/(x_{k+d+1} - x_k)N_{k,d}(x)$ whose shape depend on k, d and a knot sequence T may be viewed as a generalization of the beta density.

Meanwhile, we want to emphasize three references in a similar scope. The work Goldman (1988) investigates urn models arising from the B-splines at nonnegative integer knots and Karlin's classic book on total positivity Karlin (1968) has a blend of variation diminishing kernels, and the paper Carnicer & Peña (1994) focuses on optimality of B-splines.

CHAPTER FIVE
SIMULATION AND NUMERICAL EXAMPLES

Simulation studies and numerical examples obtained by comparing the proposed methods with the available methods in the literature are presented for each of the three different areas handled by geometric modeling.

5.1 Simulation Studies and Numerical Example for New Regression Model

In Section 5.1.1, the regression model based on the rational Bézier curve is established and the results are compared with polynomial regression.

5.1.1 Numerical Example

The prestige data set has one dependent and independent variable. The dependent variable is prestige score for occupation; independent variable is average income of incumbents. The data set, which is available in R programming, has 102 observations. Firstly, simple linear regression and modeling based on rational Bézier curves of degree 1 are applied on this data set. The weights of control points are determined as 1 and 3, respectively. Fitted values for two different models are given in Figure 5.1.

The coefficient of determination for simple linear regression and modeling based on rational Bézier curves of degree 1 are found 0.5112 and 0.5862, respectively. The proposed model is better represented to data than the simple linear regression, and it can be said that is mimicked to data by giving appropriate weight to the control point in compliance with the shape of the data set. More control over the shape of a curve is provided owing to flexible structure of models established by using rational Bézier curves.

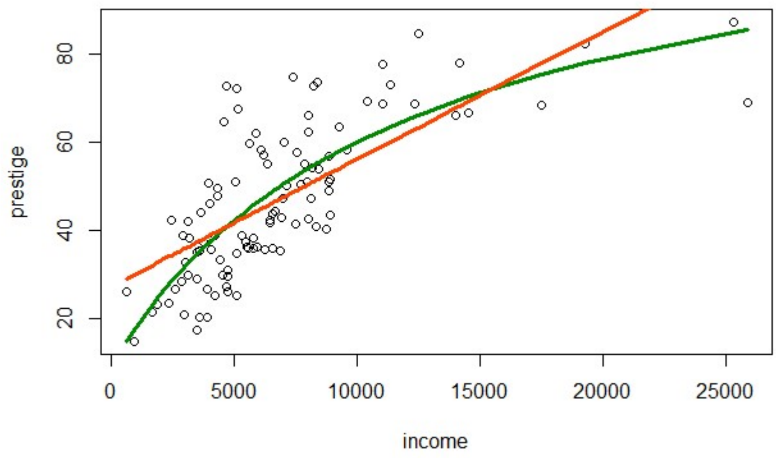

Figure 5.1 Fitted values for simple linear regression and modeling based on rational Bézier curves of degree 1

5.1.2 Simulation Studies

The performance of proposed regression method is compared with polynomial regression methods by simulation. First, dependent variable and independent variable values are derived 10000 times. Then, we generate 10000 Monte Carlo samples of size n = 15, 25, 35 for the floating variance of the error term. The value of the first control point is taken 1 in regression model based on rational Bézier Curve. The reason for this is to ensure simplicity of operation and avoid repeated results. The weights of the other control points are assigned to maximize the coefficient of determination. An algorithm is used for this case in R programming.

Relative efficiency values (MSE_{PR}/MSE_{RBC}) are given in Table 5.1 (PR: Polynomial regression, RBC: Rational Bézier curves). σ^2 denotes the variance of the error term.

The proposed model for each n and σ^2 has given the best results with regard to relative efficiency at linear, quadratic and cubic degree. All values in the table are greater than 1. Therefore, it is concluded that modelling based on rational Bézier curves is better than polynomial regression. If n increases, the relative efficiency

Table 5.1 Relative efficiency values (MSE_{PR}/MSE_{RBC}) for linear, quadratic and cubic degree respectively

		Linear	Quadratic	Cubic
n	σ^2	RE	RE	RE
	10	1.0741	1.1437	1.1665
15	30	1.0841	1.1502	1.1829
	50	1.0953	1.1588	1.1943
	100	1.1023	1.1673	1.2067
	10	1.0540	1.1058	1.1127
25	30	1.0604	1.1113	1.1205
	50	1.0720	1.1211	1.1371
	100	1.0772	1.1296	1.1485
	10	1.0447	1.0862	1.0958
35	30	1.0496	1.0919	1.1009
	50	1.0526	1.1006	1.1145
	100	1.0604	1.1085	1.1297

values close to one. But when n decreases, the relative efficiency values increase gradually. Also, the increment of the variance of the error term significantly improves the value of relative efficiency. For instance, the relative efficiency value is 1.2067 for $n = 15$ and $\sigma^2 = 100$ which is higher than 1.1023 or 1.1673. These results show us the new technique is suitable.

From the Table 5.1 we showed that, if the variance of the error term is fixed and n is increasing, the relative efficiency values decrease. If the n is fixed and the variance of the error term is increasing, the relative efficiency values increase.

5.2 Simulation Studies and Numerical Examples for New Distribution Function Estimation Method

In this section, the performance of the new distribution function estimation method under appropriate choice of the degree of the polynomials (cubic and septic) is compared with the Bernstein polynomial and the empirical distribution function method by using simulation.

The simulation study consists of eight different underlying distribution functions: standard normal distribution, skewed unimodal distribution, strongly skewed distribution, outlier distribution, which are selected from Marron & Wand (1992), and gamma distribution, exponential distribution, beta distribution, triangular distribution. These distributions cover a broad range of shapes, supports and characteristics (see Xue & Wang (2010)). The parameters for these distributions are given in Table 5.2. 2000 simulated samples for the sample sizes $n = 20, 50$ and 100 are generated for each of the eight distributions respectively. We consider five estimators of the distribution function: Cubic Bernstein polynomial (BP3), septic Bernstein polynomial (BP7), cubic rational Bernstein polynomial (RBP3), septic rational Bernstein polynomial (RBP7) and empirical cumulative distribution function (F_n). We compare the estimation accuracy of the four estimators with the empirical estimator with regard to their averaged squared errors (ASE). The ASE is defined by $ASE(\hat{F}) = (1/n)\sum_{i=1}^{n}\left[\hat{F}(X_i) - F(X_i)\right]^2$ for a given $\hat{F}$. When the ASE value is small, it means that the cdf estimation method approximates better to the actual distribution function.

Table 5.2 Distribution functions used in the simulation study

Name	Distribution function
Standard normal	$N(0,1)$
Skewed unimodal	$0.2N(0,1) + 0.2N(\frac{1}{2}, (\frac{2}{3})^2) + 0.6N(\frac{13}{12}, (\frac{5}{9})^2)$
Strongly skewed	$\sum_{l=0}^{7} \frac{1}{8}N(3[(\frac{2}{3})^l - 1], (\frac{2}{3})^{2l})$
Outlier	$0.1N(0,1) + 0.9N(0, (\frac{1}{10})^2)$
Gamma	$Gamma(2,2)$
Exponential	$Exp(4)$
Beta	$Beta(4,4)$
Triangular	$Tri(0,1)$

In Table 5.3, it is clear that the ASE values decrease as the sample size n increases. To examine the performance of the five estimators, the ASE values obtained for all models and sample size combinations in Table 5.3 are ordered by subscripts from small to large. When the ASE values are examined, it is observed that at least one of the RBP3 and RBP7 estimators has the smallest ASE value, except for four cases (standard normal

Table 5.3 Simulation results: ASE values (x10^{-3}) of three different estimation method

		degree 3		degree 7		
Distribution function	n	BP3	RBP3	BP7	RBP7	F_n
Standard normal	20	8.12_4	7.10_3	4.65_1	7.07_2	8.13_5
	50	6.04_5	3.76_3	2.66_1	3.43_2	3.94_4
	100	4.99_5	1.99_4	1.63_2	1.50_1	1.69_3
Skewed unimodal	20	14.91_5	7.26_2	8.57_3	7.18_1	8.70_4
	50	12.91_5	3.40_2	6.45_4	2.96_1	3.47_3
	100	12.03_5	2.00_3	5.58_4	1.47_1	1.67_2
Strongly skewed	20	31.76_5	6.72_1	21.27_4	7.56_2	8.56_3
	50	30.89_5	3.21_1	18.28_4	3.22_2	3.29_3
	100	30.82_5	2.11_3	17.46_4	1.93_2	1.64_1
Outlier	20	25.56_5	7.57_2	17.63_4	6.77_1	8.53_3
	50	24.88_5	4.44_3	15.54_4	2.52_1	3.21_2
	100	24.83_5	3.55_3	14.85_4	1.33_1	1.64_2
Gamma	20	14.98_5	6.68_1	10.41_4	7.06_2	8.58_3
	50	13.68_5	2.95_2	8.22_4	2.86_1	3.44_3
	100	13.11_5	1.62_2	7.33_4	1.43_1	1.67_3
Exponential	20	25.87_5	6.96_1	7.91_3	7.65_2	8.85_4
	50	25.34_5	2.59_1	6.43_4	2.86_2	3.36_3
	100	25.03_5	1.32_1	5.82_4	1.44_2	1.68_3
Beta	20	12.70_5	6.25_1	7.21_3	7.06_2	8.42_4
	50	10.77_5	2.69_1	4.90_4	2.94_2	3.48_3
	100	9.97_5	1.34_1	4.05_4	1.43_2	1.68_3
Triangular	20	8.33_4	6.73_2	5.71_1	7.48_3	8.63_5
	50	6.39_5	2.74_1	3.32_3	3.01_2	3.45_4
	100	5.53_5	1.32_1	2.35_4	1.44_2	1.65_3

$n = 20$ and 50, strongly skewed $n = 100$, triangular $n = 20$).

RBP3 always have lower ASE values compared to the BP3. RBP7 is closer to the true cumulative distribution function in comparison with BP7 except for three cases. When we compare RBP3 and BP7, it is observed that in general more efficient performance is reached with the smaller degree of our estimator. Also, the empirical distribution function only achieves the best results in strongly skewed $n = 100$.

5.2.1 Numerical Examples

5.2.1.1 Suicide Data

In this section, we illustrate the proposed method by analyzing the real data. Suicide data (see Silverman (1986)) is used to help demonstrate the performance of our proposed method and other methods. This data includes the duration of the psychiatric treatment (in days) administered to 86 patients participated in the study of suicide risk as the control group. The minimum and maximum values contained in this data are 1 and 737 ($\min x_i = 1$ and $\max x_i = 737$), respectively. Then, we analyze the original data after rescaling it to the unit interval.

Rational Bernstein estimator of degree $m = 3$, Bernstein estimator of degree $m = 3$, Bernstein estimator of degree $m = 50$ and empirical distribution function are used to estimate the underlying cumulative distribution function of F. Results obtained for four different estimators underlying the actual distribution function of the durations of psychiatric treatment are presented in Figure 5.2.

Figure 5.2 shows that the rational Bernstein estimator for degree 3 has results very similar to the empirical distribution function. On the other hand, Bernstein estimator for degree 3 does not give good results compared with the rational Bernstein estimator. Bernstein estimator requires high degree of polynomial for this real data set. According to calculation, the polynomial degree of m should be around 50 in order for the Bernstein estimator to be similar to the rational Bernstein estimator for degree 3.

5.2.1.2 Eruption Time (Old Faithful) Data

In addition to the first example, another example is given to compare the performance of the new proposed distribution function estimator with previous methods. In this example, we use a data set that specifies the duration (in minutes) of

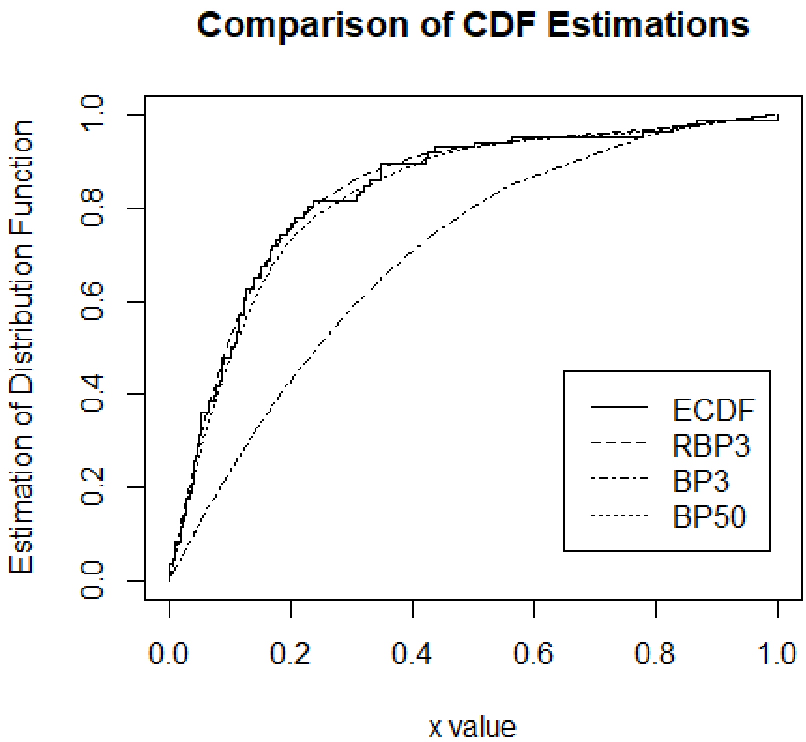

Figure 5.2 Suicide data: plots of the estimated distribution function using rational Bernstein polynomials for degree 3 (dashed), Bernstein polynomials for degree 3 (dot dash), and Bernstein polynomials for degree 50 (dotted), along with the empirical distribution function (solid)

eruption in Yellowstone National Park, Wyoming, USA, which is commonly used in the literature. This data set consists of 272 observations.

Initially, the data is rescaled as in the previous example, and the degrees of the rational Bernstein and the Bernstein estimators are chosen as quartic ($m = 4$). Then, the distribution functions obtained from different estimators for the distribution function of eruption time data are estimated and their graphs are given in Figure 5.3. Figure 5.3 shows that the distribution function obtained by the rational Bernstein estimator approaches to the empirical distribution function (F_n). In addition, the quartic Bernstein polynomial distribution estimator does not give any results close to the rational Bernstein and empirical distribution estimators. This shows that the degree of polynomial given is not sufficient for Bernstein estimation. Bernstein estimator requires high degree of polynomial for this real data set. According to these results, the polynomial degree of m should be around 50 in order for the Bernstein estimator to be similar to the rational Bernstein estimator for degree 4. The proposed new estimator gives good results without increasing the degree of the polynomial.

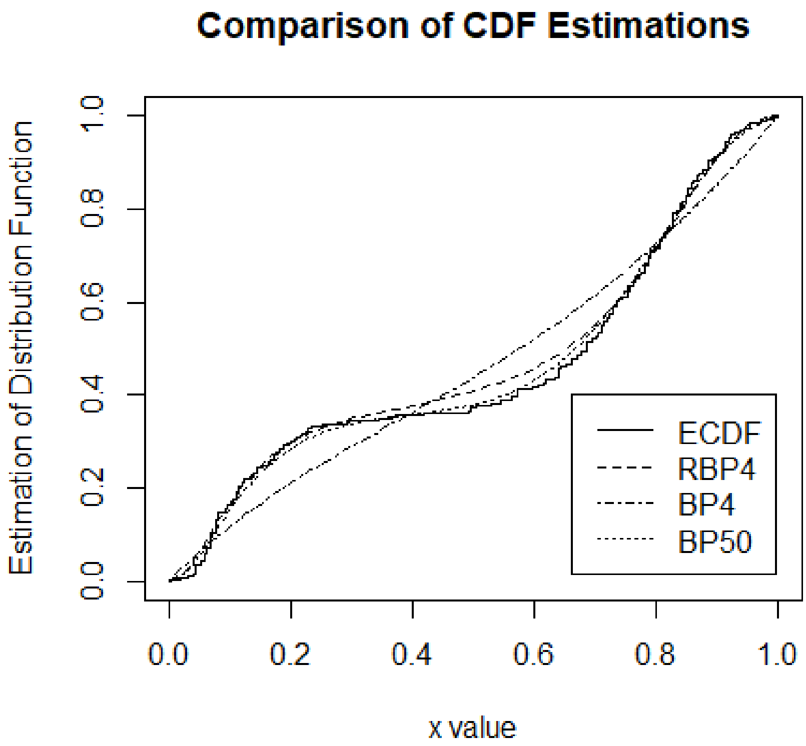

Figure 5.3 Eruption time data: plots of the estimated distribution function using rational Bernstein polynomials for degree 4 (dashed), Bernstein polynomials for degree 4 (dot dash), and Bernstein polynomials for degree 50 (dotted), along with the empirical distribution function (solid)

5.3 Numerical Examples for Unimodal Estimator Using B-Splines

In this section, we present an example using simulated data to investigate the performance of proposed unimodal density estimation method. We explore this by generating data from Beta distribution with parameters 2 and 5 for the sample size 100. Then we choose weights w_k to construct the estimated density. The software R is used to carry out computations. The results are compared to the Bernstein polynomials estimation given by Turnbull & Ghosh (2014) and the proposed estimation with the beta density. In both methods, the Bernstein polynomials and

B-splines keep the degree d fixed, here we take $d = 3$. Since the uniform knot sequence has been the most influential and widely implemented choice, $T_{n,d}$ is selected as $(-0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6)$.

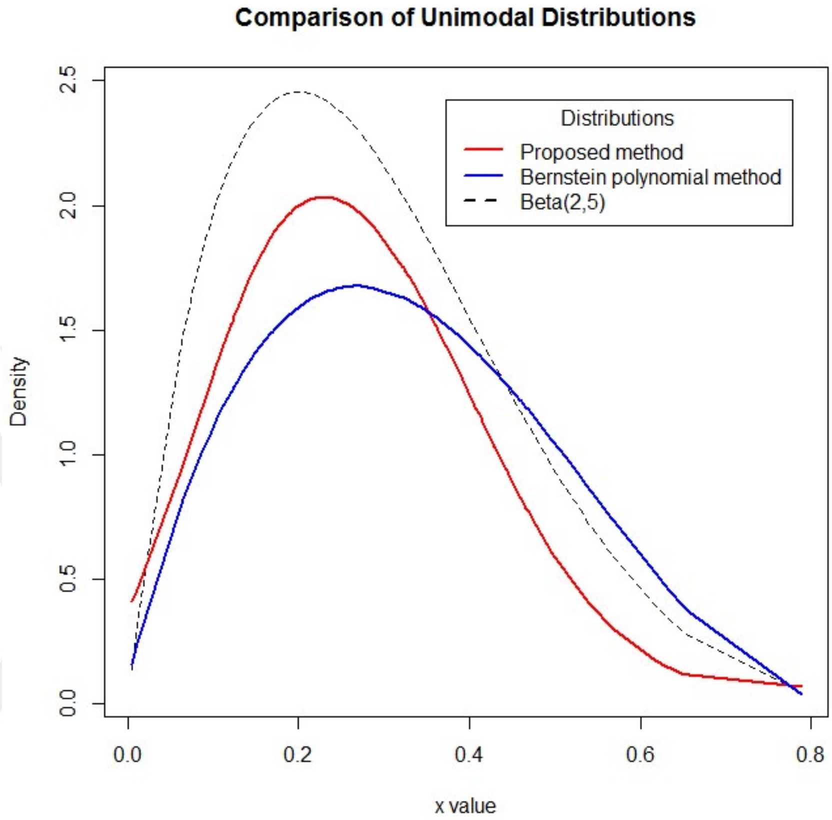

Figure 5.4 Comparison of unimodal distributions

In the Figure 5.4, we display the plot of Beta(2,5), proposed estimator and Bernstein polynomials estimator of density function for the distribution. The graphs between true and estimated density functions show that the proposed estimator for the density function presents an impressively smooth curve in the approximation of true distribution. ASE of the proposed estimation function and Bernstein polynomials estimation function are obtained as 0.1643 and 0.3554, respectively. ASE of the proposed estimation function is reduced. It means that the proposed estimator is closer to the actual density function. The advantage of the proposed method is the local property, i.e. refining the knots while keeping the degree fixed of B-splines yields better estimates. We also give another example to verify this property.

In this example, the number of knots is increased for the same data and the same degree and $T_{n,d}$ is taken as;

$$(-0.15, -0.1, -0.05, 0, 0.05, 0.1, \ldots, 0, 9, 0.95, 1, 1.05, 1.1, 1.15).$$

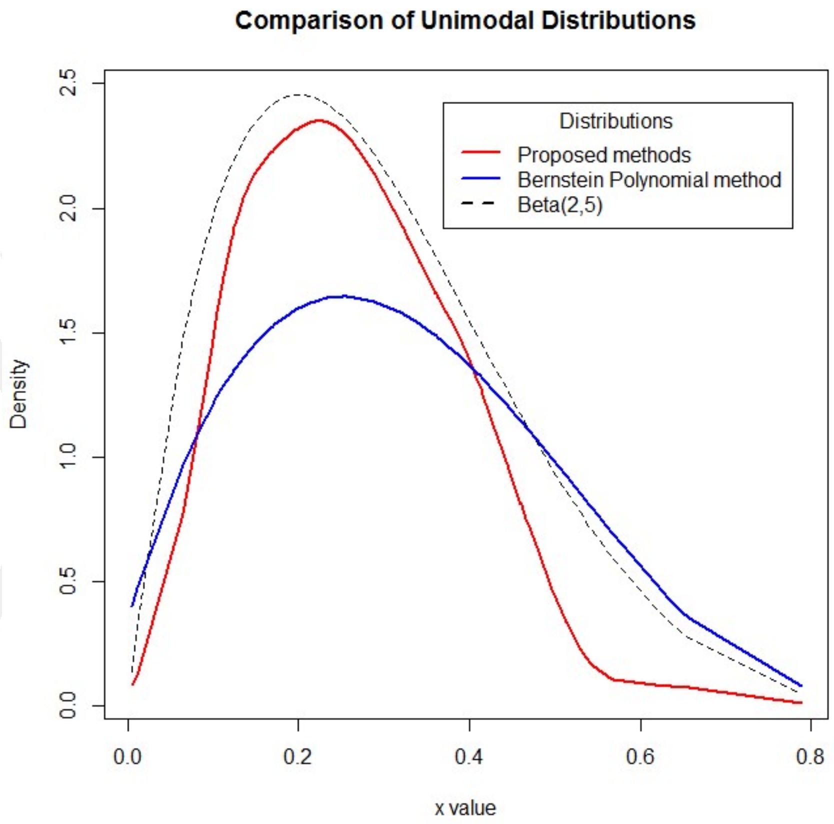

Figure 5.5 Comparison of unimodal distributions

As shown in Figure 5.5, our unimodal density estimation method achieves a good approximation of true probability density function. The averaged squared errors (ASE) of the proposed estimator value is 0.0701. When the number of knots increases, ASE of the proposed estimation function reduces. It means that the proposed estimator is closer to the actual density function. Thus, it is possible to make a better unimodal density estimate by taking increasing number of knots without changing the degree by means of our proposed method.

CHAPTER SIX
CONCLUSIONS

The contribution of this thesis is comprised of three parts. First, we propose a new polynomial regression model that uses rational Bézier curves to add adjustable weights for providing a closer approximation to the desired shape. The most important advantage of the rational Bézier curves is that they allow more control over the curve with adjustable weights. Simulation studies and numerical example demonstrate the effectiveness of our approach compared to the polynomial regression method. In the linear, quadratic, and cubic case, the proposed methods give the best results with regard to relative efficiency for each sample size and the floating variance of the error term. These results indicate that the new technique is very suitable and useful.

Secondly, we introduce a cdf estimation based on the rational Bernstein polynomial to generalize the Bernstein polynomials cdf estimation method given by Babu et al. (2002). We present some important theoretical properties about the proposed estimator. Then, we investigate the performance of our method with simulation study and two numerical examples. In general, the results show that the ASE values of the proposed estimator are smaller than the ASE values of the Bernstein polynomial estimator. Thus, it is possible to achieve a higher performance with low-degree polynomials using the rational Bernstein polynomials, instead of dealing with high-degree polynomials by preferring the Bernstein estimator. In addition, our estimator based on the rational Bernstein polynomials usually provides estimates with smaller ASE values than the empirical distribution estimator. Furthermore, determining the coefficients w_k by using the data shape and structure decreases ASE while keeping the degree of the curved unchanged. This makes our method a suitable choice to achieve a better cdf estimation. The results show that rational Bernstein polynomials present an impressive improvement for approximation to real distribution function.

Thirdly, we present a unimodal density estimator based on B-splines to generalize

Bernstein polynomials and beta density by Turnbull & Ghosh (2014). The numerical examples given in Chapter Five show the effectiveness of our approach compared to the Bernstein polynomial method. Smaller ASE values are obtained with the proposed method. In addition, by increasing the knot sequence of the B-splines, the ASE values decrease without changing the degree of polynomial. This makes our method a better option to obtain a good density estimate.

REFERENCES

Babu, G. J., Canty, A. J., & Chaubey, Y. P. (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planning and Inference*, *105*(2), 377–392.

Bae, W., Choi, H., Park, B. U., & Kim, C. (2005). Smoothing techniques for the bivariate Kaplan-Meier estimator. *Communications in Statistics - Theory and Methods*, *34*(7), 1659–1674.

Barry, P. J., Beatty, J. C., & Goldman, R. N. (1992). Unimodal properties of B-spline and Bernstein-basis functions. *Computer-Aided Design*, *24*(12), 627–636.

Bernšteın, S. (1912). Démonstration du théoreme de Weierstrass fondée sur le calcul des probabilities. *Communications of the Kharkov Mathematical Society*, *13*, 1–2.

Budakçı, G., & Oruç, H. (2012). Bernstein–Schoenberg operator with knots at the q-integers. *Mathematical and Computer Modelling*, *56*(3-4), 56–59.

Carnicer, J. M., & Peña, J. M. (1994). Totally positive bases for shape preserving curve design and optimality of B-splines. *Computer Aided Geometric Design*, *11*(6), 633–654.

Cencov, N. (1962). Evaluation of an unknown density from observations. *Soviet Mathematics*, *3*, 1559–1562.

Davis, P. J. (1975). *Interpolation and approximation*. New York: Courier Corporation.

Dutta, S. (2016). Distribution function estimation via Bernstein polynomial of random degree. *Metrika*, *79*(3), 239–263.

Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, *11*(2), 89-102.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. New York: CRC press.

Fan, J., & Gijbels, I. (1996). *Local polynomial regression*. London: Chapman and Hall.

Fix, E., & Hodges Jr, J. L. (1951). *Discriminatory analysis-nonparametric discrimination: consistency properties*. Berkeley: University of California.

Floater, M. S. (1992). Derivatives of rational Bézier curves. *Computer Aided Geometric Design*, *9*(3), 161–174.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistic*, *19*(1), 1–67.

Goldman, R. N. (1988). Urn models, approximations, and splines. *Journal of Approximation Theory*, *54*(1), 1–66.

Goldman, R. N. (2002). *Pyramid algorithms: A dynamic programming approach to curves and surfaces for geometric modeling*. USA: Elsevier Science.

Good, I., & Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika*, *58*(2), 255–277.

Good, I., & Gaskins, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *Journal of the American Statistical Association*, *75*(2), 42–56.

Green, P. J., & Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Boca Raton: CRC Press.

Grenander, U. (1956). On the theory of mortality measurement: part ii. *Scandinavian Actuarial Journal*, *1956*(2), 125–153.

Hansen, M. H., Kooperberg, C. et al. (2002). Spline adaptation in extended linear models. *Statistical Science*, *17*(1), 2–51.

Hastie, T., & Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, *46*(4), 1005–1016.

He, X., & Shi, P. (1998). Monotone B-spline smoothing. *Journal of the American statistical Association*, *93*(2), 643–650.

Huang, J. Z. (1998). Projection estimation in multiple regression with application to functional anova models. *The Annals of Statistics*, *26*(1), 242-272.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics*, *31*(5), 1600-1635.

Kakizawa, Y. (2004). Bernstein polynomial probability density estimation. *Journal of Nonparametric Statistics*, *16*(5), 709–729.

Karlin, S. (1968). *Total positivity*. California: Stanford University Press.

Kim, C. (1996). Nonparametric density estimation via the Bezier curve. *Proceeding of the ASA Section on Statistical Graphics; Chicago, IL*, *55*(2), 25–28.

Kim, C., Hong, C., & Jeong, M. (2000). Simulation-extrapolation via the Bezier curve in measurement error models. *Communications in Statistics - Simulation and Computation*, *29*(4), 1135–1147.

Kim, C., Kim, W., Park, B. U., Hong, C., & Jeong, M. (1999). Smoothing techniques via the Bezier curve. *Communications in Statistics - Theory and Methods*, *28*(7), 1577–1597.

Leblanc, A. (2010). A bias-reduced approach to density estimation using Bernstein polynomials. *Journal of Nonparametric Statistics*, *22*(4), 459–475.

Leblanc, A. (2012). On estimating distribution functions using Bernstein polynomials. *Annals of the Institute of Statistical Mathematics*, *64*(5), 919–943.

Loader, C. (1999). Bandwidth selection: classical or plug-in? *The Annals of Statistics*, *27*(2), 415–438.

Loader, C. (2006). *Local regression and likelihood*. New York: Springer Science & Business Media.

Lorentz, G. G. (2012). *Bernstein polynomials*. Chelsea: American Mathematical Society.

Marron, J. S., & Wand, M. P. (1992). Exact mean integrated squared error. *The Annals of Statistics*, *20*(2), 712–736.

Marsh, D. (2006). *Applied geometry for computer graphics and CAD*. London: Springer Science & Business Media.

Ogden, T. (2012). *Essential wavelets for statistical applications and data analysis*. New York: Springer Science & Business Media.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical statistics*, *33*(3), 1065–1076.

Phillips, G. M. (2003). *Interpolation and approximation by polynomials*. Berlin: Springer Science & Business Media.

Qu, A., & Li, R. (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*, *62*(2), 379–391.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, *27*(3), 832–837.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2009). Semiparametric regression during 2003–2007. *Electronic Journal of Statistics*, *3*, 1193.

Schoenberg, I. J. (1946). Contributions to the problem of approximation of equidistant data by analytic functions. part b. on the problem of osculatory interpolation. a second class of analytic approximation formulae. *Quarterly of Applied Mathematics*, *4*(2), 112–141.

Schoenberg, I. J. (1959). On variation diminishing approximation methods. *On Numerical Approximation*, *2*, 249–274.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. Boca Raton: Chapman and Hall/CRC.

Spivak, M. (1967). *Calculus*. New York: Benjamin.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, *13*(2), 689-705.

Tarter, M. E., & Lock, M. D. (1993). *Model-free curve estimation*. London: Chapman and Hall.

Thongjaem, P., Ghosh, S. K., & Budsaba, K. (2013). Least squares method of estimation using Bernstein polynomials for density estimation. *Thailand Statistician*, *11*(1), 45–65.

Turnbull, B. C., & Ghosh, S. K. (2014). Unimodal density estimation using Bernstein polynomials. *Computational Statistics & Data Analysis*, *72*, 13–29.

Vitale, R. A. (1975). A Bernstein polynomial approach to density function estimation. In *Statistical inference and related topics* (87–99). USA: Elsevier.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia: Series in Applied Mathematics.

Wand, M., & Jones, M. (1995). *Kernel smoothing*. London: Chapman&Hall.

Wang, L. (2009). Single-index model-assisted estimation in survey sampling. *Journal of Nonparametric Statistics*, *21*(4), 487–504.

Wasserman, L. (2006). *All of Nonparametric Statistics*. New York: Springer.

Xue, L. (2009). Consistent variable selection in additive models. *Statistica Sinica*, *19*(3), 1281–1296.

Xue, L., & Wang, J. (2010). Distribution function estimation by constrained polynomial spline regression. *Journal of Nonparametric Statistics*, *22*(4), 443–457.

Zhou, J. (2009). Robust dimension reduction based on canonical correlation. *Journal of Multivariate Analysis*, *100*(1), 195–209.

Zhou, J., & He, X. (2008). Dimension reduction based on constrained canonical correlation and variable filtering. *The Annals of Statistics*, *36*(4), 1649–1668.