Thailand Statistician
January 2013; 11(1): 45-65
http://statassoc.or.th
Contributed paper

Least Squares Method of Estimation Using Bernstein Polynomials for Density Estimation

Piyada Thongjaem [a], Sujit K. Ghosh [b] and Kamon Budsaba* [a]

[a] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, PathumThani 12120, Thailand.

[b] Department of Statistics, NC State University, Raleigh, NC USA.

*Author for correspondence; e-mail: kamon@mathstat.sci.tu.ac.th

Received: 2 November 2012

Accepted: 4 February 2013

Abstract

A novel method is used to convert the density estimation to the well-known problem of weighted least squares subject to restrictions on parameters. In turn, the problem is solved using the efficient quadratic programming method. Numerous simulation studies are performed to fast the validity of the proposed method and it is shown that mean integrated squared errors (MISE) of density estimator is smaller than standard estimator. There are various values of MISE at different degree of Bernstein polynomials, m. From our method, the MISE at m optimal will have the lowest value compared with other m. This result proved that m optimal is suitable to achieve the best density estimation. At the m optimal, comparing with Kernel method, the Bernstein polynomials can provide better (less) MISE for all simulated types of probability function.

Keywords: Bernstein polynomials, density estimation, nonparametric method, constraints least squares method.

1. Introduction

Nonparametric density estimation is undoubtedly a useful tool of data analysis for estimating the probability density function of the underlying population in the independent and identically distributed (i.i.d.) setting. This fact is certainly reflected by the abundant literature on the subject since the 1960's. One such method, which is now very popular, is the so-called kernel method originally proposed by Parzen [1]. Kernel estimators are typically obtained as the weighted average of kernel functions centered at observed values, where the average is taken with respect to the empirical cumulative distribution function (ecdf) F_n and can be expressed as

$$\hat{f}(x) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) dF_n(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \quad (1)$$

where $X_1, X_2, \ldots, X_n$ are i.i.d random variables from same density $f(\cdot)$, *K* is the kernel distribution function and *h* is the window width also called the smoothing parameter or bandwidth by some authors. In above, the ecdf is given by $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\,(X_i \leq x)$ which is a right continuous function of *X* with jumps at the observed values, $X_1, X_2, \ldots, X_n$.

For many reasons the problem of estimation of a probability density function is a fundamental aspect of any statistical inferential procedure. Rosenblatt [2] discussed remarks on some nonparametric estimates of the density of a univariate probability distribution. An estimate of the density function with any desired regularity properties can be obtained by choosing a weight function with the same regularity properties. Parzen [1] had demonstrated and discussed how one may construct a family of estimates of function and of the mode that are consistent and asymptotically Normal. Moreover, he determined conditions under which the estimated probability density function tends uniformly to the true probability density function as the sample sizes increase. Using this fact, he was able to obtain consistent estimate of the modes. Silverman [3] presented a general review of Kernel Density Estimator (KDE) and particularly so for nonparametric probability density function estimation by the kernel method which is described in his seminal book by Silverman [3]. All aspects of KDE depend critically on the choice of the bandwidth, *h*. Some of the commonly used methods are Silverman's rule of the thumb, least squares cross validation [4] and the plug-in, *h* selector proposed by Sheater and Jones [5]. It is usually agreed that a proper choice of the parameter, *h*, known as the smoothing parameter or bandwidth, is much more important than the choice of the kernel function itself. However, KDEs are known to suffer from boundary biases when the support of the density is compact.

Alternatively, linear combinations of Bernstein polynomials can be used for nonparametric density estimation. Bernstein polynomials have a long history in the mathematics literature. Studies on such polynomials began with Bernstein [6] who presented a probabilistic proof of the Weierstrass Approximation Theorem and introduced what we call today Bernstein polynomials. The Bernstein polynomials to approximate a continuous function $f(x)$ defined on a closed interval [a,b] is given by

$$B_m(x,f) = \sum_{k=1}^{m} f\left(a + \frac{k-1}{m-1}(b-a)\right)\binom{m-1}{k-1}\left(\frac{x-a}{b-a}\right)^{k-1}\left(\frac{b-x}{b-a}\right)^{m-k} ; if\ a \le x \le b. \qquad (2)$$

For smooth estimate of a density function with a finite known support, Vitale [7] first proposed a method based on the Bernstein polynomials. The idea is based on the Weierstrass Approximation Theorem, which assures that any continuous functions on a closed interval can be uniformly approximated by Bernstein polynomials as the order of the polynomial increases to infinity. In other words,

$$\|B_m(\cdot,f) - f(\cdot)\|_\infty \equiv sup_{a\le x\le b}|B_m(x,f) - f(x)| \to 0 \qquad as\ m \to \infty. \qquad (3)$$

This approach to nonparametric density estimation that naturally leads to estimators with acceptable behavior near the boundaries relies on various interesting properties of the Bernstein polynomials approximations. Interest in Bernstein polynomials stems from the fact that they are the simplest example of a polynomial approximation which has a probabilistic interpretation. More general probabilistic approximations can be found in Altomare and Campiti [8]. Many methods have used Bernstein polynomials as prior for estimating probability density function on a closed interval (e.g. Petrone [9], Ghosal [10]). Babu et al. [11] suggested the application of Bernstein polynomials for approximating a bounded and continuous density function. Moreover, Kakizawa [12] showed that the Bernstein polynomials, which can also be expressed as a mixture of Beta densities, provides a successful tool in the Bayesian context and can be used as a nonparametric prior for continuous densities. The comparison with the ordinary kernel method based on a Monte Carlo simulation has been illustrated and examined for finite sample performances. So, there were many approaches in Bernstein polynomial to get better density estimation comparing to Kernel method but the algorithms are more complicated than basic Kernel estimation. In this paper, a novel method of estimation of the true density by using Bernstein polynomials is proposed and converted the density estimation to the well-known problem of weighted least squares subject to restrictions on parameters. This new method is one of the efficient algorithms to apply in the Bernstein polynomials for density estimation.

2. Density estimation using Bernstein Polynomial

There are many ways to construct an estimator and to make inferences about the population. Especially, we often use data to make inferences about a parameter by applying the statistical functional of the ecdf. The ecdf is very useful for developing and studying a large number of estimators. 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 $F(\cdot)$ is known to be continuous. To begin with the simplified case, we consider random variables that are known to take values in the interval [0,1]. In other words, we initially assume that $F(0) = 0$ and $F(1) = 1$, later we are going to relax this assumption.

Let $X_1, X_2, \ldots, X_n$ be an i.i.d sequence of random variables generated from a cdf F where $F(\cdot)$ is continuous. We want to obtain an estimator $\widetilde{F}_n(x)$ that is continuous. The ecdf is the function of x defined by $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$. Statistical properties of the empirical distribution function are following:

i) $nF_n(x) \sim Bin(n, F(x))$ i.e, $nF(x)$ has the binomial distribution

ii) $\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} N(0, F(x)(1 - F(x)))$ as $n \to \infty$

iii) $\sup_x |(F_n(x) - F(x)| \xrightarrow{a.s.} 0$ as $n \to \infty$s

These properties demonstrate that the ecdf is a very good estimator of the true cdf. Despite these entire nice finite samples as well as asymptotic properties, $F_n(x)$ itself is not a continuous function even when $F(x)$ is continuous. Moreover, when it is known that $F(x)$ is the cdf of an unimodal density or other known properties, $F_n(x)$ will not generally satisfy these desired known properties of $F(x)$.

2.1 Density estimation using Bernstein Polynomial with support [0, 1]

Our proposed model needs to be constructed from cdf so the Bernstein polynomials are proved for the cdf properties. When such function is cdf, derivative of Bernstein polynomials can be obtained. After that, the result from differentiation using Bernstein Polynomials will be proved for probability density function (pdf) properties.

Suppose $f: [0,1] \to \mathbb{R}$ is a continuous function. Consider the Bernstein polynomial (of order m-1) to approximate $f(\cdot)$:

$$B_m(x, f) = \sum_{k=1}^{m} f\left(\frac{k-1}{m-1}\right)\binom{m-1}{k-1} x^{k-1}(1-x)^{m-k} \quad (4)$$

It is well known that, as $m \to \infty$.

$$\| B_m(\cdot, f) - f(\cdot) \|_\infty = \max_{0 \le x \le 1} |B_m(x, f) - f(x)| \to 0 \quad (5)$$

More interestingly, any qualitative known shape of f is preserved by that of B_m. The goal is to obtain a smooth estimate of $f(\cdot)$. The step of density estimation by using Bernstein polynomials are in the following.

The Bernstein Polynomial $F_m(x, w)$ can be shown to be the cdf on of all w_k if it satisfies the following properties: i) $F_m(x, w)$ is monotone non-decreasing ii) $\lim_{x \to -\infty} F_m(x, w) = 0$ and $\lim_{x \to \infty} F_m(x, w) = 1$ and iii) $F_m(x, w)$ is right continuous.

Let $f: [0,1] \to \mathbb{R}$ be a continuous function. The Bernstein polynomial of the degree m of $f(\cdot)$ on the interval $[0,1]$ is defined by Lorentz [13]. In cdf derivation, it is easier to start with the Bernstein polynomials of degree m than m-1.

Since $F_m(x, w) = \sum_{k=0}^{m} w_k \binom{m}{k} x^k (1-x)^{m-k}$; $0 \le x \le 1$.

Let $U_i = F_n(x_i)$ and $b_k(x, w) = \binom{m}{k} x^k (1-x)^{m-k}$; $k = 0,1,2,\ldots, m$, then

$$F_m(x, w) = \sum_{k=0}^{m} w_k b_k(x, w) = w_0(1-x)^m + w_1 \binom{m}{1} (1-x)^{m-1} + \ldots + w_m x^m,$$

we get, $\lim_{x \to -\infty} F_m(x, w) = 0$ implied that $F_m(x, w) \to 0$ as $n \to -\infty$ and if $x = 0$ then $F_m(x, w) = w_0 = 0$

$\lim_{x \to \infty} F_m(x, w) = 1$ implied that $F_m(x, w) \to 1$ as $n \to -\infty$ and if $x = 1$ then $F_m(x, w) = w_m = 1$.

If $w_k = F\left(\frac{k}{m}\right)$, by derivative of Bernstein polynomials, we get

$$F'_m(x, w) = \sum_{k=0}^{m} F\left(\frac{k}{m}\right) \binom{m}{k} [k x^{k-1} - (1-x)^{m-k} x^k (1-x)^{m-k-1}]$$

$$= m \sum_{k=0}^{m-1} (w_{k+1} - w_k) \binom{m-1}{k} x^k (1-x)^{m-k-1} \ge 0 \quad (6)$$

If $w_{k+1} - w_k \ge 0$ then $F'_m(x, w) \ge 0$;$k = 0,1,\ldots, m-1$ and $w_1 - w_0 \ge 0$ implied that $w_1 \ge 0$, $w_2 - w_1 \ge 0$ implied that $w_2 \ge w_1$, $w_{m-1} - w_{m-2} \ge 0$ implied that $w_{m-1} \ge w_{m-2}$, $w_m - w_{m-1} \ge 0$ implied that $1 \ge w_{m-1}$. Hence, we can conclude that $F_m(x, w)$ is the cdf of all w_k.

Consider mixture of Beta densities:

$$f_m(x,w)=\sum_{k=1}^{m} w_k \frac{x^{k-1}(1-x)^{m-k}}{B(k,m-k+1)} = m \sum_{k=1}^{m} w_k \binom{m-1}{k-1} x^{k-1}(1-x)^{m-k} \tag{7}$$

Clearly, the Bernstein Polynomial $f_m(x,w)$ can be shown to be a density on $[0,1]$ if $w_k \geq 0\ \forall k$ and $\sum_{k=1}^{m} w_k = 1$.

Since $f_m(x,w)=\sum_{k=1}^{m} w_k \frac{x^{k-1}(1-x)^{m-k}}{B(k,m-k+1)} = m\sum_{k=1}^{m} w_k \binom{m-1}{k-1} x^{k-1}(1-x)^{m-k}$

Let $b_k = \frac{x^{k-1}(1-x)^{m-k}}{B(k,m-k+1)}$ we get $f_m(x,w)=\sum_{k=1}^{m} w_k b_k$ and we know that $\binom{m-1}{k-1} x^{k-1}(1-x)^{m-k} \geq 0$.

Thus, $f_m(x,w)=\sum_{k=1}^{m} w_k \frac{x^{k-1}(1-x)^{m-k}}{B(k,m-k+1)} \geq 0$ if $w_k \geq 0$.

Next, $\int_0^1 f_m(x,w)\mathrm{d}x = \int_0^1 \sum_{k=1}^{m} w_k \frac{x^{k-1}(1-x)^{m-k}}{B(k,m-k+1)}\mathrm{d}x = \sum_{k=1}^{m} w_k \int_0^1 \frac{x^{k-1}(1-x)^{m-k}}{B(k,m-k+1)}\mathrm{d}x = 1$; if $\sum_{k=1}^{m} w_k = 1$.

Hence, $\int_0^1 f_m(x,w)\mathrm{d}x = 1$ if $w_k \geq$ for all k and $\sum_{k=1}^{m} w_k = 1$.

Thus, we can conclude that $f_m(x,w)$ is a probability density on $[0,1]$ if $w_k \geq 0\ \ \forall k$ and $\sum_{k=1}^{m} w_k = 1$.

2.2 Density estimation using Bernstein Polynomial with support [a,b]

In general, the estimators may not be observed to lie in the domain $[0,1]$. To satisfy domain restriction, we use the linear transformation defined as follows for all our empirical applications. A polynomial in Bernstein form can be defined over arbitrary domain intervals by introducing the change of variables

$$u = \frac{x-a_m}{b_m-a_m} \quad ; where\ a_m = x_{(1)} - \delta,\ b_m = x_{(n)} + \delta\ and\ \delta = sd(x) > 0 \tag{8}$$

where $x_{(1)} = min_{1\leq i\leq m} x_i,\ x_{(n)} = max_{1\leq i\leq m} x_i$ denote the minimum and maximum order statistics, and $sd(x)$ represents the sample standard deviation of $x_i;\ i = 1,2,\ldots,m$. This variable transformation from x to u when

$0 \leq u = \frac{x_i - x_{(1)} + \delta}{x_{(n)} - x_{(1)} + 2\delta} \leq 1$ map s$x \in [a,b]$ to $u \in [0,1]$ where a could approach $-\infty$ and b could approach ∞.

Then, we have

$$f_m(x,w) = \sum_{k=1}^{m} w_k \binom{m-1}{k-1} \frac{\left(\frac{x-a_m}{b_m-a_m}\right)^{k-1}\left(\frac{b_m-x}{b_m-a_m}\right)^{m-k}\frac{1}{b_m-a_m}}{B_{k,m-k+1}} \quad (9)$$

Similar to the previous approach, we use the Constraints Least Squares (CLS) method to estimate w_k. Consider the cdf based on $f_m(x,w)$,

$$F_m(u,w) = \sum_{k=1}^{m} w_k F_b((x-a)/(b-a), k, m-k+1), \quad (10)$$

where $F_b(u,k,m-k+1) = \int_0^u f_b(u,k,m-k+1)du$ and the cdf of $Beta(k,m-k+1)$ distribution. We obtain w_k by minimizing a scaled squared distance between F_m and F_n using $\sqrt{n}\left(F_n(x) - F(x)\right) \xrightarrow{d} N\left(0, F(x)\left(1-F(x)\right)\right)$.

We solve the following constrained weighted least squares problem:

$$Minimize \sum_{i=1}^{n} \frac{\sqrt{n}\left(F_n(x_i) - F_m(x_i,w)\right)^2}{\left(F_n(x_i)\right)\left(1-F_n(x_i)\right)} subject\ to\ w_k \geq 0 \quad and \sum_{k=1}^{m} w_k = 1 \quad for\ k\ 1,2,\ldots,m.$$

Next, in general case we change to simply notation for minimize:

$$Minimize : \sum_{i=1}^{n} \frac{\sqrt{n}(F_n(x_i)-\tilde{F}_n(x_i,\beta))^2}{(F_n(x_i)+\varepsilon)(1+\varepsilon-F_n(x_i))} = Minimize : \sum_{i=1}^{n} \frac{\sqrt{n}(F_n(x_i)-w_i{'}\,\beta)^2}{(F_n(x_i)+\varepsilon)(1+\varepsilon-F_n(x_i))} \quad (11)$$

So, the estimator w_k can be shown to be the solution of the following optimization problem: $Minimize : \sum_{i=1}^{n} \frac{(y_i-w_i{'}\,\beta)^2}{\sigma_i}$ subject to $R\beta \geq c$. As $F_m(x,w)$ is linear in w , the above optimization problem can easily be solved by using a quadratic programming algorithm. The estimator w_k can be calculated with the least squares method shown in the next section. A quadratic programming algorithm (e.g., quadprog in R) is proposed as a tool to solve this optimization problem.

3. The least squares method for optimization of the estimators

In the previous section, we use the Constraints Least Squares (CLS) method to estimate w_k. We obtain w_k by minimizing a scaled squared distance between F_m and F_n as follows:

Step1 : Computation the estimator w_k when the target function is cdf. Consider the Bernstein polynomials $F_m(x,\ w) = \sum_{k=0}^{m} w_k \binom{m}{k} x^k (1-x)^{m-k}$;$0 \le x \le 1$ which resembles the general linear model. We use the least squares method to estimate w_k. The estimator w_k can be shown to be the solution of the following optimization problem:

$$Minimize:\ \sum_{i=0}^{m} (F_n(x_i) - F_m(x,\ w))^2 \quad subject\ to: 0 \le w_1 \le w_2 \le \ldots \le w_{m-1} \le 1.$$

Let $U_i = F_n(x_i)$, $w = (w_1, w_2, \ldots, w_{m-1})$ and $s_{ik} = \binom{m}{k} x^k (1-x)^{m-k}$; where $k = 1,2,\ldots,m$ and $i = 1,2,\ldots,n$.

Then,

$$F_m(x, w) = \sum_{k=1}^{m} w_k s_{ik} \qquad (12)$$

$$F_n(x_i) - F_m(x,\ w) = (U_i - x_i^m) - (w_1 s_{i1} + w_2 s_{i2} + \ldots + w_{m-1} s_{im-1})$$

$$= (v_i - s'_i w) \ ; \quad where\ v_i = (U_i - x_i^m), s_i = (s_{i1}, \ldots, s_{im-1})$$

$$\sum_{i=0}^{m} (F_n(x_i) - F_m(x,\ w))^2 = (v - sw)(v - sw) \qquad (13)$$

where $v = (v_1, \ldots, v_n)$ and $s = \begin{bmatrix} s_{11} & \cdots & s_{1m-1} \\ s_{21} & \cdots & s_{2m-1} \\ \vdots & & \vdots \\ s_{n1} & \cdots & s_{nm-1} \end{bmatrix}$

Thus, we define the optimization problem on the coefficients as follows:

$Minimize:\ (v - sw)(v - sw)$

$subject\ to: Rw \ge [0 \quad 0 \quad \cdots \quad 0 \quad -1]$

where $R = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \\ 0 & 0 & 0 & \ldots & 0 & -1 \end{bmatrix}$ and $w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_{m-2} \\ w_{m-1} \end{bmatrix}$

The above optimization problem can be effectively solved by the general quadratic programming [14]. Quadratic programming has been used to impose the necessary density estimation by using Bernstein polynomials method. In this study, we use the available R package quadproc developed by Turlach and Weingessel [15] to solve quadratic programming problem.

Step 2: Computation of the estimator w_k when the target function is pdf. From the result in step1, we have two conditions; $w_k \geq 0\ \forall k$ and $\sum_{k=1}^{m} w_k = 1$. Therefore, we use CLS method to estimate w_k. Starting with the cdf, $F(x)$ by considering

$$F_m(x_i, w) = \sum_{k=1}^{m} w_k \int_0^x \frac{t^{k-1}(1-t)^{m-k}}{B(k, m-k+1)} \mathrm{d}t = \sum_{k=1}^{m} w_k F_B(x, k, m-k+1) \quad (14)$$

where $F_B(x, k, m-k+1)$ is the cdf of a $Beta(k, m-k+1)$ distribution. It is more well known to substitute w_k with $F_n\left(\frac{k}{m-1}\right) - F_n\left(\frac{k-1}{m-1}\right); k = 1,2,3, \ldots, m$. So, we get

$$F_m(x_i, w) = \sum_{k=1}^{m} w_k F_n(\frac{k}{m-1}) - F_n(\frac{k-1}{m-1}) F_B(x, k, m-k+1) \quad (15)$$

It is a consistent estimator of F_n i.e. $\sup_x |(F_m(x_i, w) - F(x)| \overset{a.s.}{\rightarrow} 0$ as $m \to \infty, n \to \infty$ and $\frac{m}{n} \to \infty$.

We obtain w_k by minimizing a scaled squared distance between F_m and F_n using $\sqrt{n}(F_n(x) - F(x)) \overset{d}{\rightarrow} N(0, F(x)(1 - F(x)))$. We solve the following constrained weighted least squares problem:

$$Minimize \sum_{i=1}^{n} \frac{\sqrt{n}(F_n(x_i) - F_m(x_i, w))^2}{(F_n(x_i))(1 - F_n(x_i))} \quad (16)$$

$$subject\ to\ w_k \geq 0 \quad for\ k = 1,2, \ldots, m\ and \sum_{k=1}^{m} w_k = 1.$$

Next, in general case we change to simply notation for minimize:

$$Minimize \sum_{i=1}^{n} \frac{\sqrt{n}\left(F_n(x_i) - \tilde{F}_n(x_i, \beta)\right)^2}{(F_n(x_i) + \varepsilon)(1 + \varepsilon - F_n(x_i))} = Minimize \sum_{i=1}^{n} \frac{(y_i - w_i{'}\ \ \beta)^2}{\sigma_i}$$

So, The estimator w_k can be shown to be the solution of the following optimization problem:

$$Minimize \sum_{i=1}^{n} \frac{(y_i - w_i{}'\ \beta)^2}{\sigma_i} \ subject\ to\ R\beta \ \geq \ c \tag{17}$$

As $F_m(x, w)$ is linear in w, the above optimization problem can easily be solved by using a quadratic programming algorithm (e.g., quadprog in R).

Step 3: Estimation of $f_m(x, w)$, in this section we develop a novel method to estimating the weights that satisfy the desired constraint and then propose a method to select m. We use the following class of estimators:

$$f_m(x, w) = \sum_{k=1}^{m} w_k f_b((x-a)/(b-a), k, m-k+1)/(b-a) \tag{18}$$

where w_k are unknown weights satisfying $w_k \geq 0$ and $\sum_{k=1}^{m} w_k = 1$, $f_b(u, k, m-k+1)$ denotes the density of $Beta(k, m-k+1)$ distribution. Thus, we would like to estimate w_k in $f_m(x, w)$ satisfying the following constraint: (i) $w_k \geq 0$, (ii) $\sum_{k=1}^{m} w_k = 1$. Finally, the smoothness of $f_m(\cdot, w)$ would be controlled by suitably selecting m. Consider the smooth cdf based on

$$F_m(x, w) = \sum_{k=1}^{m} w_k F_b((x-a)/(b-a), k, m-k+1) \tag{19}$$

where $F_b(u, k, m-k+1) = \int_0^u f_b(v, k, m-k+1)dv$ is the cdf of $Beta(k, m-k+1)$ distribution.

Notice that $F_b(\cdot)$ can be computed numerically (p Beta in R). We obtain w_k by minimizing a scaled squared distance between F_m and F_n using the fact that $\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} N(0, F(x)(1-F(x)))$. We solve the following constrained weighted least squares problem:

$$Minimize \sum_{i=1}^{n} \frac{n(F_n(x_i) - F_m(x_i, w))^2}{(F_n(x_i))(1 - F_n(x_i))} with\ respect\ to\ w \tag{20}$$

$$subject\ to\ w_k \geq 0, \qquad \sum_{k=1}^{m} w_k = 1 \ \ ; for\ k = 1,2, \ldots, m$$

where $c_k = \int_0^1 g(a + (b - a)u) f_b(u, k, m - k + 1) du$. As $F_m(x, w)$ is linear in w the above optimization problem can easily be solved by using a quadratic programming algorithm. One may replace the denominator $(F_n(x_i))(1 - F_n(x_i))$ by $(F_n(x_i) + \varepsilon_n)(1 + \varepsilon_n - F_n(x_i))$ where ε_n = 3/8n has been suggested by Anscombe and Aumann [16]. Optimal asymptotic rates for choosing m = m(n) has been derived by Vitale [7] and Tenbusch [17]. However, in practice such asymptotic rates are not very useful as the rates depend on the unknown density f. In simulation studies optimal m is obtain by minimizing the Mean Integrated Squared error (MISE).

4. Simulation studies

In this section, we present several scenarios using simulated data to validate the performance of our method to explore how well the estimated Bernstein polynomials density approximates the underlying true density. First, suppose we have n observations from the known density. Then we can estimate the estimator w_k by constrainted least squares method described in section 3 and construct the estimated density. We consider simulated samples from three kinds of distribution families with sample sizes n=50, 100, 150 and 200. The software used for computation is R. The results compare the smoothing estimation, i.e. traditional kernel estimation and Bernstein Polynomials estimation to the true density. For Monte Carlo simulation, we obtain simulated data from three kinds of distribution families to cover all data in real line. For density with support, within $\mathbb{R} = [-\infty, \infty]$, we use the standard Normal distribution to illustrate our method for this support. For density with finite support within [a,b], we use the Beta density to illustrate method. For density with semi-finite support within $(-\infty, b]$ or $[a, \infty)$, we use the Gamma density to illustrate method. Also, the algorithm for finding m optimal has several steps as shown in the following; First, generation of data set, in order to represent all standard density function, the Normal, Gamma and Beta distribution functions have been used as the input data set in the simulation. Second, solving the following constrained weighted least squares problem, first considering w_k for moment of constraints as the algorithm to calculate w_k. Third, finding the estimate density, $f_m(x, w)$, after we get the w_k, it is used in Bernstein polynomial to estimate the density function, $f_m(x, w)$, Then, the actual density function generated from the simulation is compared to the estimated density from our method. The bias, standard deviation and MISE are the parameters in the simulation that we can get the average from a thousand of Monte Carlo simulation. Fourth, finding optimal density function, the w_k from the algorithm has been determined by the number of parameters in the $F_b(u_i, k, m - k + 1)$ function, m. In our simulation, we used m that was varied from five to fifteen and we can select the optimal m from the lowest MISE situations. Then, we use the case that has the optimal m to estimate density function of sample. Finally, comparing with various constrains, because the w_k has constrains in function, we have tried our algorithm in various constrains and re-simulate the results. We use simulated data sets to explore how well the estimated Bernstein polynomials approximates the underlying true density with the criteria to select m as follows:

1. In cdf, the Root Mean Integrated Squared Error is defined as

$$RMISE = \sqrt{\int \left[Var\left(\hat{f}(x)\right) + Bias^2(f(x))\right]dx},$$

2. In pdf, the Mean Integrated Squared Error is defined as

$$MISE\left[\hat{f}\right] = \int \left[Var\left(\hat{f}(x)\right) + Bias^2(f(x))\right]dx$$

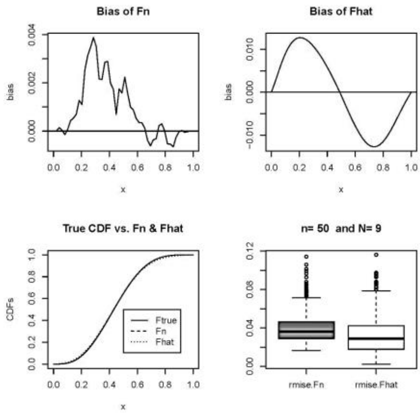
where $\hat{f}(x)$ and $f(x)$ are estimated density and true density, respectively.

5. Results of the study

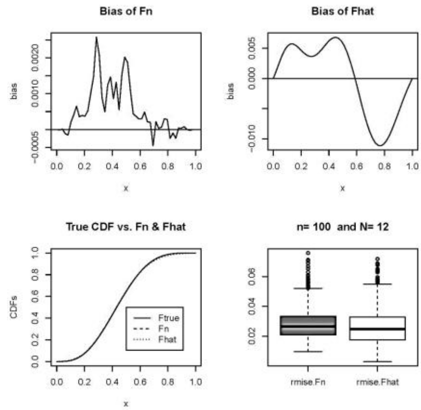
The objective of this research is to develop novel density estimation by using Bernstein polynomials. For numerical illustration, we start to simulate of the Bernstein polynomials estimator for distribution function from three different distributions, then to explore the performance of the proposed Bernstein polynomials method by using simulated data and to compare the performance of the proposed method to the kernel method. The results have two sections: Bernstein polynomials estimator for distribution function and Bernstein polynomials estimator for density function.

5.1 Results for Bernstein polynomials estimator for distribution function

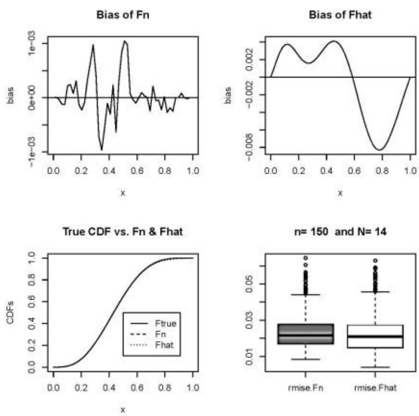
In Figures 1-3, we display the plot of true and Bernstein polynomials estimator of distribution function, the behaviour of the bias-corrected estimate, Root Mean Integrated Squared Error (RMISE) for three distributions, Normal(2,1), Beta(3.5,4.5) and Gamma(3,1). These figures give the results on 1,000 simulated samples for the sample sizes $n =$ 50, 100, 150 and 200. The graphes between True and estimated distribution function for each sample size show that the Bernstein polynomials estimator for the distribution function presents impressively smooth curve in the approximation of true distribution. Also, as presented from Figure 1 to Figure 3 when number of sample size increases, RMISE of the estimation function reduces. It means that the estimator is closer to the actual distribution function. In details it can be shown that the results of this simulation study as follows:



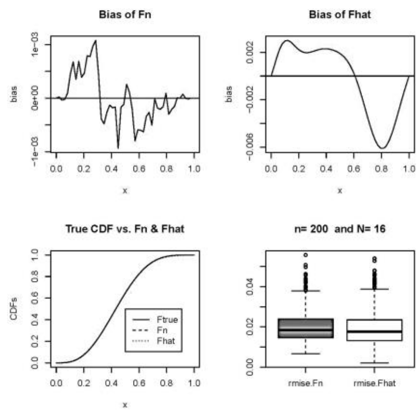
(a) n=50



(b) n=100

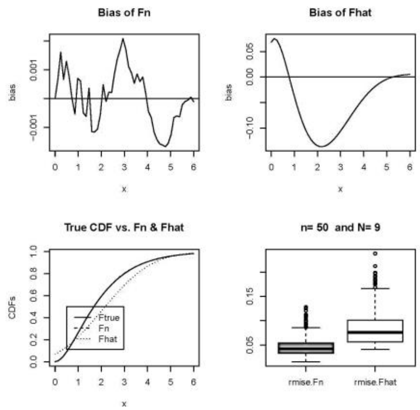


(c) n=150

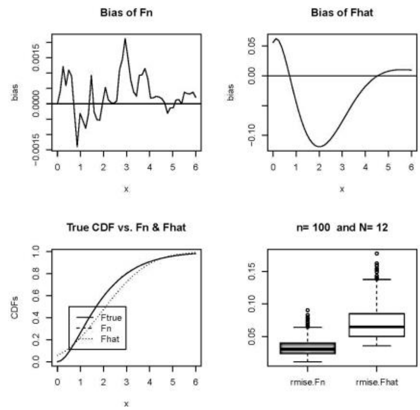


(d) n=200

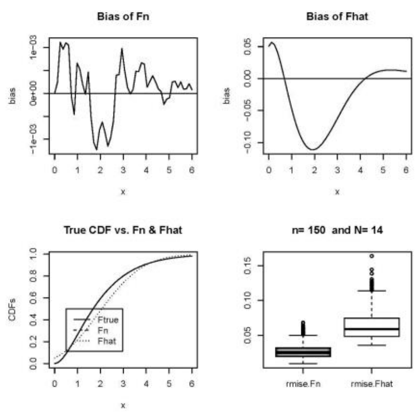
Figure 1: Bernstein polynomials estimator for distribution function from Normal distribution. The true cdf (Ftrue in solidcurve), Kernel density estimator (Fn in dash curved) and the Bernstein polynomials estimator (Fhat in dotted curve). The bias, standard deviation and RMISE for the cdf estimation method are based on 1,000 Monte Carlo repetitions.



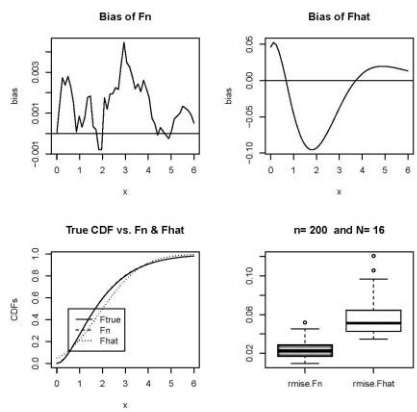
(a) n=50



(b) n=100

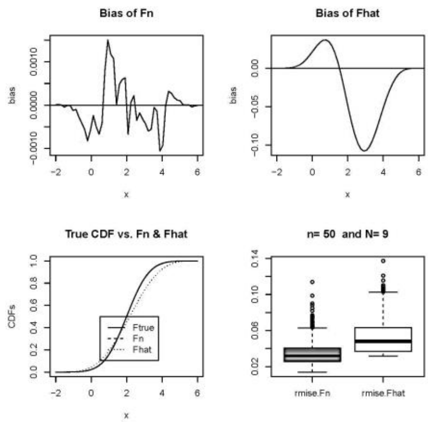


(c) n=150

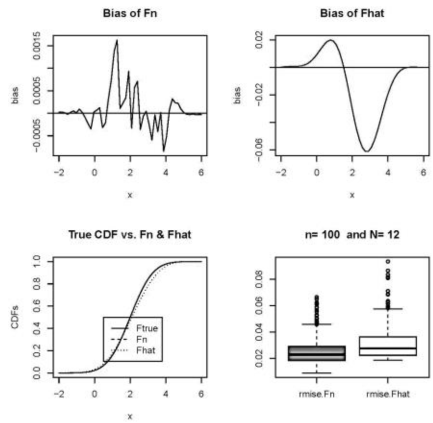


(d) n=200

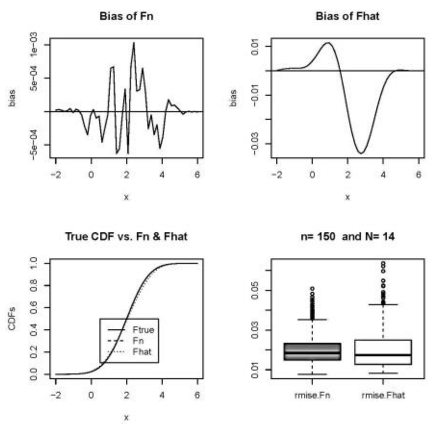
Figure 2: Bernstein polynomials estimator for distribution function from Beta distribution. The true cdf (Ftrue in solid curve), Kernel density estimator (Fn in dash curved) and the Bernstein polynomials estimator (Fhat in dotted curve). The bias, standard deviation and RMISE for the cdf estimation method are based on 1,000 Monte Carlo repetitions.



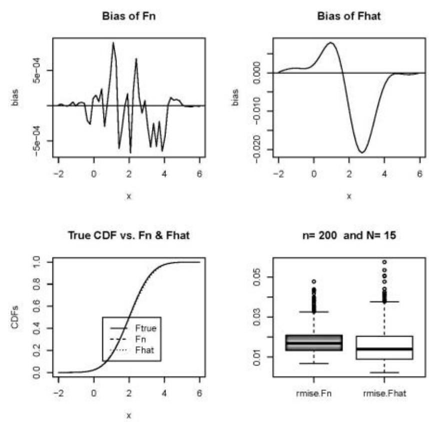
(a) n=50



(b) n=100



(c) n=150

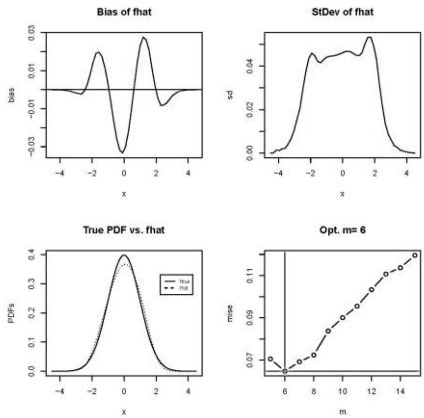


(d) n=200

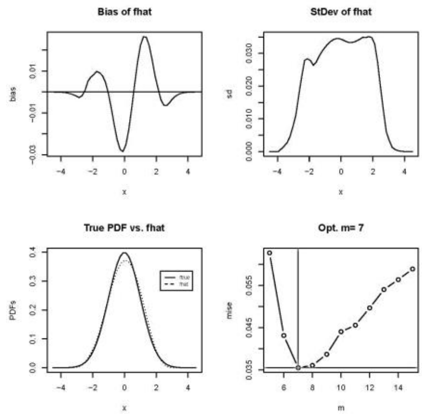
Figure 3: Bernstein polynomials estimator for distribution function from Gamma distribution. The true cdf (Ftrue in solid curve), Kernel density estimator (Fn in dash curved) and the Bernstein polynomials estimator (Fhat in dotted curve). The bias, standard deviation and RMISE for the cdf estimation method are based on 1,000 Monte Carlo repetitions.

5.2 Results for Bernstein polynomials estimator for density function

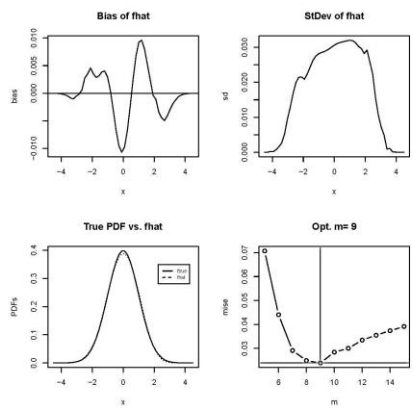
In Figures 4-6, we display the plot of true and Bernstein polynomials estimator of density function, the behaviour of the bias-corrected estimate, MISE for three distributions, Normal(0,1), Beta(3.5,4.5) and Gamma(3,1). These figures give the results on 1,000 simulated samples for the sample sizes $n = 50$, 100, 150 and 200. The graphs show that the Bernstein polynomials estimator for the distribution function presents impressively smooth curve in the approximation of true distribution.



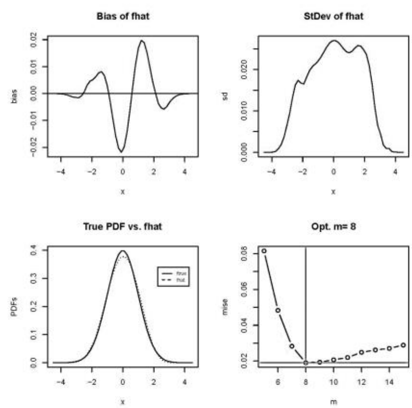
(a) n=50,m=6



(b) n=100,m=7

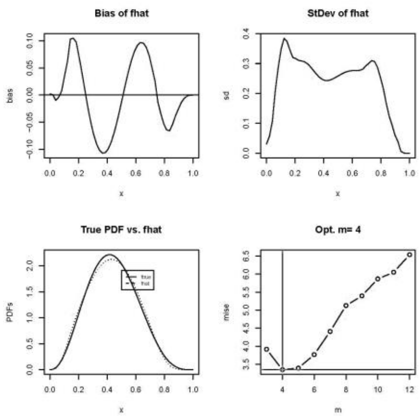


(c) n=150,m=9

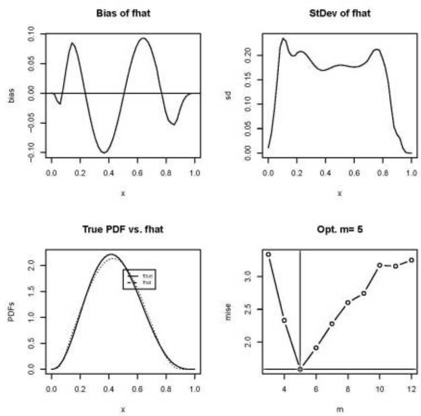


(d) n=200,m=8

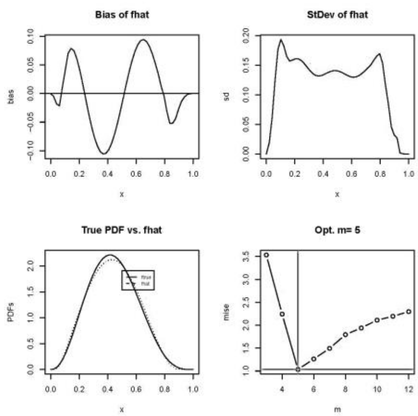
Figure 4: True and estimated density for different sample from Normal density. The true pdf (Ftrue in solid curve) and the Bernstein polynomials estimator (Fhat in dash curve). The bias, standard deviation and optimal m for the density estimation method are based on 1,000 Monte Carlo repetitions**.**



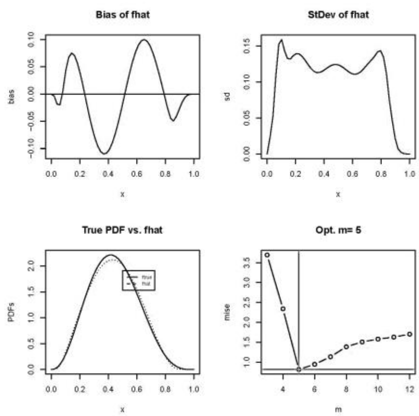
(a) n=50,m=5



(b) n=100,m=5

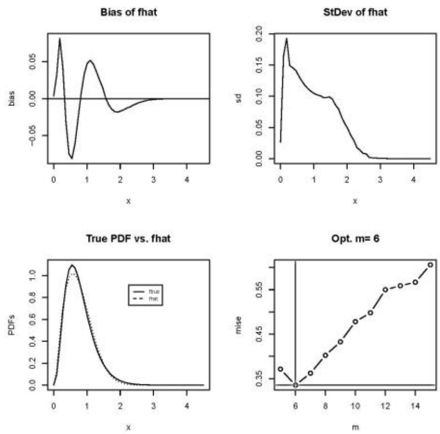


(c) n=150,m=5

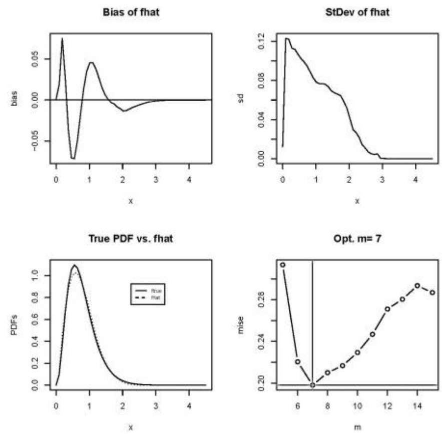


(d) n=200,m=5

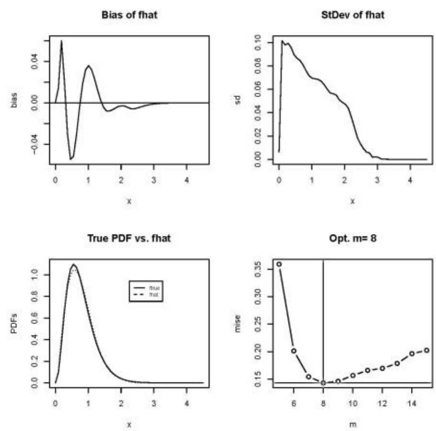
Figure 5: True and estimated density for different sample from Beta density. The true pdf (Ftrue in solid curve) and the Bernstein polynomials estimator (Fhat in dash curve). The bias, standard deviation and optimal m for the density estimation method are based on 1,000 Monte Carlo repetitions**.**



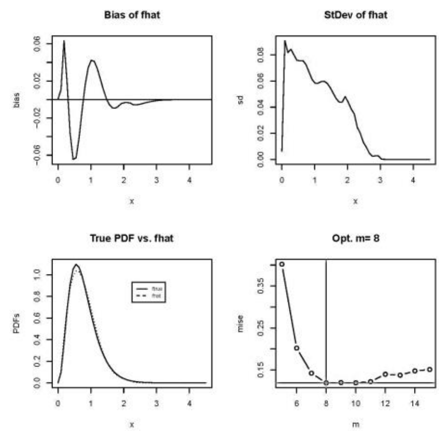
(a) n=50,m=6



(b) n=100,m=7



(c) n=150,m=8



(d) n=200,m=8

Figure 6: True and estimated density for different sample from Gamma density. The true pdf (Ftrue in solid curve) and the Bernstein polynomials estimator (Fhat in dash curve). The bias, standard deviation and optimal m for the density estimation method are based on 1,000 Monte Carlo repetitions**.**

The more detail information is presented in Table 1. The estimation used m that was varied from five to fifteen and we can select the optimal m from the lowest MISE situations. Then we use the optimal m to estimate density function of sample. At the m optimal, comparing with Kernel method, the Bernstein polynomials can provide better (less) MISE for all simulated types of probability function. Also, when number of sample size increases, MISE of the estimation function reduces. It means that the estimator is closer to the actual distribution function. Therefore, from the above results it proves that, the proposed method to choose m optimal explained in Section 4 is suitable to achieve the good approximation of probability density function.

Table 1: MISE of Bernstein polynomials method for different sample from Normal, Beta and Gamma densities. (1,000 replications).

		MISE				**Optimal**	**MISE (optimal)**	
	n	**Min**	**Median**	**Mean**	**Max**	**m**	**BP**	**Kernel**
Normal density	50	0.0647	0.0900	0.0903	0.1195	6	0.0650	0.0665
	100	0.0355	0.0456	0.0477	0.0627	7	0.0350	0.0369
	150	0.0239	0.0335	0.0361	0.0707	9	0.0244	0.0258
	200	0.0190	0.0262	0.0315	0.0815	8	0.0199	0.0207
Beta density	50	3.3500	4.7690	4.7820	6.5380	5	3.2078	3.2397
	100	1.5830	2.6720	2.6370	3.3380	5	1.4845	1.8619
	150	1.0330	2.0250	1.9900	3.5330	5	0.9912	1.3661
	200	0.8163	1.5440	1.6730	3.6910	5	0.7864	1.0851
Gamma density	50	0.3351	0.4781	0.4692	0.6059	6	0.0428	0.0509
	100	0.1980	0.2468	0.2515	0.3134	7	0.0235	0.0298
	150	0.1431	0.1696	0.1885	0.3590	8	0.0172	0.0213
	200	0.1191	0.1399	0.1639	0.4020	8	0.0145	0.0172

6. Conclusions and Discussion

The weighted parameter can be determined by the constraint least squares method for each number of order of the Bernstein polynomials, m. Because the MISE varies whenever the m is changed, the optimal m is chosen from the case that has the lowest MISE. At the optimal m, the optimal weight parameter is obtained and the optimal density is finally estimated. The method is validated by many simulated data science. The input data are considered in three different distributions i.e. Normal, Beta and Gamma and the simulation was run 1,000 times for each data set. Three scenarios with constraints and without constraint are simulated while the number of samples is varied.

There are various value of MISE at different degree of Bernstein polynomials m. From our method, the MISE at m optimal will have the lowest value compared with other m. This result proved that m optimal is suitable to achieve the best density estimation. At the m optimal, comparing with Kernel method, the Bernstein polynomials can provide better (less) MISE for all simulated types of probability function. So, by converting the Bernstein polynomial to the weighted least squares method, this new approach can provide better density estimation.

Acknowledgements

We would like to express my deep appreciation and gratitude to Statistics department at North Carolina State University, Thammasat University, Surat Thani Rajabhat University and office of the Higher Education Commission Scholarship for partial support of this research.

References

[1] Parzen, E., On estimation of a probability density function and mode, *The Annals of Mathematical Statistics*., 1962; 33(3): 1065-1076.

[2] Rosenblatt, M., Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics*., 1956; 27: 832-837.

[3] Silverman, B.W., *Density estimation for statistics and data analysis*, London: Chapman and Hall, 1986.

[4] Park, B.U., and Marron, J.S., Comparison of data-driven bandwidth selectors, *Journal of the American Statistical Association*., 1990; 85(409): 66-72.

[5] Sheater, S.J., and Jones, M.C., A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society*., 1991; 53(3): 683-690.

[6] Bernstein, S.N., D´emonstration du th´eoreme de weierstrass fond´ee sur le calcul desprobabilit´es, C*ommunications of the Kharkov Mathematical Society*., 1912; 13: 1-2.

[7] Vitale, R.A., A Bernstein polynomial approach to density function estimation, *Statistical Inference and Related Topics*., 1975; 2: 87-99.

[8] Altomare, F. and Campiti, M., *Korovkin-type Approximation Theory and its Application*, Berlin, Walter de Gruyter, 1994.

[9] Petrone, S., Random Bernstein polynomials,*Scandinavian Journal of Statistics*., 1999; 26: 373-393.

[10] Ghosal, S., Convergence rates for density estimation with Bernstein polynomials,*The Annals of Statistics*., 2001; 29: 1264-1280.

[11] Babu, G.J., Canty, A.J. and Chaubey, Y.P., Application of Bernstein polynomials for smooth estimation of a distribution and density function, *Journal of Statistical Planning and Inference*., 2002; 105: 377-392.

[12] Kakizawa, Y., Bernstein polynomial probability density estimation, *Journal of Nonparametric Statistics*., 2004; 16: 709-729.

[13] Lorentz, C.G., *Bernstein Polynomials*. Chelsea: New York, 1986.

[14] Goldfarb, D. and Idnani, A., A numerically stable dual method for solving strictly convex quadratic programs, *Mathematical Programming*, 1983; 27: 1–33.

[15] Turlach, B.A. and Weingessel, A., quadprog: Functions to solve Quadratic Programming Problems., R package version 1.5-3. S original by Berwin A. Turlach, R port by Andreas Weingessel, Available from URL: http://CRAN.R-project.org/package=quadprog[accessed Aug 1, 2010].

[16] Anscombe, F.J., and Aumann, R.J., A definition of subjective probability, *Annals of Mathematical Statistics*., 1963; 34: 199-205.

[17] Tenbusch, A., Two-dimensional Bernstein polynomial density estimators, *Metrika*., 1994; 41: 233-253.