Australian Journal of Basic and Applied Sciences, 9(13) Special 2015, Pages: 43-49

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Density Estimation with Constraints using Bernstein Polynomials

[1]Piyada Thongjaem and [2]Sujit K. Ghosh

[1]Department ofMathematicsand Statistics, Thammasat University, Pathumthani 12121, Thailand
[2]Department of Statistics, NC State University, Raleigh, NC 27695-8203, USA

ARTICLE INFO

Article history:
Received 22 February 2015
Accepted 20 March 2015
Available online 23 April 2015

Keywords:
Bernstein Polynomials,
Constrained density estimation,
Nonparametric method

ABSTRACT

The statistical inference with known constraints provides better estimate of the true density. In many applied problems, often the statistician has subject matter knowledge about the moments of the distribution. In this paper, density estimation with constraints using the Bernstein Polynomial has been proposed. A novel method is used to convert the density estimation with moment constraint 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 true and estimated densityof moment constraint of density estimator. The paper is concluded with application to real data sets.

To Cite This Article: Piyada Thongjaem and Sujit K. Ghosh, Density Estimation with Constraints using Bernstein Polynomials. ***Aust. J. Basic & Appl. Sci.,*** *9(13): 43-49, 2015*

INTRODUCTION

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 (1912) 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

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

For smooth estimate of a density function with a finite known support, Vitale (1975) first proposed a method based on the Bernstein polynomials. The idea is based on the Weierstrass Approximation Theorem which assures that any continuous function on a closed interval can be uniformly approximated by Bernstein polynomials as the order of the polynomial increases to infinity. In other words,

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

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. Many methods have used Bernstein polynomial as prior for estimating probability density function on a closed interval (e.g. Petrone, 1999; Ghosal, 2001). Babu *et al*. (2002) suggested the application of Bernstein polynomials for approximating a bounded and continuous density function. Moreover, Kakizawa (2004) showed that the Bernstein polynomial, 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. In this thesis, a method of estimation of the true density by using Bernstein polynomials is proposed when the density is known to satisfy a set of known moment constraints. It is essential to consider moment constraints in the density estimation and until now there is no research considering the moment constraints in the Bernstein polynomials method. Therefore, this research attempts to improve the method of Bernstein

Corresponding Author: Piyada Thongjaem. Thammasat University, Department of Mathematics and Statistics, Thammasat University, Pathumthani 12121, Thailand
Tel: (+66 816668516) E-mail: poystat@gmail.com

Australian Journal of Basic and Applied Sciences, 9(13) Special 2015, Pages: 43-49

polynomials in order to estimate the probability density estimation when there are moment constraints in the data sets and take the advantage of Bernstein polynomials in the density estimation of the nonparametric approach.

Methodology:

In some other application of density estimation, the constraints are more qualitative in nature when we have a priori information namely a moment of the underlying density, then we can possibly obtain a better estimate of the true density. Hall (1999) has demonstrated a general method for resolving problems of density estimation under constraints by using a weighted bootstrap, which is based on the general biased-bootstrap methods. The constraints included condition on moment, quantiles and entropy of the density estimator. The mixture of Beta densities have led to different method of density estimation for univariate data. Another approach for solving the density estimation with moment constraints problem suggested by Eloyan and Ghosh (2011) is based on using mixture of known densities by using the expectation maximization(EM) algorithm. The constraints for example may include restriction on the density $f(\cdot)$ with mean 0 and variance 1. Our method using Bernstein polynomials for estimating density with moment constraints extends the work of Eloyan and Ghosh (2011).

Density estimation with moment constraints:

Assume that $X_i \overset{iid}{\sim} f(\cdot) i = 1,2,\ldots,n$ and suppose it is known that$E_f[g_j(X)] = 0$where $g(\cdot) = \left(g_1(\cdot),\ldots,g_p(\cdot)\right)'$ is know function. The motivation of this paper is to answer the question : How can we estimate f by say $\hat{f}$ that would satisfy $E_{\hat{f}}[g(X)] = 0$. For example suppose $g_1(x) = x$ and $g_2(x) = x^2 - 1$. Then $E_f[g(X)] = 0$ implies that$E[X] = 0$ and $E[X^2 - 1] = 0$ (i.e., $E[X^2] = 1$) and hence we get $E[X] = 0$ and $Var[X] = 1$. Hence, given a sample $X_1, X_2, \ldots, X_n$ from f and it is known that $E[X] = 0$ and $Var[X] = 1$ we want to construct $\hat{f}$ such that $E_{\hat{f}}[X] = 0$ and $Var_{\hat{f}}[X] = 1$.

Let us consider the mixture of Beta densities

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

where a_m, b_md and suitable function of data $f_b(u,a,b)$ denotes the density of the Beta(a,b) distribution and is given by

$$f_b(u,a,b) = \begin{cases} \dfrac{u^{(a-1)}(1-a)^{b-1}}{B(a,b)} & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $f_m(x,w)$ is a density on $\mathbb{R}$ if $w_k \ge 0 \forall k$ and $\sum_{K=1}^{m} w_k = 1$. Next we obtain the constraints on the weights that would satisfy the moment constraints. For example if we desire to have $E_{f_m}[g(X)] = 0$ then we need $\int_{a_m}^{b_m} g(x) f_m(x,w)dx = 0$ which means that the weight $w = (w_1 \ldots w_n)'$ must satisfy $\sum_{k=1}^{m} w_k \int_{a_m}^{b_m} g(x) f_b\left(\frac{x-a_m}{b_m-a_m}, k, m-k+1\right) dx = 0$. We explore this general condition in more specific cases.

We derive the constraints on the weights that would satisfy the moment constraints $E_{f_m}[X] = 0$. Suppose $E_{f_m}[X] = 0$. Hence, $\int_{a_m}^{b_m} x f_m(x)dx = 0$, then

$$\int_{a_m}^{b_m} x \sum_{k=1}^{m} w_k \frac{b\left(\frac{x-a_m}{b_m-a_m}, k, m-k+1\right)}{b_m - a_m} dx = 0$$

$$a_m + \frac{(b_m - a_m)}{m+1} \sum_{K=1}^{m} k w_k = 0$$

$$\sum_{k=1}^{m} k w_k = -\frac{(m+1)a_m}{b_m - a_m} \text{ (Notice that } a_m < 0)$$

Thus, to estimate f that satisfy $E_f[X] = 0$ by $f_m(x,w)$ we need the following conditions

(i)$w_k \ge 0\, ;\, for k = 1,2,\ldots,m$

(ii) $\sum_{k=1}^{m} w_k = 1$

(iii) $\sum_{k=1}^{m} k w_k = -\frac{(m+1)a_m}{b_m - a_m}$

In above condition, we notice that in order to have $E[X] = 0$ we must have $a_m \le 0$. The above set of constraints can be expressed in matrix from $Rw \ge \mathbf{c}$ where $R^T = [I_m \;\; 1_m \;\; d_m]$ is $(m+2) \times m$ vector,$d_m^T = [1, \; 2, \; \ldots \; m]$ and $c^T = \left[\mathbf{0}_m \;\; 1 \;\; -\frac{(m+1)a_m}{b_m - a_m}\right]$.

Newt, we derive the constraints on the weights that would satisfy the moment constraints $E_{f_m}[X^2] = 1$, which implies $\int_{a_m}^{b_m} x^2 f_m(x)dx = 1$, then

$$\int_{a_m}^{b_m} x^2 \sum_{k=1}^{m} w_k \frac{b\left(\frac{x-a_m}{b_m-a_m}, k, m-k+1\right)}{b_m - a_m} dx = 1$$

$$a_m^2 - 2a_m^2 + \frac{(b_m - a_m)^2}{(m+1)(m+2)} \sum_{k=1}^{m} k(k+1)w_k = 1$$

$$\sum_{k=1}^{m} k(k+1)w_k = \frac{(m+1)(m+2)a_m^2}{(b_m - a_m)^2}.$$

Hence, to estimate f that satisfies $E_{f_m}[X^2] = 1$, we need the following conditions

(i) $w_k \ge 0\, ; \forall k$

(ii) $\sum_{k=1}^{m} w_k = 1$

(iii) $\sum_{k=1}^{m} k(k+1)w_k = \frac{(m+1)(m+2)a_m^2}{(b_m - a_m)^2}$

The above constraints can be expressed again in matrix for $R\underline{w} \ge \underline{c}$ where $R^T = [I_m \;\; 1^T_m \;\; d_m]$ is $(m+2) \times m$ vector ,$d_m = [1 \times 2, \;\; 2 \times 3, \;\; \ldots \;\; m(m+1)]$ and $c^T = \left[\mathbf{0}_m \;\; 1 \;\; \frac{(m+1)(m+2)a_m^2}{(b_m - a_m)^2}\right]$.

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. To begin with the Bernstein polynomials, suppose $f: [0,1] \to R$ is a continuous density function. Consider the Bernstein polynomial(of order m-1) to approximate $f(\cdot)$:

$$B_m(x, f) = \sum_{k=1}^{m} f\left(\frac{k-1}{m-1}\right)\binom{m-1}{k-1} x^{k-1}(1-x)^{m-k} \quad (2)$$

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

$$\| B_m(\cdot, f) - f(\cdot) \|_\infty = \max_{0 \le x \le 1} |B_m(x, f) - f(x)| \to 0 \quad (3)$$

More interestingly, any qualitative known shape of f is preserved by that of B_m. If instead $x \in [a, b]$ we can use $u = (x-a)/(b-a)$. Suppose $f: [a, b] \to R$ is a density which is known to satisfy $E_f(g(X)) = \int g(x)f(x)dx = 0$. We propose to use the following class of estimators:

$$f_m(x, w) = \sum_{k=1}^{m} w_k f_b((x-a)/(b-a), k, m-k+1)/(b-a), \quad (4)$$

where w_k are unknown weights satisfying $w_k \ge 0$ 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 and $f_m(x, w)$ satisfies $\int g(x)f_m(x, w)dx = 0$ if $\sum_{k=1}^{m} w_k c_k = 0$ where $c_k = \int_0^1 g(a + (b-a)u)f_b(u, k, m-k+1)du$. Thus, we would like to estimate w_k in $f_m(x, w)$ satisfying the following constraint: (i) $w_k \ge 0$, (ii) $\sum_{k=1}^{m} w_k = 1$ and (iii) $\sum_{k=1}^{m} w_k c_k = 0$. Finally, the smoothness of $f_m(\cdot, w)$ would be controlled by suitably selecting m.

Consider the smooth cdf based on

$$F_m(x, w) = \sum_{k=1}^{m} w_k F_b((x-a)/(b-a), k, m-k+1) \quad (5)$$

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 (pbeta 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 square problem:

$$Minimize \sum_{i=1}^{n} \frac{n(F_n(x_i) - F_m(x_i, w))^2}{(F_n(x_i))(1-F_n(x_i))} withrespecttow \quad (6)$$

$$subjecttow_k \ge 0, \sum_{k=1}^{m} w_k = 1 \; and \sum_{k=1}^{m} w_k c_k = 0 \; fork = 1,2,\ldots,m, \quad 7)$$

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 (e.g., quadprog in R). In order to avoid numerical instabilities in (2), one may replace the denominator $(F_n(x_i))(1-$ $Fn(xi)$ by $Fn(xi)+\varepsilon n1+\varepsilon n-Fn(xi)$ where εn = 3/8n has been suggested by Anscombe and Aumann (1963). Optimal asymptotic rates for choosing m = m(n) has been derived by Vitale (1975) and Tenbush (1997). 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 MISE.

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. For Monte Carlo simulation, we obtain simulated samples from the standard normal distribution with sample sizes n=50, 100, 150 and 200to illustrate our method for this support. The software used for computation is R. The results compare the smoothing estimation Bernstein Polynomials estimation to the true density.

For Monte carlo simulation, we obtain simulated criteria to select m.The first, In the linear regression framework, Davies *et al*. (2005) presented an adjusted predictive divergence criterion (PDCa) can be serves as the minimum variance.Consider the linear model : $y = z\beta + \varepsilon$; $\varepsilon \sim (0, \sigma^2 c)$ where $c = diag(c_1, \ldots, c_n)$, $c_i = n/(F_n(x_i) + \varepsilon_n)(1 + \varepsilon_n - F_n(x_i))y$ is an $(n \times 1)$ observation vector, z is a $(n \times p)$ design matrix of full column rank, and β is a $(p \times 1)$ parameter vectors vector. By linear transformation. Let $\tilde{y} = \tilde{z}\beta + \tilde{\varepsilon}$; $\varepsilon \sim (0, \sigma^2 I)$ where $\tilde{y}_i = y_i/\sqrt{c_i}$, $\tilde{z}_i = z_i/\sqrt{c_i}$ and $\varepsilon_i = \varepsilon_i/\sqrt{c_i} \sim N(0, \sigma^2)$. An adjusted predictive divergence criterion (PDCa) is given by

$$PDCa = n\ln\hat{\sigma}^2 + \frac{n-1}{n-p-3}\sum_{i=1}^{n}\frac{1}{h_{ii}},$$

where h_{ii} is the i^{th} diagonal element of matrix, $H = X(X'X)^{-1}X'$ and $\hat{\sigma}^2 = (\tilde{y} - \tilde{z}\beta)'(\tilde{y} - \tilde{z}\beta)/n$.

The second, the optimal choice of the density estimation by using Bernstein polynomials can be achieved by minimizing the Mean Integrated Squared Error (MISE). Let $X_1, \ldots, X_n$ are i.i.d random variables from same density $f(\cdot)$. The Mean Integrated Squared Error is defined as

$$MISE[\hat{f}] = E\left[\int \left(\hat{f}(x) - f(x)\right)^2 dx\right]$$

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

Results:

The result are summarized in the table 1 by presenting the minimum of Mean Integrated Square Error (MISE) of true and estimated densities of three contraints.

Australian Journal of Basic and Applied Sciences, 9(13) Special 2015, Pages: 43-49

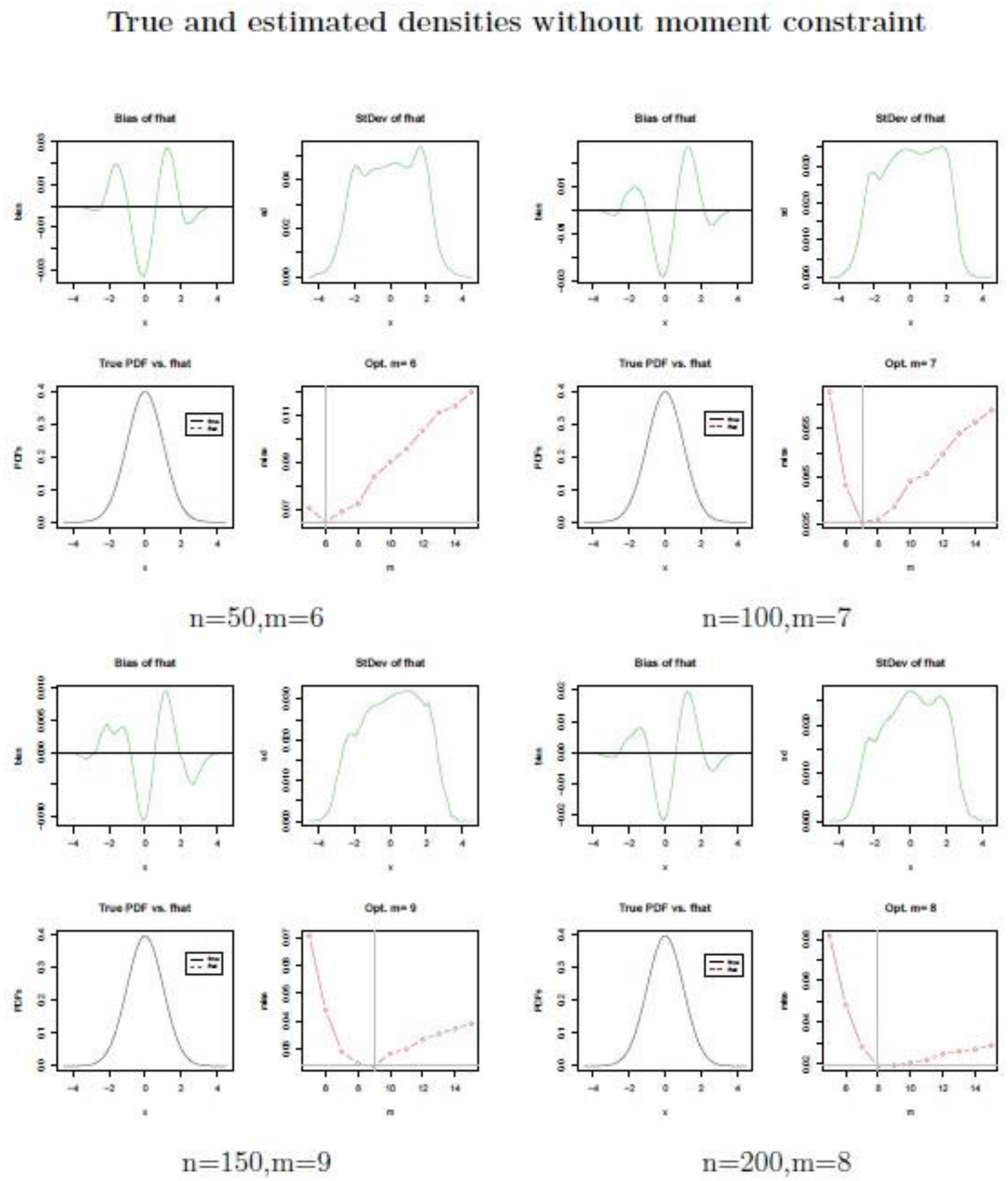


Fig. 1: True and estimated density for different sample from Normal density. The true density in solid (black) and the proposed density estimator without moment constrains in dash (red). The bias, standard deviation and optimal m for the density estimation method are besed on 1000 Monte Carlo repetitions.

Conclusion:

Whenever we know more information about a data set, the estimation of true density is expected to more accurate. In this work, we used the moment of constraints in the Bernstein polynomial approximation to estimate the density function. The way to define the constraint is guided by the adjustable weighted parameter. The weighted parameter can be determined by the constraint least square 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 were considered in three different distributions i.e. normal, beta and gamma and the simulation was run 1000 times for each data set. Three scenarios with constraints and without constraint were 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. Finally, our method clearly indicates that the consideration of more constrains can improve the errors of estimation.

Australian Journal of Basic and Applied Sciences, 9(13) Special 2015, Pages: 43-49

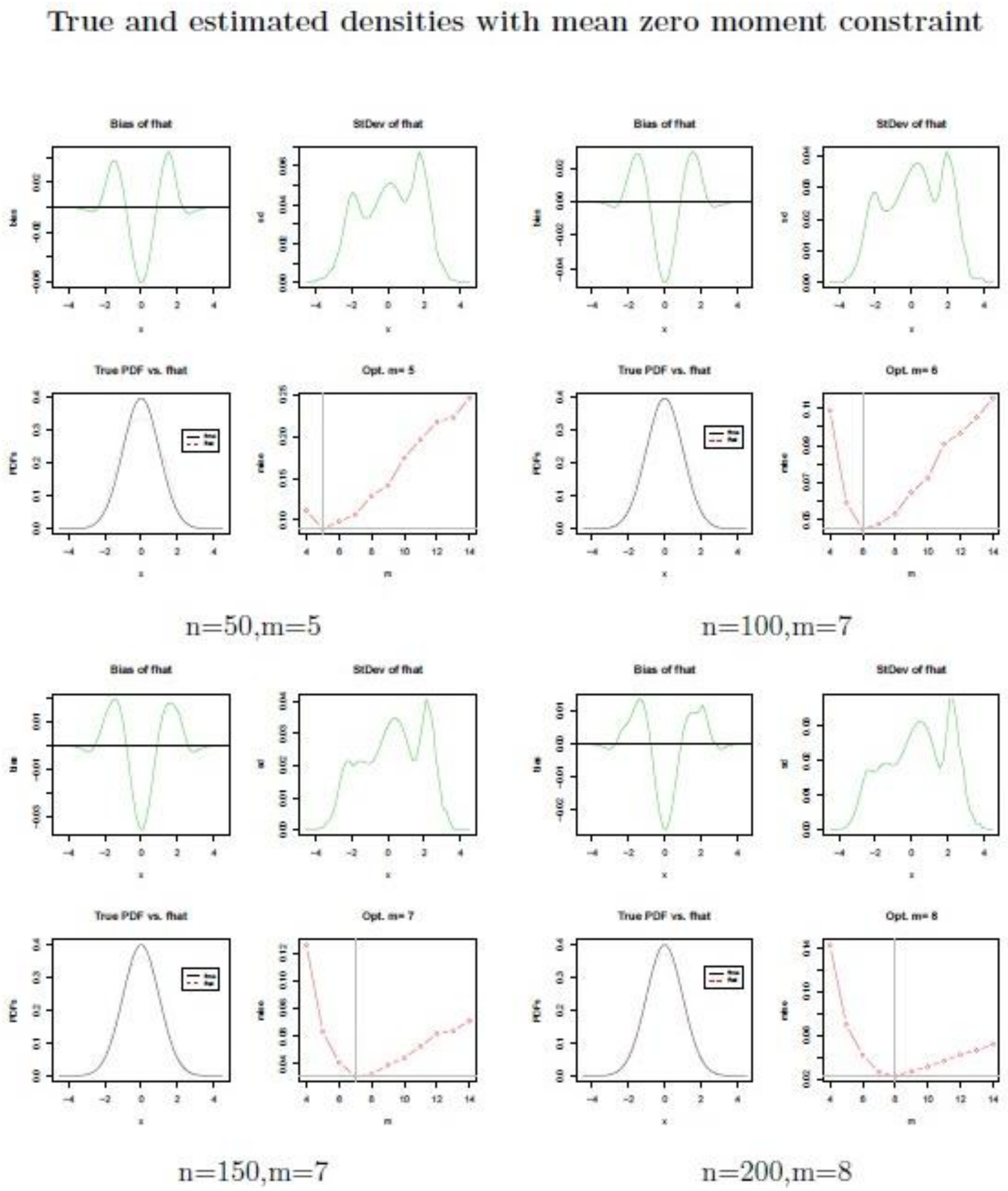
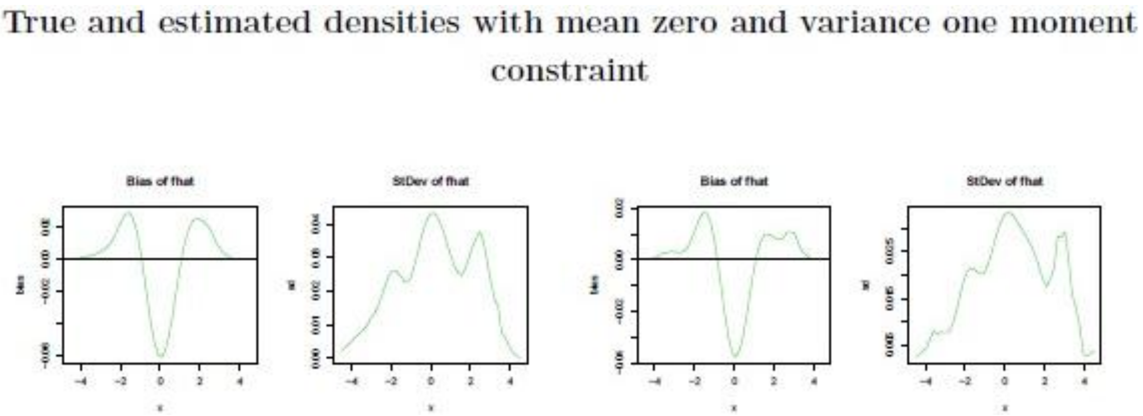


Fig. 2: True and estimated density for different sample from Normal density. The true density in solid (black) and the proposed density estimator with mean zero moment constrains in dash (red). The bias, standard deviation and optimal m for the density estimation method are besed on 1000 Monte Carlo repetitions.



Australian Journal of Basic and Applied Sciences, 9(13) Special 2015, Pages: 43-49

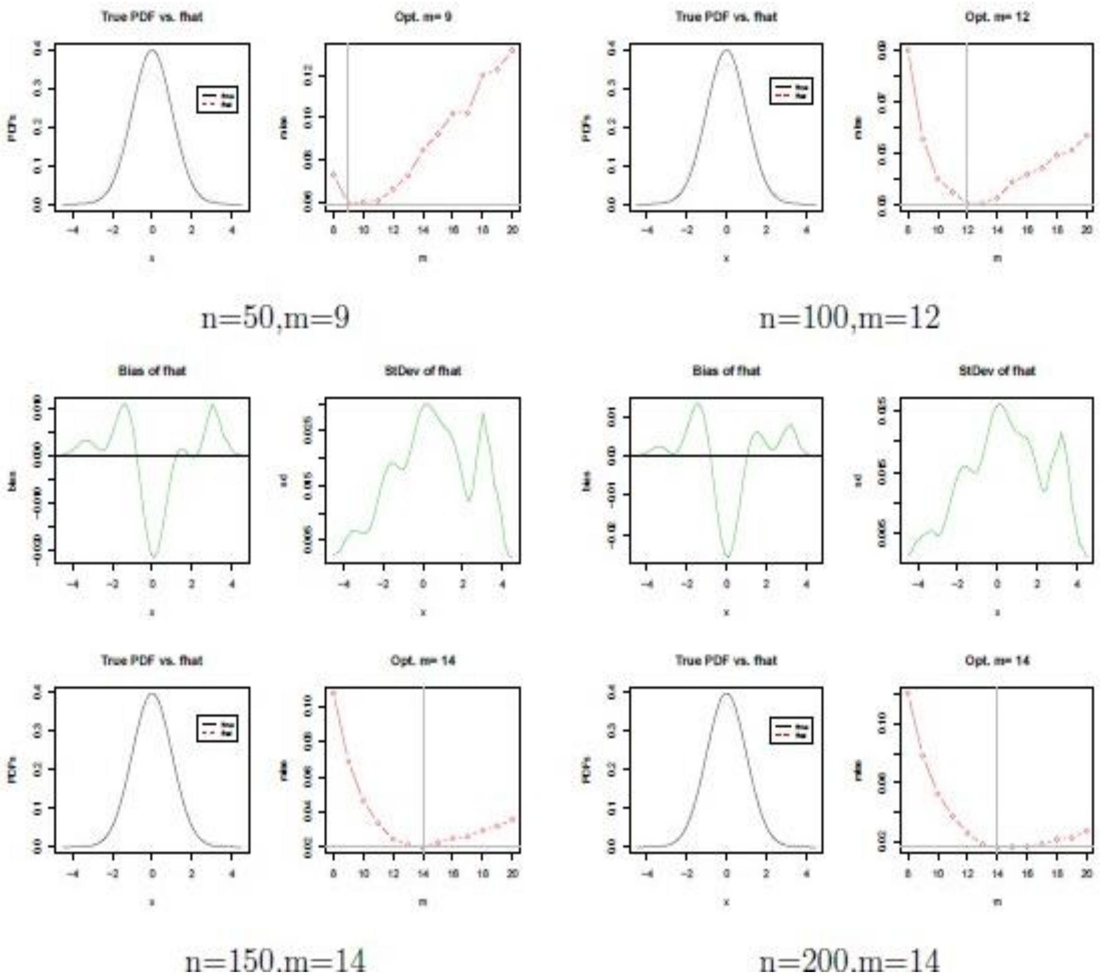


Fig. 3: True and estimated density for different sample from Normal density. The true density in solid (black) and the proposed density estimator with mean zero and variance one moment constrains in dash (red). The bias, standard deviation and optimal m for the density estimation method are besed on 1000 Monte Carlo repetitions.

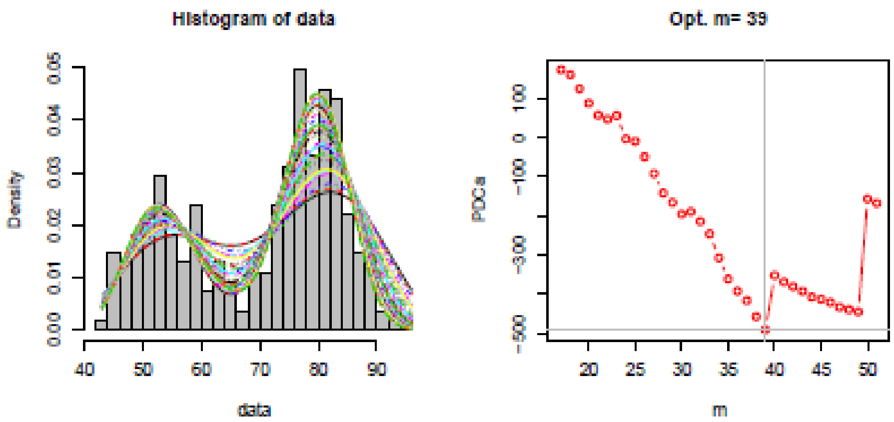


Fig. 4: By applying our method to the Old Faithful data set without moment constraint in dashed (blue) line shows the proposed estimated densities by using Bernstein polynomial method.

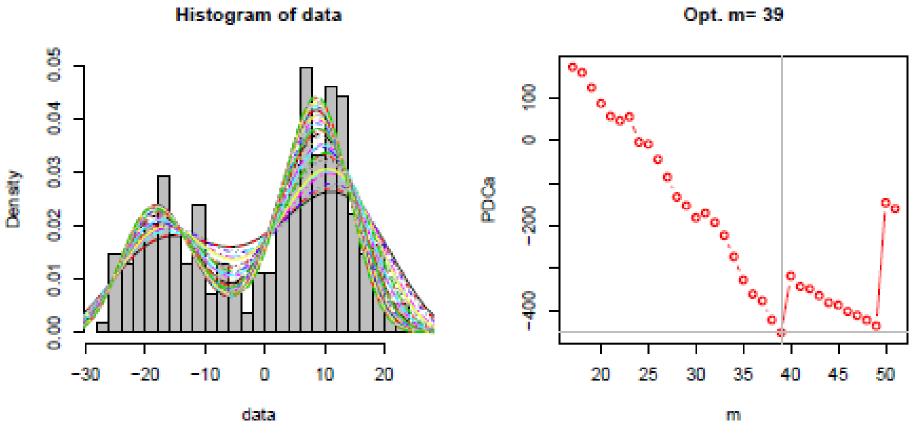


Fig. 5: By applying our method to the Old Faithful data set with mean zero moment constraint in dashed (blue) line shows the proposed estimated densities by using Bernstein polynomial method.

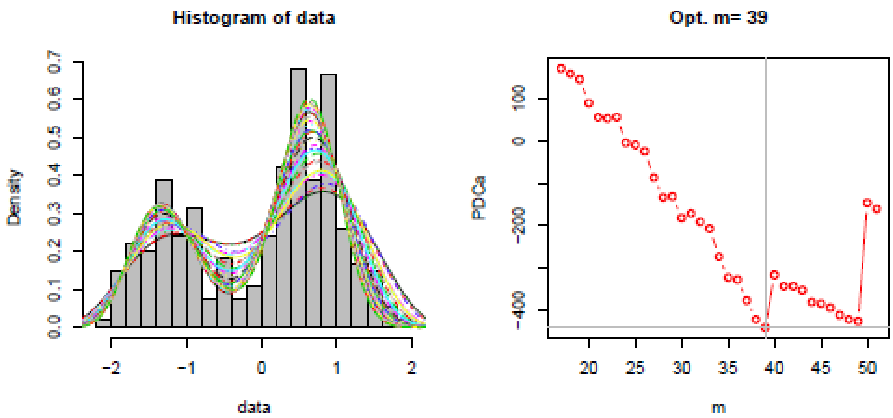


Fig. 6: By applying our method to the Old Faithful data set with mean zero and variance one moment constraint in dashed (blue) line shows the proposed estimated densities by using Bernstein polynomial method.

REFERENCES

Bernstein, S.N., 1912. Demonstration du theoreme de weierstrassfondeesurlecalculdesprobabilites. Communications of the Kharkov Mathematical Society, 13: 1-2.

Davies, S.L., A.A. Neath, J.E. Cavanaugh, 2005. *Cross validation model selection criteria for linear regression based on the Kullback-Leibler discrepancy*, Statistical Methodology, 2: 249-266.

Eloyana, K. Ghosh, 2011. *Smooth density estimation with moment constraints using mixture distributions*, Journal of Nonparametric Statistics, 23(2): 513-531.

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

Hall, P. and B. Presnell, 1999. *Density estimation under constraints*, Journal of Computational and Graphical Statistics, 8: 259-277.

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

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

Tenbusch, A., 1997. *Nonparametric curve estimation with Bernstein estimates*, Metrika, 45: 1-30.

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