International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 5 Issue 7, July-2016

A proposed nonparametric mixture density estimation using B-spline functions

Atizez Hadrich[1,2], Afif Masmoudi[1], Mourad Zribi[2]

[1]Laboratory of Probability and Statistics, Faculty of Sciences of Sfax, Sfax University.B. P. 1171, Sfax. Tunisia
[2] Laboratoire d'Informatique Signal et Image de laCôted' Opale (LISIC-EA 4491), ULCO, 50 Rue Ferdinand Buisson BP 719, 62228 Calais Cedex, France.

ABSTRACT

In this paper, we suppose that a density of probability f is expressed as a finite linear combination of second order B-spline functions. Then, we obtain a finite mixture of B-spline. We extend the Expectation Maximization (EM) algorithm in order to estimate the new mixture density. The experiments show that the proposed estimator using B-spline functions can produce a satisfactory estimation of mixture density than Gaussian classical theory.

Keywords: B-spline functions, EM algorithm, Estimator, Mixture density, Probability density function.

1. INTRODUCTION

In statistics, concerning the density estimation in the construction, there are several competitive classes of nonparametric estimators, the two most popular being the Kernel estimators introduced by Rosenblatt [8] and the Orthogonal series estimators [8]. Usually, a Kernel estimator has a greater efficiency than an estimator based on Orthogonal series, although it may be more complex to calculate and update. In case the density has its support confined to the positive half line and may not be continuous at the origin, the Kernel method will not be so attractive since the estimator may perform poorly in the neighborhood of the origin. In such a situation, an estimator based on Orthogonal series may be more appropriate. Although scarcely used, the Orthogonal series density estimator bears a striking resemblance to the B-spline density estimator [1].

However, the Orthogonal series estimator may sometimes lead to a function that is not a pdf. Again, the only reason for including this estimator is that the B-spline estimator actually has the same form, except that the basis functions are not Orthogonal. In this paper, we propose a nonparametric method to estimate mixture density using the second order B-spline and the EM algorithm [3]. The combination of the B-spline functions with the EM algorithm allows us to define a new algorithm which we denote as EM Generalized B-spline (EMGB) algorithm. The paper is organized as follows: In section 2,we describe the nonparametric density estimation methods as well as the B-spline functions. Convergence properties of the iterative Maximum Likelihood Estimation (MLE) are given in section 3.

The performance comparison based on the Mean Squared Errors (MSE) is presented in section 4 in order to show the accuracy of the proposed estimator. Finally, the conclusion appears in section 5.

2. PROPOSED ESTIMATION METHODS BY USING B-SPLINE FUNCTIONS

In the mathematical subfield of numerical analysis, a B-spline is a spline function that has a minimum support with respect to a given degree, smoothness and domain partition. It's well known that every spline density function can be represented as a finite linear combination of B-spline [1].The term B-spline stands for basis splines according to Isaac Jacob Schoenberg [2]. A B-splinenon parametric density estimator with uniformly spaced knots convenient for large data sets was discussed by Gehringer [4]. Curry and Schoenberg (1966) [2] have proved that every spline function S of degree d (d = 1, 2, …) with m knots (m = 1, 2, …) has a unique expansion

$$S(x) = \sum_{l=}^{m+d} \frac{b_l}{h_l} B_l^d(x), \text{ for a} < x < b \text{ and } h_l = \int B_l^d(x)dx, \quad (1)$$

where $a, b \in IR$ and b_l's are unknown parameters that need to be estimated. Note that $b_l \geq 0$ and $\sum_{l=1}^{m+d} b_l = 1$ is a special requirement when using B-splines in order to estimate probability densityfunctions. The B-spline of d degree are defined recursively by

$$B_l^d(x) = \frac{x - x_l}{x_{l+d} - x_l} B_l^{d-1}(x) + \frac{x_{l+d+1} - x}{x_{l+d+1} - x_{l+1}} B_{l+1}^{d-1}(x) \qquad (2)$$

where

$$B_l^0(x) = \begin{cases} 1, if \quad x \in [x_l, x_{l+1}) \\ 0, elswhere \end{cases} \qquad (3)$$

For general splines, this special requirement is not true. The case $d = 1$ corresponds to a piecewise linear approximation which is attractively simple but produces a visible roughness, unless the knots are close to each other.
In more technical terms, a spline function S of degree d with m interior knots, $a = x_1 \prec x_2 \prec ..... \prec x_{m+2} = b$ is a $(d-1)$ continuously differentiable function, such that $S \in P^d$ [2], (P^d is the class ofpolynomials of a maximum degree d in each of the intervals $(a, x_2), (x_2, x_3), ..., (x_{m+1}, b))$.

When we consider the approximation by the B-splines of d degree, the density function f is supposed to be $(d-1)$ continuously differentiable in each of the intervals $(a, x_2), (x_2, x_3), ..., (x_{m+1}, b)$. Hence, we have chosen the quadratic B-splines functions which are only continuously differentiable, that isto say, having a minimal requirement, as compared with the B-splines of degree $d \succ 2$.
In our work, we have used the second order B-splines functions $(B_l^2)_{l=1,2..,m+2}$ defined by

$$B_l^2(x) = \begin{cases} \dfrac{(x - x_l)^2}{(x_{l+1} - x_{lm})(x_{l+2} - x_{lm})}, x \in [x_l, x_{l+1}) \\ \dfrac{1}{x_{l+2} - x_{l+1}} \left[\dfrac{(x - x_l)(x_{l+2} - x)}{(x_{l+2} - x_l)} + \dfrac{(x - x_{l+1})(x_{l+3} - x)}{(x_{l+3} - x_{l+1})} \right], x \in [x_{l+1}, x_{l+2}) \\ \dfrac{(x_{l+3} - x)^2}{(x_{l+3} - x_{l+1})(x_{l+3} - x_{l+2})}, x \in [x_{l+2}, x_{l+3}) \\ 0, elswhere \end{cases} \qquad (4)$$

where l is an integer as usual. We notice that $0 \leq B_l^2 \leq 1$ is always verified.
The support of each spline covers three intervals. It is a quadratic polynomial on each support interval. Note that the peak value of $B_l^2(x)$ is less than 1. If all the B_l^2 splines could be plotted in any interval, there would be contributions from the three splines. Note that any pdf f can be approximated by the following mixture (5) of B-splines:

$$f(x) = \sum_{l=1}^{m+2} b_l B_{l,h_l}^2(x), \qquad (5)$$

with $b_l \geq 0$ and $\sum_{l=1}^{m+d} b_l = 1$, $B_{l,h_l}^2(x) = \frac{B_l^2(x)}{h_l}$ and $h_l = \int_{-\infty}^{+\infty} B_l^2(x)dx = \int_{x_l}^{x_{l+3}} B_l^2(x)dx$. The estimation of the density f reduces the estimation of the finite-dimensional parameters $(b_1, b_2, ..., b_{m+2})$ that characterize f.

2.1 Estimation of b_l by the EM algorithm

Assuming independence between the observations $X_1, X_2, ..., X_n$ we define the Log-likelihood function

$$l(X_1,....,X_n)=\sum_{i=1}^{n} Log(\sum_{l=1}^{m+2} b_l B_{l,h_l}^2(X_i)). \qquad (6)$$

To obtain the MLE of $(b_1,b_2,...,b_{m+2})$, we apply the EM algorithm [3]. We initialize the B-spline coefficients by $b_l^{(0)}=\frac{1}{n}\sum_{j=1}^{n} B_{l,h_l}^2(X_j)$.

E-Steep:

$$\tau_{l,j}^{(p)}=\frac{b_l^{(p)} B_{l,h_l}^2(X_j)}{\sum_{s=1}^{m+2} b_s^{(p)} B_{s,h_l}^2(X_j)}; l=1,....,m+2,\ j=1,....,n \qquad (7)$$

where $\pi_{l,j}^{(p)}$ is the is the conditional probability at the p^{th} iteration.

M-Step: The coefficients of B-spline at $(p+1)^{th}$ iteration are given by

$$b_l^{(p+1)}=\frac{1}{n}\sum_{j=1}^{n}\tau_{l,j}^{(p)}, l=1,...,m+2 \qquad (8)$$

After some iterations of the EM algorithm, we obtain the mixture B-spline estimator of f as

$$\hat{f}(x)=\sum_{l=1}^{m+2}\hat{b_l} B_{l,h_l}^2(x), \qquad (9)$$

where $\hat{b}_l$ is the maximum likelihood of b_l.

In what follows, we suggest to extend the second order B-spline estimator to the problem of the mixture density estimation [1]. The combination of the B-spline functions with the EM algorithm allows us to define a new algorithm denoted as EM Generalized B-spline (EMGB) algorithm.

.2.2 Description of the EMGB algorithm

Let's consider the following mixture density of mixture of B-spline

$$f(x)=\sum_{k=1}^{K}\pi_k f_k(x/b_{.,k})=\sum_{k=1}^{K}\pi_k\sum_{l=1}^{m+2} b_{l,k} B_{l,h_l}^2(x), \text{ where } \pi_k\geq 0 \text{ and } \sum_{k=1}^{K}\pi_k=1. \qquad (10)$$

$$\Theta=\{\theta=(\pi_1,.....,\pi_k,b_{.,1},.....,b_{.,K}); \pi_k\geq 0, \sum_{k=1}^{K}\pi_k=1; k=1,2,....K; b_{.,k}\in[0,1]^{m+2}; \sum_{l=1}^{m+2} b_{l,k}=1\} \qquad (11)$$

and $f(x/b_{.,k})=\sum_{l=1}^{m+2} b_{l,k} B_{l,h_l}^2(x)$. Let $X_1,X_2,...,X_n$ be n random variables with an unknown common pdf f. For each X_i, we consider a random variable $Z=(Z_{1,i},Z_{2,i},...,Z_{K,i})$ such that Z_i follows a multinomial distribution $Z=(Z_{1,i},Z_{2,i},...,Z_{K,i})$. Let $X=(X_1,X_2,...,X_n)$ and $Z=(Z_1,Z_2,...,Z_n)$, the log maximum likelihood is given by

$$l(X,Z/\theta)=\sum_{i=1}^{n}\{\sum_{k=1}^{K} Z_{k,i} Log(\pi_k f(X_i/b_{.,k}))\}. \qquad (12)$$

We define the mean of the log likelihood

$$Q(\theta,\theta^{(p)})=IE(l(X,Z/\theta)/\theta^{(p)})=\sum_{i=1}^{n}\{\sum_{k=1}^{K} IE(Z_{k,i}/X,\theta^{(p)}) Log(\pi_k f(X_i/b_{.,k}))\}. \qquad (13)$$

If $\tau_{k,i}^{(p)} = IE(l(X,Z/\theta)/\theta^{(p)})$, then

$$Q = Q(\theta,\theta^{(p)}) = \sum_{i=1}^{n}\{\ \sum_{k=1}^{K}\tau_{k,i}^{(p)} Log(\pi_k f(X_i/b_{.,k}))\ \} \qquad (14)$$

which implies

$$Q = \sum_{i=1}^{n}\{\ \sum_{k=2}^{K}\tau_{k,i}^{(p)} Log(\pi_k f(X_i/b_{.,k}))+ \tau_{1,i}^{(p)} Log((1-\sum_{k=2}^{K}\pi_k) f(X_i/b.,1)\Big\}. \qquad (15)$$

The EMGB algorithm seeks to find the MLE by applying iteratively the following two steps:

E-Step: We estimate a posterior probability $\tau_{k,i}^{(p)}$ belonging to the class k at the p^{th} iteration:

$$\tau_{k,i}^{(p)} = IE(Z_{k,i}/X_i,\theta^{(p)}) = \frac{\pi_k^{(p)} f(X_i/b_{.,k}^{(p)})}{\sum_{k=1}^{K}\pi_k^{(p)} f(X_i/b_{.,k}^{(p)})}. \qquad (16)$$

M-Step: we calculate the parameter $\theta^{(p+1)}$ that maximizes $\theta^{(p+1)} = \arg\max_{\theta\in\Theta} Q(\theta,\theta^{(p)})$.

On the one hand,

$$\frac{\partial Q}{\partial \pi_k} = \sum_{i=1}^{n}(\frac{\tau_{k,i}^{(p)}}{\pi_k} - \frac{\tau_{1,i}^{(p)}}{1-\sum_{k=2}^{K}\pi_k}) = 0. \qquad (17)$$

This means that

$$\pi_k = \frac{\sum_{i=1}^{n}\tau_{k,i}^{(p)}}{\sum_{i=1}^{n}\tau_{1,i}^{(p)}}(1-\sum_{k=2}^{K}\pi_k) \text{ which implies } \sum_{k=2}^{K}\pi_k = \sum_{k=2}^{K}\frac{\sum_{i=1}^{n}\tau_{k,i}^{(p)}}{\sum_{i=1}^{n}\tau_{1,i}^{(p)}}(1-\sum_{k=2}^{K}\pi_k). \qquad (18)$$

Therefore,

$$\pi_k^{(p+1)} = \frac{\sum_{i=1}^{n}\tau_{k,i}^{(p)}}{\sum_{k=1}^{K}\sum_{i=1}^{n}\tau_{k,i}^{(p)}}(1-\sum_{k=2}^{K}\pi_k) = \frac{1}{n}\sum_{k=1}^{n}\tau_{k,i}^{(p)}. \qquad (19)$$

On the second hand,

$$\frac{\partial\ Q}{\partial b_{l,k}} = \sum_{i=1}^{n}\tau_{k,i}^{(p)}(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i/b_{.,k})}) = 0; l = 2,....., m+2 \qquad (20)$$

and

$$\frac{\partial^2 Q}{\partial^2 b_{l,k}} = -\sum_{i=1}^{n}\tau_{k,i}^{(p)}(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i/b_{.,k})})^2; l = 2,....., m+2. \qquad (21)$$

In order to solve $\frac{\partial\ Q}{\partial b_{l,k}} = 0$, we use the Newton Raphson method [5] or the steepest descent method, we obtain

$$b_{l,k}^{(p+1)} = b_{l,k}^{(p)} + \frac{\sum_{i=1}^{n}\tau_{k,i}^{(p)}(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i/b^{(p)}{}_{.,k})})}{\sum_{i=1}^{n}\tau_{k,i}^{(p)}(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i/b_{.,k}^{(p)})})^2}; l = 2,....., m+2 \text{ and } k = 1,.....K. \qquad (22)$$

The calculation cycles are made from one step to another until we achieve the convergence. Asstarting values for the EMGB algorithm, we take the initial values of π_k and $b_{l,k}$ which have to be equal respectively to $\frac{n_k}{n}$ and $\frac{1}{n}\sum_{j=1}^{n_k} B_{l,h_l}^2(X_j)$; where n_k is the total number of observations in the k^{th} class.

3. CONVERGENCE OF THE PROPOSED ESTIMATOR

In this section, we wish to show first that the EMGB iterations converge to a value $(\hat{\pi}_k, \hat{b}_{l,k})$, second that this value $(\hat{\pi}_k, \hat{b}_{l,k})$ indeed satisfies the likelihood equations.

Proposition

Let $\theta = (\pi_1,...,\pi_K, b_{.,1},...,b_{.,K}) \in \Theta$. If $\theta^{(p+1)}$ converges to $\hat{\theta}$ as $p \to +\infty$, then $\hat{\theta}$ satisfies the likelihood equations.

Proof

On the one hand, for a fixed k, the derivative of the Log-likelihood function with respect to $b_{l,k}$ is given by

$$\frac{\partial l(X_1,...,X_n|\hat{\theta})}{\partial \hat{b}_{l,k}} = \sum_{=1}^{n} \hat{\pi}_k \left(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{\hat{f}(X_j)}\right); l = 2,...,m+2. \quad (23)$$

where $\hat{f}(X_j) = \sum_{k=1}^{K} \hat{\pi}_k \hat{f}_k(X_j)$ and $\hat{f}(X_j) = \sum_{j=1}^{n} \hat{b}_{l,k} B_{l,h_l}^2(X_j)$.

By letting $p \to +\infty$ in (22), by using the fact that $b_{l,k}^{(p)} \to \hat{b}_{l,k}$ and $\tau_{k,j}^{(p)} \to \hat{\tau}_{k,j} = \frac{\hat{\pi}_k \hat{f}_k(X_j)}{\sum_{k=1}^{K} \hat{\pi}_k \hat{f}_k(X_j)}$, one has

$$\sum_{j=1}^{n} \hat{\tau}_{k,j} \left(\frac{B_{l,h_l}^2(X_j) - B_{1,h_l}^2(X_j)}{\hat{f}_k(X_j)}\right) = 0; l = 2,...,m+2. \quad (24)$$

Therefore $\sum_{j=1}^{n} \hat{\pi}_k \left(\frac{B_{l,h_l}^2(X_j) - B_{1,h_l}^2(X_j)}{\hat{f}(X_j)}\right) = 0$. And, $\hat{b}_{l,k}$ satisfies the derivation of the likelihood equations.

On the other hand, $\hat{\pi}_k = \frac{1}{n}\sum_{j=1}^{n} \frac{\hat{\pi}_k \hat{f}_k(X_j)}{\sum_{k=1}^{K} \hat{\pi}_k \hat{f}_k(X_j)}$. This implies that

$$0 = n - \sum_{j=1}^{n} \frac{\hat{f}_k(X_j)}{\sum_{k=1}^{K} \hat{\pi}_k \hat{f}_k(X_j)} \quad (25)$$

$$= \sum_{j=1}^{n} \frac{\hat{\pi}_1 \hat{f}_1(X_j) + \hat{\pi}_2 \hat{f}_2(X_j) + ... + \hat{\pi}_{K-1} \hat{f}_{K-1}(X_j) + (1 - \hat{\pi}_1 - \hat{\pi}_2 - ... - \hat{\pi}_{K-1}) \hat{f}_K(X_j) - \hat{f}_k(X_j)}{\sum_{k=1}^{K} \hat{\pi}_k \hat{f}_k(X_j)} \quad (26)$$

$$= \sum_{j=1}^{n} \frac{\hat{\pi}_1 (\hat{f}_1(X_j) - \hat{f}_K(X_j)) + ... + \hat{\pi}_{K-1} (\hat{f}_{K-1}(X_j) - \hat{f}_K(X_j)) + \hat{f}_K(X_j) - \hat{f}_k(X_j)}{\sum_{k=1}^{K} \hat{\pi}_k \hat{f}_k(X_j)} \quad (27)$$

Let

$$H=\sum_{j=1}^{n}\frac{\hat{\pi}_1(\hat{f}_1(X_j)-\hat{f}_K(X_j))+...+\hat{\pi}_{K-1}(\hat{f}_{K-1}(X_j)-\hat{f}_K(X_j))}{\sum_{k=1}^{K}\hat{\pi}_k\hat{f}_k(X_j)}. \quad (28)$$

Then, multiplying (27) by $\hat{\pi}_k$ for $k=1,...,K-1$ and by summing the results of the $(K-1)$ successive multiplications, we obtain $(\hat{\pi}_1+\hat{\pi}_2+...+\hat{\pi}_{K-1})H=H$ Since $\sum_{k=1}^{K}\hat{\pi}_k=1$, then $(1-\hat{\pi}_K)H=H,$ and $\hat{\pi}_K H=0.$ Note that we chose $\hat{\pi}_K$ arbitrarily in (28) . So we can say that $\hat{\pi}_k H=0,\forall k=1,...,K.$ Necessarily, $H=0$ .From (27),this implies

$$\frac{\partial l(X_1,...,X_n|\hat{\theta})}{\partial\hat{\pi}_k}=\sum_{j=1}^{n}\frac{\hat{f}_k(X_j)-\hat{f}_K(X_j)}{\sum_{k=1}^{K}\hat{\pi}_k\hat{f}_k(X_j)}=0;\text{ for } k=1,...,K-1. \quad (29)$$

Therefore, $\hat{\pi}_1,\hat{\pi}_2,...,\hat{\pi}_K$ of the EMGB satisfy the likelihood equations.

4. PERFORMANCE COMPARISON

In our study, we simulate $n=1000$ observations according to different distributions (Normal (N),Beta (B) and Gamma (G)) with given true parameters. We calculate the density estimator by using Histogram, Kernel, Orthogonal and B-spline methods. Then, we compute the Mean Squared Errors for each method.

$$MSE=\frac{1}{n}\sum_{j=1}^{n}(f(x_j)-\hat{f}(x_j))^2 \quad (30)$$

Among the different methods, Table 1 shows that the B-spline method is the best one by giving the lowest MSE. This finding is valid for a unique density as well as a mixture density. In Table 2, we have first computed the MSE between the empirical distribution and the estimated mixture density by using the Gaussian EM algorithm. Second, we have computed the MSE between the empirical distribution and the estimated mixture density by using the EMBG algorithm. This work was done for two images (Lena and Boat). We notice thatthe EMGB method gives a much lower MSE for both images. In Figure 1, in the case of mixture densities, the B-spline method remains the best one in terms of being close to the real density.

Table 1: MSE of different estimations methods.

Methods / True model	Histogram	Kernel	Orthogonal	B-spline
N(0, 1)	0.143	0.04	0.0094	0.00299
B(3, 5)	0.079	0.022	0.00119	0.001115
G(2, 3)	0.072	0.019	0.0027	0.0012
Mixed of N(-2, 2), G(3, 1) and B(11, 5)	0.0022599	0.00054985	0.00051597	0.00015088

Table 2: MSE of mixture distribution obtained by EM and EMGB.

Images / Methods	MSE of Lena	MSE of Boat
Gaussian EM	0.2981	0.47
EMGB	0.0406	0.0989

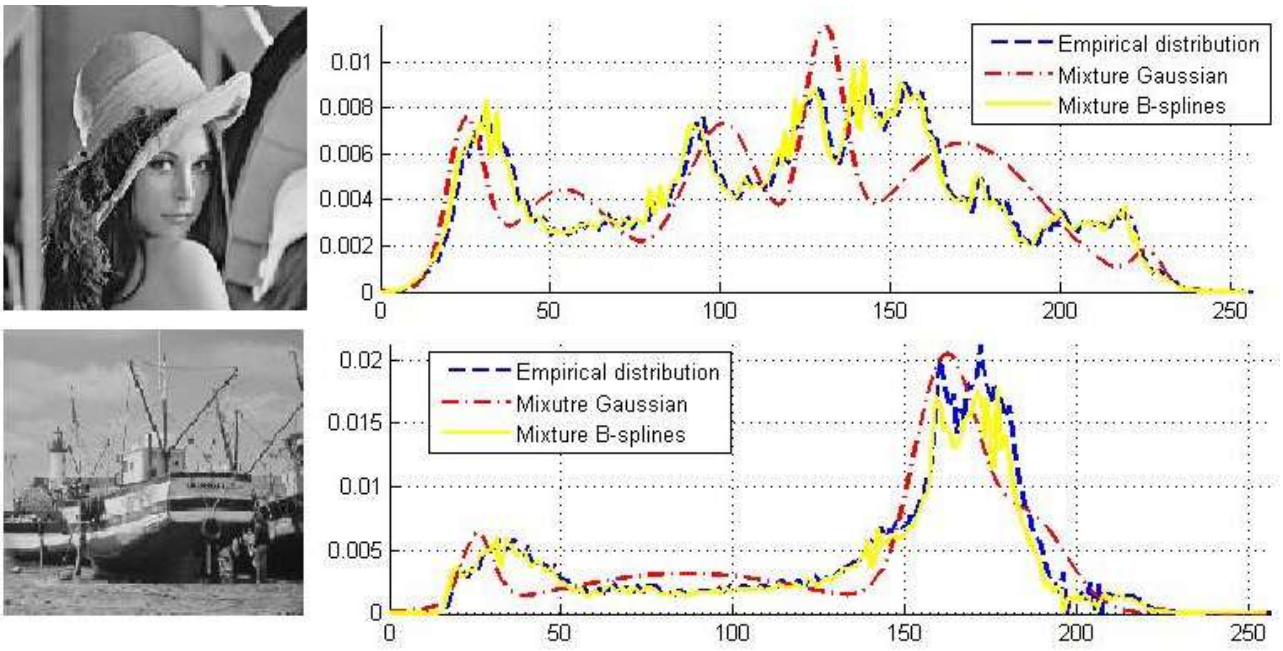

Figure 1: Estimation of mixture density using both classical EM algorithm and EMGB algorithm.

CONCLUSION

In this paper, we have introduced a new nonparametric B-spline estimator for mixture distributions. Many results presented show that the estimation of pdf density by using the proposed estimator is better than those of other methods. This comparison is obtained by the computation of the Mean Squared Errors between the estimated density and the empirical distribution.

REFERENCES

[1]. Atilgan, Bozdogan, (1992). Convergence properties of MLE's and asymptotic simultaneous confidence intervals in fitting cardinal B-spline for density estimation. Statistics and Probability Letters, Vol. 13, pp. 89-98, North-Holland.
[2]. Curry H.B, Schoenberg I.J., (1966). On poly freqency functions IV: The fundamental spline Functions and their limits. J.Anal. Math, Vol. 17, pp. 71-107.
[3]. Dempster A.P., Laird N.M., Rubin D.B, (1977). Maximum Likelihood from Incomplete Data via the E.M. Algorithm, Journal of the Royal Statistical Society , Ser. B, Vol. 39, pp. 1-38.
[4]. Gehringer, K.R. and Redner, R.A., (1992), Nonparametric density estimation using tensor product splines. Communications in Statistics-Simulation and Computation, Vol. 21, pp. 849878.
[5]. Hazewinkel, Michiel, ed. (2001), Newton method, Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4.
[6]. Kim J. and al., (2005). A Nonparametric Statistical Method for image segmentation using information theory and curve evolution. IEEE transactions on image processing, Vol. 14, pp. 1486-1502.
[7]. Peter Hall, (1983). Orthogonal series distribution function estimation, with application. J, R, Statist, Vol 48, pp. 115-122.
[8]. Rosenblatt M., (1956). Remarks on some nonparametric estimates of a density function. Ann. Math.Statist, Vol. 27, pp. 832-837.