The Korean Journal of Applied Statistics (2012) DOI: http://dx.doi.org/10.5351/KJAS.2012.25.6.1049
25*(6), 1049–1058*

On the Selection of Bezier Points in Bezier Curve Smoothing

Choongrak Kim[1] · Jin-Hee Park[2]

[1]Department of Statistics, Pusan National University
[2]Department of Statistics, Pusan National University

(Received October 11, 2012; Revised November 14, 2012; Accepted November 15, 2012)

Abstract

Nonparametric methods are often used as an alternative to parametric methods to estimate density function and regression function. In this paper we consider improved methods to select the Bezier points in Bezier curve smoothing that is shown to have the same asymptotic properties as the kernel methods. We show that the proposed methods are better than the existing methods through numerical studies.

Keywords: Kernel density estimation, mean integrated squared error, regression function.

1. Introduction

Nonparametric methods are often used as alternatives do to unrealistic and restrictive assumptions on parametric methods in density function estimation and regression function estimation. For the nonparametric estimation of density function, kernel density estimation is usually used, and three methods such as kernel method (local polynomial regression), series method (wavelet estimator), and spline method (regression spline and smoothing spline) are quite popular for the nonparametric estimation of regression function. See Silverman (1986), Eubank (1988), Fan and Gijbels (1996), Loader (1999) and Wasserman (2006), for example, among others. Bezier curve smoothing (Bézier, 1977) is regarded as one of kernel-type approaches and is also a useful nonparametric method to estimate density function and regression function. The Bezier curve is very popular smoothing technique in computational graphics, especially for computer-aided-geometric design; however, it has rarely been used in statistics. Kim (1996) first applied the Bezier curve to density estimation in the statistical area and Kim *et al.* (1999) showed that estimators using the Bezier curve smoothing in density estimation and regression function estimation have the same asymptotic properties as classical kernel estimators. Subsequent works on applications of the Bezier curve to statistics are the estimation in the measurement error model (Kim *et al.*, 2000), the smoothing of the Kaplan-Meier estimator (Kim *et al.*, 2003), and the smoothing of the bivariate Kaplan-Meier estimator (Bae *et al.*, 2005).

This work was supported for two years by a Pusan National University Research Grant.
[1]Corresponding author: Professor, Department of Statistics, Pusan National University, Jangjeon 2-dong, Geumjeong-gu, Busan 609-390, Korea. E-mail: crkim@pusan.ac.kr

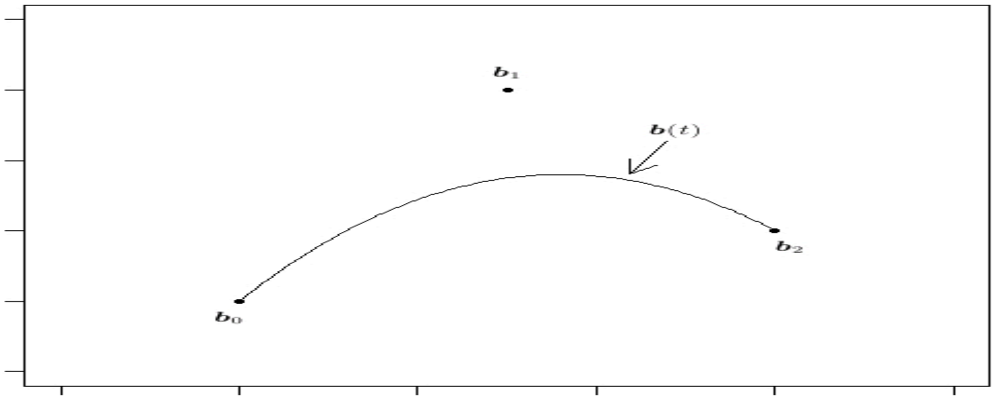

Figure 2.1. Bezier curve $\boldsymbol{b}(t)$ based on 3 Bezier points $\boldsymbol{b}_0$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$.

One of the most important and difficult problems in the Bezier curve smoothing is the selection of Bezier points because the Bezier curve is totally determined by the selected Bezier points. In fact, the selection of Bezier points in Bezier curve smoothing corresponds to the estimation of the smoothing parameter in kernel estimation. In this paper, we consider the selection of Bezier points in distribution function estimation and in regression function estimation. To estimate the distribution function, the proposed method gives similar numerical results to Kim *et al.* (1999) even though the number of Bezier points used in the proposed method is almost half the number of Bezier points used in Kim *et al.* (1999). Kim *et al.* (1999) chose the middle points of the cumulative histogram as Bezier points; however, the proposed method suggested choosing the right-most points where they are less than the median and choosing the left-most points where they are larger than the median. To estimate the regression function, we propose to choose the sample median instead of the sample mean for the bin estimator; subsequently, the proposed method gives better numerical results than Kim *et al.* (1999), especially when there are potential outliers in data.

This paper is organized as follows; In Section 2, review on the Bezier curve are introduced. Section 3 proposes methods to select the Bezier points to estimate the distribution function and regression function; subsequently, relevant numerical results are given. Concluding remarks are provided in Section 4.

2. The Bezier Curve

Consider $N+1$ points $\boldsymbol{b}_0 = (z_0, w_0)'$, $\boldsymbol{b}_1 = (z_1, w_1)', \ldots, \boldsymbol{b}_N = (z_N, w_N)'$ in R^2, then the Bezier curve based on the $N+1$ Bezier points $\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$ is defined as

$$\boldsymbol{b}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \sum_{j=0}^{n} \boldsymbol{b}_j B_{N,j}(t), \quad t \in (0,1) \tag{2.1}$$

where $B_{N,j}(t) = \binom{N}{j} t^j (1-t)^{N-j}$ is a binomial density function. See Figure 2.1 for illustration.

There are many properties about the Bezier curve. First, the Bezier curve has an endpoint interpolation property. That is, $\boldsymbol{b}_0$ and $\boldsymbol{b}_N$ are always on the curve $\boldsymbol{b}_t$. Next, $\boldsymbol{b}_t$ is symmetric. It does not matter if the Bezier points are labeled $\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$ or $\boldsymbol{b}_N, \boldsymbol{b}_{N-1}, \ldots, \boldsymbol{b}_0$. Another property is linear

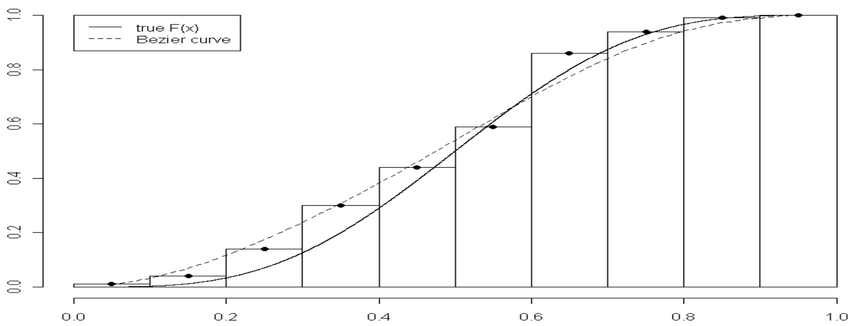

Figure 3.1. Cumulative histogram, the Bezier points at the middle of each histogram, and the corresponding Bezier curve ($m = 10$, $n = 100$ random numbers from a Beta$(4, 4)$ distribution)

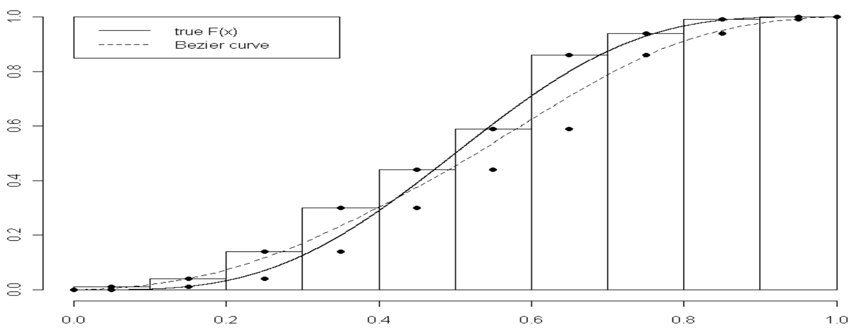

Figure 3.2. Cumulative histogram, the Bezier points, and the corresponding Bezier curve ($m = 10$, $n = 100$ random numbers from a Beta$(4, 4)$ distribution)

precision in that $\sum_{j=0}^{N}(j/N)\boldsymbol{B}_{N,j}(t) = t$ so that an initial straight line is reproduced. Finally, the first derivative of $\boldsymbol{b}_t$ with respect to t can be easily shown to be

$$\frac{d}{dt}\boldsymbol{b}(t) = N\sum_{j=0}^{N-1}(\boldsymbol{b}_{j+1} - \boldsymbol{b}_j)B_{N-1,j}(t). \tag{2.2}$$

See Farin (1990) for other properties of the Bezier curve. As an extension of the Bezier curve, if the Bezier points are in R^3, then the subsequent one is called the Bezier surface.

3. Selection of Bezier Points

3.1. Distribution function

3.1.1. Existing methods Let $X_1, \ldots, X_n$ be random sample from a distribution with density function f and distribution function F which is assumed to be continuous. Kim *et al.* (1999) proposed a Bezier curve smoothing technique to estimate F based on $X_1, \ldots, X_n$. In fact, the estimation of F is equivalent to estimating f. Let m be the number of intervals in the cumulative histogram based on $X_1, \ldots, X_n$. To estimate F, they considered locating the Bezier points at the middle of each rectangle in the cumulative histogram (see Figure 3.1); however, it underestimates (overestimates)

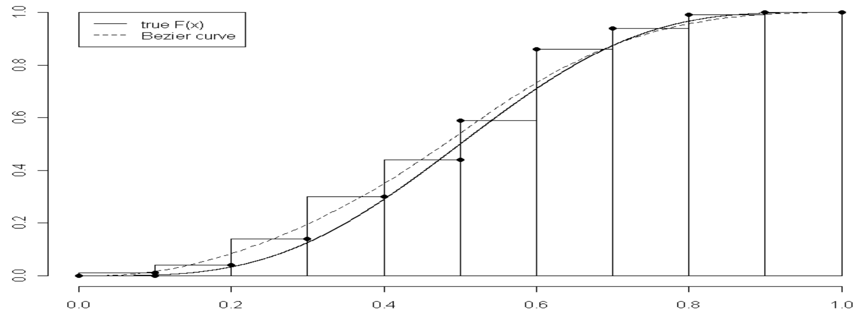

Figure 3.3. Cumulative histogram, the proposed Bezier points, and the corresponding Bezier curve ($m = 10$, $n = 100$ random numbers from a Beta$(4, 4)$ distribution)

when the Bezier points are convex (concave). Note that there are m Bezier points in this approach. To overcome the undesirable aspect, they suggested increasing the number of Bezier points by locating two points in each rectangle (see Figure 3.2), and the resulting number of the Bezier points is $2m + 4$. Let $\hat{F}_1$ and $\hat{F}_2$ be estimators of F based on m and $2m + 4$ Bezier points, respectively. Kim *et al.* (1999) argued that $\hat{F}_2$ is superior than $\hat{F}_1$ in the sense of the mean integrated square error(MISE) at the sacrifice of using more Bezier points. Also, they derived the asymptotic bias and variance of $\hat{F}_1$, and noted that the stochastic order of leading terms of the bias and the variance of $\hat{F}_2$ is the same as the kernel density estimator. Specifically, they showed that for $x \in (0, 1)$, the bias and variance of the Bezier curve density estimator $\hat{f}_B(x)$ are

$$\begin{aligned}\mathrm{Bias}\left(\hat{f}_B(x)\right) &= \frac{\left\{(1-2x)f^{'}(x) + x(1-x)f^{''}(x)\right\}}{4m} + o\left(m^{-1}\right),\\ \mathrm{Var}\left(\hat{f}_B(x)\right) &= \frac{\sqrt{m}}{n}\frac{1}{\sqrt{2\pi}}\sqrt{f(x)}\sqrt{x(1-x)} + o\left(m^{\frac{1}{2}}n^{-1}\right).\end{aligned}$$

Using this result, a theoretical choice of m can be obtained by minimizing the asymptotic mean integrated square error(AMISE) yielding $m_{opt} = (4c_1/c_2)^{2/5}n^{2/5}$ where

$$c_1 = \frac{1}{16}\int_0^1 \left\{(1-2x)f^{'}(x) + x(1-x)f^{''}(x)\right\}^2 dx$$

and

$$c_2 = \int_0^1 \frac{1}{\sqrt{2\pi}}\frac{f(x)}{\sqrt{x(1-x)}}dx.$$

3.1.2. Proposed method Recall that $\hat{F}_2$, based on $2m+4$ Bezier points, showed superior numerical performance than $\hat{F}_1$, based on m Bezier points. The main reason for using more Bezier points in $\hat{F}_2$ is to mitigate the overestimating (underestimating) aspect of $\hat{F}_1$ in the convex (concave) region. Here, we propose a method of choosing the Bezier points to remove the undesirable aspect of $\hat{F}_1$ based on the same number of Bezier points in $\hat{F}_1$. We propose a method of choosing the Bezier points that shows similar numerical performance based on half the number of the Bezier points used in $\hat{F}_2$. Based on this motivation, the proposed method is given in Figure 3.3. The number of Bezier points used here is $m + 2$, and the corresponding estimator is denoted by $\hat{F}_3$. The detailed method of locating the Bezier points for the compactly supported distribution, $[0, 1]$ is as follows.

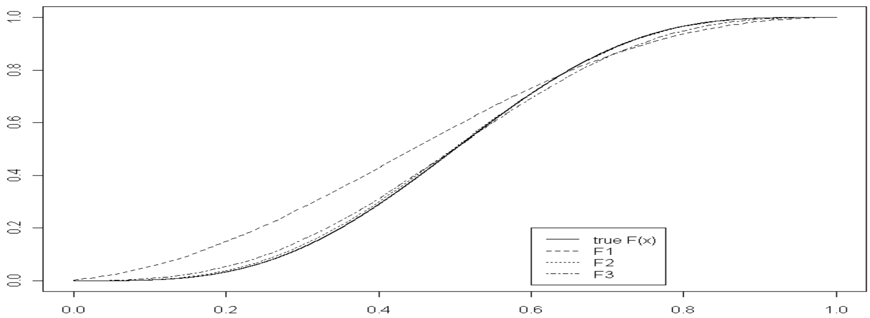

Figure 3.4. Mean of 100 replications of $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$ with the true $F(x)$ ($m = 10$, $n = 100$ random numbers from a Beta$(4,4)$ distribution)

First, consider the Bezier points in computing $\hat{F}_1$. If the cumulative histogram is less than 0.5, then move the middle point of each rectangle to the right side of the rectangle, and if the cumulative histogram is larger than 0.5, then move the middle point of each rectangle to the left side of the rectangle. Second, add two points at $(0,0)$ and $(1,1)$. Therefore, the number of Bezier points becomes $m+2$ (see Figure 3.3).

3.1.3. Numerical performance To compare the proposed method to the existing methods, we consider three distributions. (i) Beta$(4,4)$, symmetric with compact support; (ii) Beta$(3,2)$, asymmetric with compact support; and (iii) $N(0,1)$, symmetric with infinite support. When $n = 100$, the optimal number of Bezier points m are 19 and 11 for Beta$(4,4)$ and Beta$(3,2)$, respectively. We generate $n = 100$ random numbers from each distribution, and 100 replications are done. Table 3.1 lists MISE, IV (integrated variance), and ISB (integrated square bias) of 3 estimators $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$ in 3 distributions with $m = 10$ and $m = 20$ cases. We see that the MISE of the proposed estimator $\hat{F}_3$ is smaller than that of existing estimators, even though the number of Bezier points of $\hat{F}_3$ is half the number of Bezier points of $\hat{F}_2$. In addition, Table 3.1 shows that the estimators are quite sensitive to the number of Bezier points.

3.2. Regression function

3.2.1. A proposed method Consider a regression model

$$Y_i = f(x_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where the ϵ_i's with mean 0 and variance σ^2. For simplicity and computational convenience, assume that x_i's are uniformly distributed on $[0,1]$ and $n = mc$ for some positive integers m and c. Kim *et al.* (1999) suggested a method to generate Bezier points as follows; First, partition $[0,1]$ into m intervals with equal length, then compute bin estimator (sample mean of responses) based on c points in each interval. Let $\hat{f}_1(x)$ be the resulting Bezier curve, and it is an estimator of the regression function f. They showed that, for $x \in (0,1)$, the bias and variance of $\hat{f}_1(x)$ are

$$\begin{aligned}
\text{Bias}\left(\hat{f}_B(x)\right) &= \frac{1}{4m} x(1-x) f^{''}(x) + o\left(m^{-1}\right), \\
\text{Var}\left[\hat{f}_B(x)\right] &= \frac{\sqrt{m}}{n} \frac{\sigma^2}{\sqrt{2\pi x(1-x)}} + o\left(\frac{\sqrt{m}}{n}\right).
\end{aligned}$$

Table 3.1. MISE, IV (integrated variance), and ISB (integrated square bias) of 3 estimators $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$ in 3 distributions with $m = 10$ and $m = 20$ cases ($\times 10^4$).

m	Dist^n	Est.	MISE	IV	ISB
10	Beta(4, 4)	$\hat{F}_1$	70.7271	4.1207	66.6064
		$\hat{F}_2$	5.5291	5.1465	0.3826
		$\hat{F}_3$	5.5700	2.8430	2.7270
	Beta(3, 2)	$\hat{F}_1$	55.6008	7.0649	48.5359
		$\hat{F}_2$	12.3810	6.4706	5.9105
		$\hat{F}_3$	11.9429	4.2403	7.7026
	$N(0, 1)$	$\hat{F}_1$	90.8629	37.5698	53.2930
		$\hat{F}_2$	33.0930	32.2267	0.8662
		$\hat{F}_3$	31.4656	30.8253	0.6404
20	Beta(4, 4)	$\hat{F}_1$	181.3256	17.2074	164.1182
		$\hat{F}_2$	5.8053	5.7709	0.0344
		$\hat{F}_3$	5.0542	4.2858	0.7684
	Beta(3, 2)	$\hat{F}_1$	16.1021	3.0167	13.0854
		$\hat{F}_2$	10.1743	7.5687	2.6056
		$\hat{F}_3$	8.9522	6.5059	2.4463
	$N(0, 1)$	$\hat{F}_1$	158.5843	5.0732	153.5111
		$\hat{F}_2$	41.8546	40.3062	1.5484
		$\hat{F}_3$	38.9423	37.5090	1.4333

Note that the leading terms of both the bias and variance are the same as those of the local linear regression with $m = h^{-2}$, where h is a bandwidth.

As in density estimation the theoretical choice of m is obtained by minimizing the AMISE yielding $m_{opt} = (4c_1/c_2)^{2/5} n^{2/5}$, where

$$c_1 = \frac{1}{16} \int_0^1 \left\{ x(1-x) f^{''}(x) \right\}^2 dx$$

and

$$c_2 = \sigma^2 \int_0^1 \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x(1-x)}} dx = \sqrt{\frac{\pi}{2}} \sigma^2.$$

Since the bin estimator is quite sensitive to potential outliers, the estimator $\hat{f}_1(x)$ could be unstable when outliers exist. Motivated by this argument, we propose an estimator, denoted by $\hat{f}_2(x)$, using the sample median instead of the sample mean in each interval. Then, the estimator $\hat{f}_2(x)$ would be quite robust to outliers.

3.2.2. Numerical performance To compare the numerical performance of the proposed estimator $\hat{f}_2(x)$ to the existing estimator $\hat{f}_1(x)$, we consider 3 regression functions. The first function

$$f_A(x) = 2 \left\{ 20 \left(x - \frac{1}{2} \right)^3 - 3 \left(x - \frac{1}{2} \right) \right\}, \quad x \in (0, 1) \tag{3.1}$$

is sinusoidal type. The second function, given in Linhart and Zucchini (1986),

$$f_B(x) = \exp\left(0.1 + 0.02x^2\right), \quad x \in (1, 10) \tag{3.2}$$

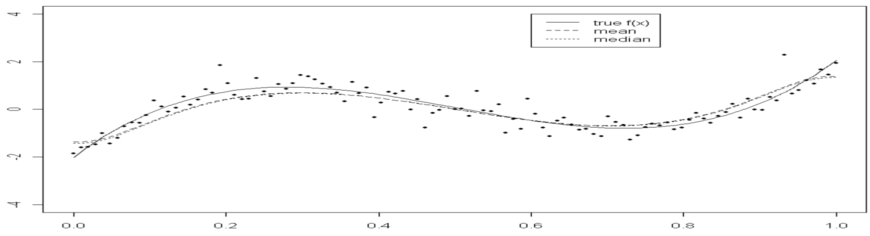

Figure 3.5. True $f_A(x)$, $\hat{f}_1$, and $\hat{f}_2$ with $n = 105$ random numbers, $m = 15$ and $\sigma = 0.5$ in the potential outlier model.

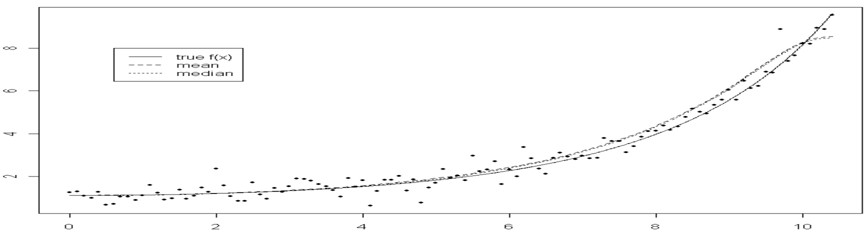

Figure 3.6. True $f_B(x)$, $\hat{f}_1$, and $\hat{f}_2$ with $n = 105$ random numbers, $m = 15$ and $\sigma = 0.5$ in the potential outlier model.

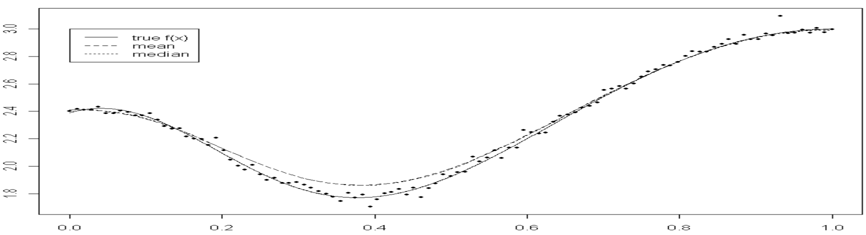

Figure 3.7. True $f_C(x)$, $\hat{f}_1$, and $\hat{f}_2$ with $n = 105$ random numbers, $m = 15$ and $\sigma = 0.5$ in the potential outlier model.

is monotone increasing. The third function, given in Wand and Jones (1995),

$$f_C(x) = 2\exp\left\{-\frac{x^2}{(0.3)^2}\right\} + 3\exp\left\{-\frac{(x-1)^2}{(0.7)^2}\right\} \tag{3.3}$$

is a convex type.

For the generation of random numbers, we first generate error terms from $N(0, \sigma^2)$ for several values of σ. By computing the optimal number of intervals m_{opt} for several values of σ and 3 types of a regression function, we consider two cases; $n = 105$ random numbers with $m = 15$ ($c = 7$) and $n = 104$ random numbers with $m = 26$ ($c = 4$).

To see the robustness of the proposed estimator, we generate random numbers with potential outliers. When we generate random numbers, we assume that $\epsilon_i \sim N(0, \sigma^2)$. Now, we replace kc-th random numbers by those from $\epsilon_i \sim N(0, 3\sigma^2)$ for $k = 1, \ldots, n/c$. Therefore, n/c number of

Table 3.2. MISE, IV (integrated variance), and ISB (integrated square bias) of $\hat{f}_1$ and $\hat{f}_2$ for 3 different regression functions $f_A(x)$, $f_B(x)$ and $f_C(x)$ when $n = 105$ ($m = 15$, $c = 7$) and $n = 104$ ($m = 26$, $c = 4$). (a) $\sigma = 0.5$, (b) $\sigma = 1$, (c) $\sigma = 1.5$, (d) $\sigma = 2$

	m	func.	Est.	MISE	IV	ISB
(a) $\sigma = 0.5$	15	$f_A(x)$	$\hat{f}_1$	6.5753	0.3024	6.2728
			$\hat{f}_2$	6.6225	0.6931	5.9295
		$f_B(x)$	$\hat{f}_1$	8.7504	0.0032	8.7472
			$\hat{f}_2$	8.6260	0.0070	8.6190
		$f_C(x)$	$\hat{f}_1$	0.2107	0.0016	0.2091
			$\hat{f}_2$	0.2069	0.0041	0.2028
	26	$f_A(x)$	$\hat{f}_1$	3.0462	0.3834	2.6628
			$\hat{f}_2$	3.0624	0.5096	2.5528
		$f_B(x)$	$\hat{f}_1$	8.0248	0.0038	8.0210
			$\hat{f}_2$	7.9800	0.0052	7.9748
		$f_C(x)$	$\hat{f}_1$	0.0768	0.0024	0.0744
			$\hat{f}_2$	0.0776	0.0031	0.0745
(b) $\sigma = 1$	15	$f_A(x)$	$\hat{f}_1$	11.2430	4.8391	6.4039
			$\hat{f}_2$	12.7983	7.1379	5.6604
		$f_B(x)$	$\hat{f}_1$	8.7313	0.0469	8.6844
			$\hat{f}_2$	8.6966	0.0762	8.6203
		$f_C(x)$	$\hat{f}_1$	0.2357	0.0263	0.2094
			$\hat{f}_2$	0.2568	0.0444	0.2124
	26	$f_A(x)$	$\hat{f}_1$	9.1878	6.1347	3.0532
			$\hat{f}_2$	10.1020	7.4779	2.6241
		$f_B(x)$	$\hat{f}_1$	8.1791	0.0613	8.1177
			$\hat{f}_2$	8.0880	0.0758	8.0122
		$f_C(x)$	$\hat{f}_1$	0.1150	0.0382	0.0768
			$\hat{f}_2$	0.1160	0.0433	0.0728
(c) $\sigma = 1.5$	15	$f_A(x)$	$\hat{f}_1$	30.5370	22.8440	7.6931
			$\hat{f}_2$	39.8663	34.7703	5.0960
		$f_B(x)$	$\hat{f}_1$	9.0867	0.2450	8.8417
			$\hat{f}_2$	8.9411	0.3571	8.5840
		$f_C(x)$	$\hat{f}_1$	0.3467	0.4221	0.2185
			$\hat{f}_2$	0.4294	0.2071	0.2223
	26	$f_A(x)$	$\hat{f}_1$	37.2780	0.1335	0.2131
			$\hat{f}_2$	40.7737	37.9493	2.8245
		$f_B(x)$	$\hat{f}_1$	8.6336	0.3410	8.2926
			$\hat{f}_2$	8.4251	0.3809	8.0442
		$f_C(x)$	$\hat{f}_1$	0.2457	0.1747	0.0710
			$\hat{f}_2$	0.2844	0.2146	0.0698
(d) $\sigma = 2$	15	$f_A(x)$	$\hat{f}_1$	82.7002	75.0986	7.6016
			$\hat{f}_2$	113.5430	109.0770	4.4660
		$f_B(x)$	$\hat{f}_1$	9.8590	0.7220	9.1370
			$\hat{f}_2$	9.6257	1.1109	8.5148
		$f_C(x)$	$\hat{f}_1$	0.6406	0.4221	0.2185
			$\hat{f}_2$	0.8705	0.6403	0.2302
	26	$f_A(x)$	$\hat{f}_1$	112.5729	107.7600	4.8130
			$\hat{f}_2$	123.1754	119.7928	3.3826
		$f_B(x)$	$\hat{f}_1$	9.4960	0.9815	8.5145
			$\hat{f}_2$	9.2900	1.2011	8.0889
		$f_C(x)$	$\hat{f}_1$	0.6360	0.5688	0.0672
			$\hat{f}_2$	0.7426	0.6754	0.0672

Table 3.3. MISE, IV (integrated variance), and ISB (integrated square bias) of $\hat{f}_1$ and $\hat{f}_2$ for 3 different regression functions $f_A(x)$, $f_B(x)$ and $f_C(x)$ when $n = 105$ ($m = 15$, $c = 7$) and $n = 104$ ($m = 26$, $c = 4$) in potential outlier models. (a) $\sigma = 0.5$, (b) $\sigma = 1$, (c) $\sigma = 1.5$, (d) $\sigma = 2$

	m	func.	Est.	MISE	IV	ISB
(a) $\sigma = 0.5$	15	$f_A(x)$	$\hat{f}_1$	7.0044	0.6556	6.3488
			$\hat{f}_2$	7.1021	0.7373	6.3649
		$f_B(x)$	$\hat{f}_1$	8.7382	0.0065	8.7318
			$\hat{f}_2$	8.5883	0.0081	8.5802
		$f_C(x)$	$\hat{f}_1$	0.2120	0.0034	0.2086
			$\hat{f}_2$	0.2071	0.0044	0.2027
	26	$f_A(x)$	$\hat{f}_1$	3.5898	1.1388	2.4511
			$\hat{f}_2$	3.3823	0.8213	2.5610
		$f_B(x)$	$\hat{f}_1$	7.9257	0.0124	7.9134
			$\hat{f}_2$	7.7736	0.0082	7.7654
		$f_C(x)$	$\hat{f}_1$	1.6861	1.5943	0.0918
			$\hat{f}_2$	0.0809	0.0045	0.0764
(b) $\sigma = 1$	15	$f_A(x)$	$\hat{f}_1$	15.1605	8.7257	6.4348
			$\hat{f}_2$	15.7447	8.6478	7.0968
		$f_B(x)$	$\hat{f}_1$	8.8420	0.1041	8.7379
			$\hat{f}_2$	8.5942	0.0861	8.5081
		$f_C(x)$	$\hat{f}_1$	0.2760	0.0544	0.2216
			$\hat{f}_2$	0.2410	0.0456	0.1953
	26	$f_A(x)$	$\hat{f}_1$	20.5873	18.2202	2.3671
			$\hat{f}_2$	14.6980	11.9568	2.7412
		$f_B(x)$	$\hat{f}_1$	8.7627	0.0065	8.7562
			$\hat{f}_2$	8.5883	0.0081	8.5802
		$f_C(x)$	$\hat{f}_1$	0.1706	0.1002	0.0704
			$\hat{f}_2$	0.1475	0.0656	0.0819
(c) $\sigma = 1.5$	15	$f_A(x)$	$\hat{f}_1$	57.3771	52.5445	4.8326
			$\hat{f}_2$	51.6538	43.6046	8.0492
		$f_B(x)$	$\hat{f}_1$	9.4263	0.5250	8.9013
			$\hat{f}_2$	8.8355	0.4336	8.4019
		$f_C(x)$	$\hat{f}_1$	0.4858	0.2725	0.2133
			$\hat{f}_2$	0.4041	0.2058	0.1983
	26	$f_A(x)$	$\hat{f}_1$	102.6580	100.3576	2.3004
			$\hat{f}_2$	64.3483	60.6602	3.6881
		$f_B(x)$	$\hat{f}_1$	9.1332	0.5277	8.6055
			$\hat{f}_2$	8.8947	0.3913	8.5034
		$f_C(x)$	$\hat{f}_1$	0.6204	0.5532	0.0672
			$\hat{f}_2$	0.4199	0.3285	0.0914
(d) $\sigma = 2$	15	$f_A(x)$	$\hat{f}_1$	167.1306	159.7076	7.4231
			$\hat{f}_2$	147.0441	137.2554	9.7887
		$f_B(x)$	$\hat{f}_1$	10.2672	1.6607	8.6065
			$\hat{f}_2$	9.7815	1.3646	8.4169
		$f_C(x)$	$\hat{f}_1$	1.1409	0.8697	0.2711
			$\hat{f}_2$	0.8975	0.6864	0.2111
	26	$f_A(x)$	$\hat{f}_1$	306.4358	301.1116	5.3242
			$\hat{f}_2$	198.5529	192.1103	6.4425
		$f_B(x)$	$\hat{f}_1$	10.9141	3.0111	7.9030
			$\hat{f}_2$	8.8649	1.9290	6.9359
		$f_C(x)$	$\hat{f}_1$	1.6861	1.5943	0.0918
			$\hat{f}_2$	1.1427	1.0352	0.1075

potential outliers are contained in the generated data.

We calculated MISE, IV (integrated variance), and ISB (integrated square bias) of $\hat{f}_1$ and $\hat{f}_2$ for 3 different regression functions $f_A(x)$, $f_B(x)$ and $f_C(x)$ based on 4 types of σ. MISE of $\hat{f}_2$ is lower than MISE of $\hat{f}_1$, especially in potential outlier models. We can identify $\hat{f}_2$ gives better numerical results than $\hat{f}_1$.

4. Concluding Remarks

Bezier curve smoothing is one of many useful and efficient nonparametric techniques to estimate the density function and the regression function. The asymptotic property of the Bezier curve smoothing as an estimator of density was shown to be the same as the kernel density estimator. Also, the asymptotic property of the Bezier curve smoothing as an estimator of regression function was shown to be the same as the local linear regression estimator. The choice of the Bezier points is very crucial in the Bezier curve smoothing because the choice of the smoothing parameter in nonparametric estimation is very important.

In this paper, we proposed novel methods for choosing the Bezier points to estimate the density function and the regression function. Through numerical studies, the proposed methods showed superior numerical performance over the existing methods of choosing Bezier points. To estimate the distribution function, we suggested choosing the right-most points where they are less than the median, and choosing the left-most points where they are larger than the median. To estimate the regression function, we proposed to choose the sample median instead of the sample mean for the bin estimator, and the proposed method provided superior numerical results than Kim *et al.* (1999) especially when there are potential outliers in data.

References

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

Bézier, P. (1977). Essay de Definition Numerique des Courbes et des Surfaces Experimentals. Ph.D. thesis, University of Paris VI.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Farin, G. E. (1990). *Curves and Surfaces for Computer Aided Geometric Design*, Academic Press Inc, London.

Kim, C. (1996). Nonparametric density estimation via the Bezier curve, ASA *Proceedings of the Section on Statistical Graphics*, 25–28.

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

Kim, C., Kim, W., Hong, C., Park, B.-U. and Jeong, M. (1999). Smoothing techniques via the Bezier curve, *Communications in Statistics - Theory and Methods*, **28**, 1577–1596.

Kim, C., Kim, W., Park, B.-U. and Lim, J. (2003). Bezier curve smoothing of the Kaplan-Meier estimator, *Annals of the Institute of Statistical Mathematics*, **55**, 359–367.

Linhart, H. and Zucchini, W. (1986). *Model Selection*, Wiley, New York.

Loader, C. (1999). *Local Regression and Likelihood*, Springer, London.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman and Hall, London.

Wasserman, L. (2006). *All of Nonparametric Statistics*, Springer, London.