Nonparametric Estimation of Distribution Function using Bezier Curve

Whasoo Bae[a], Ryeongah Kim[b], Choongrak Kim[1,b]

[a]Department of Data Science/Institute of Statistical Information, Inje University, Korea
[b]Department of Statistics, Pusan National University, Korea

Abstract

In this paper we suggest an efficient method to estimate the distribution function using the Bezier curve, and compare it with existing methods by simulation studies. In addition, we suggest a robust version of cross-validation criterion to estimate the number of Bezier points, and showed that the proposed method is better than the existing methods based on simulation studies.

Keywords: Bezier points, cross validation, mean integrated square error, smoothing techniques.

1. Introduction

Nonparametric methods are often used to estimate the density function of a distribution function because parametric methods are unrealistic and too restrictive to satisfy a prespecified distribution. For the nonparametric estimation of density function, the best references are Silverman (1986), Eubank (1988), Loader (1999) and Wasserman (2006). As nonparametric density estimation, kernel type smoothing is widely used. The Bezier curve (Bezier, 1977) smoothing (regarded as one of kernel-type approaches) is another nonparametric method to estimate density function and regression function. In computational graphics (especially for the computer-aided-geometric design) Bezier curve smoothing is popular; however, it rarely used in statistics. Kim (1996) applied a Bezier curve to density estimation for the first time in statistics and Kim *et al.* (1999) showed that estimators using Bezier curve smoothing in density estimation and regression function estimation provide similar results to kernel type smoothing as well as has the same asymptotic properties as classical kernel estimators. Subsequently, Kim *et al.* (2000) applied Bezier curve smoothing to estimation in the measurement error model. Subsequent works of the Bezier curve are the smoothing of a Kaplan-Meier estimator (Kim *et al.*, 2003), the smoothing of a bivariate Kaplan-Meier estimator (Bae *et al.*, 2005), and the selection of Bezier points in density estimation and regression (Kim and Park, 2012).

The smoothness of the Bezier curve is determined by the number of Bezier points and the position of each point. The curve will be smooth as the number of Bezier points increases and the resulting curve will be wiggly if the number of Bezier points become smaller. The choice and the number of optimal Bezier points were first considered by Kim *et al.* (1999) to estimate the density function and the regression function, however, it was a theoretical result when the true function is known. It is not easy to apply their results to a real data set. In this paper, we consider two issues in Bezier curve smoothing for the estimation of distribution function. First, we suggest an efficient method of choosing Bezier points to reduce the mean integrated squared error. This is a data-driven method
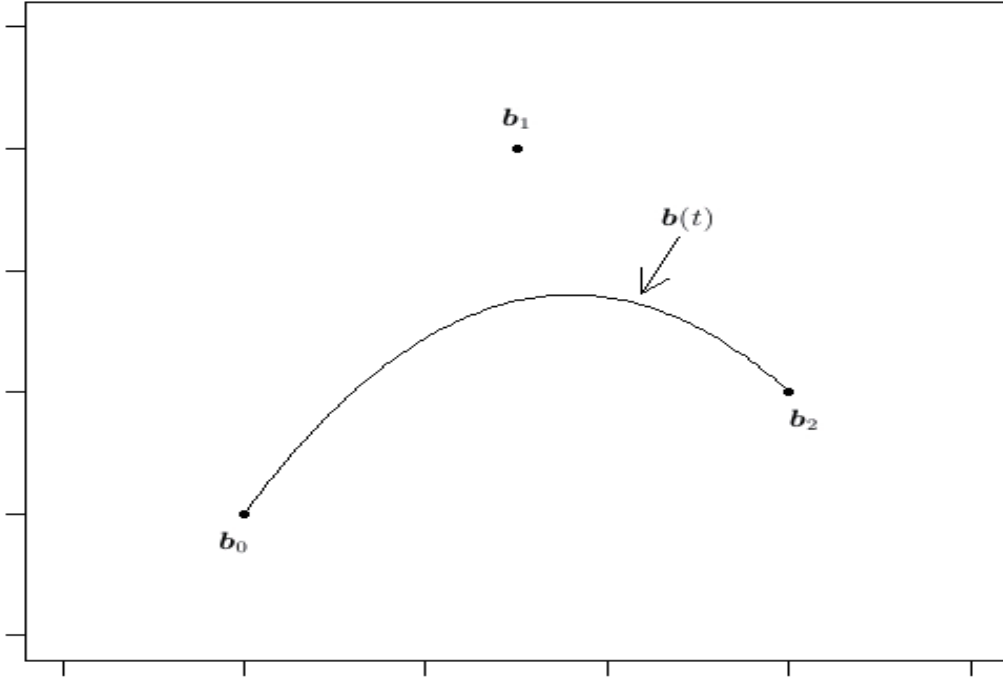
Figure 1: *Bezier curve $\boldsymbol{b}(t)$ based on 3 Bezier points $\boldsymbol{b}_0$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$.*

based an existing estimator by the convexity and concavity property. Existing methods are not useful when the true density is not symmetric, and the suggested method shows better numerical performance in symmetric and the skewed density. In addition, the suggested method outperforms recent results by Kim and Park (2012). Second, we consider estimating the number of Bezier points to estimate the distribution function. We suggest a simple leave-one-out cross-validation, and compare with it with existing cross validation methods such as least squares cross validation and the likelihood cross validation.

This paper is organized as follows; methods of selecting Bezier points in estimating distribution function are proposed in Chapter 2 and relevant numerical results are given. A leave-one-out cross-validation is proposed, and is compared with existing cross validation methods in Chapter 3. Concluding remarks are given in Chapter 4.

2. Selection of Bezier Points in Distribution Function

2.1. The Bezier curve

Let $\boldsymbol{b}_0 = (z_0, w_0)'$, $\boldsymbol{b}_1 = (z_1, w_1)', \ldots, \boldsymbol{b}_N = (z_N, w_N)'$ be $N + 1$ points 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

$$\boldsymbol{b}(t) = \left(\begin{array}{c} x(t) \\ y(t) \end{array} \right) = \sum_{j=0}^{n} \boldsymbol{b}_j B_{N,j}(t), \quad t \in (0, 1), \tag{2.1}$$
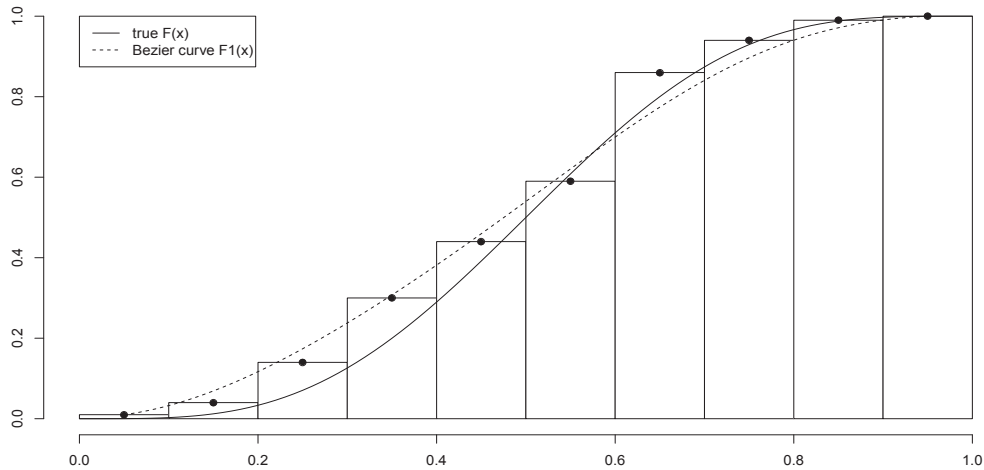
Figure 2: *Cumulative histogram, the Bezier points at the middle of each histogram, and the corresponding Bezier curve ($m = 10$, $n = 100$ random numbers from a Beta$(4, 4)$ distribution)*

where $B_{N,j}(t) = \binom{N}{j} t^j (1 - t)^{N-j}$ is a binomial density function (also called Bernstein polynomial in the computer graphics area) (see Figure 1 for illustration).

The Bezier curve has many desirable properties. First, $\boldsymbol{b}_0$ and $\boldsymbol{b}_N$ are always on the Bezier curve $\boldsymbol{b}_t$, and this property is called endpoint interpolation property. Second, $\boldsymbol{b}_t$ is symmetric, *i.e.*, it does not matter if the Bezier points are labeled $\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$ or $\boldsymbol{b}_N, \boldsymbol{b}_{N-1}, \ldots, \boldsymbol{b}_0$. Third, it has linear precision in that $\sum_{j=0}^{N} (j/N) \boldsymbol{B}_{N,j}(t) = t$, hence, 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. The Bezier curve can be extended to the higher dimensions, for example, if the Bezier points are in R^3, then the resulting one is called the Bezier surface.

2.2. 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 two Bezier curve smoothing techniques to estimate F based on $X_1, \ldots, X_n$.

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 2), however, it underestimates (overestimates) 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), and the resulting number of the Bezier points is $2m + 4$. The second point in each rectangle is located below the first point with the same height as the left rectangle. 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 much
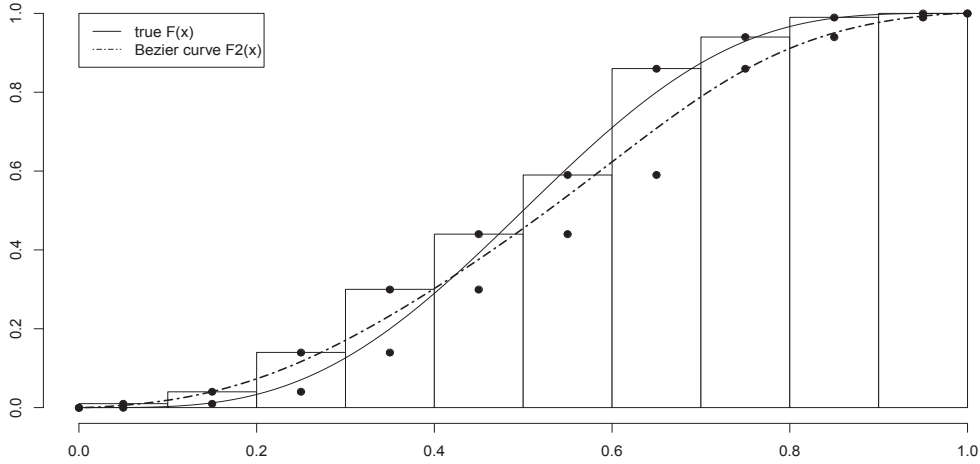
Figure 3: *Cumulative histogram, the Bezier points, and the corresponding Bezier curve ($m = 10, n = 100$ random numbers from a Beta$(4, 4)$ distribution)*

better than $\hat{F}_1$ in the sense of the mean integrated square error (MISE) at the sacrifice of using more Bezier points. Condition, they derived the asymptotic bias and variance of $\hat{F}_2$, and noted that the stochastic order of leading terms of the bias and the variance of $\hat{F}_2$ is the same as that of the kernel density estimator (Kim *et al.*, 1999).

Note that $\hat{F}_2$, based on $2m + 4$ Bezier points, showed better 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 overestimating (underestimating) aspect of $\hat{F}_1$ in the convex (concave) region. Recently, motivated by this reason, Kim and Park (2012) proposed a method to select 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 which show a similar numerical performance based on half the number of the Bezier points used in $\hat{F}_2$. The proposed method by Kim and Park (2012) is given in Figure 4. The number of Bezier points used here is $m + 2$, and the corresponding estimator is denoted by $\hat{F}_3$. The detailed method to locate the Bezier points for the compactly supported distribution $[0, 1]$ is as follows.

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 4).

2.3. A proposed method

The estimator $\hat{F}_3$, proposed method by Kim and Park (2012), performs well when the underlying distribution is symmetric because the middle point of each rectangle is moved to the right side of the rectangle if the cumulative histogram is less than 0.5 and the middle point of each rectangle is moved to the left side of the rectangle if the cumulative histogram is larger than 0.5. However, if the underlying distribution is far away from symmetric shape, then the performance of $\hat{F}_3$ is not good as mentioned in the concluding remarks of Kim and Park (2012). The reference value 0.5 in $\hat{F}_3$ should be changed to the inflection point of the distribution function to overcome this undesirable
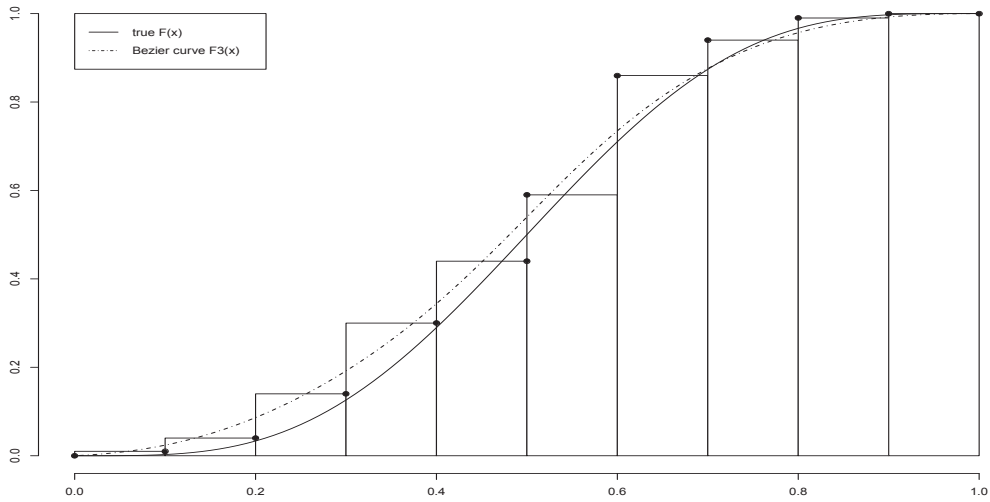
Figure 4: *Cumulative histogram, the Bezier points, and the corresponding Bezier curve (m = 10, n = 100 random numbers from a Beta(4, 4) distribution)*
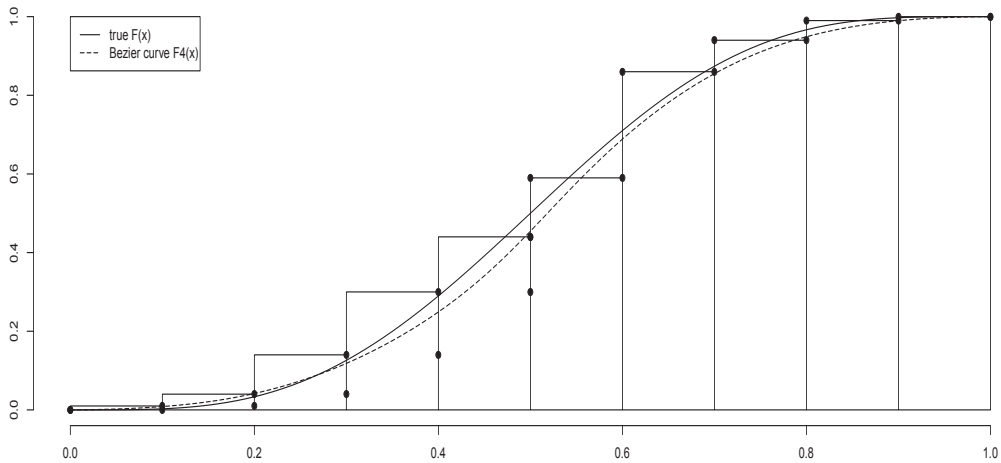

Figure 5: *Cumulative histogram, the proposed Bezier points, and the corresponding Bezier curve (m = 10, n = 100 random numbers from a Beta(4, 4) distribution)*

phenomenon. The inflection point is around 0.5 if the true underlying distribution is symmetric; however, the inflection point is also far from 0.5 if the true underlying distribution is far from a symmetric shape.

As a proposed estimator, we first estimate the inflection point using $\hat{F}_2$ and then apply both the ideas of $\hat{F}_2$ and $\hat{F}_3$. We denote the proposed estimator as $\hat{F}_4$ given in Figure 5. Note that Bezier points below the inflection point are moved to the right and those above the inflection point are moved to the left. Therefore, the number of Bezier points used in $\hat{F}_4$ is $2m + 4$.
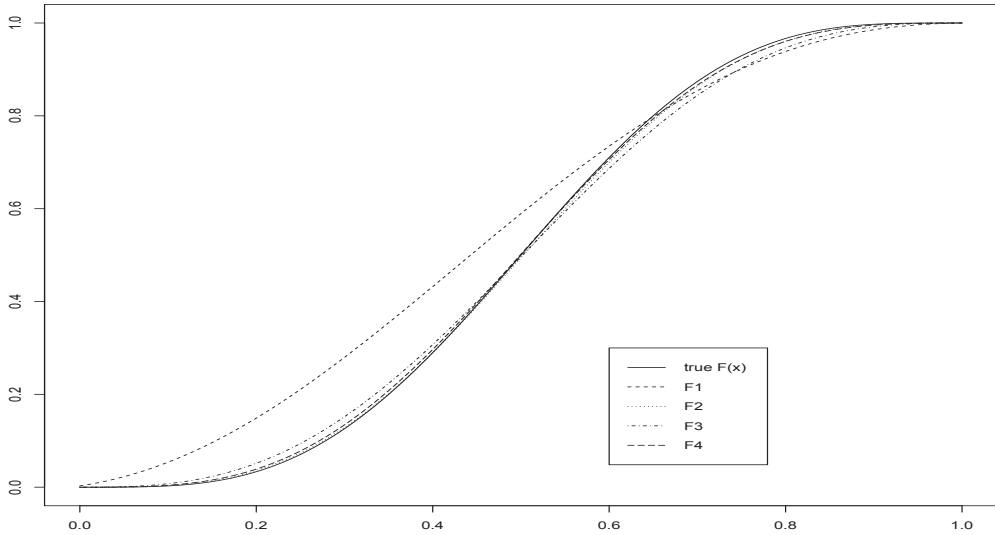
Figure 6: *Mean of 100 replications of $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$ and $\hat{F}_4$ with the true $F(x)$ ($m = 10$, $n = 100$ random numbers from a Beta$(4, 4)$ distribution)*

2.4. Numerical performance

We consider six distributions to compare the proposed method with existing methods. (i) Beta$(4, 4)$, symmetric with compact support; (ii) Beta$(5, 2)$, asymmetric with compact support; (iii) Exponential(1), heavily skewed to the right with infinite support; (iv) Chi-square(3), lightly skewed to the right with infinite support; (v) 0.6 Beta$(4, 8) + 0.4$ Beta$(10, 3)$, mixture distribution with compact support; and (vi) $0.75\, N(-1, 1) + 0.25\, N(1.5, 3)$, mixture distribution with infinite support;

We generate $n = 100$ random numbers from each distribution, and 100 replications are done. Optimal number of Bezier points m_{opt} for distributions (i)–(iv) is close to 18 and that for (v) and (vi) are close to 24. Recall that m_{opt} is determined by minimizing the MISE of $\hat{F}_1$ (Kim *et al.*, 1999). Table 1 lists MISE, IV (integrated variance), and ISB (integrated square bias) of 4 estimators $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$, and $\hat{F}_4$ in six distributions with $m = 10$ and $m = 16$ cases (smaller than m_{opt}). In addition, Table 2 list them using m_{opt}. The MISE of the proposed estimator $\hat{F}_4$ is smaller than that of existing estimators in all the distributions considered. Table 1 and Table 2 show that the proposed estimator is not sensitive to the number of Bezier points.

To compare the Bezier curve estimators of the distribution function with traditional nonparametric estimators (such as a kernel estimator), we note the work by Kim *et al.* (2006). The kernel distribution function estimator is given by

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} W\left(\frac{x - X_i}{h}\right),$$

where W is a cumulative kernel function and h is a bandwidth. Kim *et al.* (2006, p.597) showed that the MISE ($\times 10^4$) of $\hat{F}(x)$ when $n = 100$ under $N(0, 1)$ and $\chi^2(1)$ distributions are between 50 and 70. Note that these results are quite similar to $\hat{F}_1$ (see Table 1 and Table 2) and the proposed estimator $\hat{F}_4$ is superior to the kernel distribution function estimator.

Table 1: MISE, IV (integrated variance), and ISB (integrated square bias) of 4 estimators $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$ and $\hat{F}_4$ in 6 distributions with $m = 10$ and $m = 16$ cases ($\times 10^4$)

m	Dist^n	Est.	MISE	IV	ISB
10	Beta$(4, 4)$	$\hat{F}_1$	42.3145	3.9658	38.3487
		$\hat{F}_2$	19.9904	0.1570	19.8334
		$\hat{F}_3$	11.6017	0.0790	11.5226
		$\hat{F}_4$	4.5441	1.1960	3.3480
	Beta$(5, 2)$	$\hat{F}_1$	81.2447	2.4001	78.8446
		$\hat{F}_2$	9.9721	2.1875	7.7845
		$\hat{F}_3$	70.6866	1.8587	68.8278
		$\hat{F}_4$	5.5507	2.2079	3.3428
	Exp(1)	$\hat{F}_1$	22.3325	1.0256	21.3068
		$\hat{F}_2$	31.7133	0.8328	30.8804
		$\hat{F}_3$	44.3081	0.7312	43.5768
		$\hat{F}_4$	9.0389	0.1995	8.8393
	Chi(3)	$\hat{F}_1$	108.2880	25.0671	83.2207
		$\hat{F}_2$	26.7206	19.5307	7.1898
		$\hat{F}_3$	182.2250	18.2814	163.9436
		$\hat{F}_4$	21.7396	0.0411	21.6985
16	0.6 Beta$(4, 8)$ + 0.4 Beta$(10, 3)$	$\hat{F}_1$	11.3019	0.5954	10.7064
		$\hat{F}_2$	3.5562	0.5554	3.0008
		$\hat{F}_3$	15.6011	0.5049	15.0962
		$\hat{F}_4$	2.5781	0.1389	2.4391
	0.75$N(-1, 1)$ + 0.25$N(1.5, 0.3)$	$\hat{F}_1$	30.0859	1.3722	28.7137
		$\hat{F}_2$	3.4620	0.5152	2.9467
		$\hat{F}_3$	37.1376	1.1670	35.9706
		$\hat{F}_4$	1.6289	0.1679	1.4610

Table 2: MISE, IV (integrated variance), and ISB (integrated square bias) of 4 estimators $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$ and $\hat{F}_4$ in 6 distributions with $m = 18$ and $m = 24$ cases ($\times 10^4$).

m	Dist^n	Est.	MISE	IV	ISB
18	Beta$(4, 4)$	$\hat{F}_1$	17.0616	1.1176	15.9440
		$\hat{F}_2$	12.1215	1.1302	10.9913
		$\hat{F}_3$	6.0909	0.9611	5.1297
		$\hat{F}_4$	5.5910	0.3041	5.2868
	Beta$(5, 2)$	$\hat{F}_1$	27.2197	4.2307	22.9889
		$\hat{F}_2$	6.7030	4.1513	2.5516
		$\hat{F}_3$	24.1232	3.6796	20.4435
		$\hat{F}_4$	6.2914	4.0712	2.2201
	Exp(1)	$\hat{F}_1$	10.7300	1.1883	9.5416
		$\hat{F}_2$	11.9358	1.1110	10.8247
		$\hat{F}_3$	23.4198	1.0045	22.4152
		$\hat{F}_4$	5.9207	0.2974	5.6233
	Chi(3)	$\hat{F}_1$	81.4216	27.7514	53.6701
		$\hat{F}_2$	40.6532	24.7002	15.9529
		$\hat{F}_3$	123.6186	23.6800	99.9385
		$\hat{F}_4$	10.3242	0.0420	10.2821
24	0.6 Beta$(4, 8)$+ 0.4 Beta$(10, 3)$	$\hat{F}_1$	4.7377	0.5635	4.1741
		$\hat{F}_2$	2.5864	0.5498	2.0365
		$\hat{F}_3$	4.2053	0.5083	3.6970
		$\hat{F}_4$	1.1440	0.1778	0.9661
	0.75$N(-1, 1)$+ 0.25$N(1.5, 0.3)$	$\hat{F}_1$	16.0855	0.5025	15.5830
		$\hat{F}_2$	4.0278	0.5185	3.5093
		$\hat{F}_3$	18.2345	0.4510	17.7835
		$\hat{F}_4$	1.7405	0.4985	1.2419

3. Estimation of the Number of Bezier Points

The most crucial problem in nonparametric estimation is how to choose a smoothing parameter, and the number of the Bezier points corresponds to the smoothing parameter in Bezier curve smoothing. As far as we know there has been no study on choosing the number of the Bezier points. In this chapter we propose leave-one-out cross validation method and compare it with some existing methods, which are often used in kernel density estimation, such as the least squares cross validation and the likelihood cross validation.

3.1. Existing methods

The least squares cross validation (LSCV) is an automatic method of choosing smoothing parameter, and it has been studied by many authors (Rudemo, 1982; Bowman, 1984; Bowman *et al.*, 1984; Hall, 1983; Stone, 1984). Let $\hat{f}$ be an estimator of a density f, then it is reasonable to minimize the integrated square error, *i.e.*,

$$\int \left(\hat{f} - f \right)^2 = \int \hat{f}^2 - 2 \int \hat{f} f + \int f^2,$$

and equivalently, minimizing

$$R\left(\hat{f} \right) = \int \hat{f}^2 - 2 \int \hat{f} f.$$

Since f is unknown, it is not possible to minimize $R(\hat{f})$. The LSCV minimizes the estimator of $R(\hat{f})$ using the data themselves. Consequently, the estimator of $R(\hat{f})$ is given by

$$\text{LSCV}(m) = \int \hat{f}^2 - \frac{2}{n} \sum_i \hat{f}_{(i)}(X_i),$$

where $\hat{f}_{(i)}$ is the density estimate based on $n - 1$ observations after deleting X_i with m Bezier points. The idea of LSCV is to minimize LSCV(m) over the number of the Bezier points m.

On the other hand, the log-likelihood cross validation (LCV), suggested by Stone (1974) and Geisser (1975), maximizes

$$\text{LCV}(m) = \frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_{(i)}(X_i).$$

The estimate of m is given by minimizing LCV(m) over the number of the Bezier points m.

3.2. A proposed method

A leave-one-out CV (LOOCV) is a special case of leave-k-out CV, and a similar method to the LSCV. In fact, the leave-one-out CV in estimating the distribution function can be defined as

$$\text{LOOCV}(m) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{F}(X_i) - \hat{F}_{(i)}(X_i) \right)^2,$$

where $\hat{F}_{(i)}(X_i)$ is the distribution function estimate at X_i based on $n - 1$ observations after deleting X_i with m Bezier points. The estimate of m is given by minimizing LOOCV(m) over the number of the

Table 3: Estimation of the number of Bezier points by LSCV, LCV, LOOCV and ACV for Beta(5, 2) with $m_{opt} = 18$ and $0.6\,\text{Beta}(4, 8) + 0.4\,\text{Beta}(10, 3)$ with $m_{opt} = 24$ evaluated for four estimators $\hat{F}_1, \hat{F}_2, \hat{F}_3$ and $\hat{F}_4$.

Dist^n		$\hat{F}_1$	$\hat{F}_2$	$\hat{F}_3$	$\hat{F}_4$
	LSCV	25	19	19	21
	LCV	24	17	16	22
Beta(5, 2)	LOOCV	18	19	17	19
	ACV	18	19	17	18
	LSCV	27	23	17	26
$0.6\,\text{Beta}(4, 8) +$	LCV	29	29	25	27
$0.4\,\text{Beta}(10, 3)$	LOOCV	26	25	23	25
	ACV	24	23	24	24

Bezier points m. The LOOCV uses L_2 norm, and can be regarded as a sample version of LSCV. Here we propose L_1 version of LOOCV because it is robust to outlying observations. We call it absolute CV (ACV), and it can be written as

$$\text{ACV}(m) = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{F}(X_i) - \hat{F}_{(i)}(X_i) \right|.$$

3.3. Numerical performance

To see the numerical performance of four types of cross validation, we estimate the number of Bezier points m by LSCV, LCV, LOOCV, and ACV for Beta(5, 2) with $m_{opt} = 18$ and $0.6\,\text{Beta}(4, 8) + 0.4\,\text{Beta}(10, 3)$ with $m_{opt} = 24$ evaluated for four estimators $\hat{F}_1, \hat{F}_2, \hat{F}_3$ and $\hat{F}_4$. Table 3 shows that ACV is quite consistent and accurate compared to LSCV, LCV, and LOOCV.

4. Concluding Remarks

Bezier curve smoothing is a nonparametric techniques to estimate density function and the regression function, and it shares the same property with the kernel density estimator in the asymptotic properties. One of the most important issues is the estimation of the smoothing parameter in nonparametric estimation. In the same spirit, the selection of Bezier points and the estimation of the number of Bezier points in Bezier curve smoothing is very important in Bezier curve smoothing.

In this paper, we proposed a novel method to select Bezier points and a simple method to estimate the number of Bezier points when estimating the distribution function. In addition, we compared the proposed methods with existing methods. Through numerical studies, the proposed methods showed a better numerical performance than existing methods. For the choice of Bezier points, the proposed method is quite robust to different types of distributions which are either symmetric or skewed. In addition, L_1 version of the leave-one-out cross validation to estimate the number of Bezier points is quite accurate and stable for various types of distributions compared to least square cross validation, the likelihood cross validation, and L_2 version of the leave-one-out cross validation.

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.

Bezier, P. (1977). Essay de definition numerique des courbes et des surfaces experimentals. Ph.D. dissertation, University of Paris VI.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates, *Biometrika*, **71**, 353–360.

Bowman, A. W., Hall, P. and Titterington, D. M. (1984). Cross-validation in nonparametric estimation of probabilities and probability densities, *Biometrika*, **71**, 341–351.

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

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

Geisser, S. (1975). The predictive sample reuse method with applications, *Journal of the American Statistical Association*, **70**, 320–328.

Hall, P. (1983). Large sample optimality of least squares cross-validation in density estimation, *The Annals of Statistics*, **4**, 1156–1174.

Kim, C. (1996). Nonparametric density estimation via the Bezier curve, *Proceedings of the Section on Statistical Graphics of the American Statistical Association*, 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.

Kim, C., Kim, S., Park, M. and Lee, H. (2006). A bias reducing technique in kernel distribution function estimation, *Computational Statistics*, **21**, 589–601.

Kim, C. and Park, J.-H. (2012). On the selection of Bezier points in Bezier curve smoothing, *Korean Journal of Applied Statistics*, **25**, 1049–1058.

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

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators, *Scandinavian Journal of Statistics*, **9**, 65–78.

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

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates, *The Annals of Statistics*, **12**, 1285–1297.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society, Series B*, **36**, 111–147.

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