Journal of Statistical
Theory & Methods
1995, Vol. 6, pp 31~37

A Kernel Approach to the Goodness of Fit Problem[1]

Daehak Kim[2]

Abstract We consider density estimates of the usual type generated by a kernel function. By using the limit theorems for the maximum of normalized deviation of the estimate from its expected value, we propose to use data dependent bandwidth in the tests of goodness of fit based on these statistics. Also a small sample Monte Carlo simulation is conducted and proposed method is compared with Kolmogorov-Smirnov test.

Keywords : Goodness of fit, Kernel density, bandwidth

1. Introduction

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables with continuous density function $f(x)$. We consider the well known kernel density estimates known kernel density estimates

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h}) \tag{1}$$

where h is bandwidth that tends to 0 as $n \to \infty$ but $nh \to \infty$ and K is kernel function.

Usually, we have a great concern for the underlying distribution of given data. Main reason is to check the validity of given assumptions on the data. If the assumptions are not appropriate all results obtained from the statistical analysis are useless. For these goodness of fit problem, Kolmogorov(1933) suggested an α level test. It is well known and called Kolmogorov-Smirnov one sample test. It use empirical distribution function.

Nowadays, with the aids of fast modern computers, kernel density estimator can

1 This Research was supported by special grant, Catholic University of Taegu Hyosung, 1994
2 Department of Statistics, Catholic University of Taegu Hyosung, Gyungsan, Kyungbuk, Korea

be calculated without heavy computational difficulties and is widely applicable. Bickel and Rosenblatt(1973) considered global measures of how good $\hat{f}_n(x)$ is as an estimate of $f(x)$ and proved the asymptotic distribution of maximum of normalized deviation of the kernel estimates from the true density

$$\sup_{a\le x\le b}\left|\hat{f}_n(x)-f(x)\right|/(f(x))^{1/2} \tag{2}$$

By the asymptotic distribution of (2), inference containing goodness of fit problem based on kernel density estimate are possible and compared with that of Kolmogorov-Smirnov method. But bandwidth h selection problem is remained in their global measure. So we propose to use data dependent bandwidth in (2) based on cross-validation. As has been proposed by Bickel and Rosenblatt(1973), we conduct a small sample Monte Carlo simulation. Empirical powers for various alternatives are evaluated and compared with Kolmogorov-Smirnov method.

2. Approximation of global measure

Let's consider the following convenient assumptions called A1-A4 for the result on the global measure of the absolute deviation.

A.1 The kernel function $K(\cdot)$ also assigns mass 1 to the line and either

(a) vanishes outside an interval [-A,A] and is absolutely continuous with $K'(\cdot)$

(b) is absolutely continuous on $\boldsymbol{R}$ with $K'(\cdot)$ such that $\int|K(t)|^k dt<\infty, k=1,2$

A.2 The density f is continuous, positive and bounded

A.3 The function $f^{1/2}$ is absolutely continuous and its derivative is bounded in absolute value. Moreover

$$\int_{[|z|\ge 3]}|z|^{3/2}[\log\log|z|]^{1/2}[|K'(z)|+|K(z)|]dz<\infty$$

A.4 The second derivative f'' of f exists and is bounded. Moreover $K(\cdot)$ is symmetric (about 0) and $z^2K(z)$ is integrable.

Now, consider the stochastic process

$$Y_n(x)=(nh)^{1/2}[\hat{f}_n(x)-E\hat{f}_n(x)]/f(x)^{1/2} \tag{3}$$

and let

$$\tilde{M}_n=\sup\{|Y_n(t)|:0\le t\le 1\} \tag{4}$$

There is no loss of generality in considering [0,1] rather than any other interval on which the density is bounded away from 0 and ∞. We introduce the following theorem of Bickel and Rosenblatt(1973).

Theorem 1. (Bickel and Rosenblatt)
Let $K(\cdot)$ satisfy assumptions A1-A3 and $h = n^{-\delta}$ $(0 < \delta < 1/2)$. Then

$$P[(2\delta \log n)^{1/2}(\frac{\tilde{M}_n}{\lambda^{1/2}} - d_n) < x] \to e^{-2e^{-x}} \tag{5}$$

where $\lambda = \int K^2(t)dt$,

$$d_n = \begin{cases} (2\delta \log n)^{1/2} + \dfrac{1}{(2\delta \log n)^{1/2}} \{\log \dfrac{K_1}{\pi^{1/2}} - \dfrac{1}{2}[\log \delta + \log \log n]\} & \text{if } K_1 > 0 \\ (2\delta \log n)^{1/2} + \dfrac{1}{(2\delta \log n)^{1/2}} \log(K_2^{1/2} / p) & \text{otherwise} \end{cases}$$

where $K_1 = \frac{1}{2}(K^2(A) + K^2(-A)) / \lambda$ and $K_2 = \frac{1}{2} \int [K'(t)]^2 dt / \lambda$

Remark. The kernel function

$$K(t) = \begin{cases} 0.5 & |t| \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

falls under the first case (a), while the optimal kernel function of Epanechnikov

$$K(t) = \begin{cases} 0.75(1 - t^2/5)/\sqrt{5} & |t| \le \sqrt{5} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

(1969) falls under the second case (b).

3. Applications

To test the null hypothesis $H: f = f_0$, it is natural to compute $\tilde{M}_n$ with $f = f_0$ and reject for large values of the statistics. According to the theorem 1, to obtain approximate level α we should use as cutoff point,

$$c(\alpha) = -[\log|\log(1-\alpha)| - \log 2] \frac{\lambda^{1/2}}{(2\delta \log n)^{1/2}} + d_n \lambda^{1/2} \tag{8}$$

Under some assumptions the same cutoff point may be used for testing composite hypothesis of the form $H: f = f_0(\cdot, \theta)$ where θ is unknown parameter.

When computing $\tilde{M}_n$, we propose to use a data dependent bandwidth $\hat{h}$ which replaces $h = n^{-\delta}$ in $\tilde{M}_n$. Data dependent bandwidth selection methods are crucial problem in modern kernel estimation area. Least squares cross-validation method suggested by Rudemo(1982) and Bowman(1984) is a strong candidate. The idea of cross-validation is to minimize the score $M_0(h)$ over h

$$M_0(h) = \int \hat{f}^2 - 2n^{-1} \sum_i \hat{f}^2(X_i)$$

where $f_{-i} = (n-1)^{-1} h^{-1} \sum_{j \neq i} K\{h^{-1}(x - X_j)\}$. Stone(1984) had shown that cross-validatory choice of bandwidth achieves the best possible choice of smoothing parameter in the sense of minimizing the integrated squared error. By using $\hat{h}$ based on cross-validation method, the cutoff point (7) becomes

$$c*(\alpha) = -[\log|\log(1-\alpha)| - \log 2] \frac{\lambda^{1/2}}{(-2\log \hat{h})^{1/2}} + d_n \lambda^{1/2} \quad (9)$$

4. Monte carlo Simulation

In order to compare the performance of proposed method based on kernel density estimator with Kolmogorov-Smirnov test based on empirical distribution function, Monte Carlo simulation was carried out for small sample cases. We consider standard normal distribution as a null distribution in all cases. Various alternative distributions such as cauchy, uniform, t distribution with 1 degree of freedom were considered. Though the shape of distribution is not identical to normal distribution, the mean and variance are close to the null hypothesis. Normal distribution with mean μ and standard deviation σ was also considered as an alternative distribution ($\mu = 0, 0.5, 0.75, 1.0$ and $\sigma^2 = 1,2,3$ respectively). It will be interesting to compare the two tests when a small changes happen in mean and variance.

For the reason of computational efficiency we used the Epanechnikov kernel. Empirical powers were computed based on 1000 replications with 4 sample size $n = 10, 20, 30, 40$. All computation was carried out by Sun-10 Workstation and random numbers were generated through IMSL(1991). The numerical integration nedeed to compute

$$E\hat{f}_n(x) = \int K(\frac{x-y}{h}) f(y) dy$$

was done using evenly spaced grid of 100 point for the appropriate range by using Simpson's method.

Table 1. Empirical power with $n = 10,\ \alpha = 0.05$

$n = 10$		$\mu = 0$	$\mu = 0.5$	$\mu = 0.75$	$\mu = 1$
δ	σ^2	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol
0.05	1	.09 .08 .04	.47 .55 .28	.72 .79 .54	.89 .92 .78
	2	.76 .93 .12	.89 .96 .41	.97 .98 .72	.99 .99 .88
	3	.91 .98 .16	.97 .99 .51	.99 .99 .77	.99 1.0 .91
0.1	1	.08 .06 .04	.51 .62 .26	.76 .86 .52	.90 .94 .74
	2	.80 .96 .10	.89 .97 .44	.97 .99 .66	.99 .99 .88
	3	.94 .99 .17	.97 .99 .51	.99 .99 .75	1.0 1.0 .93
0.2	1	.08 .08 .06	.52 .69 .28	.71 .86 .53	.89 .97 .77
	2	.77 .97 .12	.89 .99 .41	.96 .99 .67	.99 1.0 .88
	3	.91 1.0 .19	.98 1.0 .49	.99 .99 .77	.99 1.0 .92

Results are appeared in Table 1 to 5. From Table 1 to Table 4, each value represent the empirical power when samples were drawn from $N(\mu,\sigma^2)$. We choose delta 0.05,0.1 and 0.2 for $\tilde{M}_n$ where $h = n^{-\delta}$. Kolmogorov-Smirnov results are denoted by "kol". The proposed method was denoted $\hat{h}_{cv}$.

Table 2. Empirical power with $n = 20,\ \alpha = 0.05$

$n = 20$		$\mu = 0$	$\mu = 0.5$	$\mu = 0.75$	$\mu = 1$
δ	σ^2	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol
0.05	1	.08 .07 .06	.62 .76 .48	.91 .94 .84	.99 .99 .97
	2	.91 .99 .14	.98 .99 .69	.99 1.0 .93	1.0 1.0 .99
	3	.99 1.0 .30	.99 1.0 .79	1.0 1.0 .95	1.0 1.0 .99
0.1	1	.07 .06 .05	.70 .85 .48	.92 .97 .83	.99 .99 .97
	2	.93 .99 .17	.98 1.0 .66	1.0 1.0 .94	1.0 1.0 .99
	3	.99 1.0 .26	.99 1.0 .80	1.0 1.0 .96	1.0 1.0 .99
0.2	1	.08 .05 .06	.66 .89 .50	.91 .98 .83	.99 1.0 .96
	2	.90 .99 .17	.98 .99 .65	.99 1.0 .93	1.0 1.0 .98
	3	.98 1.0 .27	.99 1.0 .77	1.0 1.0 .98	1.0 1.0 .99

Empirical powers of proposed method and Bickel and Rosenblatt's method were uperior than Kolmogorov-Smirnov method in all cases. Particularly, when small changes happen in variance, empirical power of Kolmogorov-Smirnov method was very low, but the other two methods were very high in all cases. It is notable result.

Table 3. Empirical power with $n = 30,\ \alpha = 0.05$

$n = 30$		$\mu = 0$	$\mu = 0.5$	$\mu = 0.75$	$\mu = 1$
δ	σ^2	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol
0.05	1	.05 .06 .05	.81 .90 .65	.98 .99 .95	.99 1.0 .99
	2	.96 .99 .17	.99 1.0 .84	1.0 1.0 .99	1.0 1.0 .99
	3	.99 .99 .43	1.0 1.0 .94	1.0 1.0 .99	1.0 1.0 1.0
0.1	1	.06 .05 .05	.79 .95 .66	.98 .99 .95	1.0 1.0 .99
	2	.95 .99 .18	.99 1.0 .85	1.0 1.0 .98	1.0 1.0 1.0
	3	.99 1.0 .42	1.0 1.0 .94	1.0 1.0 .99	1.0 1.0 1.0
0.2	1	.06 .06 .05	.77 .94 .66	.97 .99 .94	.99 1.0 .99
	2	.94 .99 .19	.99 1.0 .84	1.0 1.0 .98	1.0 1.0 .99
	3	.99 1.0 .38	.99 1.0 .93	1.0 1.0 .99	1.0 1.0 1.0

Table 5 represent empirical powers when samples were drawn from the three distributions which are different from the null distribution. We considered uniform$(-\sqrt{3},\sqrt{3})$, student's t distribution with 1 degrees of freedom and cauchy distribution. In this case, the proposed method and Bickel and Rosenblatt's method are similar in their perfomance and results were better than Kolmogorov-Smirnov method. In cauchy distribution, proposed method does not work well as well as Komogorov-Smirnov method but Bickel and Rosenblatt's method works well. It seems that large bandwidth is selected in cauchy distribution due to the large data variation.

Table 4. Empirical power with $n = 40,\ \alpha = 0.05$

$n = 40$		$\mu = 0$	$\mu = 0.5$	$\mu = 0.75$	$\mu = 1$
δ	σ^2	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol
0.05	1	.05 .06 .05	.87 .96 .79	.99 .99 .98	1.0 1.0 .99
	2	.98 .99 .23	1.0 1.0 .93	1.0 1.0 .99	1.0 1.0 1.0
	3	1.0 1.0 .52	1.0 1.0 .97	1.0 1.0 1.0	1.0 1.0 1.0
0.1	1	.06 .06 .05	.89 .98 .78	.99 1.0 .99	1.0 1.0 1.0
	2	.99 1.0 .21	1.0 1.0 .92	1.0 1.0 .99	1.0 1.0 1.0
	3	1.0 1.0 .52	1.0 1.0 .98	1.0 1.0 1.0	1.0 1.0 1.0
0.2	1	.05 .05 .05	.87 .98 .76	.99 .99 .98	1.0 1.0 1.0
	2	.96 .99 .23	1.0 1.0 .93	1.0 1.0 .99	1.0 1.0 1.0
	3	.99 1.0 .51	1.0 1.0 .98	1.0 1.0 1.0	1.0 1.0 1.0

Table 5. Empirical power with $\alpha = 0.05$
(H_0:$N(0,1)$ vs H_1:Not H_0)

distri-bution	N	n =10	n = 15	n = 20
	δ	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol	h $\hat{h}_{cv}$ kol
uniform $(-\sqrt{3},\sqrt{3})$	0.05	.86 .77 .77	.92 .78 .80	.92 .78 .78
	0.1	.95 .91 .93	.96 .96 .94	.97 .97 .93
	0.2	.98 .97 .98	.99 .96 .98	.99 .98 .98
t(1)	0.05	.39 .29 .13	.50 .45 .12	.58 .57 .12
	0.1	.41 .37 .15	.58 .54 .14	.59 .61 .14
	0.2	.53 .47 .19	.63 .63 .19	.66 .69 .20
cauchy	0.05	.36 .05 .13	.47 .09 .12	.57 .15 .11
	0.1	.43 .13 .16	.58 .14 .17	.60 .17 .16
	0.2	.52 .14 .20	.64 .18 .21	.69 .22 .19

References

Bickel, P. J. and Rosenblatt, M.(1973) On some global measures of deviations of density function estimates. *The Annals of Statistics, vol. 1, No 6, 1071-1095.*

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

Epanechnikov, V. A.(1969) Nonparametric estimation of multidimensional probability density. *Theory of Probability and its applications, 14, 153-158.*

Kolmogorov, A. N.(1993) Sulla determinazion empirica di una egge di distribuzione. *Giorn. Inst. Ital. Att., 4, 83-91.*

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

Silverman, B.W.(1985) *Density estimation for statistics and data analysis.* Chapman and Hall, London.

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