Testing Normality and Bandwith Estimation Using Kernel Method For Small Sample Size

Netti Herawati & Khoirin Nisa
Department of Mathematics FMIPA University of Lampung

ABSTRACT

This article aimed to study kernel method for testing normality and to determine the density function based on curve fitting technique (density plot) for small sample sizes. To obtain optimal bandwith we used Kullback-Leibler cross validation method. We compared the result using goodness of fit test by Kolmogorof Smirnov test statistics. The result showed that kernel method gave the same performance as Kolmogorof Smirnov for testing normality but easier and more convinient than Kolmogorof Smirnov does.

Keywords: Normality, kernel method, bandwith, Kolmogorof Smirnov test

INTRODUCTION

In many statistical inference the statistics test is ussually based on assumption of normal distribution. A check of normality assumption could be made by plotting a histogram of the residual. If the NID(0, σ^2) assumption on the errors is satisfied, this plot should look like a sample from normal distribution centered at zero. Unfortunately with small samples, considerable fluctuation in shape of histogram often occurs, so the appearance of moderate departure from normality does not necessarily imply a serious violation of assumption. However, gross deviations from normality are potentially serious and required further analysis (Montgomery 2005).

Another way to test normality assumption in parametric method can be done by using normal probability plot of residulas. In nonparametric method there are also some procedures to test normality such as Shapiro-Wilk tests, Locke-Spurrier test, etc. However such tests require special tables.

Kernel method is considered as nonparameteric method. In kernel method, the idea is based on density estimator by more fairly spreading out the probability mass of each observation, not arbitrarily in a fixed interval, but smoothly around the observation, typically symmetric way (Kvam & Vidakovic 2007). In order to smooth around the observation, it is important to choose what is called smoothing function h_n or bandwiths which analogous to the bin width in a histogram. The problem of choosing the bandwith or how much to smooth is of crucial importance in density estimation. A natural method is to plot out several curves and choose the estimate that is most in accordance with one's prior ideas about the density (Silverman 1986). And according to Wand and Jones (1984), bandwith is scale factor to control the spread out of point observation in the curve.

In this article, we tested normality assumption using kernel method and estimated the optimal bandwith using Kullback-Leibler cross validation method for small sample sizes. In order to get the satisfying result we also use Kolmogorof-Smirnov goodness of fit test to compare with result obtained from kernel method.

KERNEL METHOD

Let $X_1,X_2,\ldots,X_n$ be a sample, we write the density estimator

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \qquad (2.1)$$

for Xi= xi, i=1,2…,n. The kernel function K represents how the probaility mass assigned, so for histogram it is just a constant interval, wich satistisfied $\int K(x)dx$. The smoothing funtion hn is a positive sequence of bandwiths (Kvam & Vidakovic 2007)..

Let $X_1,X_2,\ldots,X_n$ be a random sample of density function F with density f. and let $u_{ii^*} = X_i + X_{i^*}$ with density h_1 and $v_{ii^*} = X_i + X_{i^*}$ with density h_2 furthermore, let $h(u, v)$ as joint density of u_{ii^*} and v_{ii^*}, for all $i \neq i^* = 1, 2, ..., n$. The kernel estimates of h_1, h_2 and h respectively are:

$$\hat{h}_1(u) = \frac{1}{n(n-1)b}\sum_{i\neq i^*} w\left(\frac{u-u_{ii^*}}{b}\right) \qquad (2.2)$$

$$\hat{h}_2(v) = \frac{1}{n(n-1)b}\sum_{i\neq i^*} w\left(\frac{v-v_{ii^*}}{b}\right) \qquad (2.3)$$

and

$$\hat{h}(u,v) = \frac{1}{n(n-1)b^2}\sum_{i\neq i^*} w\left(\frac{u-u_{ii^*}}{b}\right) w\left(\frac{v-v_{ii^*}}{b}\right)$$

(2.4) with $b = b_n$ is a positive constant called bandwith and $w(.)$. is a known symmetric bounded density called kernel. Therefore, we can assume w has mean 0 and a finite variance $\mu_2(w)$, and that b $\rightarrow$ 0 as n $\rightarrow \infty$. If we want to test normality assumption such as: H_0 : f is $N(\mu,\sigma_2)$ agints H_1: f isnot $N(\mu,\sigma_2)$, for $\mu \in R$ and $\sigma_2 > 0$. A measure of departure from H_0 is:

$$\delta = \int_{-\infty}^{\infty}\int_{\infty}^{\infty}[h(u,v) - h_1(u)h_2(v)]^2\, dudv. \quad (2.5)$$

This is so, since H0 is equivalent to H_0 : $h(u, v) = h_1(u)h_2(v)$ u and v are independent. Using the random sample X_1, X_2, …, X_n and using estimates $\hat{h}_1, \hat{h}_2$ and $\hat{h}$ above, we can performe the normality test based on the assumption of:

$$\hat{\delta} = \int_{-\infty}^{\infty}\int_{\infty}^{\infty}[\hat{h}(u,v) - \hat{h}_1(u)\hat{h}_2(v)]^2\, dudv.$$

(Ahmad & Mugdadi 2003.)

Kullback-Leibler cross validation method

Suppose that an independent observation X_1, X_2, …, X_n from f were available. Then the likelihood of f as the density underlying the observation X would be log f(X), regarded as a function of h as the log likelihood of the smoothing parameter h. Likelihood Cross Validation (LCV) is average over each choice of ommited Xi, to give the score:

$$LCV = n^{-1}\sum \log \hat{f}(X_i) \quad (2.6)$$

From (2.6) can be seen that the value of h maximized LCV(h). The maximum LCV(h) can be obtained from Kullback Leiebler information distance, defined by:

$$d_{KL}\left(f, \hat{f}_h\right) = \int \log\left(f / \hat{f}\right)(x) f(x) dx \quad (2.7)$$

To estimate the optimal *bandwidth* can be done by minimized h_{opt} and h_{os} , where h_{opt} is the value of h which maximized the Kullback-Leibler information distance and h_{os} is *h oversmoothing bandwidth*.

For instance, we would like to test independent random variable *Xi*, *Likelihood* of *Xi* is $\prod_i \hat{f}_h(X_i)$. Statistics value for different *h* will guide us to get better *h,*, because the algorithn of this statistics approximately close to $d_{KL}(f, \hat{f}_h)$, so that with counting $\hat{f}_h$ from $\{X_j\}_{j\neq i}$ is the same as getting likelihoof function for *Xi* (Hardle,1991).

$$\hat{f}_{h,i}(X_i) = \frac{1}{(n-1)h}\sum_{i\neq j}^{n} K\left(\frac{X_i - X_j}{h}\right) \quad (2.8)$$

This estimate is called cross validation defined by:

$$L(X_i) = \prod_i^n \hat{f}_{h,i}(X_i) = \frac{1}{(n-1)h}\prod_{i=1}^{n}\sum_{i\neq j}^{n} K\left(\frac{X_i - X_j}{h}\right) \quad (2.9)$$

times $1/n$, we get Kullback-Leibler cross validation (CV_{KL}) :

$$CV_{KL}(h) = \frac{1}{n}\sum_{i=1}^{n}\log\left[\hat{f}_h(X_i)\right] = \frac{1}{n}\sum_{i=1}^{n}\log\left[\sum_{j\neq i} K\left(\frac{X_i - X_j}{h}\right)\right] - \log[(n-1)h] \quad (2.10)$$

According to Hardle (1991), the optimal bandwidth h is h which maximized (CV_{KL})

$h_{KL} = h_{\max} = \max CV_{KL}(h)$

$$CV_{KL}(h) = \frac{1}{n}\sum_{i=1}^{n}\log\left[K\left(\frac{X_i - X_j}{h}\right)\right] - log[(n-1)h] \quad (2.11)$$

Kolmogorov-Smirnov

Kolmogorov dan Smirnov (1948) goodness of fit test is used to test:

$H_0 : F(x) = F_0(x), (\forall x)$

$H_1 : F(x) \neq F_0(x)$

We reject H_0 if

$D_n = \max | F_n(x) - F(x)| > D_\alpha$

RESULTS AND DISCUSSION

To demonstre the method introduced we simulated the data from N(0,1), N(5,10), exponential ditribution, and Gamma distribution with the size of samples are 10 and 25. The estimating optimal *bandwidth* using Kullback-Leibler cross validation method was done by using S-Plus software. We compare the result with the Kolmogorov-Smirnov Goodness of Fit test statistics..

Normal distribution ((N(0,1)) with n = 10

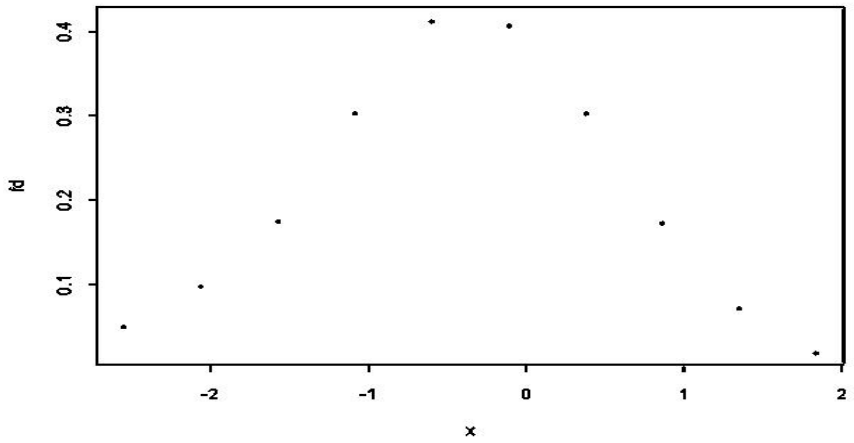
We obtained h oversmoothing bandwidth or h_{os} = 0.53 and $CV_{KL}(h)$ maximum = 0.6. Using the estimating curve with bandwith (h) 0.3 to 0.6, it showed that the optimal bandwith (h_{opt}) is is the same as $CV_{KL}(h)$ maximum= 0.6, and the estimated curve is shown in Figure 1.

Normal distribution ((N(0,1)) with n = 25

We obtained h oversmoothing bandwidth or h_{os} = 0.65 and $CV_{KL}(h)$ maximum = 0.7. Using the estimating curve with bandwith (h) 0.45 to 0.7, it showed that the optimal bandwith (h_{opt}) is the same as $CV_{KL}(h)$ maximum = 0.7, and the estimated curve is shown in Figure 2.

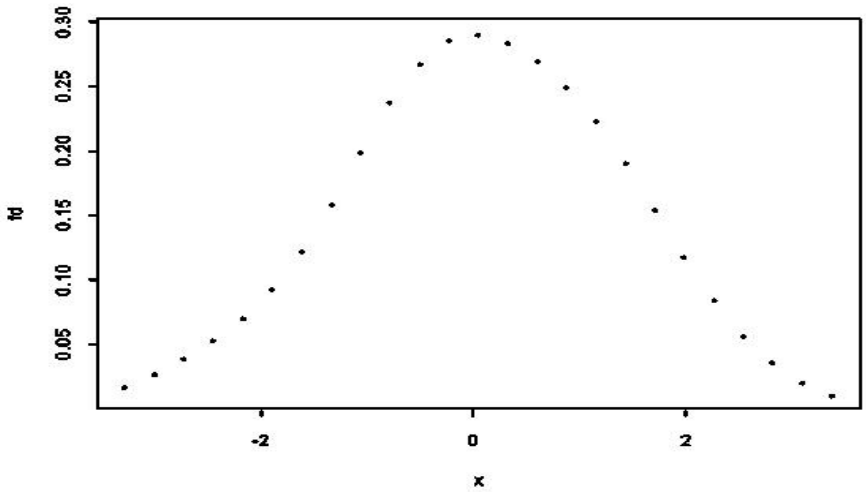
Normal distribution ((N(5,10)) with n = 10
We obtained h oversmoothing bandwidth (h_{os}) = 4.4 and $CV_{KL}(h)$ maximum = 5. By using the estimating curve with bandwith (h) 3 to 5, it showed that the optimal bandwith (h_{opt}) is the same as $CV_{KL}(h)$ maximum = 5, and the estimated curve is shown in Figure 3.

Normal distribution ((N(5,10)) with n = 25
We obtained h oversmoothing bandwidth (h_{os}) = 5.3 and $CV_{KL}(h)$ maximum = 6. Using the estimating curve with bandwith (h) 3 to 6, it showed that the optimal bandwith (h_{opt}) is the same as $CV_{KL}(h)$ maximum = 6, and the estimated curve is shown in Figure 4.



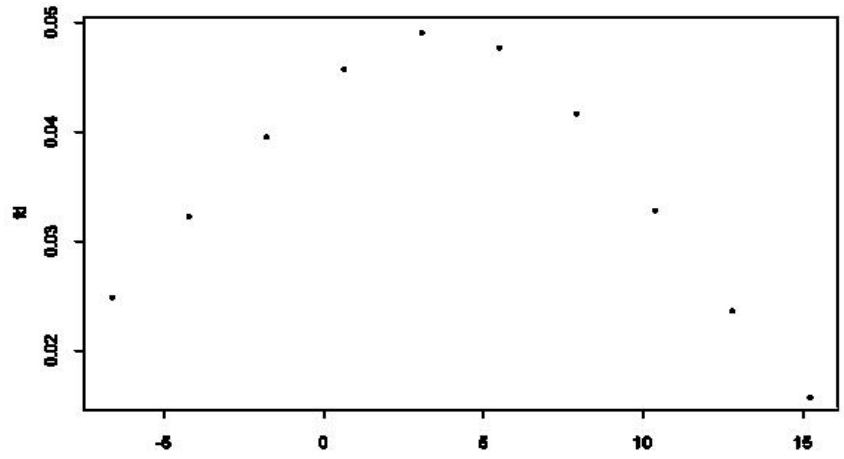
(d) h_{opt}=0.6

Figure 1. Normal density curve (N(0,1)) with n=10.



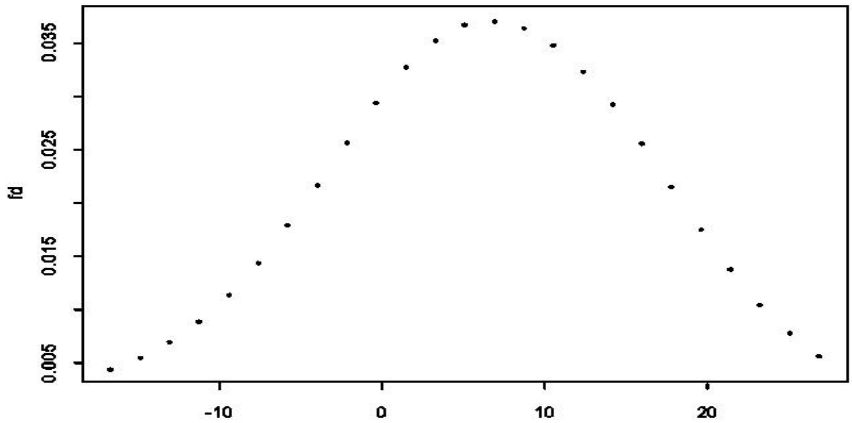
(d) h_{opt}=0.7

Figure 2. Normal density curve (N(0,1)) with n=25.



(d) $h_{opt}^{x} = 5$

Figure 3. Normal density curve (N(5,10)) with n=10.



(d) $h_{opt}^{x} = 6$

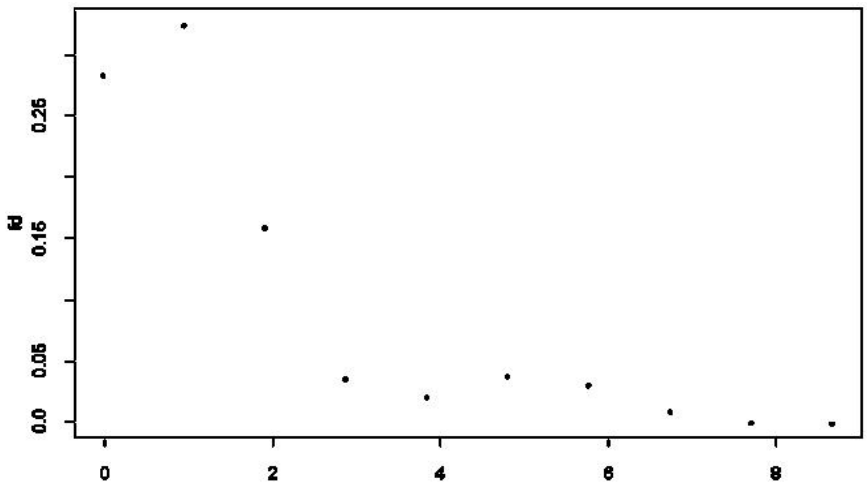
Figure 4. Normal density curve (N(5,10)) with n=25.

Exponential distribution ((E(1)) with n = 10
We obtained h oversmoothing bandwidth (h_{os}) = 0,97 and $CV_{KL}(h)$ maximum = 1. Using the estimating curve with bandwith (h) 0.5 to 1, it showed that the optimal bandwith (h_{opt}) is the same as $CV_{KL}(h)$ maximum = 1, and the estimated curve is shown in Figure 5.

Exponential distribution ((E(1)) with n = 25
We obtained *h* oversmoothing bandwidth (h_{os}) = 0.49 and $CV_{KL}(h)$ maximum = 0.5. Using the estimating curve with bandwith (h) 0.2 to 0.5, it showed that the optimal bandwith (h_{opt}) is the same as $CV_{KL}(h)$ maximum = 0.5, and the estimated curve is shown in Figure 6.

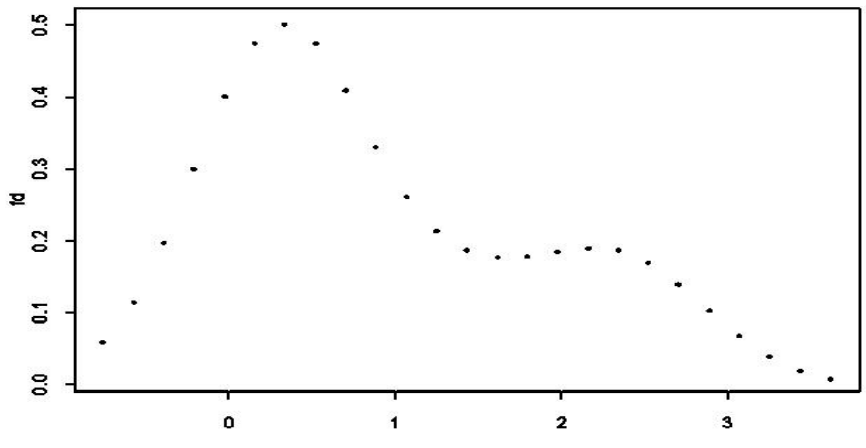
Gamma distribution (G(1,1)) with n = 10
We obtained h oversmoothing bandwidth (h_{os}) = 1.52 and $CV_{KL}(h)$ maximum = 0.5. Using the estimating curve with bandwith (h) 1 to 2, it showed that the optimal bandwith (h_{opt}) is the same as $CV_{KL}(h)$ maximum = 2, and the estimated curve is shown in Figure 7.

Gamma distribution (G(1,1)) with n = 25
We obtained h oversmoothing bandwidth (h_{os}) = 0.63 and $CV_{KL}(h)$ maximum = 0.7. Using the estimating curve with bandwith (h) 0.3 to 0.7, it showed that the optimal bandwith (h_{opt}) is the same as $CV_{KL}(h)$ maximum = 0.7, and the estimated curve is shown in Figure 8.



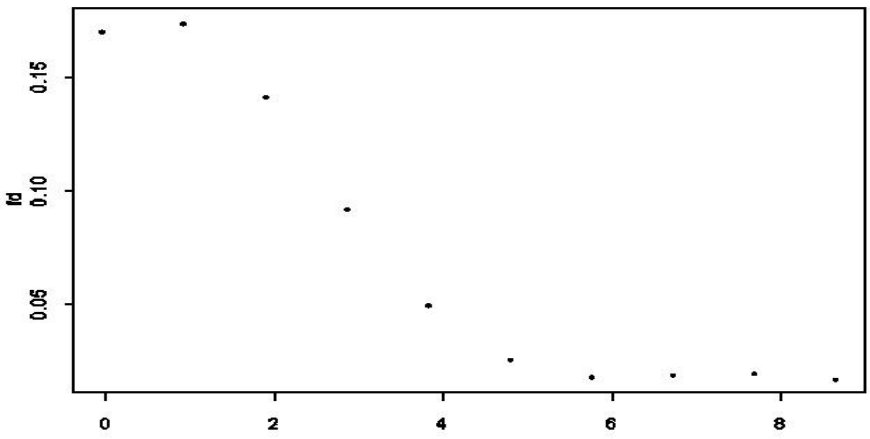
(d) $h_{opt}=1$

Figure 5. Exponential density curve (E(1)) with n=10.



(d) $h_{opt}=0.5$

Figure 6. Exponential density curve (E(1)) with n = 25.



(d) $h_{opt}=2$

Figure 7. Gamma density curve (G(1,1)) with n = 10.

Table 1. Kolmogorov-Smirnov Goodness of Fit Test

	Distribution							
	N(0,1)		N(5,10)		Exponential (1)		Gamma (1,1)	
Sample size	D_n	$D_{0.05}$	D_n	$D_{0.05}$	D_n	$D_{0.05}$	D_n	$D_{0.05}$
n=10	0.1443^{ns}	0.369	0.1081^{ns}	0.369	0.3920^{*}	0.369	0.4438^{*}	0.369
n=25	0.0879^{ns}	0.283	0.0776^{ns}	0.283	0.2843^{*}	0.283	0.2772^{*}	0.283

Note: ns= nonsignificant at α=0.05
*=significant at α=0.05

Kolmogorov-Smirnov goodness of Fit test

The result of Kolmogorov-Smirnov Goodness of Fit Test can be seen in Table 1. We reject hipotesis nul when $D_n > D_{0.05}$.

By comparing the figures obtained by Kernel Method with Kolmogorof Smirnov test, we can say that both methods gave the same preformance. Figure 1, Figure 2, Figure 3, and Figure 4 showed that the data were normally distributed which were the same as Kolmogorof Smirnov's (Table 1). The same result were given by Figure 5 and Figure 6 (data from exponential distribution) as well as Figure 7 and Figure 8 (data from Gamma distribution) when comparing with Kolmogorof Smirnov's (Table 1).

The result is also the same as the result by Ahmad & Mugdadi (2003) which showed that kernel method gave the same performance with the result of Locke & Spurrier (1976) when simulated from distribution different than normal such as from the Chi-Square, the Cauchy and the Beta Distributions.

CONCLUSION

The simulation study illustrates that kernel method is useful for testing normality for n=10 and n=25. This study also reveals that severe departure from normality can be detected easily using kernel method. By comparing the results from Kernel Method and Kolmogorov Smirnov test, we conclude that the two test gave the same performance

REFERENCES

Ahmad IA & Mugdadi AR. 2003. Testing Normality using Nonparametric Kernel Methods. *Jurnal of Nonparametrics Statistics*. **15**(3)**:** 273-288.

Fan YQ. 1994. Testing Goodness of Fit Tests of a Parametric Density Function by Kernel Method. *Econometric Theory*. **10**: 316-356.

Kvam PH & Vidakovic B. 2007. *Nonparametric Statistics with Apllications to Science and Engineering.* John Wiley and Sons, New Jersey.

Montgomery DC. 2005. *Design and Analysis of Experiment*. John Wiley and Sons, New Jersey.

Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall Ltd, New York.

Shapiro SS & Wilk MB. 1965. An Analysis of Variance Test for Normality. *Biometrika*. **52**: 591-611