International Journal of Pure and Applied Mathematics

Volume 42 No. 4 2008, 569-573

A PRACTICAL IMPLEMENTATION FOR BANDWIDTH SELECTION IN KERNEL DISTRIBUTION FUNCTION ESTIMATORS

Y.L. Lio

Department of Mathematical Sciences
University of South Dakota
414, East Clark Street, Vermillion, SD 57069-2390, USA
e-mail: Yuhlong.Lio@usd.edu

Abstract: Two kernel smooth distribution estimators for a distribution function, $F(x)$ are considered. It is shown that the bandwidth $h_0 = O(n^{-1/3})$ is asymptotically optimal in the sense

$$\left[\hat{\phi}(x,n,h_0) - F(x)\right]^2 / \inf_h \left[\hat{\phi}(x,n,h) - F(x)\right]^2 \to 1$$

in probability as $n \to \infty$, where $\hat{\phi}(x,n,h)$ indicates both kernel smooth estimators. To implement the optimal bandwidth selection, a Bayesian bandwidth is proposed. The computer simulation study shows that the proposed smooth distribution estimators with Bayesian bandwidth could have mean squared errors that are better than the smooth estimators with the optimal bandwidth h_0 in terms of mean squared errors.

AMS Subject Classification: 62A13
Key Words: bandwidth selection, Bayesian estimator, kernel estimation

1. Introduction

Let $X_1, ..., X_n$ be a random sample from the distribution $F(x)$ which has a density function $f(x)$. Nadaraya [3] proposed smooth estimator for the distribution function $F(x)$, that is defined as $F_n(x) = \int_{-\infty}^{x} f_n(t)dt$, where $f_n(x)$ is the smooth kernel density estimator which is defined as $f_n(x) = \frac{1}{nh}\sum_{i=1}^{n} k((x-$

Received: August 17, 2007

$X_i)/h)$, and $k(x)$ is some density function. One of the important issue is the selection of smooth bandwidth parameter h in the estimator. Azzalini [1] investigated $F_n(x)$ and the inverse function of F_n assuming that the kernel density function has compact support and obtained the asymptotically optimal bandwidth in terms of minimizing the mean squared error. Next, the distribution function can also be estimated by $\tilde{F_n}(x) = \frac{1}{nh}\int_{-\infty}^{\infty} \hat{F_n}(t)k((t-x)/h)dt$, where $\hat{F_n}(x)$ is the empirical distribution function and $k(x)$ is some density function. It is noticed that $\tilde{F_n}(x)$ is defined differently from $F_n(x)$. For $\tilde{F_n}(x)$, the empirical distribution is smoothed; however, for $F_n(x)$, the smooth density function is used. When the kernel function $k(.)$ is symmetric about 0, it can be shown that $\tilde{F_n}(x) = F_n(x)$. In this paper, we want to extend the study of bandwidth selection for kernel smooth distribution function estimators with kernel densities which could have infinite support and develop a data base bandwidth selection for the smooth estimators via Bayes criterion.

2. Main Results

Assume that the kernel density function $k(x)$ and distribution function $K^*(x)$ satisfies the following:

1. $k(x)$ is symmetric about zero.
2. $\lim_{x\to\pm\infty} |xk(x)| = 0$.
3. $\int_{\frac{x-c}{h}}^{\infty} t^i k(t)K^*(t)dt = o(h^2)$, for i =1, 2 and $x > c$.
4. $\int_{\frac{x-c}{h}}^{\infty} t^i k(t)dt = o(h^2)$, for i =1, 2 and $x > c$.
5. $\int t^2 k(t)dt = \kappa > 0$.

It is noticed that the conditions given in this section for kernel function are different from the conditions for kernel functions from those papers mentioned in Section 1.

2.1. Classic Bandwidth Selection

Definition 1. A bandwidth sequence $\tilde{h}$ is an asymptotically optimal if $[\hat{\phi}(x,n,\tilde{h}) - \phi(x)]^2/\inf_h[\hat{\phi}(x,n,h) - \phi(x)]^2 \to 1$ in probability as $n \to \infty$, where $\phi(x)$ is distribution function and $\hat{\phi}(x,n,h)$ be the corresponding kernel smooth estimator.

Lemma 1. *Suppose the density function $f(x)$ has uniformly continuous second derivative and $k(x)$ is continuous density function which satisfies properties 1. 2. 3. 4. and 5. mentioned above. Then*

$$E\left[(F_n(x) - F(x))^2\right] = I_0 + O(n^{-1}h^2) + o(h^4)\,, \tag{2.1}$$

where

$$I_0 = n^{-1}\left[p(1-p) - 2hf(x)\int uk(u)K^*(u)du\right] + h^4(\frac{\kappa}{4})^2(f'(x))^2. \tag{2.2}$$

It can be shown that when $\int uk(u)K^*(u)du > 0, I_0$ is minimized by

$$h_0 = \left[\frac{2f(x)\int uk(u)K^*(u)du}{n\kappa^2(f'(x))^2}\right]^{\frac{1}{3}}. \tag{2.3}$$

Let $H_n = [n^{-1+\varepsilon}, n^{-\varepsilon}]$ for some $\varepsilon > 0$, we have the following lemma.

Lemma 2. *If $f(x)$ is uniform continuous, $k(x)$ satisfies the conditions mentioned above and $W(x, X_i, h) = K^*((x - X_i)/h) - E(K^*((x - X_i)/h))$, then we have the following results*

1. $\sup_{h\in H_n} [\sum_{i=1}^n n^{-2}[W(x, X_i, h)]^2 - \mathrm{Var}\,[F_n(x)]]/I_0 \to 0$, *with probability one.*

2. $\sup_{h\in H_n} [n^{-2}\sum\sum_{i<j}^n n^{-2}[W(x, X_i, h)][W(x, X_j, h)]/I_0 \to 0$, *with probability one.*

3. $\sup_{h\in H_n} [F_n(x) - E(F_n(x))][E(F_n(x)) - F(x)]/I_0 \to 0$, *in probability.*

Combining Lemma 1 and Lemma 2 gives the following result.

Theorem 1. *If, in addition to the conditions of Lemma 1, $f(x) > 0$, $\int tk(t)K^*(t)dt > 0$, and $k(x)$ is Hölder continuous density function, then*

$$[\hat{\phi}(x, n, h_0) - \phi(x)]^2 / \inf_{h\in H_n} [\hat{\phi}(x, n, h) - \phi(x)]^2 \to 1$$

in probability as $n \to \infty$.

Using Assumptions 3 and 4 for kernel function to take care of both tails in the integrals and following the same structure of arguments by Lio and Padgett [2], Lemma 1, Lemma 2 and Theorem 1 can be proved. However, it should be mentioned that the conditions for Lemma 1, Lemma 2 and Theorem 1 are different from the ones of Lio and Padgett [4]. Here the kernel function could have unbounded support. For example, $k(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$ for $-\infty < x < \infty$ satisfies all the conditions mentioned for kernel function.

2.2. Bayesian Bandwidth Section

The standard normal kernel function and a h conjugate prior which is defined as $\theta(h) = \frac{1}{\Gamma(\gamma)\lambda^{\gamma}h^{2\gamma+1}}\exp(-1/(\lambda h^2))$ are used. Then the posterior for h is given as

$$\hat{\pi}(h|\vec{X}, x) = \frac{F_h(x)\theta(h)}{\int F_h(x)\theta(h)dh}\,, \tag{2.4}$$

where $\vec{X}$ is data given. Therefore, $\hat{\pi}(h|\vec{X}, x)$ can be presented as

$$\hat{\pi}(h|\vec{X}, x) = \frac{(1/n)\sum_{i=1}^{n} K^*((x - X_i)/h)\theta(h)}{(1/n)\int \sum_{i=1}^{n} K^*((x - X_i)/h)\theta(h)dh}\,. \tag{2.5}$$

Let the posterior mean be the local Bayes bandwidth. It can be shown that the local Bayes bandwidth, $h^*(x)$, is given by

$$h^*(x) = \frac{\Gamma(\gamma)\sum_{i=1}^{n}\int_{-\infty}^{x}[\frac{1}{\lambda[(t-X_i)^2+2]}]^{\gamma}dt}{\Gamma(\gamma+0.5)\sqrt{2\lambda}\sum_{i=1}^{n}\int_{-\infty}^{x}[\frac{1}{\lambda[(t-X_i)^2+2]}]^{\gamma+0.5}dt}\,. \tag{2.6}$$

Hence a kernel estimation can be based on smooth parameter $h^*(x)$ which does not contain unknown parameters from distribution.

3. Empirical Studies

In this section, a simulation study will be conducted to compare effects from the optimal bandwidth selection, $h_0(x)$, with from the Bayes bandwidth, $h^*(x)$, in terms of mean squared errors. For each distribution and sample size ($n = 100$), the simulation is repeated for 1000 replications to find 1000 kernel estimates with h_0 and 1000 kernel estimates with $h^*(x)$. Then the ratio of mean squared errors of both types estimators can be estimated as following:

$$\rho(h_0, h^*, x) = \sum_{1}^{1000}\left[\hat{\phi}(x, n, h_0) - F(x)\right]^2 / \sum_{1}^{1000}\left[\hat{\phi}(x, n, h^*) - F(x)\right]^2\,, \tag{3.1}$$

where $\hat{\phi}(x, n, h_0)$ is the kernel estimator based on $h_0(x)$ and $\hat{\phi}(x, n, h^*)$ is the kernel estimator based on $h^*(x)$. Given a true distribution, kernel density function, sample size and a input x, $h_0(x)$ can be evaluated exactly. However, $h^*(x)$, which is based on sample and kernel density, involves two parameters γ and λ from the prior. In the simulation, the γ parameter is given as 3 and λ is running from a wide range (from 0.5 to 30) to control the $h^*(x)$. For saving space consideration, parts of simulation results to select a better range of λ which make $\rho(h_0, h^*, x) \geq 1$ or close to 1 for the kernel estimates of distribution function with smooth parameter $h^*(x)$ are reported in Table 1.

λ	$F(t) = 0.1$	$F(t) = 0.25$	$F(t) = 0.50$	$F(t) = 0.75$
Weibull distribution, $shape = 0.75$ and $scale = 1.0$				
Range	$25 < \lambda$	$1 \geq \lambda$	$8 < \lambda < 21$	$3 < \lambda$
Weibull distribution, $shape = 1.0$ and $scale = 1.0$				
Range	$14 < \lambda$	$2 < \lambda < 11$	$5 < \lambda < 13$	$5 < \lambda < 12$
Weibull distribution, $shape = 1.25$ and $scale = 1.0$				
Range	$8 < \lambda$	$3 < \lambda < 12$	$5 < \lambda < 9$	$7 < \lambda < 16$
Normal distribution, $\mu = 5.0$ and $\sigma = 2.0$				
Range	$1 < \lambda$	$0.5 < \lambda \leq 11$	non λ	$0.5 < \lambda \leq 11$

Table 1: The range selection for λ parameter of prior distribution. Shape parameter for prior $\gamma = 3.0$ and sample size $n = 100$.

Acknowledgments

The author would like to thank the conference committee for inviting him as a speaker.

References

[1] A. Azzalini, A note on the estimation of a distribution function and quantiles by a kernel method, *Biometrika*, **68** (1981), 326-328.

[2] Y.L. Lio, W.J. Padgett, A note on the asymptotically optimal bandwidth for Nadaraya's quantile estimator, *Statistics and Probability Letters*, **11** (1991), 243-249.

[3] E.A. Nadaraya, Some new estimates for distribution functions, *Theory of Probability and its Applications*, **9** (2002), 497-500.