Choice of smoothing parameters in boosting kernel density estimates

*C. C. Ishiekwene, S. M. Ogbonmwan and F. O. Oyegue, Department of Mathematics, University of Benin, Nigeria.

Abstract

This paper provides a scheme for choosing the smoothing parameter in boosting kernel density estimates. Several boosting algorithms are implemented using different choices of smoothing parameters and the "best" choice is found to be at order – four after considering different sets of data.

Keywords: Smoothing parameter, boosting, kernel density estimates, boosting algorithms, order-four.

1.0 Introduction

The choice of the smoothing parameter has always been the "Cross" of kernel density estimation (Silverman (1986 [16]), Wand and Jones (1995 [17]), Jones and Signorini (1997 [9])). The choice of the smoothing parameter is not easily found and as a result of this, the idea of boosting in KDE was introduced by Schapire in 1990. Other contributors include Freund (1995 [4]), Freund and Schapire (1996 [15]) and Schapire and Singer (1999 [14]).

Boosting is a means of improving the performance of a "Weak learner" in the sense that given sufficient data, it would be guaranteed to produce an error rate which is better than "random guessing".

Different boosting algorithms are now in existence. These include Mazio and Taylor's (2004[11]), Adaboost (1997) and Ishiekwene et.al (2007a [7] and 2007b [8]) just to mention a few. All these boosting algorithms use suitably re-weighting of Data. This re-weighting is done by placing a weight on the kernel estimator. Boosting has been shown by all these authors to be a bias reduction technique.

In section two, we show different boosting algorithms and how boosting is a bias reduction technique. Section three clearly shows how the choice of the smoothing parameter is chosen and the effects on the boosting algorithms stated in section two. Numerical examples are used to validate our claims in section four and the findings discussed with recommendation.

2.0 Different boosting algorithms in kernel density estimates and bias reduction

Three different boosting algorithms which are bias reduction techniques are stated in this paper. The three algorithms are numbered 1 - 3.

```
2.1 Algorithm 1 (Mazio and Taylor (2004 [11] ))
Step 1
```

Given
$$\{x_i, i = 1, 2, ..., n\}$$
, initialize $W_1(i) = \frac{1}{n}$

Step 2

Select *h* (the smoothing parameter).

Step 3

For m = 1, 2, ..., M, obtain a weighted kernel estimate,

*e-mail: cycigar@yahoo.co.uk

$$\hat{f}_m(x) = \sum_{i=1}^n \frac{W_m(i)}{h} k\left(\frac{x - x_i}{h}\right)$$
(2.1)

and then update the weights according to

$$W_{m+1}(i) = W_m(i) + \log\left\{\frac{\hat{f}_m(x_i)}{\hat{f}_m^{(-1)}(x_i)}\right\}$$
(2.2)

Step 4

Provide output as

 $\prod \hat{f}_m(x)$ renormalized to integrate to unity

2.2 Algorithm 2 (Ishiekwene et.al 2007a [7])

Step 1

m=1

Given
$$\{x_i, i = 1, 2, ..., n\}$$
, initialize $W_1(i) = \frac{1}{n}$

Step 2

Select h (the smoothing parameter).

Step 3: For m = 1, 2, ..., M,

(i) Get
$$\hat{f}_m(x) = \sum_{i=1}^n \frac{W_m(i)}{h} k \left(\frac{x - x_i}{h}\right)$$

(ii) Update $W_{m+1}(i) = W_m(i) + mesh$

Step 4

Provide output
$$\prod_{m=1}^{M} \hat{f}_m(x)$$
 and normalized to integrate to unity

Here, we observe that the weight function uses a meshsize instead of the leave-one-out log ratio function of Mazio & Taylor (2004).

2.3 Algorithm 3 (Ishiekwene et.al 2007 b [8]) Step 1

Given
$$\{x_i, i = 1, 2, ..., n\}$$
, initialize $W_1(i) = \frac{1}{n}$

Step 2

Step 3

Select h (the smoothing parameter).

For m = 1, 2, ..., M,

Update

$$\hat{f}_{m}(x) = \sum_{i=1}^{n} \frac{W_{m}(i)}{h} k \left(\frac{x - x_{i}}{h} \right)$$
$$W_{m+1}(i) = W_{m}(i) + \log \left\{ \frac{\hat{f}_{m}(x_{i})}{\hat{f}_{m}^{B}(x_{i})} \right\}$$
(2.3)

(iv)

Provide output
$$\prod_{m=1}^{M} \hat{f}_m(x)$$
 and normalized to integrate to unity

Here, we also observe that the weight function uses a bootstrap estimate instead of the leave-one-out log ratio function of Mazio and Taylor (2004 [11]).

2.4 Boosting as a bias reduction in kernel density estimation

Suppose we want to estimate f(x) by a multiplicative estimate. We also suppose that we use only "weak" estimates which are such that *h* does not tend to zero as $n \rightarrow \infty$. Let us use a population version instead of sample in which our weak learner, for h > 0 is given by

$$\hat{f}_m(x) = \int \frac{1}{h} W_m(y) K\left(\frac{x-y}{h}\right) f(y) dy$$
(2.4)

where $W_1(y)$ is taken to be 1. We shall take our kernel function to be Gaussian (since all distributions tend to normal distribution as n - the sample size, becomes large through central limit theory). The first approximation in the Taylor's series, valid for h < 1 provided that the derivatives of f(x) are properly behaved, is

$$\hat{f}_{(1)}(x) = f(x) + \frac{h^2 f''(x)}{2}$$
 and so we observe the usual bias of order $O(h^2)$ of Wand and Jones (1995)

[17]). If we now let $W_2(x) = \hat{f}_{(1)}^{-1}(x)$, the boosted estimator at the second step is

$$\hat{f}_{2}(x) = \int k(z) \left\{ f(x+zh) + h^{2} \frac{f''(x+zh)}{2} + 0(h^{4}) \right\}^{-1} f(x+zh) dz = 1 - \frac{h^{2} f''(x)}{2 f(x)} + 0(h^{4})$$
(2.5)

This gives an overall estimator at the second step as

$$\hat{f}_{(1)}(x) \cdot \hat{f}_{(2)}(x) = f(x) \left\{ 1 + h^2 \frac{f''(x)}{2f(x)} + 0(h^4) \right\} \left\{ 1 - \frac{h^2 f''(x)}{2f(x)} + 0(h^4) \right\} f(x) + 0(h^4)$$
(2.6)

which is clearly of order four and so we can see a bias reduction from order two to order four (for further details, see Mazio and Taylor 2004 [11]).

3.0 Choice of smoothing parameter in boosting KDE's

We present an expression which is to generate different values of the smoothing parameter, h. This gives the advantage of seeing how any data set behave when different h-values are used in boosting KDE. These choices of h would be seen to give "good" density estimates when used in boosting unlike when h is chosen subjectively, it can result in either oversmoothing or undersmoothing the density function.

The expression for h is given as

$$h^{(p)} = \left\{ \frac{(p!)^2}{2p} \right\}^{\frac{1}{2p+1}} V_p^{-\frac{2}{2p+1}} \left\{ \int_{-\infty}^{\infty} k(t)^2 dt \right\}^{\frac{1}{2p+1}} \cdot \left\{ \int_{-\infty}^{\infty} f^{(p)}(x)^2 dx \right\}^{\frac{1}{2p+1}} \cdot n^{-\frac{1}{2p+1}}$$
(3.1)

where $p = 2, 4, 6, ..., < \infty$. And provided *k* is a kernel function satisfying

$$\int_{-\infty}^{\infty} k(t)dt = 1$$

$$\int_{-\infty}^{\infty} t k(t)dt = \int_{-\infty}^{\infty} t^{2}k(t)dt = \dots = \int_{-\infty}^{\infty} t^{p-1}k(t)dt = 0$$

$$\int t^{p}k(t) = V_{p} \neq 0$$
(3.2)

where $p = 2, 4, 6, ..., < \infty$ is the order of the kernel function which is symmetric and continuously differentiable.

For further details on (3.1) and (3.2), you can see Ishiekwene and Afere (2001 [5]), Osemwenkhae (2003 [12]), Ishiekwene et.al (2006a [[6]).

From the three different boosting algorithms stated above we can see that the kernel density estimator represented by equation (2.1), is dependent on *h* since the quantity $\frac{W_m(i)}{h} \rightarrow \infty$ as $h \rightarrow 0$

this means that a heavier weight is placed on the kernel function. This produces "noises" on the boosted density estimates. Also, $\frac{W_m(i)}{h} \rightarrow 0$ as $h \rightarrow \infty$, thereby placing a lighter weight on the kernel function and resulting in smooth boosted density estimates. Thus, the value of *h* has a prominent role to play in all boosting algorithms in KDE. It is also a fact that *h* in turn determines the quantity – bias squared which translates to affecting the AMISE which measures the discrepancy in our estimates (Wand and Jones 1995 [17], Duffy and Helmbold 2000 [1].

4.0 Numerical results

We used three sets of data (Data 1 is the lifespan of a car battery in years, Data 2 is the number of written words without mistakes in every 100 words by a set of students and Data 3 is the scar length of patients randomly sampled in millimeters. See Ishiekwene and Osemwenkhae, 2006 [6]) which were suspected to be normal to illustrate algorithms 1, 2 and 3 and bring out clearly the effect of *h* on all three algorithms listed in this paper. BASIC programming language is used to generate the results and their corresponding graphs shown in figures 4.1 - 4.6.

Table 4.1 shows the values of the smoothing parameters obtained using equation 3.1 above for three different data sets. Table 4.2 shows the values of the bias squared, variance and AMISE for the three different data sets respectively.

Table 4.1:	Values of	smoothing	parameter	obtained	for	different	higher	order
-------------------	-----------	-----------	-----------	----------	-----	-----------	--------	-------

Method	Data 1 ($n = 40$)	Data 2 $(n = 64)$	Data 3 ($n = 110$)
H^2	0.3559702	3.899261	0.1542693
H^4	0.476858	5.4463023	0.2261031
H^{6}	0.536033	6.2213213	0.2631049
H^8	0.5707907	6.681324	0.2853423
H^{10}	0.5936022	7.059361	0.300129
H^{12}	0.6089225	7.1998385	0.3105749
H^{14}	0.62104455	7.3607483	0.31853044
H^{16}	0.63028343	7.48493897	0.32463876
H^{18}	0.6373304	7.58388756	0.32951451

 $H^{20} \qquad 0.643619145 \qquad 7.6645779 \qquad 0.3334964$



Figure 4.1: Probability density estimate using the Leave-One-Out Scheme for Data 1 (Order 2)



Figure 4.3: Probability density estimate using the leave-one-out scheme for Data 1 (Order 18)





Figure 4.2: Probability density estimate using the leave-one-out scheme for Data 1 (Order 4)



Figure 4.4: Probability density estimate using the leave-one-out scheme for Data 1 (Order 20)



Journal of the Nigerian Association of Mathematical Physics Volume 12 (May, 2008), 443 - 454 Smoothing parameters in kernel density estimates Ishiekwene, Ogbonmwan and Oyegue *J of NAMP*













Figure 4.10: Probability density estimate using the leave-one-out scheme for Data 2 (Order 4)



Journal of the Nigerian Association of Mathematical Physics Volume 12 (May, 2008), 443 - 454 Smoothing parameters in kernel density estimates Ishiekwene, Ogbonmwan and Oyegue *J of NAMP*



Journal of the Nigerian Association of Mathematical Physics Volume 12 (May, 2008), 443 - 454 Smoothing parameters in kernel density estimates Ishiekwene, Ogbonmwan and Oyegue *J of NAMP*



















Journal of the Nigerian Association of Mathematical Physics Volume 12 (May, 2008), 443 - 454 Smoothing parameters in kernel density estimates Ishiekwene, Ogbonmwan and Oyegue *J of NAMP*



Journal of the Nigerian Association of Mathematical Physics Volume 12 (May, 2008), 443 - 454 Smoothing parameters in kernel density estimates Ishiekwene, Ogbonmwan and Oyegue *J of NAMP*

5.0 Conclusion and remarks

The values of *h*- the smoothing parameter obtained using equation (3.1) for m = 1,4,6, ... 20, are displayed in Table 4.1 and are used for the different boosting algorithms for all three data set . Some selected smoothing parameter choices (namely, orders 2,4,18 and 20) are used in plotting estimates for the kernel density function for the three different data sets, and the density estimates are displayed in figures 4.1 - 4.36. In all, the order two choices showed some noises at the 'peaks' of all three density estimates. Thus, we recommend the use of higher orders greater than or equal to four smoothing parameter choices when using boosting algorithms in kernel density estimates. This also supports Jones and Signorini's [9] order four choice of smoothing parameter arguably the "best". The estimates for the density curves as shown in Figures 4.1 - 4.36, are done for all three algorithms namely; Algorithm 1 (Mazio and Taylor [11]), Algorithm 2 (Ishiekwene et.al 2007a [8]) and Algorithm 3 (Ishiekwene et.al 2007b [7]). The results showed that apart from the order two choice of the smoothing parameter, all other smoothing parameter choices are appropriate for any of the boosting algorithm employed. We therefore recommend the use of these higher order smoothing parameter choices in boosting in kernel density estimation.

Table 4.2:	Showing	bias	reduction
------------	---------	------	-----------

		Data 1	Data 2	Data 3
Normal (conventional order two)	Base 2	0.005276637	0.000293946	0.004697604
	Variance	0.019811685	0.001130402	0.016623515
	AMISE	0.250883225	0.001424348	0.021321119
H4 (as in Jones and Signorini 1997 [9] and Ishiekwene and Osemwenkhae 2006 [6])	Base 2	0.002071803	0.000108591	0.001767617
	Variance	0.014789245	0.000809307	0.011342162
	AMISE	0.016861048	0.000917898	0.013109779
BOOSTED (proposed scheme)	Base 2	0.002078009	0.000108715	0.001768218
	Variance	0.015168456	0.000822153	0.011446219

References

- [1] Duffy, N. and Hemlbold, D. (2000) Potential bosters: Advances in Neural info. Proc. Sys. 12, 258 264.
- [2] Freund and Schapire (1996) Experiments with a new boosting algorithm. In Mach. Learn. Proceedings of 12th Intl Conf. Ed. 1 Saite 148- 156.
- [3] Freund, Y. and Schapire, R. (1997) A decision theoretic genralisation of online learning and application to boosting. J. of comput. Sys. Sc. 55, 119 139.
- [4] Freund, Y. (1995). Boosting a weak learning algorithm by majority. Info. Comp. 121, 256 285.
- [5] Ishiekwene, C. C. and Afere, B. A. E. (2001) Higher Order Window Width Selector for Empirical Data, JSAN 14, 69 – 82.
- [6] Ishiekwene, C. C. and Osemwenkhae, J. E. (2006) A comparison of fourth order window width selectors in KDE (A univariate Case). ABACUS 33 # 2A, 14 – 20.
- [7] Ishiekwene, C.C; Ogbonmwan, S.M. and Odiase, J.O. (2007b) Bootstrap in boosting kernel density estimates. Accepted in Int. J. of Natural & Applied Sciences (F.U.T.O), Vol 3(3) Indexed in AJOL.
- [8] Ishiekwene, C.C; Ogbonmwan, S.M; and Osemwenkhae, J.E. (2007a) A meshsize boosting algorithm in kernel density estimation. Accepted in J.U.S.T, Ghana
- Jones, M. C. & Signorini, d. F. (1997) A comparison of Higher order Bias Kernel Density Estimators. JASA 92, 1063 – 1072.
- [10] Mannor, S.; Meir, R. & Mendelson, S. (2001) On consistency of boosting algorithms; submitted to Adv. In Neural info. Proc. Sys.
- [11] Mazio, D. M. & Taylor Boosting Kernel Density estimates: A bias reduction technique? Biometrika 91, 226 233.
- [12] Osemwenkhae, J. E. (2003) Higher Order Forms In Kernel Density Estimation. Unpublished Ph.D thesis submitted to School of Postgraduate Studies, University of Benin, Nigeria.
- [13] Rohatgi, V.J (1984) Statistical Inference. John Wiley & sons, New York.
- [14] Schapire, R. & Singer, Y. (1999) Improved boosting algorithm using confidence rated prediction. Machine learn. 37, 297 – 336.
- [15] Schapire, R. (1990) The strength of weak learnability. Machine learn. 5, 313 321.
- [16] Silverman, B. W. (1986) Density estimation for statistics and Data analysis. Chapman & Hall, London.
- [17] Wand, M. P. & Jones, M. C. (1995) Kernel smoothing. Chapman & hall/CRC, London.