Bootstrap Bandwidth and Kernel Order Selection for Density Weighted Averages

Y. Nishiyama

Kyoto Institute of Economic Research, Kyoto University, Kyoto, Japan.

Abstract: Density weighted average is a nonparametric quantity expressed by expectation of a function of random variables with density weight. It is associated with parametric components of some semiparametric models, and we are concerned with an estimator of this quantity. Asymptotic properties of semiparametric estimators have been studied in econometrics since the end of 1980's and it is now widely recognized that they are $\sqrt{n}-$consistent in many cases. Many of them involve nonparametric estimates of unknown density or regression function but they are biased estimators for the true functions. Because of this, we typically need to use some bias reduction technique in the nonparametric estimates for $\sqrt{n}-$ consistency of the semiparametric estimators. When we use a kernel estimator, a standard way is to take a higher order kernel function. For density estimation, the higher the kernel order is, the less becomes the bias without changing the order of variance in theory. However, it is also known that higher order kernels can inflate the variance which may cause the result that the mean squared error with very high order kernel becomes larger than that with low order kernel in small sample. This paper propose to select the bandwidth and kernel order by minimizing bootstrap mean squared error for a plug-in estimator of density weighted averages. We show standard bootstrap does not work at all for bias approximation as in density estimation, but smoothed bootstrap is useful in our problem if suitably transformed.

Keywords: *Bandwidth selection; Kernel order selection; bootstrap; Density weighted averages*

1. INTRODUCTION

Nonparametric kernel density estimator involves two user-determined components, one is bandwidth and the other is kernel function. One principle of determining them is such that the mean squared error (MSE) or mean integrated squared error is minimized. When we take kernel function from the class of density functions, Epanechnikov kernel is known to be optimal in the sense it minimizes the MISE (see e.g. Silverman (1986)), and the resulting MISE is of exact order $O(n^{-4/(d+4)})$, where d is the dimension of random vector of interest and n is the sample size. If we do not restrict ourselves to the class of densities in choosing the kernel function, we can attain a better MISE result by using a higher order kernel function stated below. It decreases the order of bias without affecting variance order and thus we can make MISE smaller asymptotically. However it is also known higher order kernels tend to inflate the variance through its multiplicative constant. Then higher order kernels may increase MISE in total depending on the relative effect of bias reduction and variance inflation in small sample. Also the MISE depends on the true density, and thus it is hard to say if we should use a higher order kernel or not practically. Furthermore, as claimed in Scott [1992, p.140], "the search for optimal higher order kernel is not so fruitful…Choosing among higher order kernels is quite complex and it is difficult to draw guidelines", so that higher order kernel may not be a common tool in density estimation.

In semiparametric framework however, there are some cases where we need to use higher order kernels. It is well known that we can estimate parametric components of semiparametric models $\sqrt{n}$ -consistently and asymptotically normally in many cases. In order for the $\sqrt{n}$ -consistency, we typically need to use higher order kernel because $\sqrt{n}$ times the bias inherited from the nonparametric (density) estimator must converge to zero. This paper considers this situation for the estimate of density-weighted averages associated with some semiparametric models, and proposes how to determine the bandwidth and kernel order. We take a standard approach that we do it such that they minimize the MSE of the estimator. MSE typically involves unknown functions so that we approximate it by bootstrap. It is known in kernel density estimation that standard bootstrap does not work to evaluate the bias, but smoothed bootstrap works for it as shown in Taylor [1989]. We observe a similar phenomenon in our problem, but, unlike density estimation, smoothed bootstrap bias itself does not well approximate the true bias. We show we need to make a linear transformation to it. We also show that the bootstrap MSE is a second order approximation to the true MSE.

The following section describes density weighted derivatives and illustrate some related semiparametric models and their estimators. Section 3 give MSE and bootstrap MSE of the estimator explained in Section 2. Section 4 states the main result, while Section 5 proposes how to calculate bootstrap moments.

2. DENSITY WEIGHTED AVERAGES

Density weighted average (abbreviated to DWA hereafter) considered in Powell and Stoker [1996] is a nonparametric quantity expressed as

$$\delta = E[k(X,W)f(X)]$$

where X is a $d \times 1$ vector of random variables with unknown joint density $f(x)$, W is a vector of some other random variables, and $k: R^d \to R^q$ is a function. We are concerned with the case when δ is estimated by a statistic with the following U-statistic form given a random sample $Z_i = (X_i, W_i),\ i = 1,\dots,n$.

$$\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} U_{ij}\ ,\ U_{ij} = U(Z_i, Z_j; h) \qquad (1)$$

where $U(x,y;h)$ is a symmetric function with respect to x and y, and h is a parameter decaying to zero as $n \to \infty$. The simplest example is averaged density $E[f(X)]$ where $k(x,w)$=1. Given a random sample $X_1,\dots,X_n$, let

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

be leave-one-out kernel estimator of the density $f(x)$, where $K(.)$ is a symmetric kernel function which integrates to unity and h is the bandwidth. An estimator for $E[f(X)]$ is

$$\frac{1}{n}\sum_{i=1}^{n} \hat{f}(X_i) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{h^d} K\left(\frac{X_j - X_i}{h}\right).$$

Also δ is associated with parametric components of some semiparametric models. For example, density weighted conditional covariances

$$\alpha_{1y} = E[(X_1 - E(X_1 \mid X_2))(Y - E(Y \mid X_2)) f(X_2)]$$

and

$$\alpha_{11} = E[(X_1 - E(X_1 \mid X_2))(X_1 - E(X_1 \mid X_2))^\tau f(X_2)],$$

τ denoting transpose, are useful to estimate the parametric component of partly linear regression model:

$$Y = \beta^\tau X_1 + \theta(X_2) + \varepsilon\ ,\ E(\varepsilon \mid X_1, X_2) = 0$$

where $\theta(.)$ is an unknown function. It is easily seen $\beta = \alpha_{11}^{-1}\alpha_{1y}$. Thus if we have consistent estimates of α_{11} and α_{1y} , we obtain a consistent estimate of β. Natural estimators for α_{11} and α_{1y} are respectively

$$\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2h^d} K\left(\frac{X_{2j} - X_{2i}}{h}\right)(X_{1j} - X_{1i})(Y_j - Y_i)$$

and

$$\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2h^d} K\left(\frac{X_{2j} - X_{2i}}{h}\right)(X_{1j} - X_{1i})(X_{1j} - X_{1i})^\tau$$

Another example is density weighted averaged derivatives for a semiparametric index model

$$Y = G(\beta^\tau X) + \varepsilon$$

where $G(.)$ is an unknown function and β is the parameter vector of interest. Putting $g(X) = G(\beta^\tau X)$, we have

$$\beta = cE[g'(X)f(X)] = -2cE[f'(X)Y]$$

for some constant c. The second equality requires some conditions on f and g (see e.g. Powell, Stock and Stoker [1989]). The expectation on the right is estimated by

$$\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{h^{d+1}} K'\left(\frac{X_j - X_i}{h}\right)(Y_j - Y_i)\ . \qquad (2)$$

In this paper we are concerned with the following estimator for δ ,

$$\hat{\delta} = \frac{1}{n}\sum_{i=1}^{n} k(X_i, W_i)\hat{f}(X_i) \qquad (3)$$

which is shown to have a representation (1) with

$$U_{ij} = \frac{1}{2h^d}\{k(X_i, W_i) + k(X_j, W_j)\} K\left(\frac{X_i - X_j}{h}\right)$$

(2) has a slightly different form from (3) because we take derivatives of density. We note, however, (2) also has a U-statistic form and can be handled similarly. (3) has a standard U-statistic form, though the kernel (in U-statistic sense) depends on n through h. The variance of the kernel is indeed infinite asymptotically and thus standard asymptotic theory for U-statistics does not apply. Nevertheless we can show $\hat{\delta}$ is consistent for δ and $\sqrt{n}(\hat{\delta} - \delta) \to^d N(0,V)$ for some positive definite matrix V under some regularity conditions. Heuristically, these results are explained as follows. Writing

$$\hat{\delta}-\delta=\frac{2}{n}\sum_{i=1}^{n}V_i+\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}W_{ij}+\{E(\hat{\delta})-\delta\}$$
$$=(I) \quad + \quad (II) \quad + \quad (III)$$

where

$$U_i=E(U_{ij}\mid W_i,X_i),\ V_i=U_i-E(\hat{\delta}),$$

$$W_{ij}=U_{ij}-E(\hat{\delta})-V_i-V_j,$$

we can show that all of (*I*), (*II*), (*III*) converge to zero in probability, while

$$\sqrt{n}\times(I)\to^{d} N(0,V),$$

$$\sqrt{n}\times(II)\to^{p} 0,\ \sqrt{n}\times(III)\to 0. \tag{4}$$

In order to show (4), we need to use the following higher order kernel and restrict the order of bandwidth. When *K*(.) satisfies

$$\int u_1^{i_1}\cdots u_d^{i_d}K(u)du=\begin{cases}=1, & i_1=\cdots=i_d=0\\ =0, & 0<i_1+\cdots+i_d\le L-1\\ \neq 0, & i_1+\cdots+i_d=L\end{cases} \tag{5}$$

for positive integers $i_1,\cdots,i_d$, it is called a higher order kernel and *L* is the kernel order. Supposing we use this kernel function in $\hat{\delta}$, the bandwidth must satisfy

$$n^{-1}h^{-d}+\sqrt{n}h^{L}\to 0 \tag{6}$$

for $\sqrt{n}$ – consistency. The first and second terms correspond to the variance of $\sqrt{n}\times(II)$ and $\sqrt{n}\times(III)$ respectively. A necessary condition for the existence of *h* satisfying (6) is *L*>*d*/2. Therefore if $d\ge 2$, we need to use a higher order kernel.

3. MSE AND BOOTSTRAP MSE FOR DWA

We obtain the leading terms of MSE of DWA in Section 3.1 and those for its bootstrap version in Section 3.2. We show that standard bootstrap does not work for bias estimation, but smoothed bootstrap approximates the bias after certain modification.

3.1. MSE

In order to simplify expressions, we describe the case of *d*=1, *q*=1 and $k(x,w)=k(x)$. For $d\ge 2$, we obtain qualitatively the same results. Note

$$\delta=\int k(x)f(x)^2dx.$$

Decomposing the MSE as

$$E(\hat{\delta}-\delta)^2=E\{\hat{\delta}-E(\hat{\delta})\}^2+\{E(\hat{\delta})-\delta\}^2,$$

we evaluate the variance and bias separately. Since $U_{12}=U_{21}$ and *K*(.) is symmetric, we have

$$E(\hat{\delta})=E(U_{12})=E[\frac{1}{h^d}k(X_1)K(\frac{X_1-X_2}{h})]$$
$$=\iint\frac{1}{h^d}k(x)K(\frac{x-y}{h})f(x)f(y)dxdy$$
$$=\iint k(x)K(u)f(x)f(x-hu)dxdu\ . \tag{7}$$

Thus letting $f^{(L)}(x)=d^Lf(x)/dx^L$ and $C_K=\int u^LK(u)du$, we have

$$E(\hat{\delta})=\int k(x)f(x)^2dx$$
$$+\frac{C_Kh^L}{L!}\int k(x)f(x)f^{(L)}(x)dx+o(h^L)\ ,$$

using (5). Therefore the bias is

$$E(\hat{\delta})-\delta=\frac{C_Kh^L}{L!}\int k(x)f(x)f^{(L)}(x)dx+o(h^L). \tag{8}$$

As is standard in U-statistic theory, we have

$$Var(\hat{\delta})=\frac{4}{n}Var(V_1)+\binom{n}{2}^{-1}Var(W_{12}).$$

Write $Var(V_1)=E(U_1^2)-\{E(U_1)\}^2$. Since

$$U_i=E(U_{ij}\mid X_i)$$
$$=\frac{1}{2h^d}\int\{k(X_i)K(\frac{X_i-x}{h})+k(x)K(\frac{x-X_i}{h})\}f(x)dx$$
$$=\frac{1}{2}\int\{k(X_i)+k(X_i-hu)\}K(u)\}f(X_i-hu)du,$$

we have

$$E(U_1^2)$$
$$=\int\{\int\frac{1}{2}\{k(x)+k(x-hu)\}f(x-hu)K(u)du\}^2f(x)dx.$$
$$=\int k(x)^2f(x)^3dx$$
$$+\frac{C_Kh^L}{L!}\int\{e^{(L)}(x)+k(x)f^{(L)}(x)\}k(x)f(x)^2dx+o(h^L)$$

where $e(x)=k(x)f(x)$. Writing

$$Var(W_{12})=E(U_{12}^2)-\{E(U_1)\}^2-2Var(V_1),$$

the second and third terms on the right are derived above. Taking iterated expectations, the first term equals to

$$\int\int\frac{1}{4h^d}\{k(x)+k(x-hu)\}^2K(u)^2f(x-hu)^2dudx$$
$$=\frac{1}{h^d}\int\int k(x)^2K(u)^2f(x)^2dudx+o(h^{-d})$$

Therefore, the MSE is

MSE($\hat{\delta}$)=$E(\hat{\delta}-\delta)^2$

$$=\frac{1}{n}\int k(x)^2 f(x)^3\,dx$$
$$+\frac{C_K h^L}{L!\,n}\int \{e^{(L)}(x)+k(x)f^{(L)}(x)\}k(x)f(x)^2\,dx$$
$$-\frac{1}{n}\{\int k(x)f(x)^2\,dx\}^2-\frac{C_K h^L}{L!\,n}\int k(x)f(x)^2\,dx$$
$$\times\int \{e^{(L)}(x)+k(x)f^{(L)}(x)\}f(x)dx$$
$$+\frac{1}{n^2h^d}\int\int k(x)^2K(u)^2f(x)^2\,dudx$$
$$+\frac{C_K^2h^{2L}}{(L!)^2}\{\int k(x)f(x)f^{(L)}(x)dx\}^2$$
$$+o(n^{-1}h^L+n^{-2}h^{-d}+h^{2L}). \qquad (9)$$

It is an optimal way to determine h and L such that leading terms of (9) is minimized. We need to check out theoretically if they exist, however, it is not a simple question, and without restricting the class of kernel function as in e.g. Hall and Marron [1987] in the context of density estimation, it seems virtually impossible. So we do not treat this problem in this work.

We implemented a small Monte Carlo study for the case of averaged density and found a result that MSE is smooth in h for fixed L, but not necessarily smooth in L for fixed h.

3.2. Bootstrap MSE

Section 3.1 shows leading terms of MSE of $\hat{\delta}$. We would like to select h and kernel order L such that it is minimized. However, it is infeasible because the MSE involves unknown density f. We employ a bootstrap method to approximate the MSE.

Given a random sample $X_1,\ldots,X_n$ from a distribution $H(x)$, suppose we would like to make a statistical inference on a quantity $\theta(H)$. A consistent estimator for this quantity is $\theta(\hat{H})$ where $\hat{H}$ is the empirical distribution function. Random sample from $\hat{H}$ given $X_1,\ldots,X_n$ is called a (standard) bootstrap sample. Let $X_1^*,\ldots,X_n^*$ be a bootstrap sample and $\hat{H}^*$ be its empirical distribution function conditional on the original sample. Under some conditions, the distribution of $\theta(\hat{H})-\theta(H)$ is close to that of $\theta(\hat{H}^*)-\theta(\hat{H})$, then we can make a statistical inference on $\theta(\hat{H})-\theta(H)$ using $\theta(\hat{H}^*)-\theta(\hat{H})$. This method is called standard bootstrap. When we know H has a smooth density, we may replace $\hat{H}$ in the above by $\tilde{H}$, a smooth estimator of H. It is called smoothed bootstrap.

Our interest is to obtain a feasible estimator of the MSE for DWA by bootstrapping. It is known, in density estimation, standard bootstrap fails to approximate the bias, but smoothed bootstrap well approximates the bias (see Taylor [1989]). We observe a similar but slightly different phenomenon in the case of DWA. For smoothed bootstrap, we use a kernel density estimate.

Let $X_1^*,\ldots,X_n^*$ be a standard bootstrap sample, and $X_1^+,\ldots,X_n^+$ be a smoothed bootstrap sample from $\tilde{f}(x)$ where

$$\tilde{f}(x)=\frac{1}{n-1}\sum_{i=1}^{n}\frac{1}{h^d}K(\frac{X_i-x}{h})$$

and $K(.)$ is an L-th order kernel. Also define

$$\hat{\delta}^*=\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}U_{ij}^*,\quad \hat{\delta}^+=\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}U_{ij}^+$$

where

$$U_{ij}^*=U(X_i^*,X_j^*),\ U_{ij}^+=U(X_i^+,X_j^+),$$
$$U(x,y)=\frac{1}{2h^d}\{k(x)+k(y)\}K(\frac{x-y}{h}).$$

Letting $E^*(.)=E(.\mid X_1,\ldots,X_n)$, we investigate smoothed bootstrap expectation $E^*(\hat{\delta}^+)$ and variance $E^*\{\hat{\delta}^+-E^*(\hat{\delta}^+)\}^2$. We also calculate standard bootstrap expectation for comparison.

$$E^*(\hat{\delta}^+)=E^*\{U(X_1^+,X_2^+)\}$$
$$=\iint U(x,y)\tilde{f}(x)\tilde{f}(y)dxdy$$
$$=\iint k(x)K(u)\tilde{f}(x)\tilde{f}(x-hu)dudx$$

Comparing this expression with (7), we immediately find the similarity, however, if we take the standard bootstrap expectation, we obtain

$$E^*(\hat{\delta}^*)=\frac{n-1}{n}\hat{\delta}+\frac{1}{n^2}\sum_{i=1}^{n}U(X_i,X_i),$$

so that $E^*(\hat{\delta}^*)-\hat{\delta}$ obviously does not work for estimating the bias. Straightforward calculation gives, using (5),

$$E[E^*(\hat{\delta}^+)]=\int k(x)[f(x)+\frac{3C_Kh^L}{L!}f^{(L)}(x)]f(x)dx$$
$$+\frac{1}{nh^d}\iint K(u)K(v)K(u+v)dudv\int k(x)f(x)dx$$
$$+o(h^L+n^{-1}h^{-d}),$$

and therefore

$$\frac{1}{2}E[E^*(\hat{\delta}^+)-\hat{\delta}-\frac{\gamma}{nh^d}]$$

$$=\frac{C_K h^L}{L!}\int k(x)f(x)f^{(L)}(x)dx+o(h^L+n^{-1}h^{-d})$$

$$=E(\hat{\delta}-\delta)+o(h^L+n^{-1}h^{-d}) \;, \qquad (10)$$

using (8), where

$$\gamma=\frac{1}{nh^d}\iint K(u)K(v)K(u+v)dudv\int k(x)f(x)dx.$$

Also we have

$$Var[E^*(\hat{\delta}^+)]=Var(\hat{\delta})+o(n^{-1})=O(n^{-1})\,. \quad (11)$$

(10) and (11) yield

$$\frac{1}{2}[E^*(\hat{\delta}^+)-\hat{\delta}-\frac{\gamma}{nh^d}]$$

$$=E(\hat{\delta}-\delta)+o_p(n^{-1/2}+n^{-1}h^{-d}+h^L)\,. \quad (12)$$

(12) shows that a linear transformation of the bootstrap bias well approximates the bias of $\hat{\delta}$. γ is an unknown quantity due to unknown $f(x)$, but we have

$$\hat{\gamma}=\frac{1}{nh^d}\iint K(u)K(v)K(u+v)dudv\frac{1}{n}\sum_{i=1}^{n}k(X_i)$$

$$\to^p \gamma\,. \qquad (13)$$

On the bootstrap variance, we write as is standard in U-statistic theory

$$Var^*(\hat{\delta}^+)=E^*[\hat{\delta}^+-E^*(\hat{\delta}^+)]^2$$

$$=\frac{2}{n}E^*({V_1^+}^2)+\binom{n}{2}^{-1}E^*({W_{12}^+}^2)$$

where

$$U_i^+=E^*(U_{ij}\mid X_i),\; V_i^+=U_i^+-E^*(\hat{\delta}^+),$$

$$W_{ij}^+=U_{ij}^+-E(\hat{\delta})-V_i^+-V_j^+\,.$$

Tedious but straightforward calculation gives

$$E[E^*({V_1^+}^2)-E(V_1^2)]^2=o(1),$$

$$E[E^*({W_{12}^+}^2)-E(W_{12}^2)]^2=o(1)\,.$$

Therefore, we have

$$Var^*(\hat{\delta}^+)=Var(\hat{\delta})+o_p(n^{-1/2}+h^L)\,. \qquad (14)$$

Thus combining (12), (13) and (14), we have the following proposition.

PROPOSITION

Under some regularity conditions on smoothness, boundedness and integrability of k, f and their derivatives, we have

$$E^*[\hat{\delta}^+-E^*(\hat{\delta}^+)]^2+[\frac{1}{2}\{E^*(\hat{\delta}^+)-\hat{\delta}-\frac{\hat{\gamma}}{nh^d}\}]^2$$

$$=E(\hat{\delta}-\delta)^2+o_p(n^{-1/2}+n^{-1}h^{-d}+h^L)\,. \quad (15)$$

The left hand side is not exactly the bootstrap MSE of $\hat{\delta}$, in the sense that we subtract $\hat{\gamma}/nh^d$ from the bias and multiply it by 1/2 which adjusts the bias of the bootstrap bias estimate of $\hat{\delta}$. But (15) claims it gives a sensible approximate to the MSE $E(\hat{\delta}-\delta)^2$.

Also we would like to remark that there is a possibility depending on $f(x)$ and $k(x)$ that standard bootstrap variance approximates the true variance better than the smoothed bootstrap.

4. BANDWIDTH AND KERNEL ORDER SELECTION

Previous section shows that the bootstrap MSE approximates the true MSE in probability up to second order. This asymptotically justify the method of selecting bandwidth and kernel order such that they minimize the bootstrap MSE, or more precisely,

$$\min_{h,L}E^*[\hat{\delta}^+-E^*(\hat{\delta}^+)]^2+\frac{1}{4}\{E^*(\hat{\delta}^+)-\hat{\delta}-\frac{\hat{\gamma}}{nh^d}\}^2$$

$$=\min_{h,L}E(\hat{\delta}-\delta)^2,$$

asymptotically.

There are two open questions practically. One is that how to generate smoothed bootstrap samples from a nonparametric kernel density with higher order kernel. If we use a density for the kernel, we easily obtain a smoothed bootstrap sample by drawing a random sample from the kernel multiplying it by h and adding it to the ordinary bootstrap sample, namely

$$X_i^+=X_i^*+h\varepsilon_i,\; i=1,2,... \qquad (16)$$

where ε_i, $i=1,2,...$ is a random sample from density $K(.)$ (see e.g. Davison and Hinkley [1997], p.p.79). It is a common way of generating a smoothed bootstrap sample, but higher order kernel functions does not satisfy $K(.)>0$ and thus we cannot generate a random sample from $K(.)$. In our case however, we do not need to generate a smoothed bootstrap sample because what we need is not the sample itself, but its conditional expectation. We explain it in the next section.

The second problem is how to determine the class of kernel function. There are some ways of constructing higher order kernels from a density. See for example, Robinson [1988], Hall and Marron [1989], Jones and Foster [1993] and Wand and Jones [1995]. Practitioners will take one of these methods, but there still remains a problem of choosing the base density. As often claimed in density estimation problem, kernel shape itself may not be a too critical matter. It can be true here. We leave this topic for future research.

5. SMOOTHED BOOTSTRAP MOMENTS WITH HIGHER ORDER KERNELS

As stated in the above, when we use a higher order kernel density estimate, we have a technical problem in generating a smoothed bootstrap sample. However, in our problem we need not generate a smoothed bootstrap sample because what we need is not the sample itself, but the conditional first and second order moments $E^*(\hat{\delta}^+)$ and $E^*[\hat{\delta}^+ - E^*(\hat{\delta}^+)]^2$. In the case of averaged density for instance, the former is

$$E^*(\hat{\delta}^+) = E^*[\frac{1}{h}K(\frac{X_1^+ - X_2^+}{h})]$$

$$= \iint \frac{1}{h}K(\frac{x-y}{h})\tilde{f}(x)\tilde{f}(y)dudv$$

$$= \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n}\iint K(u)K(v)K(\frac{X_i - X_j}{h} + u + v)dudv. \tag{17}$$

Formula for the second order moment is omitted, but it is also similarly obtained. We simply need to work out the integral. However, for relatively higher order kernels, it could be quite hard to calculate the double integral in (17). There is a way to avoid it for certain higher order kernels, as proposed in Nishiyama and Robinson [2003] in the context of averaged derivative estimation. For example, if we use e.g. a higher order kernel as in Robinson [1989], L-th order kernel function is represented by

$$K(u) = m_L(u)g(u)$$

where $g(u)$ is a density and $m_L(u)$ is a (L-2)-th polynomial whose coefficients depend on the moments of $g(u)$. For example, in the case when $g(u) = \phi(u)$,

$$m_4(u) = \frac{3-u^2}{2},\quad m_6(u) = \frac{15-10u^2+u^4}{8},\ \cdots$$

Then in the case of 4-th order kernel, for a given x,

$$\int K(u)K(v)K(x+u+v)dudv$$

$$= \int m_4(u)m_4(v)K(x+u+v)\phi(u)\phi(v)dudv$$

$$= E[m_4(U)m_4(V)K(x+U+V)] \tag{18}$$

where U and V are independent standard normal variates. Thus given a random sample $(U_1,V_1),....,(U_m,V_m)$ for some sufficiently large integer m, we can approximate (18) by its sample analogue

$$\frac{1}{m}\sum_{k=1}^{m} m_4(U_k)m_4(V_k)K(x+U_k+V_k),$$

therefore, a natural approximate to (17) is

$$\frac{1}{n^2 mh}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} m_L(U_k)m_L(V_k)K(\frac{X_i - X_j}{h} + U_k + V_k).$$

We note that without $m_L(.)$ factors, this is quite similar to smoothed bootstrap expectation with a second order kernel conditional on the additive smoother ε_i in (16). The difference is how X is contaminated.

6. REFERENCES

Davison, A.C. and D.V. Hinkley, Bootstrap Methods and their Application, Cambridge University Press, 1997.

Hall, P. and J.S. Marron, Choice of Kernel Order in Density Estimation, Annals of Statistics, 16, 1, 161-173, 1987.

Jones, M.C. and P.J. Foster, Generalized jackknifing and higher order kernels, Journal of nonparametric statistics, 3, 81-94.

Nishiyama, Y. and P.M. Robinson, The Bootstrap and Edgeworth Correction for Semiparametric Averaged Derivatives, mimeo, 2003.

Powell,J.L., J.H. Stock and T.M. Stoker [1989], Semiparametric Estimation of Index Coefficients, Econometrica, 57, 1403-1430.

Powell,J.L. and T.M. Stoker [1996], Optimal Bandwidth Choice for Density-Weighted Averages, Journal of Econometrics, 75, 291-316.

Robinson, P.M., Root-N Consistent Semiparametric Regression, Econometrica, 56, 4, 931-954, 1988.

Scott, D.W., Multivariate Density Estimation, John Wiley, 1992.

Silverman, B.W., Density Estimation, Chapman & Hall, London, 1986

Wand, M.P. and M.C. Jones, Kernel Smoothing, Chapman & Hall, 1995.