THE DELTA METHOD
FOR NONPARAMETRIC KERNEL FUNCTIONALS

Yacine AÔt-Sahalia
Graduate School of Business
University of Chicago
1101 E 58th St
Chicago, IL 60637

January 1992
Revised, August 1994

1 This paper is part of my Ph.D. dissertation at the MIT Department of Economics. I am very indebted to
Richard Dudley, Jerry Hausman and Whitney Newey for helpful suggestions and advice. The comments of
Peter Robinson and three anonymous referees have considerably improved this paper. I am also grateful to
seminar participants at Berkeley, British Columbia, Carnegie-Mellon, Chicago, Columbia, Harvard,
Northwestern, Michigan, MIT, Princeton, Stanford, Washington University, Wharton, Yale and the Yale-
NSF Conference on Asymptotics for Infinite Dimensional Parameters as well as Emmanuel Guerre, Pascal
Massart and Jens Praestgaard for stimulating conversations. Part of this research was conducted while I was
visiting CREST-INSEE in Paris whose hospitality is gratefully acknowledged. Financial support from
Ecole Polytechnique and MIT Fellowships is also gratefully acknowledged. All errors are mine.
The Delta Methods for Nonparametric Kernel Functionals

by Yacine AÔt-Sahalia

Abstract

This paper provides under weak conditions a generalized delta method for functionals of nonparametric kernel estimators, based on (possibly) dependent and (possibly) multivariate data. Generalized derivatives are allowed to permit the inclusion of virtually any functional, global or pointwise, explicitly or implicitly defined. It is shown that forming the estimator with dependent data modifies the asymptotic distribution only if the functional is more irregular than some threshold level. Variance estimators and rates of convergence are derived. Many examples are provided.

Keywords: Delta Method, Nonparametric Kernel Estimation, Dependent Data, Rate of Convergence, Functional Differentiation, Generalized Functions.
1. Introduction

The delta method is a simple and widely used tool to derive the asymptotic distribution of nonlinear functionals of an estimator. Most parametric estimators converge at the familiar root-n case, and so does the functional. When the estimator is nonparametric, however, some functionals will converge at a rate slower than root-n while others will retain the root-n rate. The slower than root-n functionals require some form of smoothing to be estimated, the most popular being the kernel method. A delta method has long been available for the class of root-n functionals that can be estimated without smoothing (von Mises [1947], Reeds [1976], Huber [1981], Dudley [1990]).

In all cases, the essence of the delta method is a first order Taylor expansion of the functional. The problem is that the slower-than-root-n functionals are not differentiable in the usual sense. Therefore the examples of slower-than-root-n functionals studied in the literature have been tackled without the systematic "plug-and-play" feature that made the delta method attractive in the settings where it was available. This paper proposes a simple delta method that covers also slower-than-root-n functionals, and under conditions that equal and often relax those used in the previous "case-by-case" work. To address the problem of non-differentiability, the paper allows generalized functions as functional derivatives. An example of a generalized function is the Dirac delta function, and its derivatives. With generalized functions, the familiar delta method approach based on differentiating the functional is shown to be easily implemented for non-trivial examples. The results are valid even when the data are serially correlated, with independent data as a special case.

The main contribution of the paper is to show how to linearize systematically slower-than-root-n functionals, and then to provide a general yet simple result yielding their asymptotic distribution. The purpose of the examples is two-fold: first, some classical
examples (regression function, etc.) are included to show how the method of this paper significantly beats the previous approaches; second, new distributions are derived for cases where they were not previously available (dependent censored least absolute deviation, quantiles, mode, stochastic differential equations, etc.). The paper is organized as follows: Section 2 derives the generalized delta method. Consistent estimators of the asymptotic variances are proposed in Section 3. Section 4 discusses the rates of convergence of the estimators. Section 5 illustrates the application of the result through many examples. Section 6 concludes. Proofs are in the Appendix.

2. The Delta Method with Generalized Derivatives

2.1 Assumptions

Consider $\mathbb{R}^d$-valued random variables $X_1, X_2, \ldots, X_n$ identically distributed as $f(.)$, an unknown density function with associated cumulative density function $F(x) \equiv \int_{-\infty}^{x} f(t)dt$ where $x \equiv (x_1, x_2, \ldots, x_d)$. The following regularity conditions are imposed:

Assumption A1: The sequence $\{X_i\}$ is a strictly stationary $\beta$-mixing sequence satisfying:

$$k^\delta \beta_k \xrightarrow{k \to \infty} 0 \text{ for some fixed } \delta > 1.$$ 

$\beta_k = 0, \forall k \geq 1$ corresponds to the independence case. As long as $\beta_k \xrightarrow{k \to \infty} 0$, the sequence is said to be absolutely regular.

Assumption A2: The density function $f(.)$ is continuously differentiable on $\mathbb{R}^d$ up to order $s$. Its successive derivatives are bounded and in $L^2(\mathbb{R}^d)$.

Let $C^s$ be the space of density functions satisfying A2. To estimate the density function $f(.)$, a Parzen-Rosenblatt kernel function $K(.)$ will be used. The kernel will be required to satisfy:
**Assumption A3:**

(i) $K$ is an even function integrating to one;

(ii) The kernel is of order $r = s$, an even integer:

1) $\forall \lambda \in \mathbb{N}^d / |\lambda| = \lambda + \ldots + d \lambda \in \{1, \ldots, r-1\}, \int_{-\infty}^{\infty} \frac{x^{i\lambda} \ldots d x^{i\lambda} K(x)dx}{\lambda_{\lambda_{\lambda_{\lambda}}}} = 0$;

2) $\exists \lambda \in \mathbb{N}^d / |\lambda| = r$ and $\int_{-\infty}^{\infty} \frac{x^{i\lambda} \ldots d x^{i\lambda} K(x)dx}{\lambda_{\lambda_{\lambda_{\lambda}}}} \neq 0$;

3) $\int_{-\infty}^{\infty} \|x\|^r |K(x)|dx < +\infty$.

(iii) $K$ is continuously differentiable up to order $s+d$ on $\mathbb{R}^d$, and its derivatives of order up to $s$ are in $L^2(\mathbb{R}^d)$.

The last assumption indicates how the bandwidth $h_n$ in the kernel density estimator should be chosen. The statement of the assumption depends upon an exponent parameter $e > 0$ and an integer $m$, $0 \leq m \leq e/2$.

**Assumption A4(e,m):** As $n \to \infty$: $n^{1/2} h_n^e + \left(n^{1/2} h_n^{2m}\right)^{-1} \to 0$.

Each specific application requires that A4(e,m) be satisfied for some particular e and m. Finally, construct the kernel cumulative density function (KCDF) estimator: $\hat{F}_n(x) \equiv \int_{-\infty}^{x} \hat{f}_n(t)dt$. Consider also the empirical cumulative distribution function (ECDF) $F_n(x) \equiv \frac{1}{n} \sum_{i=1}^{n} 1(x - x_i)$.

### 2.2. Functional Differentiation

Consider a functional $\Phi[.]$ defined on an open subset of $C^s$ with the $L^2$ norm, and taking values in $\mathbb{R}$. Say that $\Phi$ is be $L(2,m)$-differentiable at $F$ in $C^s$ if it admits a first order Taylor expansion:

$$\Phi[F + H] = \Phi[F] + \Phi^{(1)}[F](H) + R_{\Phi}[F,H], \text{ with } |R_{\Phi}[F,H]| = O\left(\left\|H\right\|_{L(2,m)}^2\right)$$
where $\Phi^{(1)}[G](\cdot)$ is a continuous linear (in $H$) functional and $L(2,m)$ is the sum of the $L^2$ norm of the all derivatives of $H$ up to order $m$. If this holds uniformly on $H$ in any compact subset $K$ of $C^s$, and \[ |\Phi^{(1)}[G](H)| \leq C(K) \|H\|_{L(2,s)}, \]
then $\Phi$ is said to be $L(2,m)$-Hadamard-differentiable at $F$. In what follows it will always implicitly be assumed that the linear term $\Phi^{(1)}[F](\cdot)$ is not degenerate. If it were then the asymptotic distribution would be given by a term of higher order in the Taylor expansion.

By the Riesz Representation Theorem (see e.g., Schwartz [1966]), there exists a distribution $\varphi[F]: \mathbb{R}^d \mapsto \mathbb{R}$ such that $\Phi^{(1)}[F](H) = \int \varphi[F](x) dH(x)$. Call $\varphi[F](\cdot)$ the functional derivative\(^2\) of $\Phi$ at $F$. The standard delta method is applicable only if $\varphi[F](\cdot)$ is a regular function, i.e., at least cadlag (right-continuous, left-limit). For some functionals $\Phi$, the functional derivative will indeed be a regular function. For example, let $\Phi[F] \equiv \int_{-\infty}^{+\infty} f(x)^2 \, dx$. Then:

$$
\Phi[F + H] = \int_{-\infty}^{+\infty} \{f(x) + h(x)\}^2 \, dx = \int_{-\infty}^{+\infty} f(x)^2 \, dx + \int_{-\infty}^{+\infty} 2f(x)h(x) \, dx + \int_{-\infty}^{+\infty} h(x)^2 \, dx
$$

so $\mathcal{R}_\varphi[F,H] = \int_{-\infty}^{+\infty} h(x)^2 \, dx$ and $\Phi^{(1)}[F](H) = \int_{-\infty}^{+\infty} \varphi[F](x) \, dH(x) = \int_{-\infty}^{+\infty} 2f(x)h(x) \, dx$. Thus its functional derivative is: $\varphi[F](\cdot) = 2f(\cdot)$, a function in $C^s$.

Unfortunately, many functionals of interest in econometrics do not have "regular" functional derivatives, that is $\varphi[F](\cdot)$ will not be a cadlag function. Instead, it will be a

\(^2\) Although $\varphi[F]$ is not unique, the respective asymptotic distributions given by Theorem 3 are independent of the choice of $\varphi[F]$. One way to make the representation unique would be to impose that \[ \int \varphi[F](x) \, dF(x) = 0. \]
Any $\varphi[F]$ given by the Riesz Theorem will be called "the" derivative, even though "a" derivative would be more appropriate.
generalized function. The existing delta method cannot treat such functionals. The main point of the paper is that the same familiar delta approach will work provided that one includes generalized functions as functional derivatives. This method turns out to be very simple as well as powerful. Many examples will be provided below. To get the flavor of the result immediately, consider the hazard rate function \( \Phi[F] \equiv \frac{f(y)}{1-F(y)} \) evaluated at some \( y \). It will be shown below that it is differentiable in the extended sense of this paper, with functional derivative: \( \varphi[F](x) = \frac{1}{1-F(y)} \delta_{(0)}(x) + \frac{F^{(1)}(y)}{[1-F(y)]^2} 1(y \geq x) \). This functional derivative is a linear combination of a Dirac mass at \( y \) (a generalized function) and an indicator function (a regular function). The asymptotic distribution of the kernel estimator of the hazard rate will be driven by the Dirac term in the linear expansion, only the most unsmooth term counts.

2.3. Generalized Functions

The concept of generalized function, or distribution, was formally introduced by Schwartz [1954,1966]. Simply put, any function \( g \), DEFINED AS?? no matter how unsmooth, can be differentiated. Its derivative \( g^{(1)} \) is defined by its cross-products against smooth functions \( f \): \( \int_{-\infty}^{\infty} g^{(1)}(x)f(x)dx \equiv (-1)^{q} \int_{-\infty}^{\infty} g(x)f^{(q)}(x)dx \) in the univariate case, where \( f^{(1)} \) is a standard derivative. Of course, by integration by part, this reduces to the common definition of differentiability if \( g \) turns out to be a regular function.

For example, a Dirac function at 0 is defined by: \( \int_{-\infty}^{\infty} \delta_{(0)}(x)f(x)dx = f(0) \), and its derivative is given by \( \int_{-\infty}^{\infty} \delta^{(1)}_{(0)}(x)f(x)dx = -\int_{-\infty}^{\infty} \delta_{(0)}(x)f^{(1)}(x)dx = -f^{(1)}(0) \). Successive differentiation of the Dirac function is possible up to the number of derivatives that the functions \( f \) admit, to yield \( \int_{-\infty}^{\infty} \delta^{(q)}_{(0)}(x)f(x)dx = (-1)^{q} \int_{-\infty}^{\infty} \delta_{(0)}(x)f^{(q)}(x)dx = (-1)^{q} f^{(q)}(0) \). Besides Dirac functions, many other generalized functions can be constructed; see Schwartz [1954,1966] or Zemanian [1965] for examples.
This paper allows functional derivatives $\varphi[F](\cdot)$ to be generalized functions. I will define an increasing sequence of spaces of generalized functions. Each space will contain functions of a given "level of unsmoothness." It will be first shown that the asymptotic distribution of the plug-in depends on the particular space containing $\varphi[F](\cdot)$ and that the more unsmooth $\varphi[F](\cdot)$ is the slower the rate of convergence. Furthermore when $\varphi[F](\cdot)$ is more unsmooth than cadlag (i.e., when it is a generalized function instead of a regular function) it will be shown that constructing the estimator on correlated data does not affect its asymptotic variance.

Start by defining the space $C^{-1}$ of bounded cadlag functions from $[0,1]^d$ to R. $C^{-1}$ contains all the usual spaces $C^0$, $C^1$, etc. of continuous, continuously once-differentiable functions, etc. The regular functions are the elements of $C^{-1}$. Now define $C^{-2}$ to be the space of linear combinations of Dirac functions and functions of $C^{-1}$, $C^{-3}$ to be the space of linear combinations of derivatives of Dirac functions and functions of $C^{-2}$, etc. When $\varphi[F](\cdot)$ belongs to the generalized function space $C^{-q}$, $q \geq 2$, but not to the space immediately smaller $C^{-q+1}$, write $\varphi[F] \in C^{-q} \setminus C^{-q+1}$. $q$ can readily be interpreted as an "order of unsmoothness" of $\varphi[F](\cdot)$. Moving up the following scale, the functions become more unsmooth, and conversely:
II.4. The Generalized Delta Method

The result is stated in dimension $d=1$. An extension to the multivariate case is provided in the Appendix. $\varphi[F] \in C^{-q} \setminus C^{-q+1}$ has the form:

$$\varphi[F](x) = \sum_{i=1}^{L} \alpha_i[F](x) \delta_{(y_i)}^{(q-2)}(x) + B[F](x)$$

where each $y_i$ is a fixed point, $\alpha_i[F](\cdot) \in C^{-1}$ and $B[F](\cdot) \in C^{-q+1}$ (see Section 5 for examples).

The following delta method characterizes the asymptotic distribution of the plug-in functional as a function of the particular space $C^{-q}$ where the functional derivative $\varphi[F](\cdot)$ lies:
**Theorem:** Suppose that $\Phi$ is $L(2,m)$-Hadamard-differentiable at the true cdf $F$ with functional derivative $\phi[F](\cdot)$. Then under A1-A3:

(i) If $\phi[F] \in C^{-1}$, then under $A4(r,m)$:

$$n^{1/2} \left\{ \Phi(\hat{F}_n) - \Phi(F) \right\} \xrightarrow{d} N(0, V_\phi[F])$$ with asymptotic variance:

$$V_\phi[F] = \text{VAR} \left( \phi[F](x_i) \right) + 2 \sum_{k=1}^{\infty} \text{COV} \left( \phi[F](x_i), \phi[F](x_{i+k}) \right)$$

(ii) If $\phi[F] \in C^{-q} \setminus C^{-q+1}$ for some $q \in \mathbb{Z}$, then under $A4(r+1/2, m)$:

$$h_n^{(2q-3)/2} \left\{ \Phi(\hat{F}_n) - \Phi(F) \right\} \xrightarrow{d} N(0, V_\phi[F])$$ with asymptotic variance:

$$V_\phi[F] = \int_{-\infty}^{\infty} \left| K^{(q)}(x) \right|^2 \, dx \sum_{l=1}^{L} \left\{ \alpha_l[F](y_l) \right\}^2.$$ The asymptotic variance is the same whether or not the data are serially dependent.

When the functional derivative is a regular function (that is $\phi[F] \in C^{-1}$), the result does not depend on how smooth it is beyond being cadlag. On the other hand, when it is a generalized function, the result depends on the exact degree of unsmoothness of the functional derivative (belonging to the space $C^{-q}$, but not the one immediately smoother, $C^{-q+1}$). Many (but not all, e.g. the integrated squared density) of the functionals with regular derivatives can be estimated without smoothing. In that case, the result is exactly the same whether a kernel or empirical cdf is plugged-in, and there is no reason to smooth.

For functionals with unsmooth derivatives however, smoothing is essential, as the plug-in cannot even be defined at the empirical cdf. And the asymptotic distribution is driven exclusively by the "most unsmooth" component of the functional derivative: the smoother component $B[F](\cdot) \in C^{-q+1}$ of $\phi[F] \in C^{-q} \setminus C^{-q+1}$ does not appear in the asymptotic variance. Such a functional (asymptotically) behaves essentially like a linear combination of the density or its derivatives that are not integrated upon. When $\phi[F] \in C^{-q} \setminus C^{-q+1}$, it can also be noted that not only the asymptotic variance contains no
time-series term, but also has no cross-covariances across the L terms in the functional derivative. For example, it is known that the kernel density evaluated at a point \( y_1 \) and at a different point \( y_2 \) are asymptotically uncorellated.

This brings the following remark. The slower-than-root-n functionals have a "local" character, such as the density evaluated at a point, or the mode of the density function. Consider for example the density function \( f(\cdot) \) and the local functional \( \Phi_y[F] \equiv f(y) \) (real-valued) as opposed to the global functional \( \Phi[F] \equiv f(\cdot) \) (\( C^s \)-valued).

Drawing from the experience of root-n functionals, it may be tempting to try to obtain weak convergence to a Gaussian process of \( \Phi[F] \equiv f(\cdot) \). Unfortunately, no such result holds for slower-than-root-n functionals. Indeed if it existed a limiting process for the normalized kernel density estimator, this process would have to take independent values \( W(t) \) and \( W(s) \) for every \( t \ldots s \).

The delta method derived here has an intuitive duality interpretation. The asymptotic distribution of an unsmooth functional \( \Phi \) is driven by the inner product \( \int_{-\infty}^{+\infty} \phi[F](x) \mathrm{d}H(x) \). When \( \phi[F] \) is a generalized function (in \( C^{-q}, q \geq 2 \)) then \( H \) must belong to \( C^{+q-1} \). Therefore one needs to have a sufficiently regular nonparametric estimator and unknown cdf. to plug-in as \( H = \hat{F}_n - F \). This is the role played by the kernel smoothing. If one uses the empirical distribution \( F_n \) instead of the KCDF \( \hat{F}_n \) then \( H = F_n - F \) will be in \( C^{-1} \) only, and therefore the only functionals that can be plugged-into must have derivatives in \( C^{-1} \).

3. **Consistent Estimation of the Asymptotic Variances**

The asymptotic variances given by the delta method can be consistently estimated in each case:
(i) If \( \varphi[F] C^{-1} \), then under \( A4(r,m) \) and the technical regularity condition \( A5 \) (given in the Appendix, and designed to guarantee that the truncated sum in the variance estimator will effectively approximate the infinite sum) the asymptotic variance \( V_\varphi[F] \) can be consistently estimated by:

\[
\hat{V}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi(\hat{F}_n)(x_i)^2 - \left( \frac{1}{n} \sum_{i=1}^{n} \varphi(\hat{F}_n)(x_i) \right)^2 + \sum_{k=1}^{G_n} \frac{2}{G_n + 1 - k} \left\{ \frac{1}{n} \sum_{i=1}^{n-k} \varphi(\hat{F}_n)(x_i) \varphi(\hat{F}_n)(x_{i+k}) - \left( \frac{1}{n} \sum_{i=1}^{n} \varphi(\hat{F}_n)(x_i) \right)^2 \right\}
\]

where \( G_n \) is a truncation lag chosen such that \( \lim_{n \to \infty} G_n = +\infty \) and \( G_n = O(n^{1/3}) \). This is an estimator of the spectral density at zero (see Newey-West [1987], Robinson [1989,1991]). The choice of the truncation lag \( G_n \) and the Bartlett kernel is subject to the same provisions and can be improved upon as in Andrews [1991] in the parametric case.

(ii) If \( \varphi[F] C^{-q} \setminus C^{-q+1} \), then under \( A4(r+1/2, m) \) the asymptotic variance \( V_\varphi[F] \) can be consistently estimated by:

\[
\hat{V}_n = \left\{ \int_{-\infty}^{+\infty} \left| K^{(q)}(x) \right|^2 dx \right\} \sum_{i=1}^{L} \{ \alpha_i \{ \hat{F}_n \}(y_i) \}^2.
\]

The appropriate estimate of the asymptotic variance makes it possible to construct confidence intervals on \( \Phi[\hat{F}_n] \) and carry out tests of general hypotheses regarding \( \hat{F}_n \). For example, to test the hypothesis \( H_0 : \Phi[F] = 0 \) versus \( H_1 : \Phi[F] \neq 0 \) one could simply use the following Wald-type test statistics:

\[
W_n \equiv \lambda(n) \Phi(\hat{F}_n) \hat{V}_n^{-1} \Phi(\hat{F}_n) \overset{d}{\underset{H_0}{\to}} \chi^2_{(1)}, \text{ where } \lambda(n) = \begin{cases} \frac{n}{n^{h_{(2q-3)}^2}} & \text{in (i)} \\ \frac{n}{n^{h_{(2q-3)}^2}} & \text{in (ii)} \end{cases}
\]

4. Rates of Convergence

The speed of decrease of the bandwidth to zero as the sample size increases is constrained by \( A4 \). The bandwidth can be chosen within the bounds allowed by \( A4 \) in
order to generate the fastest possible rate of convergence $\beta$ (the speed of convergence being $n^{-\beta}$).

(i) If $\varphi[F] \in C^{-1}$, then the plug-in will converge at rate $\beta = 1/2$. The root-n rate is achieved by kernel plug-ins under A3 no matter how $h_n$ is chosen within A4, and will produce an asymptotic distribution centered at zero.

(ii) If $\varphi[F] \in C^{-q} \setminus C^{-q+1}$ for some $q \in [2,s]$, then the rate of convergence is at best $\beta = (r - (q - 2))/(2r + 1)$. It can be achieved by kernel plug-ins under A3 when choosing $h_n$ of the order $n^{-\alpha}$, with $\alpha = 1/(2r + 1)$. The resulting asymptotic distribution of the plug-in will not be centered at zero. For any $\varepsilon > 0$, the rate of convergence $\beta - \varepsilon$ however can be achieved with a resulting asymptotic distribution centered at zero by choosing $h_n$ of the order $n^{-\alpha}$, with $\alpha = \lfloor 1/(2r + 1) \rfloor + \varepsilon \{1/(2q - 3)\}$. This choice is admissible under $A4(r+1/2,m)$. Given the optimal rates of Stone [1980] and Goldstein and Messer [1992], it therefore turns out that the kernel-type estimators can achieve the optimal rate (but if one insists on getting the optimal rate then the limiting distribution is not centered at zero).

5. Examples and Applications

Classical examples as well as new distributions are provided in this section to both show how the method can very easily yield classical results and provide new results.

Example 1: Ordinary Least Squares

The following trivial example illustrates the method in a very simple case, recovering the asymptotic distribution of classical parametric estimators. Consider a simple linear model: $y_i = x_i \beta + \varepsilon_i, \ E[\varepsilon_i | x_i] = 0$. Although at first sight a quintessentially parametric model, the linear regression model in fact makes no assumptions whatsoever regarding the distribution of the disturbances (other than uncorrelatedness with the
regressors). In that sense, the OLS estimator can be treated as a nonparametric estimator.

OLS estimates the functional:

\[
\beta = \frac{E[XY]}{E[X^2]} = \Phi(F) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x,y)dx\,dy \quad \text{and} \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f(x,y)dx\,dy
\]

by plugging-into this expression the empirical cdf \( F_n \):

\[
\hat{\beta}_{OLS} \equiv \Phi(F_n) = \frac{1}{n} \sum_{i=1}^{n} y_i x_i \sqcup \frac{1}{n} \sum_{i=1}^{n} x_i^2
\]

Now compute the functional derivative of \( \Phi \):

\[
\Phi[F + H] = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x y \{ f + h \}}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 \{ f + h \}} = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x y f}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f} + \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x y h}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f} \quad \text{and} \quad \{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f \}
\]

\[
= \Phi[F] + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi[F](x,y) h(x,y) \, dx\,dy + O\left( \|H\|_{L(2,1)}^2 \right)
\]

So the functional \( F \mapsto \Phi(F) = \beta \) is \( L(2,1) \)-Hadamard-differentiable at \( F \), and its derivative is \( \phi[F](u,v) = \frac{uv}{E[X^2]} - \frac{E[XY]u^2}{\{E[X^2]\}^2} \in C^{-1} \). Theorem (i) gives the asymptotic distribution of the plug-in (using either the empirical or the kernel estimator of \( F \)) for \( \phi[F] \in C^{-1} : n^{1/2} \{ \Phi(F_n) - \Phi(F) \} \xrightarrow{d} N(0, \text{VAR}_F) \), with asymptotic variance given by:

\[
\text{VAR}_F[\phi[F]] = 2 \sum_{k=1}^{+\infty} \text{COV}(\phi[F](y_1,x_1),\phi[F](y_{1+k},x_{1+k})
\]

Replacing the functional derivative \( \phi[F] \) by its expression and \( y_t-x_i \beta \) by \( \epsilon_t \), it is easy to check that this expression is equal to the classical OLS asymptotic variance

\[
V_{OLS} = \left( E[x_i^2]\right)^{-2} \left\{ E[\epsilon_i^2x_i^2] + 2 \sum_{k=1}^{+\infty} E[\epsilon_i \epsilon_{i+k} x_i x_{i+k}] \right\}
\]
Example 2: Least Absolute Deviations

Consider again the simple linear model \( y_t = x_t \beta + \varepsilon_t \) where \( x \) is \( K \)-variate, \( \beta \) is an unknown \( K \)-dimensional parameter vector. The identification assumption on \( \{\varepsilon_t / t = 1, \ldots, T\} \) consists now of being independent of \( \{x_t / t = 1, \ldots, T\} \) and having zero median. Let \( f_\varepsilon \) be the marginal density of the disturbances \( \varepsilon \). The least absolute deviation (LAD) estimator is defined by (with the minimum taken over a compact set):

\[
\hat{\beta}_{LAD} = \arg \min_{\beta} \frac{1}{T} \sum_{t=1}^{T} |y_t - x_t \beta|.
\]

The first order condition is:

\[
\frac{1}{T} \sum_{t=1}^{T} \text{sign}(y_t - x_t \hat{\beta}_{LAD}) x_t = 0,
\]

where \( \text{sign}(a) = -1 \) if \( a < 0 \) and 1 if \( a \geq 0 \). Because the \( \varepsilon \) process has zero median and is independent of the \( x \) process, \( E[\text{sign}(y - x \beta) x] = 0 \). Let \( \Phi(F) \equiv \beta \) (this is a \( \mathbb{R}^K \)-valued functional), where \( F \) is the joint cdf of \( (x,y) \), be defined by \( E[\text{sign}(y - x \Phi(F)) x] = 0 \). This provides an example of an implicitly defined functional. By construction \( \hat{\beta}_{LAD} = \Phi(F_n) \). Its asymptotic variance was obtained by Pollard [1990] and Phillips [1991].

Here the functional derivative is easy to compute:

\[
\phi_{LAD}[F](x,y) = \left( \frac{1}{2 f_\varepsilon(0)} \right) A^{-1} \text{sign}(y - x \beta) x' \in \mathbb{C}^{-1} \text{ where } A \equiv E[x', x], \text{ so:}
\]

\[
n^{1/2} \left\{ \hat{\beta}_{LAD} - \beta \right\} \overset{d}{\longrightarrow} \mathcal{N} \left( 0, \left( 2 f_\varepsilon(0) \right)^{-2} A^{-1} V A^{-1} \right), \text{ where:}
\]

\[
V \equiv \text{VAR} \left( \text{sign}(\varepsilon_i) x_i \right)
\]

\[
+ \sum_{k=1}^{\infty} \left\{ \text{COV} \left( \text{sign}(\varepsilon_i) x_i, \text{sign}(\varepsilon_{i+k}) x_{i+k} \right) + \text{COV} \left( \text{sign}(\varepsilon_{i+k}) x_{i+k}, \text{sign}(\varepsilon_i) x_i \right) \right\}
\]
Example 3: Censored Least Absolute Deviations

Powell [1984] extended the LAD estimator to the case where the dependent variable is censored, i.e., only $y_i = \max \{0, x_i \beta + \epsilon_i\}$ is observed. This case is typical of situations arising in a labor supply context. In that case, consider the CLAD estimator:

$$\hat{\beta}_{\text{CLAD}} \equiv \arg \min_{\beta} \frac{1}{T} \sum_{t=1}^{T} |y_t - \max \{0, x_t \beta\}|$$

Regularity conditions guaranteeing the identifiability of $\beta$, and the existence and unicity (asymptotically) of a solution are given by Powell [1984], and also assumed here. The population first order condition is: $E \left[ \text{sign} (y - \max \{0, x \beta\}) I(x \beta) x \right] = 0$ and let $\Phi(F) \equiv \beta$

so $\hat{\beta}_{\text{CLAD}} = \Phi(F_t)$. Now: $\Phi_{\text{CLAD}}(F)(x, y) = \left( \frac{1}{2 f_\epsilon(0)} \right) B^{-1} \text{sign}(y - \max \{0, x \beta\}) I(x \beta) x \in C^{-1}$

with $B \equiv E[I(x, \beta) x, x]$. Thus $n^{1/2} \left\{ \hat{\beta}_{\text{CLAD}} - \beta \right\} \overset{d}{\longrightarrow} N \left( 0, \left( 2 f_\epsilon(0) \right)^{-2} B^{-1} W B^{-1} \right)$,

$W \equiv \text{VAR}(\text{sign}(\epsilon_i) I(x, \beta) x_i) + \sum_{k=1}^{\infty} \left\{ \text{COV}(\text{sign}(\epsilon_i) I(x, \beta) x_i, \text{sign}(\epsilon_{i+k}) I(x_{i+k}, \beta) x_{i+k}) + \text{COV}(\text{sign}(\epsilon_{i+k}) I(x_{i+k}, \beta) x_{i+k}, \text{sign}(\epsilon_i) I(x, \beta) x_i) \right\}$

This asymptotic distribution for dependent data appears to be new.

Example 4: Integrated Functionals

Consider next the family of real-valued functionals of the following form, where $\omega(.)$ is a trimming function: $\Phi(F) \equiv \int_{-\infty}^{\infty} \omega(x) \psi(x, F^{(1)}(x), F^{(2)}(x), ..., F^{(m)}(x)) \, dx$. This class includes the information matrix giving the asymptotic variance of maximum likelihood estimators, the entropy measure, the average derivative estimators of Powell, Stock and Stoker [1989] and Robinson [1989], the integral of the squared density, etc. The functional derivative is: $\varphi[F](x) = \sum_{q=1}^{m} (-1)^{q-1} \frac{\partial^{q-1}}{\partial x^{q-1}} \left( \omega(x) \frac{\partial \psi}{\partial F^{(q)}(x)} (x, F^{(1)}(x), ..., F^{(m)}(x)) \right) C^{-1}$, so the plug-in will converge at rate root-n and have an asymptotic distribution sensitive to dependent data.
Example 5: Pointwise Estimation

Consider the classical example $\Phi_q[F] \equiv F^{(q)}(y)$, a derivative of the cdf evaluated at $y$. Then if $q=0$, $\Phi_q[F](x) = 1(y-x) \in C^{-1}$, while $\Phi_q[F](x) = \delta^{(q-1)}(y) \in C^{-q-1} \setminus C^{-q}$, hence for $q > 1:

\begin{align*}
\frac{h_n^{(2q-1)/2} n^{1/2}}{\sqrt{V_F[F]}} \left\{ \hat{F}_n^{(q)}(y) - F^{(q)}(y) \right\} \xrightarrow{d} N \left( 0, \int_{-\infty}^{+\infty} |K^{(q-1)}(x)|^2 \, dx \right) f(y).
\end{align*}

The extension to multivariate data is immediate given the multivariate result in the Appendix.

Example 6: Smooth Quantiles

Take $\Phi[F] = F^{-1}[y]$ for some $y$. In the independent case, smooth estimation of quantiles has been studied e.g., by Parzen [1979] and Silverman and Young [1987]. Here the functional derivative can be computed as: $\Phi[F](x) = -\frac{1}{F^{(1)}[F^{-1}(y)]} \left( F^{-1}(y) - x \right) \in C^{-1}$ so the asymptotic distribution will converge at rate root-n and have time-dependent terms. Letting $f$ be the joint density of observations at lag $k$, the asymptotic variance is:

\begin{align*}
V_\Phi[F] &= \left\{ y(l-y) + 2 \sum_{k=1}^{\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \{ f_k(s,t) - f(s)f(t) \} ds dt \right\} \\
&= \left( F^{(1)}[F^{-1}(y)] \right)^2
\end{align*}

This result also appears to be new. Weak convergence of the quantile process to a Gaussian process is proved using the same method (see AÔt-Sahalia [1993]).

Example 7: Mode

The mode of a unimodal univariate density, studied by Parzen [1962] for i.i.d. data, can be obtained by the following functional: $\Phi(F) = [F^{(2)}]^{-1}(0)$, that is the point at which the derivative of the density is zero. The functional derivative can be computed here:

\begin{align*}
\Phi[F](x) = -\frac{1}{F^{(3)}[F^{(2)}]^{-1}(0)} \delta^{(1)}_{\{F^{(2)}]^{-1}(0)} \in C^{-3} \setminus C^{-2},
\end{align*}

so it follows that:
\( h_n^{3/2} \left\{ \frac{1}{\sqrt{n}} \left( \widehat{F}_n^{(2)}(0) - F^{(2)}(0) \right) \right\} \overset{d}{\longrightarrow} N \left( 0, \frac{\int_{-\infty}^{\infty} |K^{(1)}(x)|^2 \, dx \cdot \frac{F^{(1)} \left[ \left( \left( F^{(2)} \right)^{-1}(0) \right) \right]}{\left( F^{(3)} \left[ \left( F^{(2)} \right)^{-1}(0) \right) \right]^2} \right) \).

This result appears to be new for dependent data.

**Example 8: Hazard Rate**

Consider \( \Phi[F] \equiv \frac{F^{(1)}(y)}{1 - F(y)} \) for some fixed \( y \). Its kernel estimation has been studied by Roussas [1990]. Hazard rates are typically useful in unemployment studies. Here the derivative can easily be computed:

\[
\varphi[F](x) = \frac{1}{1 - F(y)} \delta_{y}(x) + \frac{F^{(1)}(y)}{[1 - F(y)]^2} \mathbb{I}(y \geq x) \in C^{-2} \setminus C^{-1}, \text{ and therefore:}
\]

\[
h_n^{1/2} \left\{ \frac{\hat{f}_n(y)}{1 - \hat{F}_n(y)} - \frac{f(y)}{1 - F(y)} \right\} \overset{d}{\longrightarrow} N \left( 0, \frac{\int_{-\infty}^{\infty} |K^{(1)}(x)|^2 \, dx \cdot \frac{f(y)}{(1 - F(y))^2} \right) .
\]

**Example 9: Regression Function**

The Nadaraya-Watson method relies on a kernel plug-in to estimate \( \Phi[F] \equiv E[Z | Y = y] = \int_{-\infty}^{\infty} z f(y, z) \, dz / \int_{-\infty}^{\infty} f(y, z) \, dz \). The asymptotic distribution of Robinson [1983] and Bierens [1985] can be recovered by computing the functional derivative: \( \varphi[F](x) = \left( \frac{Z - a[F]}{b[F]} \right) \delta_{y}(w) \in C^{-2} \setminus C^{-1} \) where \( a[F] \equiv E[Z | Y = y] \) and \( b[F] \equiv \int_{-\infty}^{\infty} f(y, z) \, dz \). Hence for \( \varepsilon \equiv Z - E[Z | Y = y] \) and \( k \) regressors in \( Y \):

\[
h_n^{k/2} \left\{ \Phi[\hat{F}_n] - \Phi(F) \right\} \overset{d}{\longrightarrow} N \left( 0, \frac{\mathbb{E}\left[ \varepsilon^2 | Y = y \right] \left\{ \int_{-\infty}^{\infty} \left[ K^{(1)}(w) \right]^2 \, dw \right\}}{\int_{-\infty}^{\infty} f(y, z) \, dz} \right) .
\]
6. Conclusions

This paper has extended the delta method to nonparametric estimators of unsmooth functionals. The regularity conditions are simple, easily verifiable and generally equal or beat conditions used in case-by-case studies. Generalized derivatives were allowed to permit the inclusion of virtually any functional, global or pointwise, explicitly or implicitly defined. It was found here that both the rate of convergence to the asymptotic distribution and the asymptotic variance were functions of the unsmoothness of the functional derivative. Basing the estimator on dependent data modifies the asymptotic distribution only if the functional is more irregular than some threshold level (cadlag). New functional derivatives were computed for a variety of practical estimators used in econometrics, and used to obtain straightforwardly their asymptotic distribution.

Compared to the case-by-case approach, the generalized delta method has another advantage. It isolates the computation of the functional derivative, which is computed once and for all. When considering dependent sequences, or nonparametric estimation strategies other than kernel-based, the exact same functional derivative will be needed. The kernel results of this paper could therefore potentially be extended to cover other nonparametric methods. Many popular nonparametric procedures for density estimation are indeed of the form \( \hat{f}_n(u) = \frac{1}{n} \sum_{i=1}^{n} K_n(u, x_i) \). For example, the kernel method sets \( K_n(u, x_i) = \frac{1}{h_n^d} K \left( \frac{u - x_i}{h_n} \right) \) with a fixed function \( K(\cdot) \), while the orthogonal function method is based on \( K_n(u, x_i) = \sum_{j=1}^{h_n^{-d}} p_j(u)p_j(x_i) \) where \( \{p_i\} \) is a system of orthogonal functions in \( L^2(\mathbb{R}^d) \).
References


Appendix

The statement of the Theorem in the multivariate case is the following:

(i) If $\varphi[F] \in C^{-1}$, the result reads the same in dimension $d$;

(ii) If $\varphi[F] \in C^{-q} \setminus C^{-q+1}$: let $\varphi[F](x) = \sum_{i=1}^{L} \alpha_i[F](x) \delta^{(\ell)}(\gamma)(x) + B[F](x)$ where $|\ell| = q - 2$, $\alpha_i[F](\cdot) \in C^{-1}$ and $B[F](\cdot) \in C^{-q+1}$. $\gamma \in \mathbb{R}^{d(\ell)}$ contains $d(\ell)$ components, and $x = (\gamma^T, x^T)$ is partitioned accordingly. The maximal number of variables affected by the Dirac mass is $d^* \equiv \max\{d(\ell) : \ell \in \{1, \ldots, L\}\}$ and is attained at $\ell$ in $L^* \equiv \{\ell \in \{1, \ldots, L\} / d(\ell) = d^*\}$.

Then under $A4(r + d^*/2, m)$: $h_n^{[d^{r+2q-4}]\downarrow 1/2} \{\Phi(\hat{F}_n) - \Phi(F)\} \overset{d}{\to} N(0, \Sigma_{\phi}[F])$

where: $V_{\phi}[F] = \sum_{\ell \in L} \left\{\int_{\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} \left(\alpha_i[F](\gamma, -\ell t)d\gamma\right)^2 f(\gamma, -t)d\gamma\right)\right\}$ and $K(\cdot) \equiv \int_{-\infty}^{+\infty} K(\cdot, -t)d\gamma$.

Proof of Theorem: The following two lemmas will be used:

Lemma 1 (Central Limit Theorem): Under $A1-A4(0,0)$, $\hat{A}_n \equiv n^{1/2}(\hat{F}_n - E[\hat{F}_n])$ converges in law to a Gaussian $C^0$-stochastic distribution $G_{\Phi}$ in the space $(C^0, \| \cdot \|_{L(\omega, 0)})$, with finite-dimensional covariances given below. If $A4(0,0)$ is replaced by the more stringent requirement $A4(r,0)$, then the preceding statements hold for the centered process $\hat{A}_n \equiv n^{1/2}(\hat{F}_n - F)$ instead of $\hat{A}_n \equiv n^{1/2}(\hat{F}_n - E[\hat{F}_n])$. The covariance kernel of the generalized Brownian Bridge $G_{\Phi} \equiv \hat{B} \circ F$ is given by (where $F_k$ is the joint cdf of observations at lag $k$):
\[ E\left[ \tilde{B}(s) \tilde{B}(t) \right] = F(\min(s,t))(1 - F(t)) + \sum_{k=1}^{\infty} \{ F_k(s,t) + F_k(t,s) - 2F(s)F(t) \}. \]

**Lemma 2 (Bounds for Remainder Term):** Under A1-A3, for \( q = 0, \ldots, d+s \):

\[
\left\| \hat{F}_n - E[\hat{F}_n] \right\|_{L(2,q)} = O_p\left( n^{-1/2} h_n^{-q} \right) \quad \text{and} \quad \left\| E[\hat{F}_n] - F \right\|_{L(2,q)} = O\left( h_n^{1-(q-d)} \right).
\]

To prove Lemma 1, show that the class of functions \( \Gamma \equiv \{ W_{(y,h)} / y \in R^d, h \in R_+^* \} \) where \( W_{(y,h)}(t) = \frac{1}{h^d} K\left( \frac{t-y}{h} \right) \) forms a subgraph VC class. But such a class is a Euclidean class (from Lemma (2.12) in Pakes and Pollard [1989]); conclude with Theorem 1 of Arcones and Yu [1992]. Alternatively one could use the U-statistics approach of Robinson [1989]. Lemma 2 is easy given A2. Details are in AÔt-Sahalia [1993].

(i) Consider now the first part of the Theorem. By differentiability of the functional \( \Phi \) at \( F \): \( n^{1/2} \left\{ \Phi[\hat{F}_n] - \Phi[F] \right\} = \int_{-\infty}^{+\infty} \varphi[F](x) d\hat{A}_n(x) + R_{\Phi}[F, \hat{A}_n] \), where:

\[
\left\| R_{\Phi}[F, \hat{A}_n] \right\|_X = O\left( n^{1/2} \left\| \hat{F}_n - F \right\|_{L(\infty, m)}^2 \right). \quad \text{First,} \quad \left\| R_{\Phi}[F, \hat{A}_n] \right\|_X = o_p(1) \quad \text{follows from Lemma 2 since:} \quad n^{1/2} \left\| \hat{F}_n - F \right\|_{L(\infty, m)}^2 = O_p \left( n^{-1/2} h_n^{-2m} + n^{1/2} h_n^{2r-2(m-d)} \right) \quad \text{is} \quad o_p(1) \quad \text{under A4(r,m) as} \ r>2(m-d). \text{ Then by Slutsky's Theorem, the distribution of} \ n^{1/2} \left\{ \Phi[\hat{F}_n] - \Phi[F] \right\} \text{is given by that of} \ \int_{-\infty}^{+\infty} \varphi[F](x) d\hat{A}_n(x). \]

But by the continuous mapping theorem (e.g., Proposition 9.3.7 in Dudley [1989]), \( \int_{-\infty}^{+\infty} \varphi[F](x) d\hat{A}_n(x) \) converges in law to \( \int_{-\infty}^{+\infty} \varphi[F](x) d(\tilde{B}(F(x))) \) since from Lemma 1 \( \hat{A}_n \) converges in law to the process \( G_F \equiv \tilde{B} \circ F \). \( \int_{-\infty}^{+\infty} \varphi[F](x) d(\tilde{B}(F(x))) \) is the Itô integral of the real-valued, non-random function \( \varphi[F] \) with respect to the Gaussian stochastic process \( \tilde{B} \circ F \) and is therefore normally distributed. The asymptotic variance of the generic Itô integral \( \int_{-\infty}^{+\infty} \varphi(x) d(\tilde{B}(F(x))) \) can be computed as:
\[ E \left( \int_{-\infty}^{+\infty} \omega(x) d(\mathcal{B} \circ F(x)) \right)^2 = E \left( \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \omega(x) \omega(y) d(\mathcal{B} \circ F(y))d(\mathcal{B} \circ F(x)) \right) \]

\[ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} E \left[ (\mathcal{B} \circ F(y))(\mathcal{B} \circ F(x)) \right] \frac{\partial^d \omega(x)}{\partial x_1 \ldots \partial x_d} \frac{\partial^d \omega(y)}{\partial y_1 \ldots \partial y_d} dy dx \]

Thus:

\[ V_{\phi}[F] = E \left[ \phi(F)(x_i) \right]^2 - E \left[ \phi(F)(x_i) \right]^2 + 2 \sum_{k=1}^{\infty} \left\{ E \left[ \phi(F)(x_i) \phi(F)(x_{i+k}) \right] - E \left[ \phi(F)(x_i) \right] E \left[ \phi(F)(x_{i+k}) \right] \right\} \]

(ii) Consider now the case where:

\[ \phi[F](x) = \alpha_r[F](x) \partial^{(|\Delta|)} \delta_{(\cdot,y)}(\cdot x) \in C^{-q} \setminus C^{-q-1}, \text{ for } |\Delta| = q - 2, 2 \leq q \leq s. \]

The remainder term in the expansion of the functional is bounded as in (i). Then by Slutsky's Theorem the asymptotic distribution is given by the linear term (scaled for now at the rate \( n^{1/2} \)):

\[ \int_{-\infty}^{+\infty} \partial^{(\Delta)} \delta_{(\cdot,y)}(\cdot x) \alpha_r[F](x) d\hat{A}_n(x) \]

\[ = \left\{ \int_{-\infty}^{+\infty} \frac{1}{h_n} \int_{-\infty}^{+\infty} \partial^{(\Delta)} \left\{ \alpha_r[F](\cdot x, \cdot x') K \left( \frac{\cdot x - x - t}{h_n}, \frac{\cdot x' - x - t}{h_n} \right) \right\} d\cdot x' \right\} a_n(t) dt \]

\[ + n^{1/2} \left\{ \int_{-\infty}^{+\infty} \frac{1}{h_n^d} \int_{-\infty}^{+\infty} \partial^{(\Delta)} \left\{ \alpha_r[F](\cdot x, \cdot x') K \left( \frac{\cdot x - x - t}{h_n}, \frac{\cdot x' - x - t}{h_n} \right) \right\} d\cdot x' \right\} f(t) dt \]

\[ - \int_{-\infty}^{+\infty} \partial^{(\Delta)} \left\{ \alpha_r[F](\cdot x, \cdot x') f(\cdot x, \cdot x') \right\} d\cdot x' \]
\[
= \int_{-\infty}^{+\infty} \left\{ \frac{1}{h_n} \int_{-\infty}^{+\infty} \partial^{(\Delta)} \left[ \alpha_n[F](\tau, x, -\tau t) \right] \right. \\
\left. \cdot \left( \frac{\tau x - \tau t}{h_n}, \frac{\tau x - \tau t}{h_n} \right) \right\} \omega_n(t) \ dt + O\left( n^{1/2} h_n^{-1/2} \right)
\]

Write the leading term as \[ \int_{-\infty}^{+\infty} \omega_n(t) \omega_n(t) \ dt, \] where: \( \omega_n(t) = a_n(t) \) and \( A_n \equiv n^{1/2} (F_n - F) \). Then let \( v_n(t) \equiv h_n^{(d(\Delta) + 2)} \omega_n(t) \). Next show that:

\[
E \left\{ \int_{-\infty}^{+\infty} \left[ h_n^{(d(\Delta) + 2)} \omega_n(t) \ dA_n(t) \right]^2 \right\}
\]

\[
= \left\{ \int_{-\infty}^{+\infty} \left[ \partial^{(\Delta)} \right](\tau u) \right\}^2 d\tau d\tau \left\{ \int_{-\infty}^{+\infty} \left[ \alpha_n[F](\tau y, -\tau t) \right]^2 f(\tau y, -\tau t) \ d\tau \right\} + o(1)
\]

But:

\[
\int_{-\infty}^{+\infty} \omega_n(x) dA_n(x) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left\{ \omega_n(x_i) - \int_{-\infty}^{+\infty} \omega_n(x) f(x) \ dx \right\} = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left\{ \omega_n(x_i) - E[\omega_n(x)] \right\}, \text{ so:}
\]

\[
E \left[ \left( \int_{-\infty}^{+\infty} \omega_n(t) dA_n(t) \right)^2 \right] = \int_{-\infty}^{+\infty} \left[ \omega_n(t) \right]^2 f(t) \ dt - \left( \int_{-\infty}^{+\infty} \omega_n(t) f(t) \ dt \right)^2
\]

\[ + \sum_{k=1}^{n-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} 2 \left\{ f_k(t,s) - f(s)f(t) \right\} \omega_n(t) \omega_n(s) \ dt \ ds \]

The first of the three terms above is \( O\left( h_n^{-d+2} \right) \), while the other two are \( o\left( h_n^{-d+2} \right) \). In particular the "time-series" term containing the sum over time lags is of lower order than the first term. The computations are very similar for all three terms; for example for the first term:
\[
\int_{-\infty}^{+\infty} [\omega_n(t)]^2 f(t) dt = \int_{-\infty}^{+\infty} \frac{1}{h_n^d} \int_{-\infty}^{+\infty} \partial^{(\ell)} \left\{ \alpha_n[F]'(x, -t) \right\} K \left( \frac{x - t}{h_n}, \frac{-x - t}{h_n} \right) d'x f(t) dt
\]

\[
= \int_{-\infty}^{+\infty} \frac{1}{h_n^d} \sum_{\ell \in |\alpha|} \int_{-\infty}^{+\infty} \partial^{(\ell)} \left\{ \alpha_n[F]'(x, -t) \right\} \left( K \left( \frac{x - t}{h_n}, \frac{-x - t}{h_n} \right) \right) d'x f(t) dt
\]

Because \( \partial^{(\ell-\ell')} K \left( \frac{x - t}{h_n}, \frac{-x - t}{h_n} \right) = \frac{1}{(h_n)^{|\Delta| - |\ell|}} \left( \partial^{(\ell-\ell')} K \left( \frac{x - t}{h_n}, \frac{-x - t}{h_n} \right) \right) \), the term of highest order corresponds to \( |\ell'\ell| = 0 \), and is therefore given by:

\[
\int_{-\infty}^{+\infty} \frac{1}{h_n^d} \left\{ \int_{-\infty}^{+\infty} \partial^{(\ell)} K(t) d'x \right\} \left( \int_{-\infty}^{+\infty} \partial^{(\ell)} K(t) d'x \right) f(t) dt
\]

Hence:

\[
E \left\{ \int_{-\infty}^{+\infty} n^2 \int_{-\infty}^{+\infty} \partial^{(\ell)} K(t) d'x \right\} \left( \int_{-\infty}^{+\infty} \partial^{(\ell)} K(t) d'x \right) f(t) dt + o(1)
\]

\[
= V_{\varphi} [F] + o(1)
\]

Therefore

\[
\int_{-\infty}^{+\infty} v_n(t) dA_n(x) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left\{ v_n(x_i) - E[v_n(x)] \right\} \quad \rightarrow^d N(0, V_{\varphi} [F])
\]

by the Central Limit Theorem. But the first order Taylor expansion of the functional yields:

\[
n^{1/2} \left\{ \Phi(\hat{F}_n) - \Phi(F) \right\} = \int_{-\infty}^{+\infty} \omega_n(t) dA_n(t) + O_p \left( n^{1/2} h_n^{r-(q-2)} \right) + O_p \left( n^{1/2} || F_n - F ||_{L^2(-m)}^2 \right)
\]

Suppose that \( d(\ell) = d^* \) (when there are more than one \( \ell \) the only terms that matter are those corresponding to \( \ell \in L^* \)). Then under Assumption A4(r+d*/2,m), \( h_n^{d(\ell)+2|\Delta|/2} O_p \left( n^{1/2} h_n^{r-(q-2)} \right) = o_p(1) \) and the remainder term is also \( o_p(1) \). Therefore:

\[
h_n^{d(\ell)+2|\Delta|/2} n^{1/2} \left\{ \Phi(\hat{F}_n) - \Phi(F) \right\} = \int_{-\infty}^{+\infty} v_n(t) a_n(t) dt + o_p(1)
\]

\[
\rightarrow^d N \left( 0, \left\{ \int_{-\infty}^{+\infty} \partial^{(\ell)} K(t) d'x \right\} \left\{ \int_{-\infty}^{+\infty} \partial^{(\ell)} K(t) d'x \right\} f(t) dt \right)
\]
Under Assumption A4(r,m) only, this will be the asymptotic distribution of \( h_n^{(q+2)/2} \{ \Phi(\hat{F}_n) - \Phi(F) \} \) instead of \( h_n^{(q+2)/2} \{ \Phi(\hat{F}_n) - \Phi(F) \} \). Indeed: 
\[
E[\Phi(\hat{F}_n)] - \Phi(F) = O(h_n^{-(q-2)}). 
\]
Under A4(r+d*/2,m), this asymptotic bias term once multiplied by \( h_n^{(q+2)/2} \{ \Delta \} \) is \( o(1) \).

The absence of a covariance at the same order \( O(h_n^{-(q+2)/2}) \) between terms associated with different \( \ell \) yields:

\[
h_n^{(d+2)(q-2)/2} n^{1/2} \{ \Phi(\hat{F}_n) - \Phi(F) \} \xrightarrow{d} N \left( 0, \sum_{\ell \in L'} \int_{-\infty}^{+\infty} (\sigma_\ell^n K_\ell u) \left( (y,-t) f(y,-t) d(-t) \right) \right)
\]

If present, the term \( B[F](\cdot) \) \( C^{-q+1} \) only contributes terms of higher order in powers of \( h_n^{-1} \) and therefore does not change the asymptotic distribution.

**Estimation of the Asymptotic Variance:** The additional technical regularity condition is A5: Let \( V_i \equiv V_i[F] \equiv \varphi[F](x_i) \) and \( \hat{V}_i \equiv \varphi[\hat{F}_n](x_i) \). Assume:

1. \( E \left[ |V_i|^{3+\delta} \right] < \infty \) where \( \delta > (3+\varepsilon)(3+\varepsilon/2)/\varepsilon \) for some \( \varepsilon > 0 \);
2. \( E \left[ \sup_{G \in N} |V_i[G]|^2 \right] < \infty \) where \( N \) is a neighborhood of the true cdf \( F \);
3. \( E \left[ |V_i[G] - V_i[H]|^2 \right] \leq C \| G - H \|^2_{L(\infty,m)} \) for \( G \) and \( H \) in \( N \).