The Annals of Statistics
2002, Vol. 30, No. 5, 1460–1479

NEW METHODS FOR BIAS CORRECTION AT ENDPOINTS AND BOUNDARIES[1]

BY PETER HALL AND BYEONG U. PARK

Australian National University and Seoul National University

We suggest two new, translation-based methods for estimating and correcting for bias when estimating the edge of a distribution. The first uses an empirical translation applied to the argument of the kernel, in order to remove the main effects of the asymmetries that are inherent when constructing estimators at boundaries. Placing the translation inside the kernel is in marked contrast to traditional approaches, such as the use of high-order kernels, which are related to the jackknife and, in effect, apply the translation outside the kernel. Our approach has the advantage of producing bias estimators that, while enjoying a high order of accuracy, are guaranteed to respect the sign of bias. Our second method is a new bootstrap technique. It involves translating an initial boundary estimate toward the body of the dataset, constructing repeated boundary estimates from data that lie below the respective translations, and employing averages of the resulting empirical bias approximations to estimate the bias of the original estimator. The first of the two methods is most appropriate in univariate cases, and is studied there; the second approach may be used to bias-correct estimates of boundaries of multivariate distributions, and is explored in the bivariate case.

1. Introduction. Many boundary or endpoint estimation problems in statistics are closely related to problems involving nonparametric curve estimation. The methods used are generally biased, and in fact the sign or direction (in the case of spatial problems) of bias is generally known. However, the relative error of bias estimators is typically of larger order than would be found in the related curve estimation setting, owing to marked asymmetries inherent to boundary estimation. In the present paper we suggest two new methods for overcoming these types of difficulty. Both techniques are translation-based, and they use translations in ways that have not been considered before.

The first method, which seems most appropriate in univariate settings, is introduced in Section 2. It incorporates a translation-based correction *inside* the kernel, and thereby respects the sign of the bias that is being estimated. More traditional approaches to high-order bias estimation, such as the jackknife,

Received June 2000; revised November 2001.

[1]Supported in part by KOSEF, the Korean–Australian Cooperative Science Program 1999 and by the Brain Korea 21 Project.

AMS 2000 subject classifications. Primary 62G07; secondary 62G20.

Key words and phrases. Bias estimation, bootstrap, curve estimation, free disposal hull estimator, frontier estimation, kernel methods, nonparametric density estimation, productivity analysis, translation.

incorporate the correction in a directly additive way, with the result that stochastic fluctuations can render the sign of the bias estimator incorrect. High-order kernel approaches to bias estimation can be viewed as examples of the jackknife, and their tendency to reverse the sign, too, is one aspect of this difficulty, which our approach overcomes. Our technique also has application to density estimation, where it allows the boundary bias of conventional kernel estimators to be greatly reduced without using kernels that take negative values. We briefly discuss this application.

Our second approach to bias correction has versions in any number of dimensions, and we illustrate it in Section 3 in the bivariate case. It is a new form of the bootstrap, and might be termed the translation bootstrap because of its reliance on averaging over repeated empirical translations. Operationally, it involves temporarily taking the true boundary $\mathcal{B}$ to equal to its estimator, $\widehat{\mathcal{B}}$, say, and moving the latter steadily into the body of the data, recomputing the boundary estimator from data below $\widehat{\mathcal{B}}$ as we go. The bias of the latter estimate is approximated by the difference between the estimate and the respective translated position of $\widehat{\mathcal{B}}$; and the average of these bias approximations over the translations of $\widehat{\mathcal{B}}$ is an estimator of the bias of $\widehat{\mathcal{B}}$ as an estimator of $\mathcal{B}$.

The translation bootstrap can be viewed as a competitor with subsampling-bootstrap approaches to boundary-bias estimation; see Bickel, Götze and van Zwet (1997) and Politis, Romano and Wolf (1999) for discussion of subsampling. The two methods have similar theoretical properties, although giving different numerical results, in the univariate case. It is in bivariate settings that the translation bootstrap comes into its own. More conventional bootstrap techniques for inference at the edge or boundary of a distribution do not perform well, not least because relatively conventional resampling approaches do not capture the relationships among extremes of a resample drawn by resampling in conventional ways. See, for example, Athreya (1987a, b), Knight (1989) and Hall (1990) for accounts of aspects of this issue, and Simar and Wilson (1998) for discussion of bootstrapping in frontier models.

The subsampling approach to estimating bias at an endpoint has links to techniques for estimating a quantile density, discussed by, for example, Siddiqui (1960), Bloch and Gastwirth (1968), Bofinger (1975), Reiss (1978), Csörgő [(1983), page 32], Falk (1986), Welsh (1988), Jones (1992), Cheng (1995) and Cheng and Parzen (1997). Recent work on boundary estimation problem includes Hall, Park and Stern (1998), Gijbels, Mammen, Park and Simar (1999) and Hall, Park and Turlach (1998). They are closely related to estimation of density support, considered earlier by, for example, Chevalier (1976), Ripley and Rasson (1977), Mammen and Tsybakov (1995), Korostelev, Simar and Tsybakov (1995) and Härdle, Park and Tsybakov (1995). Some particular methods for estimating monotone (concave) boundary have been analysed theoretically by Kneip, Park and Simar (1998) and Park, Simar and Wiener (2000).

A variety of bias reduction methods, usually not applicable to estimation at the boundary, has been suggested for nonparametric curve estimation. See, for example, the survey by Jones and Signorini (1997), and in particular the data adjustment technique of Samiuddin and El-Sayyad (1990) and the multiplicative adjustment approach of Linton and Nielsen (1994). In one-dimensional problems our method is in the spirit of the former, which has been termed “data sharpening” by Choi and Hall (1999). In effect we add a data adjustment term, $\hat{\alpha}$, to the argument of the kernel.

2. One dimension: Bias reduction at endpoints. Let $\mathcal{X} = \{X_1, \dots, X_n\}$ denote a random sample from a univariate distribution F of which the upper endpoint or boundary, a, say, is finite but unknown. Write $X_{(1)} \le \cdots \le X_{(n)}$ for the ordered sample values. We take $\hat{a} = X_{(n)}$ to be our estimator of a. Of course, $\hat{a}$ is biased downwards; let $\beta = E(a - \hat{a}) \ge 0$ be the negative of the bias. Then, under mild regularity conditions, β is asymptotic to $\{nf(a)\}^{-1}$, where f denotes the density of the sampled distribution. Therefore, estimating β is tantamount to estimating $f(a)$, at least to first order, and so bias estimation has some of the features of density estimation. See pages 46–49 of Wand and Jones (1995) for discussion of kernel methods for estimating densities at endpoints.

One estimator of β is $\tilde{\beta}_m = m^{-1}(X_{(n)} - X_{(n-m)})$, which may variously be interpreted as a nonparametric inverse-density estimator [Siddiqui (1960), Bloch and Gastwirth (1968)] or as the m-out-of-n subsampling bootstrap estimator of β. It has a more general, kernel form,

$$\tilde{\beta}_m = \frac{\sum_i (X_{(n-i+1)} - X_{(n-i)}) K(i/m)}{\sum_i K(i/m)}, \tag{2.1}$$

where K denotes a nonnegative function supported on the positive half-line.

The estimator at (2.1) suffers from poor convergence rate, resulting from its inherent asymmetries. The main effects of the asymmetries may be removed by incorporating an empirical translation correction, $\hat{\alpha}$, into the kernel at (2.1), giving

$$\hat{\beta}_m = \frac{\sum_i (X_{(n-i+1)} - X_{(n-i)}) K\{(i + \hat{\alpha})/m\}}{\sum_i K(i/m)}. \tag{2.2}$$

There are several different ways of defining an appropriate $\hat{\alpha}$, but in this paper we consider only one:

$$\hat{\alpha} = -m \frac{\sum_i (X_{(n-i+1)} - 2X_{(n-i)} + X_{(n-i-1)}) i K(i/m)}{\sum_i (X_{(n-i+1)} - X_{(n-i)}) K'(i/m)}. \tag{2.3}$$

Note particularly that, without regard for the sign of $\hat{\alpha}$, we have $\hat{\beta} \ge 0$. Therefore, our bias-corrected bias estimator $\hat{\beta}_m$ respects the sign of bias. We may regard $\hat{\alpha}$ as a data perturbation constructed specifically to adjust bias; it is in effect an estimator of the ratio

$$\frac{m^2}{n} \frac{G''(1-)}{G'(1-)} \frac{\int_{u>0} u K(u)\, du}{K(0)},$$

where $G = F^{-1}$ and F is the sampled distribution function.

As we shall show in Section 5, for distributions with three derivatives the bias estimator defined by (2.2) and (2.3) achieves the optimal convergence rate $n^{-2/5}$, in terms of relative L^p error for any $1 \le p < \infty$, when m is chosen to be of size $n^{4/5}$. This is of course an order of magnitude faster than the relative rate $n^{-1/3}$ for the estimator $\tilde{\beta}_m$ that is obtained by choosing m to be of size $n^{2/3}$. However, the simulations in Section 4 show that our bias estimator does not outperform the simpler technique (2.1). The superior convergence rate of the method (2.2) with (2.3) seems to take effect only for very large sample sizes.

The problem of estimating f at or near an endpoint is related to that of estimating the endpoint, and may be addressed using ideas similar to those suggested above. To this end, note that a conventional, reweighted kernel estimator of $f(x)$, for $x \le a$, is given by

$$\tilde{f}(x) = \left\{ n^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\right\} \Big/ \left\{ \int_{-\infty}^{a} K\left(\frac{x-y}{h}\right) dy \right\}, \tag{2.4}$$

where the integral in the denominator adjusts for a deficit of probability mass in the neighborhood of the endpoint. Without the adjustment, $\tilde{f}(a)$ would not be consistent for $f(a)$.

However, even with the adjustment the estimator $\tilde{f}$ converges slowly to f at points that are close to a. For example, the optimal convergence rate of $\tilde{f}(a)$ to $f(a-)$ is generally only $n^{-1/3}$. As in the case of the estimator $\tilde{\beta}_m$, this difficulty may be removed by making a simple, empirical translation correction, leading to the estimator

$$\hat{f}(x) = \left\{ n^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i + \hat{\alpha}(x)}{h}\right)\right\} \Big/ \left\{ \int_{-\infty}^{a} K\left(\frac{x-y}{h}\right) dy \right\}, \tag{2.5}$$

where on the present occasion,

$$\hat{\alpha}(x) = h^2 \frac{\bar{f}'(x)}{\tilde{f}(x)} \rho\left(\frac{x-a}{h}\right), \tag{2.6}$$

$\bar{f}'$ is an estimator of f', and $\rho(u) = K(u)^{-1} \int_{v \le u} v K(v)\, dv$; here we define $0/0$ to be 0. We shall show in Section 5 that $\hat{f}$ converges to f at the optimal second-order rate $n^{-2/5}$, in a left-hand neighborhood of a, provided h is of size $n^{-1/5}$. Away from this neighborhood, $\hat{f}$ is identical to the conventional density estimator based on the kernel K and bandwidth h.

Note particularly that, provided only that K is nonnegative, $\hat{f} \ge 0$. Therefore the translation correction has preserved positivity. More traditional techniques for kernel density estimation at boundaries, based on boundary kernels, exhibit increased oscillation due to side lobes in the kernel and need to be spliced to conventional estimators away from the boundary, in addition to suffering from negativity.

The kernel K in (2.4)–(2.6) would be a smooth, symmetric, compactly supported, probability density that was nondecreasing and nonincreasing on the negative and positive half-lines, respectively. Thus, it would be a kernel of the type that is conventionally used for nonparametric density estimation. In this case the denominator at (2.4) is of course identically 1 if K is supported on the interval $[-1, 1]$ and $x < a - h$. Given that K is of this type, the function ρ in (2.6) is bounded on the negative half-line (which is the only region where it is required), and vanishes on $(-\infty, -1)$. The same kernel type is appropriate in the problem of estimating a, which is why we have not altered our notation.

In the definitions of $\tilde{f}$ and $\hat{f}$ we have assumed a to be known, which is usually the case in this class of problems. For example, the distribution of the breaking strength of a fibre would be known to be supported on the positive half-line, and in particular would not take negative values, even though the corresponding density might not vanish at the origin. The convergence rate does not deteriorate, however, if we replace a by $\hat{a} = X_{(n)}$ in (2.4) and (2.5).

3. Two dimensions: Translation bootstrap.

3.1. *Basic estimators.* Suppose we observe data $\mathcal{P} = \{(X_1, Y_1), (X_2, Y_2), \dots\}$ from a Poisson process in the plane, and that the intensity function of $\mathcal{P}$ is supported in the half-plane below the curve represented by the equation $y = g(x)$. We wish to estimate the boundary function, g. A well-known estimator is the *free disposal hull*, or FDH, introduced by Deprins, Simar and Tulkens (1984) in the context of measuring the efficiency of enterprises. There, g is generally a monotone increasing function, and the FDH is designed for that setting. Geometrically, the FDH is the lowest monotone step function above all the data (X_i, Y_i). An obvious enhancement of the FDH is its linearly interpolated form, which we shall call the LFDH estimator. It is constructed by connecting the left-hand ends of the steps of the FDH. Below, we define these two estimators concisely.

Given a point (x, y), let $\mathrm{SE}(x, y)$ denote the "south-eastern" quadrant of the plane with its vertex at (x, y),

$$\mathrm{SE}(x, y) = \{(u, v) : u \geq x,\ v \leq y\};$$

and similarly define $\mathrm{NW}(x, y)$ to be the "north-western" quadrant. Then, the FDH estimator $\hat{g}_{\mathrm{FDH}}$ is defined to be the upper boundary of the set $\bigcup_i \mathrm{SE}(X_i, Y_i)$:

$$\hat{g}_{\mathrm{FDH}}(x) = \max\left\{ y : (x, y) \in \bigcup_i \mathrm{SE}(X_i, Y_i) \right\}.$$

Obviously, the FDH is biased downward since it never exceeds g. Theoretical properties of the FDH were investigated by Park, Simar and Weiner (2000) in the multidimensional case.

Next we define the LFDH estimator. A data point (X_i, Y_i) is called a *boundary point* of the FDH if there is no other point in the region $\mathrm{NW}(X_i, Y_i)$ other than itself. The LFDH estimator is obtained by linearly interpolating these boundary points. Formally, define for a given x,

$$\begin{aligned} X_1^{\mathrm{L}} &= X_1^{\mathrm{L}}(x) = \max\{X_i \le x : (X_i, Y_i) \text{ is a boundary point}\}, \\ X_{k+1}^{\mathrm{L}} &= X_{k+1}^{\mathrm{L}}(x) = \max\{X_i < X_k^{\mathrm{L}} : (X_i, Y_i) \text{ is a boundary point}\}, \end{aligned} \tag{3.1}$$

for $k = 1, 2 \dots$ and let Y_k^{L} be the concomitant of X_k^{L}, for $k = 1, 2, \dots$. Likewise, define $(X_l^{\mathrm{R}}, Y_l^{\mathrm{R}})$, for $l = 1, 2, \dots$ on the right-hand side of x, with the ordering $x < X_1^{\mathrm{R}} < X_2^{\mathrm{R}} < \cdots$. The LFDH estimator is then given by

$$\hat{g}(x) = Y_1^{\mathrm{L}} + (Y_1^{\mathrm{R}} - Y_1^{\mathrm{L}})(x - X_1^{\mathrm{L}})/(X_1^{\mathrm{R}} - X_1^{\mathrm{L}}). \tag{3.2}$$

Note that the definition of the boundary points $(X_k^{\mathrm{L}}, Y_k^{\mathrm{L}})$ and $(X_l^{\mathrm{R}}, Y_l^{\mathrm{R}})$ depends on the point x where we wish to estimate g. Also, with this definition we can write $\hat{g}_{\mathrm{FDH}}(x) = Y_1^{\mathrm{L}}$. The boundary points $(X_k^{\mathrm{L}}, Y_k^{\mathrm{L}})$ and $(X_l^{\mathrm{R}}, Y_l^{\mathrm{R}})$, for $k, l = 2, 3, \dots$, are not required for the definition at (3.2); they are introduced here for later use. At the left- and right-hand edges of the dataset the LFDH estimator is not defined, and there one may use the FDH estimator instead.

Theoretical properties of the LFDH estimator will be derived in Section 4. In asymptotic terms, if the point process is Poisson with intensity $n\lambda$, where λ is a fixed function in the plane and strictly positive at the boundary, and if the boundary is differentiable, then as n diverges to infinity, the LFDH estimator converges to g at rate $O_p(n^{-1/2})$ in a pointwise sense. It may be shown, as in Härdle, Park and Tsybakov (1995), that this is the minimax-optimal rate for boundaries that satisfy a Lipschitz condition of order 1. If the boundary is continuously differentiable then bias of the LFDH estimator is also of size $n^{-1/2}$.

The LFDH estimator has less asymptotic bias and variance than its FDH counterpart. Indeed, it can be shown that the ratios of the bias and the variance of the LFDH estimator to those of its FDH counterpart are

$$\tfrac{5}{16}\big/\tfrac{1}{2} = 0.625 \quad \text{and} \quad \left(\tfrac{8}{9} - \tfrac{25}{128}\pi\right)\big/\left(2 - \tfrac{1}{2}\pi\right) \simeq 0.641,$$

respectively, independent of λ and g. [Asymptotic bias and variance of $\hat{g}_{\mathrm{FDH}}$ can be deduced from results of Park, Simar and Weiner (2000). Analogous results for the LFDH estimator may be derived similarly.] Because of its theoretical superiority and apparent practical advantages, we shall apply our bias correction method to the LFDH estimator rather than to its FDH counterpart.

3.2. *Bias correction of LFDH estimator.* Define $\hat{g}_y(x) = \hat{g}(x) - y$, and let $\widehat{\gamma}_y$ be the LFDH estimator of $\hat{g}_y$ computed from $\mathcal{P}_y = \{(X_i, Y_i) \in \mathcal{P} : Y_i \le \hat{g}_y(X_i)\}$. We estimate the bias of $\hat{g}$ as the average, over values y in an interval (u, v), of the empirical bias of $\widehat{\gamma}_y$ as an estimator of $\hat{g}_y$:

$$\hat{\beta}_{uv}(x) = (v - u)^{-1} \int_u^v \{\hat{g}_y(x) - \widehat{\gamma}_y(x)\}\, dy. \tag{3.3}$$

Our bias-corrected estimator of g is

$$\check{g} = \hat{g} + \kappa \hat{\beta}_{uv}, \tag{3.4}$$

where κ, an absolute constant, is a correction factor. The integrand at (3.3) is piecewise linear in g, and the integral is piecewise quadratic. This property is exploited to construct the estimator and implement the method.

Unlike the one-dimensional case encountered in Section 2, the bias correction $\kappa \hat{\beta}_{uv}$, divided by the true bias, does not converge to 1. Nevertheless, the expected value of the ratio converges to 1, and so the bias correction does effectively reduce bias. Moreover, the stochastic fluctuations of the bias correction are of the same order as those of the uncorrected estimator $\hat{g}$, and so they do not degrade the rate of convergence. The reason for this behavior is that the estimator $\hat{g}$, which is used as the "template" against which repeated bias approximations are computed and then averaged in the translation bootstrap algorithm, does not have the same smooth structure as the true function g; the stochastic fluctuations of the template manifest themselves as random variation of the bias correction.

If desired, this feature may be substantially removed by passing a second-order smoother through $\hat{g}$ before implementing the argument leading to (3.3) and (3.4). This is extremely easy to do; a standard local linear regression routine with a relatively small bandwidth, and passed through the continuum of "data pairs" $(x, \hat{g}(x))$, is suitable for this purpose. This substantially reduces the variability of $\hat{\beta}_{uv}$ without significantly impairing its bias reduction abilities. Subsequently, to ensure the monotone increasing character of $\hat{g}$ is retained in passing to $\check{g}$, we start at one point on the estimator (say at the left-hand end) and, computing $\check{g}$ on a grid of points $x_1 < x_2 < \cdots$, take

$$\check{g}(x_i) = \max\bigl\{\hat{g}(x_j) + \kappa \hat{\beta}_{uv}(x_j) : j \le i\bigr\}. \tag{3.5}$$

An alternative way of monotonizing suggested by one of the referees is to use $\check{g}(x_i) = \frac{1}{2}[\max\{\hat{g}(x_j) + \kappa \hat{\beta}_{uv}(x_j) : j \le i\} + \min\{\hat{g}(x_j) + \kappa \hat{\beta}_{uv}(x_j) : j \ge i\}]$.

Next we define κ. Consider the special case where $y = g(x)$ is the equation of a straight line, $\mathcal{L}$, say, of unit slope passing through the origin, and the point process in the half-plane below $\mathcal{L}$ is homogeneous and Poisson, of unit intensity. Call this Poisson process $\mathcal{Q}_0$. Write $\bar{g}$ for the LFDH estimator of g in this special case, and conditional on $\bar{g}$, let $\mathcal{Q}$ be another Poisson process of unit intensity, independent of $\mathcal{Q}_0$, in the region $\{(x, y) : y \le \bar{g}(x)\}$. Write $\bar{\gamma}$ for the LFDH estimator of $\bar{g}$ computed from $\mathcal{Q}$, and let $\bar{\beta} = \bar{g} - E(\bar{\gamma}|\bar{g})$ denote the translation bootstrap estimator of bias. It is clear that there exists a unique constant $\kappa > 0$ such that in this special case, $\bar{g}_{\mathrm{bc}}(0) = \bar{g}(0) + \kappa \bar{\beta}(0)$ has zero bias: $E\{\bar{g}_{\mathrm{bc}}(0)\} = g(0)$. We claim that this value of κ is appropriate in (3.4).

Exact computation of κ seems out of reach, but we can calculate a Monte Carlo approximation. To this end we worked out the exact joint distribution of $\bar{\beta}$ and $\bar{\gamma}$; it is given in a longer version of this paper, obtainable from the authors. From that result we calculated an exact formula for $E\{\bar{\gamma}(0)|\bar{g}\}$. The formula depends on the boundary points of the process $\mathcal{Q}_0$, denoted by $(V_k^{\mathrm{L}}, W_k^{\mathrm{L}})$ and $(V_l^{\mathrm{R}}, W_l^{\mathrm{R}})$ which are defined as at (3.1) with $x = 0$. We simulated 5000 realizations of these boundary points according to their joint distribution. The numbers of boundary points in each realization were 500 and 1000. For each realization we calculated $\bar{\beta}(0)$. The 5000 values of $\bar{\beta}(0)$ based on the same number of boundary points were then averaged to give an approximation to $E\{\bar{\beta}(0)\}$. There were only minor differences between the two different numbers of boundary points, and we took that for 1000 boundary points, 0.7325, as our final approximation to $E\{\bar{\beta}(0)\}$. Now, $E\{\bar{g}(0)\}$ can be computed exactly from the distribution of $\bar{g}(0)$; it equals $-5\sqrt{2\pi}/16$. Therefore, our approximation to the value of $\kappa = -E\{\bar{g}(0)\}/E\{\bar{\beta}(0)\}$ is 1.0694.

4. Numerical properties.

4.1. *Endpoint estimation.* We present the result of a numerical experiment demonstrating the effectiveness of the translation-based bias correction in univariate settings. We simulated 10,000 data sets of size $n = 100$, distributed on the interval $(0, 1)$ according to the density $f(x) = (3/2) - x$ there. They were used to approximate the mean squared error of the conventional estimator $\hat{a} = X_{(n)}$, and those of the bias-corrected estimators $\hat{a} + \tilde{\beta}_m$ and $\hat{a} + \hat{\beta}_m$ with $\tilde{\beta}_m$ as defined at (2.1), of the true endpoint a (in our case $a = 1$). The biweight kernel $K(u) = (15/16)(1 - u^2)^2 I(|u| \le 1)$ was used.

Figure 1 illustrates how the mean squared errors of the bias-corrected estimators varies with m. The horizontal straight line in the figure indicates the logarithm of mean squared error of the uncorrected estimator, $\hat{a} = X_{(n)}$, and corresponds to a mean squared error of 6.91×10^{-4}. By way of comparison the mean squared errors of the bias-corrected estimators, $\hat{a} + \tilde{\beta}_m$ and $\hat{a} + \hat{\beta}_m$, are minimized when $m = 29$ and $m = 18$ with minimal values 3.69×10^{-4} and 3.82×10^{-4}, respectively, representing efficiency gains with respect to $\hat{a}$ by the factors 1.87 and 1.81. The results also show that our bias estimator $\hat{\beta}_m$ does not outperform the simpler one $\tilde{\beta}_m$, indicating that the superior convergence rate of $\hat{\beta}_m$ may not take effect for moderate sample sizes.

4.2. *Bias-correcting the LFDH estimator.* To see the effect of applying the translation bootstrap in the bivariate case, 1,000 datasets of size $n = 400$ were generated. The boundary function was $g(x) = x + 3$. The model for the simulated data (X_i, Y_i), $i = 1, \ldots, 400$, was $Y_i = g(X_i)V_i$, where the X_i's were uniformly distributed on the interval $(-1, 1)$ and were independent of the V_i's, which were distributed on the interval $(0, 1)$ with density $f(v) = (13/9) - (4/3)v^2$ there. This means that the joint density of (X_i, Y_i) was given by

$$\lambda(x, y) = \{2(x + 3)\}^{-1}\bigl\{(13/9) - \bigl(4y^2/3(x + 3)^2\bigr)\bigr\} I_{(-1,1)}(x) I_{(0,x+3)}(y).$$

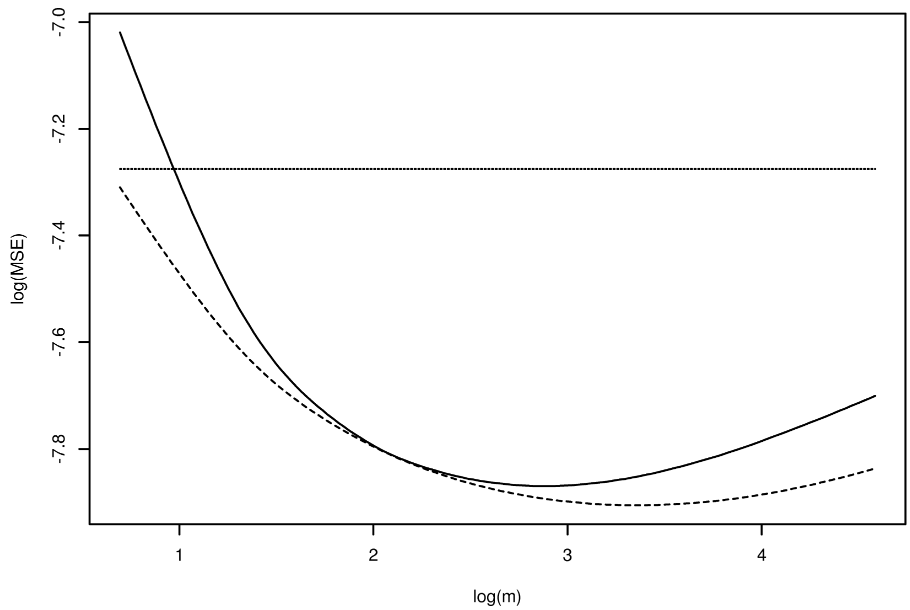

FIG. 1. *Mean squared errors of the endpoint estimators* (*one-dimensional case*). *The logarithm of the mean squared error of the uncorrected estimator is represented by the horizontal dotted line. Those of the bias-corrected estimators as functions of* $\log m$, *with* $\tilde{\beta}_m$ *given by* (2.1) *and* $\hat{\beta}_m$ *given by* (2.2) *are represented, respectively, by the dashed and solid curves.*

For the presmoothing of $\hat{g}$, the local linear regression was employed with the Epanechinikov kernel $K(u) = (3/4)(1 - u^2)I_{(-1,1)}(u)$ and bandwidth $h = 1.0$.

For each dataset the LFDH estimate, the bias-corrected estimate defined at (3.4), and its monotonized version at (3.5) with presmoothing noted in the third paragraph of Section 3.2, were calculated at the point $x = 0$. (Bias, variance and mean squared error are similar at other points, provided one is not too close to the boundaries at $x = \pm 1$.) Figure 2 summarises the results. In particular, panel (a) shows the logarithms of mean squared errors of the three estimators as functions of v. The horizontal straight dotted line corresponds to the uncorrected LFDH estimator. The corresponding mean squared error is 6.834×10^{-2}. The bias-corrected estimator and its smoothed, monotonized version have their minimal mean squared errors 3.980×10^{-2} and 1.622×10^{-2}, for $v = 0.75$ and $v = 0.90$, respectively. Their relative efficiencies with respect to the uncorrected one are thus 1.717 and 4.213, respectively. It is seen that the presmoothing and monotonization procedure improves the mean squared error of the bias-corrected estimator for a wide range of v. This is largely because the smoothed, monotonized bias correction reduces variance and yet retains good bias correction properties, as illustrated by panel (b) of Figure 2.

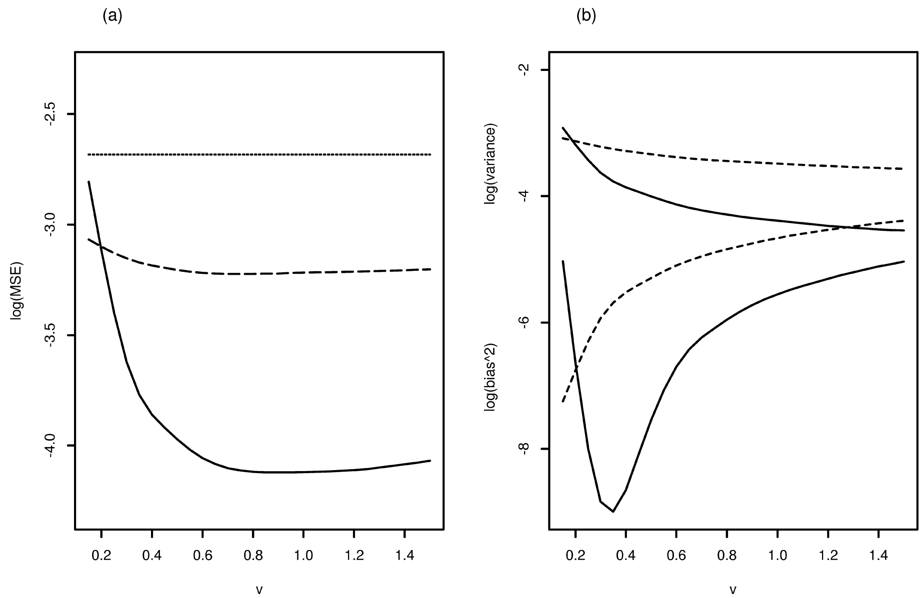

FIG. 2. *Mean squared errors, squared biases and variances of the boundary estimators* (*two-dimensional case*). *Panel* (a) *shows the logarithms of mean squared error, plotted against* v, *of the LFDH estimator* (*dotted line*), *of the bias-corrected estimator defined at* (3.4) (*dashed line*), *and of the pre-smoothed, monotonized bias-corrected estimator* (*solid line*). *The top two curves in panel* (b) *show the logarithms of the variances of the bias-corrected estimator* (*dashed line*) *and of its smoothed, monotonized bias-corrected counterpart* (*solid line*). *The bottom two curves show the logarithms of squared squared biases for these two estimators* (*in the same respective line types*).

5. Theoretical properties.

5.1. *Endpoint estimation.* We begin by describing properties of the bias-corrected endpoint estimator $\hat{\beta}_m$, defined at (2.2). Assume $E(|X|^{\varepsilon}) < \infty$ and F has three continuous derivatives on $[a-\varepsilon, a)$, both results holding for some $\varepsilon > 0$. Suppose too that $F(a-) = 1$ and $F'(a-) > 0$, and put $f = F'$ and $G = F^{-1}$. Let K denote a continuously differentiable function on $(-\infty, \infty)$, with support contained within $(-\infty, 1]$ and satisfying $\int_{[0,1]} K > 0$ and $K(0) \neq 0$. Finally, assume that $m = m(n) \to \infty$ and $m/n \to 0$ as $n \to \infty$. Call these conditions (C_1). Define

$$\tau^2 = \int_0^{\infty} \{2K(u) + uK'(u)\}^2 \, du,$$

$$\kappa_j = \int_0^{\infty} u^j K(u) \, du,$$

$$\kappa_j' = \int_0^{\infty} u^j K'(u) \, du.$$

[Thus, $\kappa_0' = -K(0)$ and $\kappa_1' = -\kappa_0$.]

THEOREM 5.1. *Assume conditions* (C_1). *Then*, $\beta = E(a - X_{(n)}) = n^{-1} f(a-)^{-1} + O(n^{-2})$ *and*

$$\frac{\hat{\beta}_m}{\beta} = 1 + \frac{\tau}{\kappa_0 m^{1/2}} N_n + \left(\frac{m}{n}\right)^2 \left(\frac{G''(a-)^2}{G'(a-)^2} \frac{\kappa_1 \kappa_1'}{\kappa_0 \kappa_0'} - \frac{G'''(a-)}{2G'(a-)} \frac{\kappa_2}{\kappa_0}\right) + o_p\{(m/n)^2\}, \tag{5.1}$$

where the random variable N_n *is asymptotically normal* $N(0, 1)$.

We can write (5.1) in the form $\hat{\beta}_m/\beta = 1 + c_1 m^{-1/2} N_n + (m/n)^2 c_2$, plus negligible terms, where c_1, c_2 are constants. Assuming $c_2 \neq 0$ we may deduce from this formula that the value of m that minimizes asymptotic L^p relative error, for any given $1 \le p < \infty$, equals a constant multiple of $n^{4/5}$; and that the minimum L^p error is asymptotic to a constant multiple of $n^{-2/5}$. When F has three bounded derivatives in a neighborhood of a this is the optimal minimax rate, and is achieved by $\hat{\beta}_m$ with $m \asymp n^{4/5}$, as may be seen from the following theorem. There (5.2) shows that the relative rate $n^{-2/5}$ cannot be improved upon for distributions with only three derivatives, and (5.3) demonstrates that the estimator $\hat{\beta}$ achieves this optimal rate. We take $a = 0$ for simplicity.

Given $B > 0$, let $\mathcal{F}(B)$ denote the class of distributions F that are supported on $[-B, 0]$, have three bounded derivatives on $[-B^{-1}, 0)$, and satisfy $|F^{(j)}(x)| \le B$ and $F'(x) \ge B^{-1}$ for $x \in [-B^{-1}, 0)$ and $j = 1, 2, 3$. For each $F \in \mathcal{F}(B)$ put $\beta(F) = \beta(F, n) = E_F(-X_{(n)})$. Let (C_2) denote the conditions imposed on K as part of (C_1), write $\hat{\beta}^0$ for the version of $\hat{\beta}_m$ in which m is taken as the integer part of any fixed, positive multiple of $n^{4/5}$, and let $\mathcal{G}$ be the set of all measurable functions $\check{\beta}$ of the data. (Each $\check{\beta}$ may be regarded as an estimator of β.)

THEOREM 5.2. *Assume conditions* (C_2) *and that* B *is so large that* $\mathcal{F}(B)$ *is nonempty*. *Then*

$$\lim_{\varepsilon \to 0} \liminf_{n \to \infty} \inf_{\check{\beta} \in \mathcal{G}} \sup_{F \in \mathcal{F}(B)} P_F\{|(\check{\beta}/\beta) - 1| > \varepsilon n^{-2/5}\} = 1, \tag{5.2}$$

$$\lim_{C \to \infty} \limsup_{n \to \infty} \sup_{F \in \mathcal{F}(B)} P_F\{|(\hat{\beta}^0/\beta) - 1| > C n^{-2/5}\} = 0. \tag{5.3}$$

Result (5.3) may be proved using methods from our derivation of Theorem 5.1, and (5.2) follows using standard techniques for proving minimax bounds in Hölder spaces. See, for example, Hall (1989) or Chapter 2 in Korostelev and Tsybakov (1993).

5.2. *Density estimation.* Here we describe properties of the bias-corrected density estimator $\hat{f}$, defined at (2.5). Recall from (2.6) that we require an estimator $\bar{f}'$ of f'. Simply differentiating the estimator $\tilde{f}$, or differentiating a standard kernel density estimator and correcting for a deficit of probability mass, as at (2.4), does not produce a consistent estimator of f' at or near the endpoint a. A consistent estimator can be obtained in a variety of other ways, however, for example by differentiating a one-sided, second-order kernel density estimator that uses bandwidth h. Such an estimator, and a wide range of others, have the property

$$\bar{f}'(x) = f'(x) + O_p\{h + (nh^3)^{-1/2}\} \tag{5.4}$$

for each x.

Assume F has three bounded derivatives on $[b, a]$, for some $b < a$. Suppose too that $F(a-) = 1$ and $F'(a-) > 0$, and put $f = F'$. Let K denote a twice-differentiable, symmetric probability density on $(-\infty, \infty)$, with support contained in $[-1, 1]$ and nondecreasing on $[-1, 0]$. Finally, assume $h = h(n) \to 0$ and $nh \to \infty$ as $n \to \infty$. Call these conditions (C_3).

THEOREM 5.3. *Assume conditions* (C_3), *and that* $\bar{f}'(x)$ *in the definition of the translation correction* (2.6) *satisfies* (5.4). *Then for each* $x = a - ch$ *and any* $c \ge 0$, *and also for each* $x \in (b, a]$,

$$\hat{f}(x) = f(x) + O_p\{h^2 + (nh)^{-1/2}\}. \tag{5.5}$$

Furthermore, $\hat{f}(x)$ *is identically equal to the conventional kernel estimator*,

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

for all $x < a - h$.

Condition (5.5) is the usual second-order statement about performance of $\hat{f}$ as an estimator of f. By way of contrast, the uncorrected estimator $\tilde{f}$ satisfies only $\tilde{f}(x) = f(x) + O_p\{h + (nh)^{-1/2}\}$ for x close to a. The larger order of bias in the latter expansion has been removed by the translation correction.

5.3. *Bias-correcting the LFDH estimator.* Let $\mathcal{P} = \mathcal{P}(n)$ be a Poisson process with intensity $n\lambda$, where $\lambda > 0$ is a fixed function in the plane and satisfies $\lambda(x, y) = 0$ for each pair (x, y) with $g(x) < y$. We shall take n to diverge to ∞. As in Section 3, let $\hat{g}$ be the LFDH estimator of g, computed from $\mathcal{P}$ and define $\check{g}$ as at (3.4).

In this setting, provided g has a continuous derivative, and λ is continuous and bounded away from 0 and infinity, $\hat{g}$ converges to g at rate $n^{-1/2}$:

$$n^{1/2}\{\hat{g}(0) - g(0)\} \to \{g'(0)/\lambda(0, g(0))\}^{1/2} Q_1$$

in distribution as $n \to \infty$, where the random variable $Q_1 = \bar{g}(0)$ has a continuous distribution supported on the negative half-line, not depending on unknowns such as g or λ. The latter property reflects the fact that $\hat{g}$ is consistently biased downward for g. [Here and below we assume, without loss of generality, that we are estimating $g(x)$ at $x = 0$.] Our next result shows that the bias-corrected estimator $\check{g}$ satisfies a similar limit theorem, also with rate $n^{-1/2}$, but, reflecting the efficacy of bias correction, the limit distribution now has mean 0.

THEOREM 5.4. *Assume g is nondecreasing and has a continuous derivative in a neighborhood of* 0; *that* $g'(0) > 0$; *that* $\lambda(0, g(0)) > 0$; *that* λ *is continuous on the set* $\{(x, y) : |x| \le \varepsilon,\ 0 \le g(x) - y \le \varepsilon\}$ *for some* $\varepsilon > 0$; *and that* u, v *in definition* (3.3) *of* $\hat{\beta}_{uv}$ *are functions of* n *satisfying* $0 \le u = o(v)$, $v \to 0$ *and* $n^{1/2}v \to \infty$. *Then*

$$n^{1/2}\{\check{g}(0) - g(0)\} \to \{g'(0)/\lambda(0, g(0))\}^{1/2} Q_2$$

in distribution as $n \to \infty$, *where the random variable* $Q_2 = \bar{g}(0) + \kappa\bar{\beta}(0)$ *has zero mean*.

The convergence rate, $n^{-1/2}$, evinced by Theorem 5.4 is optimal for boundaries with one derivative; see, for example, Härdle, Park and Tsybakov (1995).

6. Outline technical details.

6.1. *Proof of Theorem* 5.1. We shall derive only (5.1). It is notationally convenient to estimate a lower endpoint, and so we take that route in the proof. Without loss of generality the lower endpoint is the origin, and so we assume F has three continuous derivatives on $(0, \varepsilon]$ for some $\varepsilon > 0$, with $F'(0+) > 0$. Also without loss of generality, F is continuous. Put $G = F^{-1}$, and let $U_{(1)} \le \cdots \le U_{(n)}$ be the order statistics corresponding to the sequence $U_i = F(X_i)$ of independent uniform random variables. Note too that $U_{(i+1)} - U_{(i)} = Z_i/S_n$, where $S_n = Z_0 + \cdots + Z_n$ and $Z_0, Z_1, \ldots$ are independent, exponentially distributed random variables with $E(Z_i) = 1$. Arguing thus, and Taylor expanding, we may prove that

$$\begin{aligned}
&X_{(i+1)} - X_{(i)} \\
&\qquad = (U_{(i+1)} - U_{(i)})\{G'(0) + U_{(i)}G''(0) + \tfrac{1}{2}U_{(i)}^2 G'''(0)\} + R_1(i), \\
&X_{(i+2)} - 2X_{(i+1)} + X_{(i)} \\
&\qquad = (U_{(i+2)} - 2U_{(i+1)} + U_{(i)})G'(0) \\
&\qquad\quad + \{(U_{(i+2)} - U_{(i+1)})U_{(i+1)} - (U_{(i+1)} - U_{(i)})U_{(i)}\}G''(0) \\
&\qquad\quad + \tfrac{1}{2}\{(U_{(i+2)} - U_{(i+1)})U_{(i+1)}^2 - (U_{(i+1)} - U_{(i)})U_{(i)}^2\}G'''(0) + R_2(i),
\end{aligned}$$

where $R_1(i) = O_p\{(n^{-1}\log n)^2\}$ uniformly in i, $m^{-1}\sum_{i \le m} R_2(i) = o_p(m^{-3/2} + mn^{-3})$ uniformly in $m \le m_0$, and we write $G^{(j)}(0)$ to denote $G^{(j)}(0+)$.

Let Z_i be as in the previous paragraph, and define

$$T_1 = m^{-1}\sum_{i=1}^{m}(Z_i - 1)K(i/m), \qquad T_2 = m^{-1}\sum_{i=1}^{m}(Z_{i+1} - Z_i)(i/m)K(i/m),$$

$$V_1 = m^{-1}\sum_{i=1}^{m} E\{(U_{(i+2)} - U_{(i+1)})U_{(i+1)} - (U_{(i+1)} - U_{(i)})U_{(i)}\}(i/m)K(i/m),$$

$$V_2 = \tfrac{1}{2}m^{-1}\sum_{i=1}^{m} E\{(U_{(i+2)} - U_{(i+1)})U_{(i+1)}^2 - (U_{(i+1)} - U_{(i)})U_{(i)}^2\}(i/m)K(i/m),$$

$$V_1' = m^{-1}\sum_{i=1}^{m} E\{(U_{(i+1)} - U_{(i)})U_{(i)}\}K'(i/m),$$

$$T_1' = m^{-1}\sum_{i=1}^{m}(Z_i - 1)K'(i/m), \qquad T' = m^{-1}\sum_{i=1}^{m} Z_i K'(i/m).$$

In this notation,

$$\begin{aligned}
&m^{-1}\sum_{i=1}^{m}(X_{(i+1)} - X_{(i)})K'(i/m)\\
&\qquad = S_n^{-1}T'G'(0) + G''(0)V_1' + o_p(m^{-1/2}n^{-1} + mn^{-2}),\\
&m^{-1}\sum_{i=1}^{m}(X_{(i+2)} - 2X_{(i+1)} + X_{(i)})(i/m)K(i/m)\\
&\qquad = S_n^{-1}T_2G'(0) + G''(0)V_1 + G'''(0)V_2 + o_p(m^{-3/2}n^{-1} + mn^{-3}).
\end{aligned}$$

Now, $E(U_{(i+1)} - U_{(i)}|U_{(i)}) = (1 - U_{(i)})/(n - i + 1)$,

$$E\{U_{(i)}(1 - U_{(i)})\} = \frac{i(n - i + 1)}{(n + 1)(n + 2)},$$

$$E\{U_{(i)}^2(1 - U_{(i)})\} = \frac{i(i + 1)(n - i + 1)}{(n + 1)(n + 2)(n + 3)}.$$

Therefore, $V_1 = n^{-2}v_1 + O(n^{-3})$, $V_2 = (m/n^3)v_2 + O(mn^{-4})$ and $V_1' = (m/n^2)v_1' + O(mn^{-3})$, where

$$v_j = m^{-1}\sum_{i=1}^{m}(i/m)^j K(i/m), \qquad v_j' = m^{-1}\sum_{i=1}^{m}(i/m)^j K'(i/m).$$

Furthermore, $mT_2 = -T_3 + o_p(m^{-1/2})$, where

$$T_3 = m^{-1}\sum_{i=1}^{m}(Z_i - 1)\{K(i/m) + (i/m)K'(i/m)\},$$

and $T' = v_0' + O_p(m^{-1/2})$, where the $O_p(m^{-1/2})$ term has zero mean. Hence,

$$-\hat{\alpha}/m \equiv \frac{\sum_i (X_{(i+2)} - 2X_{(i+1)} + X_{(i)}) i K(i/m)}{\sum_i (X_{(i+1)} - X_{(i)}) K'(i/m)}$$
$$= -\frac{T_3}{v_0'} + \frac{m}{n}\frac{G''(0)}{G'(0)}\frac{v_1}{v_0'} + \left(\frac{m}{n}\right)^2 \left(\frac{G'''(0)}{G'(0)}\frac{v_2}{v_0'} - \frac{G''(0)^2}{G'(0)^2}\frac{v_1 v_1'}{(v_0')^2}\right)$$
$$+ o_p\{m^{-1/2} + (m/n)^2\}$$
$$= O_p\{m^{-1/2} + (m/n)^2\}.$$

Moreover,

$$m^{-1}\sum_{i=1}^{m} (X_{(i+1)} - X_{(i)}) K\{(i+\hat{\alpha})/m\}$$
$$= m^{-1}\sum_{i=1}^{m} (U_{(i+1)} - U_{(i)})\{G'(0) + U_{(i)} G''(0) + \tfrac{1}{2} U_{(i)}^2 G'''(0)\}$$
$$\times \{K(i/m) + (\hat{\alpha}/m) K'(i/m)\} + o_p(m^{-1/2} n^{-1} + m^2 n^{-3})$$
$$= S_n^{-1}\{(v_0 + T_1) + (\hat{\alpha}/m)(v_0' + T_1')\} G'(0) + (m/n^2) v_1 G''(0)$$
$$+ \tfrac{1}{2}(m^2/n^3) v_2 G'''(0) + o_p(m^{-1/2} n^{-1} + m^2 n^{-3}).$$

Hence,

$$n\hat{\beta}_m \equiv n \frac{\sum_i (X_{(i+1)} - X_{(i)}) K\{(i+\hat{\alpha})/m\}}{\sum_i K(i/m)}$$
$$= (1 + T_1 v_0^{-1}) G'(0) + (\hat{\alpha}/m)(v_0'/v_0) G'(0) + (m/n)(v_1/v_0) G''(0)$$
$$+ \tfrac{1}{2}(m/n)^2 (v_2/v_0) G'''(0) + o_p\{m^{-1/2} + (m/n)^2\}$$
$$= G'(0) + (T_1 + T_3) v_0^{-1} G'(0)$$
$$+ (m/n)^2 [-\tfrac{1}{2} G'''(0)(v_2/v_0) + G''(0)^2 G'(0)^{-1} (v_1 v_1'/v_0 v_0')]$$
$$+ o_p\{m^{-1/2} + (m/n)^2\}.$$

The theorem follows from this expansion on noting that $v_j = \kappa_j + o(1)$, $v_j' = \kappa_j' + o(1)$, $n\beta = G'(0) + O(n^{-1})$, and $m^{1/2}(T_1 + T_3)$ is asymptotically normally distributed with zero mean and variance τ^2.

6.2. *Proof of Theorem* 5.3. By Taylor's expansion,

$$A(x) \equiv \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i + \hat{\alpha}(x)}{h}\right)$$
$$= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) + h\frac{\bar{f}'(x)}{\tilde{f}(x)}\rho\left(\frac{x-a}{h}\right)\frac{1}{nh}\sum_{i=1}^{n} K'\left(\frac{x - X_i}{h}\right)$$
$$+ O_p\{h^2 + (nh)^{-1/2}\}.$$

For $j = 0, 1$, $(nh)^{-1} \sum_i K^{(j)}\{(x - X_i)/h\}$ equals its mean, which we shall denote by $A_j(x)$, plus $O_p\{(nh)^{-1/2}\}$. Furthermore,

$$\bar{f}'(x)/\tilde{f}(x) = f'(x)/f(x) + O_p\{h + (nh^3)^{-1/2}\},$$

and so, defining $t = (x - a)/h < 0$, we have

$$A(x) = A_0(x) + hf'(x)f(x)^{-1}\rho(t)A_1(x) + O_p\{h^2 + (nh)^{-1/2}\}. \tag{6.1}$$

Put $I_j(t) = \int_{u>t} u^j K(u)\,du$. In this notation,

$$A_0(x) = I_0(t)f(x) - hI_1(t)f'(x) + O(h^2), \qquad A_1(x) = -f(x)K(t) + O(h).$$

Note, too, that $\rho(t) = -I_1(t)/K(t)$. Hence,

$$A_0(x) + hf'(x)f(x)^{-1}\rho(t)A_1(x) = I_0(t)f(x) + O(h^2). \tag{6.2}$$

Combining (6.1) and (6.2) we deduce that $A(x)/I_0(t) = f(x) + O_p\{h^2 + (nh)^{-1/2}\}$, which establishes (5.5).

6.3. *Proof of Theorem* 5.4. Define

$$\lambda_1(x, y) = \begin{cases} \lambda(x, y), & \text{if } y \le g(x), \\ \lambda(x, g(x)), & \text{otherwise,} \end{cases}$$

and let $\mathcal{P}_1$ be a Poisson process with intensity $n\lambda_1$ in the plane. Without loss of generality, $\mathcal{P}$ is the restriction of $\mathcal{P}_1$ to the region below the curve defined by $y = g(x)$. Let $\mathcal{L}$ denote the straight line passing through $(0, g(0))$, with gradient $g'(0)$; let g_1 be the linear function such that $\mathcal{L}$ has equation $y = g_1(x)$; let $\mathcal{P}_2$ be the Poisson process obtained from $\mathcal{P}_1$ by deleting each point that lies above $\mathcal{L}$; and let $\hat{g}_1$ and $\check{g}_1$ denote the versions of $\hat{g}$ and $\check{g}$, respectively, computed using the data $\mathcal{P}_2$ instead of $\mathcal{P}$. Then, the probability that $\hat{g}_1(0) = \hat{g}(0)$ and $\check{g}_1(0) = \check{g}(0)$ converges to 1 as $n \to \infty$. Therefore, we may suppose without loss of generality that the boundary is linear, and of course also that $g(0) = 0$; these assumptions will be made throughout the work below.

Next we change scale by the factor $n^{-1/2}$ on both axes, altering the difference between neighboring points in $\mathcal{P}$ from $O(n^{-1/2})$ to $O(1)$. In a slight abuse of notation we shall continue to refer to the Poisson process as $\mathcal{P}$, and to define the bias correction by integration over (u, v). [In the original notation this would have been $(n^{-1/2}u, n^{-1/2}v)$, so we are in effect changing the definitions of u and v. The conditions imposed on these quantities in the theorem change directly to those at (6.6) below.] On the new scale, $\mathcal{P}$ has intensity ν_n, say, where ν_n is a bounded, continuous function on the half-plane $\mathcal{H}$ below the line, and satisfies

$$\sup_{|x| \le \varepsilon_n n^{1/2},\, -\varepsilon_n n^{1/2} \le y < \infty,\, (x,y) \in \mathcal{H}} |\nu_n(x, y) - \nu| \to 0 \tag{6.3}$$

for each sequence $\varepsilon_n \downarrow 0$, where $\nu = \lambda(0, g(0))$. We shall use "tilde" rather than "hat" notation for function estimators, however, for example, writing $\tilde{g}$ for the

LFDH estimator of g computed from $\mathcal{P}$ in this setting, taking $\widetilde{\gamma}_y$ to be the version of $\widehat{\gamma}_y$ after rescaling, and putting $\tilde{g}_y(x) = \tilde{g}(x) - y$ and

$$\tilde{\beta}_{uv}(x) = (v-u)^{-1} \int_u^v \{\tilde{g}_y(x) - \widetilde{\gamma}_y(x)\}\, dy.$$

After this change of scale, the assertion in the theorem is equivalent to

$$\tilde{g}(0) + \kappa \tilde{\beta}_{uv}(0) \to \{g'(0)/\nu\}^{1/2} Q_2 \tag{6.4}$$

in distribution, for the same Q_2 as in the theorem.

Let $\mathcal{P}^{\#}$ be the Poisson process of intensity ν in the plane, obtained by thinning points from $\mathcal{P}$, independently and with probability $(\nu_n - \nu)/\nu_n$, in regions where $\nu_n - \nu > 0$, and adding additional points to $\mathcal{P}$, from an independent Poisson process with intensity $\nu - \nu_n$, in places where $\nu - \nu_n > 0$. Write $\tilde{g}^{\#}$ and $\tilde{\beta}^{\#}_{uv}$ for the versions of $\tilde{g}$ and $\tilde{\beta}_{uv}$, respectively, that are obtained using the data $\mathcal{P}^{\#}$ rather than $\mathcal{P}$. Then, noting (6.3), it may be proved from properties of Poisson processes that the probability that $\tilde{g}^{\#}(0) = \tilde{g}(0)$ converges to 1, and $\tilde{\beta}^{\#}(0) - \tilde{\beta}(0) \to 0$ in probability, as $n \to \infty$. Therefore, we may suppose without loss of generality that $\nu_n \equiv \nu$; we shall make this assumption below.

If g has slope $s > 0$, then by stretching the vertical axis by the factor s^{-1} we obtain the case $s = 1$, altering ν to $s\nu$ in the process. Therefore, we may assume without loss of generality that $s = 1$, in which case the assertion in the theorem is equivalent to $\tilde{g}(0) + \kappa \tilde{\beta}_{uv}(0) \to \nu^{-1/2} Q_2$ in distribution. Given $c > 0$, let $\mathcal{E}_c$ denote the event that $\tilde{g}(0)$ is completely determined by those points of $\mathcal{P}$ that fall within the region $\mathcal{R}_c = [-c, c]^2$; and that the configuration of points within $\mathcal{R}_c$ is such that, no matter what the configuration outside $\mathcal{R}_c$, it cannot influence the value of $\tilde{g}(0)$. Let $\mathcal{F}_c$ be the the sigma field generated by the points within $\mathcal{R}_c$.

Define $\mathcal{B}_c$ to be the (fragment of an) LFDH boundary estimate constructed solely from points in $\mathcal{R}_c$. For $\mathcal{E}_c$ to hold it is sufficient that the following event, $\mathcal{E}'_c$, say, hold: the extrapolation to the right of some line segment $\mathcal{L}_1$ of $\mathcal{B}_c$ that lies to the right of the axis $\mathcal{A}$ represented by $x = 0$, and has gradient greater than 1, cuts the line $\mathcal{L}$ (with equation $y = x$) at a point lying within $\mathcal{R}_c$; and the extrapolation to the left of some line segment $\mathcal{L}_2$ of $\mathcal{B}_c$ that lies to the left of $\mathcal{A}$, and has gradient less than 1, cuts $\mathcal{L}$ at a point lying within $\mathcal{R}_c$. (The event $\mathcal{E}'_c$ includes the assertion that $\mathcal{L}_1$ and $\mathcal{L}_2$ both exist.)

Result (6.5) below holds if $\mathcal{E}_c$ there is replaced by $\mathcal{E}'_c$. Therefore, it holds for $\mathcal{E}_c$,

$$\lim_{c\to\infty} \liminf_{n\to\infty} P(\mathcal{E}_c) = 1. \tag{6.5}$$

Define $\tilde{g}_{(c)} = \tilde{g}$ if $\mathcal{E}_c$ holds, and $\tilde{g}_{(c)} = C$ otherwise, where C is an arbitrary constant. Now, $\mathcal{E}_c \in \mathcal{F}_c$, and so $\tilde{g}_{(c)}(0)$ is $\mathcal{F}_c$-measurable. Put $\tilde{g}_{cy}(x) = \tilde{g}_{(c)}(x) - y$.

Let $\mathcal{P}^{\dagger}_c = \{(X^{\dagger}_i, Y^{\dagger}_i) : i = 1, 2, \ldots\}$ denote the Poisson process with intensity ν constructed from $\mathcal{P}$ by removing all points of the latter that lie within $\mathcal{R}_c$, and substituting points of a completely independent Poisson process with intensity ν. Define $\widetilde{\gamma}_{cy}^{\dagger}$ to be the LFDH estimator of $\tilde{g}_{cy}$ computed from

$$\mathcal{P}_{cy}^{\dagger} = \{(X_i^{\dagger}, Y_i^{\dagger}) \in \mathcal{P}_c^{\dagger} : Y_i^{\dagger} \le \tilde{g}_{cy}(X_i^{\dagger})\},$$

and put $\widetilde{\gamma}_c^{\dagger} = \widetilde{\gamma}_{c0}^{\dagger}$,

$$\tilde{\beta}_{cuv}^{\dagger}(x) = (v-u)^{-1} \int_u^v \{\tilde{g}_{cy}(x) - \widetilde{\gamma}_{cy}^{\dagger}(x)\}\, dy, \qquad \tilde{\beta}_c = \tilde{g}_{(c)} - E(\widetilde{\gamma}_c^{\dagger} | \mathcal{F}_c),$$

where $u = u(n)$, $v = v(n)$. Assume that as $n \to \infty$,

$$0 \le u = o(v), \qquad n^{-1/2} v \to 0 \quad \text{and} \quad v \to \infty. \tag{6.6}$$

It may be proved that if (6.6) holds,

$$I(\mathcal{E}_c) |\tilde{\beta}_{cuv}^{\dagger}(0) - \tilde{\beta}_c(0)| \to 0 \tag{6.7}$$

in probability, where $I(\mathcal{E}_c)$ denotes the indicator function of the event $\mathcal{E}_c$. An outline of the derivation is given two paragraphs below.

Let $\mathcal{P}^*$ be a new Poisson process, totally independent of $\mathcal{P}$ on this occasion and with intensity ν. Let $\widetilde{\gamma}^*$ be the LFDH estimator of $\tilde{g}$ — a version of $\widetilde{\gamma}_{c0}$ — that is obtained if, in the construction of $\widetilde{\gamma}_{c0}$, $\mathcal{P}_c^{\dagger}$ is replaced by $\mathcal{P}^*$. Put $\tilde{\beta} = \tilde{g} - E(\widetilde{\gamma}^* | \tilde{g})$. Note that, with probability 1,

$$\tilde{\beta}_c(0) = \tilde{\beta}(0) \qquad \text{on the event } \mathcal{E}_c. \tag{6.8}$$

Formal derivation of (6.7) may proceed by proving convergence in L^2. Less formally, note that conditional on the sigma field $\mathcal{F}_c$, and on the event $\mathcal{E}_c$ holding, the sequence $\{\tilde{g}_{cy}(0) - \widetilde{\gamma}_{cy}^{\dagger}(0) : u < y < v\}$ has the same finite-dimensional distributions as the sequence $\{\tilde{g}_{cy}(0) - \gamma_{cy}^*(0) : u < y < v\}$ (where γ_{cy}^* has the definition of $\gamma_{cy}^{\dagger}$ except that $\mathcal{P}^*$ replaces $\mathcal{P}^{\dagger}$). Call this result (R_1). Since the numbers and distributions of points of a Poisson process that lie in disjoint sets are independent, then by the law of large numbers, the difference between $I(\mathcal{E}_c)\tilde{\beta}_{cuv}^{\dagger}(0)$ and $I(\mathcal{E}_c)E\{\tilde{\beta}_{cuv}^{\dagger}(0)|\mathcal{F}_c\}$ converges in probability to 0, for each fixed $c > 0$; call this result (R_2). By definition of $\tilde{\beta}_c$, $I(\mathcal{E}_c)E\{\tilde{\beta}_{cuv}^{\dagger}(0)|\mathcal{F}_c\}$ equals $I(\mathcal{E}_c)\tilde{\beta}_c(0)$, with probability 1, for each c, u, v; call this result (R_3). By (R_1), $I(\mathcal{E}_c)E\{\tilde{\beta}_{cuv}^{\dagger}(0)|\mathcal{F}_c\}$ also equals $I(\mathcal{E}_c)\tilde{\beta}(0)$ with probability 1; call this result (R_4). Results (R_2) and (R_3) imply (6.7), and (R_3) and (R_4) imply (6.8).

A similar law of large numbers argument, using (6.6), shows that for each fixed $c > 0$,

$$I(\mathcal{E}_c) |\tilde{\beta}_{cuv}^{\dagger}(0) - \tilde{\beta}_{uv}(0)| \to 0 \tag{6.9}$$

in probability. Results (6.7)–(6.9) imply that

$$I(\mathcal{E}_c) |\tilde{\beta}_{uv}(0) - \tilde{\beta}(0)| \to 0$$

in probability. From this property and (6.5) we deduce that the bias-corrected estimator $\tilde{g}_{\mathrm{bc}}(0) = \tilde{g}(0) + \kappa\tilde{\beta}_{uv}(0)$ satisfies

$$\tilde{g}_{\mathrm{bc}}(0) = \tilde{g}(0) + \kappa\tilde{\beta}(0) + o_p(1). \tag{6.10}$$

Recall that the Poisson process with which we are presently working is homogeneous, with an intensity ν that does not depend on n, and the function g is linear and of unit slope, passing through the origin. Therefore, $\tilde{g}(0) + \kappa\tilde{\beta}(0)$ has exactly the same distribution as $\nu^{-1/2}\{\bar{g}(0) + \kappa\bar{\beta}(0)\} = \nu^{-1/2}Q_2$, where $\bar{g}(0) + \kappa\bar{\beta}(0)$ was introduced in Section 3. Hence, (6.4) follows from (6.10). The claimed property $E(Q_2) = 0$ follows directly from the definition of κ.

Acknowledgments. We are grateful for the helpful comments of two referees and an Associate Editor.

REFERENCES

ATHREYA, K. B. (1987a). Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15** 724–731.

ATHREYA, K. B. (1987b). Bootstrap of the mean in the infinite variance case. In *Proceedings of the First World Congress of Bernoulli Society* **2** 95–98. VNU Scientific Press, Utrecht.

BICKEL, P., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than *n* observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31.

BLOCH, D. A. and GASTWIRTH, J. L. (1968). On a simple estimate of the reciprocal of the density function. *Ann. Math. Statist.* **39** 1083–1085.

BOFINGER, E. (1975). Estimation of a density function using order statistics. *Austral. J. Statist.* **17** 1–7.

CHENG, C. (1995). Uniform consistency of generalized kernel estimators of quantile density. *Ann. Statist.* **23** 2285–2291.

CHENG, C. and PARZEN, E. (1997). Unified estimators of smooth quantile and quantile density functions. *J. Statist. Plann. Inference* **59** 291–307.

CHEVALIER, J. (1976). Estimation du support et du contenu du support d'une loi de probabilité. *Ann. Inst. H. Poincaré Probab. Statist.* **12** 339–364.

CHOI, E. and HALL, P. (1999). Data sharpening as a prelude to density estimation. *Biometrika* **86** 941–947.

CSÖRGŐ, M. (1983). *Quantile Processes with Statistical Applications.* SIAM, Philadelphia.

DEPRINS, D., SIMAR, L. and TULKENS, H. (1984). Measuring labor efficiency in post offices. In *The Performance of Public Enterprises*: *Concepts and Measurements* (M. Marchand, P. Pestieau and H. Tulkens, eds.) 243–267. North-Holland, Amsterdam.

FALK, M. (1986). On the estimation of the quantile density function. *Statist. Probab. Lett.* **4** 69–73.

GIJBELS, I., MAMMEN, E., PARK, B. U. and SIMAR, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94** 220–228.

HALL, P. (1989). On convergence rates in nonparametric problems. *Internat. Statist. Rev.* **57** 45–58.

HALL, P. (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions. *Ann. Probab.* **18** 1342–1360.

HALL, P., PARK, B. U. and STERN, S. E. (1998). On polynomial estimators of frontiers and boundaries. *J. Multivariate Anal.* **66** 71–98.

HALL, P., PARK, B. U. and TURLACH, B. A. (1998). Rolling-ball method for estimating the boundary of the support of a point-process intensity. *Ann. Inst. H. Poincaré Probab. Statist.* To appear.

HÄRDLE, W., PARK, B. U. and TSYBAKOV, A. B. (1995). Estimation of nonsharp support boundaries. *J. Multivariate Anal.* **55** 205–218.

JONES, M. C. (1992). Estimating densities, quantiles, quantile densities and density quantiles. *Ann. Inst. Statist. Math.* **44** 721–727.

JONES, M. C. and SIGNORINI, D. F. (1997). A comparison of higher-order bias kernel density estimators. *J. Amer. Statist. Assoc.* **92** 1063–1073.

KNEIP, A., PARK, B. U. and SIMAR, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econom. Theory* **14** 783–793.

KNIGHT, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Ann. Statist.* **17** 1168–1175.

KOROSTELEV, A. P., SIMAR, L. and TSYBAKOV, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23** 476–489.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*. Springer, New York.

LINTON, O. and NIELSEN, J. P. (1994). A multiplicative bias reduction method for nonparametric regression. *Statist. Probab. Lett.* **19** 181–187.

MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotically minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524.

PARK, B. U., SIMAR, L. and WEINER, C. (2000). The FDH estimator for productivity efficiency scores: Asymptotic properties. *Econom. Theory* **16** 855–877.

POLITIS, D., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York.

REISS, R.-D. (1978). Approximate distribution of the maximum deviation of histograms. *Metrika* **25** 9–26.

RIPLEY, B. D. and RASSON, J. P. (1977). Finding the edge of a Poisson forest. *J. Appl. Probab.* **14** 483–491.

SAMIUDDIN, M. and EL-SAYYAD, G. M. (1990). On nonparametric kernel density estimates. *Biometrika* **77** 865–874.

SIDDIQUI, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *Journal of Research of the National Bureau Standards* **64B** 145–150.

SIMAR, L. and WILSON, P. W. (1998). A general methodology for bootstrapping in nonparametric frontier models. Research Report DP 9811, Institut de Statistique, Univ. Catholique de Louvain, Belgium.

WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

WELSH, A. H. (1988). Asymptotically efficient estimation of the sparsity function at a point. *Statist. Sinica* **6** 427–432.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 0200
AUSTRALIA

DEPARTMENT OF STATISTICS
SEOUL NATIONAL UNIVERSITY
SEOUL 151-747
KOREA
E-MAIL: bupark@stats.snu.ac.kr