ON THE ESTIMATION OF A SUPPORT CURVE OF INDETERMINATE SHARPNESS

Peter Hall[1,2], Michael Nussbaum[1,3], Steven E. Stern[1,4]

ABSTRACT. We propose nonparametric methods for estimating the support curve of a bivariate density, when the density decreases at a rate which might vary along the curve. Attention is focussed on cases where the rate of decrease is relatively fast, this being the most difficult setting. It demands the use of a relatively large number of bivariate order statistics. By way of comparison, support curve estimation in the context of slow rates of decrease of the density may be addressed using methods that use only a relatively small number of order statistics at the extremities of the point cloud. In this paper we suggest a new type of estimator, based on projecting onto an axis those data values lying within a thin rectangular strip. Adaptive univariate methods are then applied to the problem of estimating an endpoint of the distribution on the axis. The new method is shown to have theoretically optimal performance in a range of settings. Its numerical properties are explored in a simulation study.

KEYWORDS. Convergence rate, curve estimation, endpoint, order statistic, regular variation, support.

SHORT TITLE. Support curve

AMS (1991) SUBJECT CLASSIFICATION. Primary 62G07, 62H05; Secondary 62H10.

[1] Centre for Mathematics and its Applications, Australian National University, Canberra, Australia.
[2] CSIRO Division of Mathematics and Statistics, Sydney, Australia.
[3] Weierstrass Institute, Berlin, Germany.
[4] Department of Statistics, Australian National University, Canberra, Australia.

1. INTRODUCTION

The problem of estimating the endpoint of a distribution has received considerable attention, not least because of its roots in classical statistical inference. In estimation of the upper extremity of the Uniform distribution on $(0, \theta)$, the largest order statistic is a sufficient statistic for θ. It has an optimal convergence rate in a minimax sense, among distributions with densities that are bounded away from zero in a left-neighbourhood of θ. However, if the density decreases to zero at θ then, depending on the rate of decrease, faster convergence rates may be achieved by taking as the estimator an appropriate function of an increasingly large number of large order statistics. That function depends on at least first-order characteristics of the rate of decrease. These and related issues have been discussed in a parametric setting by Polfedt (1970), Woodroofe (1972, 1974) and Akahira and Takeuchi (1979), among others; and in a nonparametric context by Cooke (1979, 1980), Hall (1982), Smith (1987) and Csörgő and Mason (1989), among others.

In the case of a bivariate density, the role of an endpoint is played by the support curve, being the smallest contour within which the support of the density is contained. Alternatively, the support curve may be defined as the zero-probability contour of a density. Motivated partly by applications to pattern recognition and to boundary detection in image analysis, estimation of support curves and density contours has been considered by Devroye and Wise (1980), Mammen and Tsybakov (1992), Härdle, Park and Tsybakov (1993), Korostelev and Tsybakov (1993) and Tsybakov (1994). In that work it is typically assumed that as the support curve $\mathcal{C}$ is approached from within, the density decreases to zero at a constant, known rate. As in the univariate case, the performance of the curve estimator depends significantly on the rate at which the probability density decreases to zero as the boundary is approached. If that rate is sufficiently slow then optimal estimation may be based on a relatively small number of bivariate order statistics at the extremities of the data set. However, if the rate is unknown and fast then optimal estimation can be significantly more difficult, and may have to be based on an increasingly large number of bivariate order statistics.

Our paper addresses precisely this context. We assume that at a particular point P on $\mathcal{C}$, the density decreases to zero at rate u^α as P is approached from a location distant u from P and inside $\mathcal{C}$. (In this context, distance may be interpreted as perependicular displacement, although displacement in any direction that is not tangential to $\mathcal{C}$ results in the same exponent α, by virtue of the continuity assumed of that function.) The exponent α may be a function of the location of P, and should be estimated either implicitly or explicitly from data, as a prelude to estimating the locus of $\mathcal{C}$. In the context of the previous paragraph, the case where $\alpha < 1$ corresponds to a "slow" rate of decrease of the density. We are interested in the "fast" rate case, where $\alpha > 1$ and is an unknown function of location on the curve. Our approach to the problem is nonparametric in character, in that we assume that unknowns such as the function α and the function describing the locus of $\mathcal{C}$ are known only up to smoothness conditions, not parametrically.

Even in the one-dimensional case, of estimating the endpoint θ of a distribution, the form of the estimator of θ depends critically on the value of α. In the case of known α, Hall (1982) proposed a uniquely but implicitly defined estimator. Csörgő and Mason (1989) suggested an explicitly defined estimator whose first-order performance was identical to that of Hall's approach. Hall extended his method to the case of unknown α, and Csörgő and Mason proposed a plug-in estimator there: first estimate α using methods such as those of Hill (1975), and then substitute the estimate for the true value of α in the formula for their estimator of θ. This method is not entirely satisfactory, however, not least because application of the method of Hill to estimate α does require knowledge of the value of θ. There are ways around this problem, but they involve the use of pilot estimators and, if one seeks optimal convergence rates, iteration of the plug-in procedure.

The difficulties of following this two-stage route are even greater in the bivariate case, where the unknown α is a function. In the present paper we have chosen to use a modified version of the method proposed by Hall (1982). It involves implicit rather than explicit estimation of α. The modification is based on sliding a thin rectangular window through the data. The window is centered on an axis through a point P at which the curve is to be estimated, and those points within the window are projected onto the axis. The estimate at P is then obtained by applying adaptive univariate methods to the univariate distribution on the axis. We shall show that this approach produces consistency whenever $\alpha > 1$, and optimal convergence rates in a range of settings when $\alpha > 2$, although not when $1 < \alpha \leq 2$. Alternative procedures will produce optimal rates in the latter range, and also in other settings. But in the case where α varies with location, which is the subject of this paper, they are awkward to implement and so are not addressed here.

Härdle, Park and Tsybakov (1993) treated the case of fixed $\alpha \geq 0$, but employed estimators based on only a small number of extreme order statistics. Their definition of optimality is somewhat different from ours, being based on function classes that provide bounds only to first-order behaviour at the boundary. By way of contrast, our function classes are based on bounds to second-order behaviour. The different convergence rates of estimators that use differing numbers of extreme order statistics do not emerge from Härdle, Park and Tsybakov's (1993) approach to the problem.

Section 2 will introduce our methods and describe their main theoretical and numerical properties. Optimal bounds for convergence rates will be presented and derived in Section 3, and shown to coincide in many instances with the rates achieved by the estimators suggested in Section 2. Section 4 will present technical arguments behind the main result in Section 2.

2. MAIN PROPERTIES OF THE ESTIMATION PROCEDURE

We now present our estimator and discuss some of its basic properties. Section 2.1 discusses the basic methodology and describes in detail the actual estimation procedure. Section 2.2 then presents the main theoretical results regarding the asymptotic properties of the estimator. Finally, Section 2.3 contains two simulation studies that examine the numerical properties of the estimator.

2.1 METHODOLOGY. Let $y = g(x)$ represent the locus of a curve in the plane, below which n independent random points (X_i, Y_i) are generated according to a distribution with density f. The density is zero above the curve, and decreases to zero as the curve is approached from below. We wish to estimate g.

We assume that the decrease in density is no more than algebraically fast, perhaps with a varying rate that depends on position. Specifically, we suppose that for univariate functions a, b, α and β, and a bivariate function c,

$$f(x,y) = a(x)\,\{g(x)-y\}_+{}^{\alpha(x)} + b(x)\,\{g(x)-y\}_+{}^{\beta(x)} + c(x,y)\,\{g(x)-y\}_+{}^{\beta(x)}, \text{ for } x \in \mathcal{I}\,, \tag{2.1}$$

where $\mathcal{I}$ is a compact interval,

$$\begin{gathered} a > 0,\ |b| > 0,\ \alpha > 1,\ \beta > \alpha;\ \sup_{x\in\mathcal{I}} |c\{x, g(x) - y\}| \to 0 \text{ as } y \downarrow 0; \\ \text{the derivatives } a',\ g' \text{ and } \alpha' \text{ exist and are Hölder continuous} \\ \text{with exponent } t, \text{ where } 0 \le t \le 1; \text{ and } b,\ \beta \text{ are Hölder continuous.} \end{gathered} \tag{2.2}$$

We suppose too that

$$\begin{gathered} \text{the marginal density } e \text{ of } X \text{ is differentiable, and the} \\ \text{derivative is Hölder continuous with exponent } t. \end{gathered} \tag{2.3}$$

Next we suggest an estimator of g. Without loss of generality, suppose we wish to calculate $g(0)$, and that 0 is an interior point of $\mathcal{I}$. Given $h > 0$, let (X_i', Y_i'), for $1 \le i \le N$, denote those data pairs (X_i, Y_i) such that $X_i \in (-h, h)$, indexed in random order. Write $Y_{(1)}' \le \ldots \le Y_{(N)}'$ for the corresponding order statistics, and following Hall (1982), define

$$\xi_i(\theta) = \big(Y'_{(N-i+1)} - Y'_{(N-r+1)}\big)\big/\big(\theta - Y'_{(N-i+1)}\big)\,.$$

Our estimator $\hat{g}(0)$ is based on the r largest order statistics, $Y'_{(N-i)}$ for $0 \le i \le r-1$. It is defined to equal the largest solution, θ, of the equation

$$\left[\sum_{i=1}^{r-1} \log\{1 + \xi_i(\theta)\}\right]^{-1} - \left\{\sum_{i=1}^{r-1} \xi_i(\theta)\right\}^{-1} = r^{-1}, \tag{2.4}$$

or to equal $Y_{(N)}$ if no solution exists. One may show that as either $\theta \to \infty$ or $\theta \to Y_{(N)}$, the left-hand side of (2.4) converges to a limit that is strictly less than r^{-1}. Therefore, since the left-hand side is continuous, (2.4) must have an even number of solutions.

2.2 THEORETICAL RESULTS. Our first theorem describes large-sample properties of $\hat{g}(0)$. It provides an expansion of the difference $\hat{g}(0) - g(0)$ into bias and error-about-the-mean terms, and describes the sizes of the dominant contributions to each. As a prelude to stating the theorem, put $A \equiv 1/\{\alpha(0)+1\}$, $\gamma \equiv \beta - \alpha$,

$$\sigma \equiv \alpha(0)\,\{\alpha(0)-1\}^{1/2}\{\alpha(0)+1\}^{A-(1/2)}\,\{e(0)/a(0)\}^{A}\,,$$

$$c_1 \equiv -\tfrac{2}{3}\,\alpha(0)^2\,\{\alpha(0)-2\}^{-1}\,\{\alpha(0)+1\}^{-A}\,\{a(0)/e(0)\}^{A}\,g'(0)^2\,,$$

$$c_2 \equiv -\alpha(0)\,\{\alpha(0)-1\}\{\alpha(0)+1\}^{A\{\gamma(0)+1\}}\,\beta(0)^{-1}\,\{\beta(0)+1\}^{-2}$$

$$\times\,\gamma(0)^2\,a(0)^{-A\{\beta(0)+2\}}\,b(0)\,e(0)^{A\{\gamma(0)+1\}}\,,$$

$$c_3 \equiv -\tfrac{1}{6}\,\alpha(0)^4\,\{\alpha(0)-1\}\,\{\alpha(0)+1\}^{-(A+1)}\,\{a(0)/e(0)\}^{A}\,g'(0)^2\,.$$

Let Q_1 denote a random variable with the Standard Normal distribution. In the case $1 < \alpha(0) < 2$, define

$$Q_2 \equiv \sum_{i=1}^{\infty} i^{-2A}\left(\sum_{j=1}^{i} Z_j\right)^{-A},$$

where $Z_1, Z_2, \ldots$ are independent exponential random variables with unit mean, independent of Q_1. Recall that N is of size nh, indeed $N/nh \to c$ where $c \equiv 2e(0)$. We may replace N by cnh in the theorem below, without affecting its validity.

Theorem 2.1. *Assume that the bivariate density f and marginal density e satisfy conditions (2.1)–(2.3), and that $e(0) > 0$. Suppose too that for some $0 < \epsilon < 1/4$ and all sufficiently large n,*

$$n^{-(1/2)+\epsilon} \le h \le n^{-\epsilon}, \quad n^{\epsilon} \le r \le n^{1-\epsilon}\,h\,. \tag{2.5}$$

Then if $\alpha(0) > 2$ and $n\,h^{\alpha(0)+2}/r \to 0$,

$$\hat{g}(0) - g(0) = (N/r)^A\,h^2\,c_1 + (r/N)^{A\{\gamma(0)+1\}}\,c_2 + (r/N)^A\,r^{-1/2}\,\sigma\,Q^{(1)}$$

$$+\,O_p\big(h^{t+1}\big) + o_p\big\{(N/r)^A\,h^2 + (r/N)^{A\{\gamma(0)+1\}} + (r/N)^A\,r^{-1/2}\big\}\,;$$

if $\alpha(0) = 2$,

$$\hat{g}(0) - g(0) = (N/r)^A\, h^2\, \log r\, c_3 + (r/N)^{A\{\gamma(0)+1\}}\, c_2 + (r/N)^A\, r^{-1/2}\, \sigma\, Q^{(1)} + O_p\big(h^{t+1}\big) + o_p\big\{(N/r)^A\, h^2\, \log r + (r/N)^{A\{\gamma(0)+1\}} + (r/N)^A\, r^{-1/2}\big\}\,;$$

and if $1 < \alpha(0) < 2$ *and* $n\, h^{\alpha(0)+2} \to \infty$,

$$\hat{g}(0) - g(0) = (r/N)^{A\{\gamma(0)+1\}}\, c_2 + (r/N)^A\, r^{-1/2}\, \sigma\, Q^{(1)} + r^{2A-1}\, N^A\, h^2\, Q^{(2)}\, c_3 + O_p\big(h^{t+1}\big) + o_p\big\{(r/N)^{A\{\gamma(0)+1\}} + (r/N)^A\, r^{-1/2} + r^{2A-1}\, N^A\, h^2\big\}\,,$$

where $Q^{(1)}$ *is asymptotically distributed as* Q_1 *and, when* $\alpha(0) < 2$, $(Q^{(1)}, Q^{(2)})$ *is asymptotically distributed as* (Q_1, Q_2).

The remarks below describe the main implications of the theorem. If $p(n)$, $q(n)$ are sequences of positive numbers, the notation $p(n) \asymp q(n)$ indicates that $p(n)/q(n)$ is bounded away from zero and infinity as $n \to \infty$.

Remark 2.1: *Sign of bias terms.* Since the constants $c_1, \ldots, c_3$ are all negative then the dominant contributions to the bias of $\hat{g}$ are also negative. In this sense, $\hat{g}$ tends to underestimate g.

Remark 2.2: *Optimal choice of* h *and* r *when* $\alpha(0) > 2$. In this range of α there are two deterministic bias terms, of sizes $(N/r)^A\, h^2$ and $(r/N)^{A\{\gamma(0)+1\}}$ respectively, and one stochastic term describing the error about the mean, of size $(r/N)^A\, r^{-1/2}$. Recalling that $N \sim cnh$ we see that these three sources of error are of identical size when

$$h \asymp n^{-(\gamma+2)/(2\alpha+5\gamma+4)} \text{ and } r \asymp n^{4\gamma/(2\alpha+5\gamma+4)}\,. \tag{2.6}$$

If $t \geq \gamma/(\gamma+2)$ then, with this choice of h and r, the theorem implies that $\hat{g} - g = O_p(\delta_n)$ where $\delta_n \equiv n^{-2(\gamma+1)/(2\alpha+5\gamma+4)}$. It also follows from the theorem that for this choice of h and r, and for t strictly greater than $\gamma/(\gamma+2)$, the limiting distribution of $(\hat{g}-g)/\delta_n$ is Normal N(μ,τ^2), where $\mu < 0$ and $\tau > 0$. Observe too that when r and h satisfy (2.6), the conditions (2.5) and $n\, h^{\alpha+2}/r \to 0$ (imposed in the theorem) are both satisfied.

In the context $\alpha(0) > 2$, at least one special case is of particular interest. For large γ, where the model (2.1) is essentially $f(x,y) \equiv a(x)\,\{g(x)-y\}_+{}^{\alpha(x)}$, the optimal sizes of h and r are essentially $n^{-1/5}$ and $n^{4/5} \approx N$, respectively. This bandwidth formula may be recognised as the optimal one for estimation for a twice-differentiable curve. The root mean square convergence rate, of approximately $n^{-2/5}$ when γ is large, is also familiar from that setting. Note particularly that, since $1 \le t \le \gamma/(\gamma+2)$, then $t \to 1$ as $\gamma \to \infty$, and so for large γ we effectively require $t+1=2$ derivatives of g.

For values of t that do not exceed $\gamma/(\gamma+2)$, the optimal convergence rate is achieved not so much by balancing the terms in $(N/r)^A\, h^2$, $(r/N)^{A\{\gamma(0)+1\}}$ and $(r/N)^A\, r^{-1/2}$ on the right-hand side of the expansion of $\hat{g}-g$, but by balancing the terms in h^{t+1}, $(r/N)^{A\{\gamma(0)+1\}}$ and $(r/N)^A\, r^{-1/2}$. Indeed, the theorem implies that when $\alpha(0) > 0$ and we choose

$$h = n^{-(\gamma+1)/\{(t+1)(\alpha+2\gamma+1)+\gamma+1\}} \text{ and } \quad r = n^{2(t+1)\gamma/\{(t+1)(\alpha+2\gamma+1)+\gamma+1)\}} \tag{2.7}$$

then $\hat{g}-g = O_p(\delta_n)$, where

$$\delta_n \equiv n^{-(t+1)(\gamma+1)/\{(t+1)(\alpha+2\gamma+1)+\gamma+1\}}\,. \tag{2.8}$$

Remark 2.5 will address such results in detail.

Remark 2.3: *Optimal choice of h and r when $\alpha(0) = 2$.* This case is similar to that in the previous remark, with the optimal sizes of h and r differing only by a logarithmic factor from what they were there:

$$h \asymp \left\{n^{-(\gamma+2)}\,(\log n)^{-(\alpha+2\gamma+1)}\right\}^{1/(2\alpha+5\gamma+4)}, \text{ and } r \asymp \left(n^2/\log n\right)^{2\gamma/(2\alpha+5\gamma+4)}.$$

If $t > \gamma/(\gamma+2)$ then for these choices of h and r, $\hat{g}-g = O_p(\delta_n)$ where $\delta_n \equiv (n^2/\log n)^{-(\gamma+1)/(2\alpha+5\gamma+4)}$. Indeed, the limiting distribution of $(\hat{g}-g)/\delta_n$ is Normal with negative mean and nonzero variance. If $t < \gamma/(\gamma+2)$ then, for h and r chosen according to (2.7), result (2.8) holds.

Remark 2.4: *Optimal choice of h and r when $1 < \alpha(0) < 2$.* The situation here is distinctly different from that when $\alpha(0) \geq 2$, in that a new stochastic term with a non-Normal asymptotic distribution is introduced into the expansion of $\hat{g} - g$. The optimal sizes of h and r are now

$$h \asymp n^{-(2\alpha+5\gamma+2-\alpha\gamma)/(2\alpha^2+3\alpha\gamma+6\alpha+9\gamma+4)} \text{ and}$$
$$r \asymp n^{4\gamma(\alpha+1)/(2\alpha^2+3\alpha\gamma+6\alpha+9\gamma+4)} .$$

If t is sufficiently far from 0 then for such h and r we have $\hat{g} - g = O_p(\delta_n)$, where $\delta_n \equiv n^{-2(\alpha+1)(\gamma+1)/(2\alpha^2+3\alpha\gamma+6\alpha+9\gamma+4)}$. The asymptotic distribution of $(\hat{g} - g)/\delta_n$ is well-defined and representable as a mixture of the distributions of Q_1 and Q_2, together with a location constant.

Remark 2.5: *Optimal convergence rates.* The "optimality" discussed in Remarks 2.2–2.4 is of course with respect to choice of tuning parameters for the specific estimator $\hat{g}$, and not necessarily with respect to performance of $\hat{g}$ among all possible approaches. It will turn out, however, that when $\alpha(0) > 2$ the convergence rates derived in Remark 2.2 are optimal in the problem of estimating g when the derivative of that function satisfies a Lipschitz condition with exponent $t \leq \gamma/(\gamma+2)$. This and related results will be elucidated in the next section.

Indeed, the techniques that we shall employ to derive Theorem 2.1 may be used to obtain the result below, which provides an upper bound to complement the lower bound that will be derived in Section 3. It describes convergence rates of the estimator $\hat{g}$ uniformly over a class of densities more general than those satisfying (2.1)–(2.3). (These stronger conditions are necessary to derive concise expressions for bias and error-about-the-mean terms in Theorem 2.1. However, if only an order-of-magnitude version of that theorem is required then milder assumptions are adequate.) Let $C > 1$ denote a large positive constant, put $\mathcal{J} = [-1/C, 1/C]$, and assume that for univariate functions a, α and β, and a bivariate function b, the following conditions hold: the density f of (X,Y) satisfies

$$f(x,y) = a(x)\,\{g(x)-y\}_+{}^{\alpha(x)} + b(x,y)\,\{g(x)-y\}_+{}^{\beta(x)} \text{ for } x \in \mathcal{J}\,,$$

where $C^{-1} \leq a \leq C$, $|b| \leq C$, $2 + C^{-1} \leq \alpha \leq C$, $\alpha + C^{-1} \leq \beta \leq C$; the derivatives a', g' and α' exist and, denoted by l, satisfy $|l(0)| \leq C$ and $|l(u)-l(v)| \leq C|u-v|^t$ for $u, v \in \mathcal{J}$, where $0 \leq t \leq 1$; $|\beta(u) - \beta(v)| \leq C|u-v|^{1/C}$ for $u, v \in \mathcal{J}$; the marginal density e of X is differentiable, $e(0) \geq C^{-1}$, $|e'(0)| \leq C$, and $|e'(u) - e'(v)| \leq C|u-v|^{1/C}$ for $u, v \in \mathcal{J}$. Let $\mathcal{F}(t, C)$ denote the class of all such f's.

Theorem 2.2. *Let h and r be given by (2.7), and define δ_n by (2.8), in which formulae the functions α and $\gamma = \beta - \alpha$ should be evaluated at the origin. Fix $t \in (0,1)$. Then, for all C's which are so large that $\mathcal{F}(t, C)$ contains at least one element for which $\gamma(0)/\{\gamma(0)+2\} \geq t$, we have*

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \sup_{f\in\mathcal{F}(t,C):\, \gamma(0)/\{\gamma(0)+2\}\geq t} P\{|\hat{g}(0) - g(0)| \geq \lambda\delta_n\} = 0.$$

Remark 2.6: *Alternative estimators of g.* There are several estimators of g alternative to those treated here. In the case where α is known and fixed, estimation may be based on fitting, by maximum likelihood, local or piecewise polynomials to a and g in the fictitious model $f(x,y) = a(x)\,\{g(x)-y\}_+{}^{\alpha}$. This approach is feasible when the polynomials are linear, but is not as attractive from a computational viewpoint as the reduction-to-one-dimension method studied in the present paper. The case of second or higher degree polynomials is particularly cumbersome. When α is allowed to vary, a local or piecewise polynomial approximation to that function may be introduced, although this does make the methods very awkward.

The performance of such methods under the more plausible model (2.1) may be described using arguments similar to those developed in Section 4. They attain optimal convergence rates in a wide range of settings, but at the price of significantly increased complexity.

Remark 2.7: *Generalizations to Poisson point processes.* It is straightforward to generalize Theorems 2.1 and 2.2, and also the results in the next section, to the case where the data (X_i, Y_i) originate from a bivariate Poisson processes with intensity λf, where λ is a positive constant. The function f need not be a density, but the only change which that demands is that f need not integrate to 1. The role of n is now played by λ; in particular, the theorems are valid for high-intensity Poisson processes. In all other respects the conditions required for the theorems remain unchanged. The constants σ, $c_1, \ldots, c_3$ defined prior to Theorem 2.1 need to be adjusted, although the c_i's remain negative. With these alterations, Theorems 2.1 and 2.2 hold as before.

2.3 NUMERICAL RESULTS. We present two numerical studies that examine the performance of our estimation procedure for relatively large samples ($n = 5000$ and $n = 7500$). The first study addresses the estimator's properties when the boundary is relatively non-linear. The second examines the estimator's capabilities in distinguishing between a constant boundary with changing exponent function α and a non-constant boundary with constant exponent α. In each case, we focus on the case where $\alpha(x) > 2$. Data are generated such that the marginal distribution of the abscissa values is uniform between 0 and 1, and such that the function $\beta(x)$ is equal to twice $\alpha(x)$. In this situation, $\gamma = \alpha$, so that Remark 2.2 following Theorem 2.1 implies that the optimal sizes of the bandwidth and the number of order statistics included in the estimation procedure are $h \asymp n^{-(\alpha+2)/(7\alpha+4)}$ and $r \asymp n^{4\alpha/(7\alpha+4)}$. Using the fact that $N \asymp nh$, it is easily seen that $r \asymp N^{2\alpha/(3\alpha+1)}$, and therefore in each of the simuations we choose $r(x)$ to be proportional to $\{N(x)\}^{2/3}$.

Simulation Study I. Here we set the boundary curve to be $g(x) = 2+4x-18x^2+16x^3$ and the exponent function to be $\alpha(x) = 2 + 3x$, for $x \in [0, 1]$. We chose a sample size of $n = 5000$ points and set $r(x) = 4\{N(x)\}^{2/3}$. Figure 1 shows the results of the new estimation procedure for three different choices of the bandwidth, $h = 0.025, 0.05, 0.1$. The three plots clearly demonstrate the trade-off in variance versus bias as the bandwidth increases. For comparison, each of the plots presents a boundary estimate based solely on the maximum order statistic. As can be seen, particularly in Figure 1b, the new estimate provides a noticeable improvement over the estimate based only the maximal order statistic, particularly in the range where the abscissa value is large, which corresponds to the region with very large exponent α.

One obvious feature of the new estimation procedure is that it produces boundary estimates which are quite “rough” and prone to “spikes”. To alleviate this problem it may be useful to consider a variable bandwidth. Alternatively, we might smooth the boundary estimate. Figure 2a presents a LOWESS smooth of the boundary estimate shown in Figure 1b, as well as boundary estimates using the same bandwidth, h, and number of order statistics, r, for four additional datasets each of size $n = 5000$. Again, for comparison, a LOWESS smooth of the boundary estimates based on the maximal order statistic is presented in Figure 2b. While the smoothed estimates in Figure 2b capture the basic shape of the boundary, they are significantly biased. The smoothed version of our new boundary estimate not only captures the shape but also the location of the boundary.

Simulation Study II. Here we compare two situations. First, we set the boundary function to be constant, in fact $g(x) = 2$, and the exponent function to be quadratic, $\alpha(x) = 2 + 24x - 24x^2$ for $x \in [0, 1]$. By way of contrast, in the second situation it is the boundary which is quadratic, $g(x) = 2 - 4x + 4x^2$, while the exponent function is constant at $\alpha(x) = 2$. For samples of size $n = 7500$, each of these two situations produces data which have similar appearances at the upper extremity of the point clouds, despite the difference in boundary curves. This implies that the simple estimator which uses only the largest order statistic within the chosen bandwidth will not be able to easily distinguish between the two situations. However, our estimator, by virtue of its construction using the r largest order statistics, can make the distinction much more readily. Figure 3a presents a plot of the new estimator as well as the estimator based on the maximal order statistic, in the case of the constant boundary and quadratic exponent function $\alpha(x)$. For this plot the bandwidth was $h = 0.1$, while the number of order statistics used was $r(x) = 8\{N(x)\}^{2/3}$. In contrast, Figure 3b presents the same estimation procedures in the case of an underlying quadratic boundary with a constant exponent function α. Again, the chosen bandwidth and number of order statistics used are $h = 0.1$ and $r(x) = 8\{N(x)\}^{2/3}$, respectively. As with the previous simulation study, the new estimation procedure provides quite "ragged" curves, though again this may be mitigated somewhat by the choice of a more flexible $r(x)$ function or a variable bandwidth. In addition, smoothing may be employed as in the previous example. Figure 4 presents LOWESS smooths of the estimates presented in Figure 3. Figure 4a shows that the new estimator distinguishes between the two cases to some degree, while Figure 4b shows that the estimator based solely on the maximal order statistic does not distinguish between the two cases at all.

3. BEST ATTAINABLE CONVERGENCE RATE

In this section we will assume that the support curve g is of general smoothness $\tau > 0$. More specifically, let $\lfloor \tau \rfloor$ be the largest nonnegative integer $< \tau$ and assume that the derivative $g^{\lfloor \tau \rfloor}$ exists and satisfies $\left| g^{\lfloor \tau \rfloor}(u) - g^{\lfloor \tau \rfloor}(v) \right| \leq C \, |u - v|^{\tau - \lfloor \tau \rfloor}$ for $u, v \in \mathcal{J}$. The class of such g's will be denoted by $\Lambda^\tau(C)$. For the lower risk bound, we will assume that the functions a, α and β are known. The assumptions constituting the class $\mathcal{F}(t, C)$ in section 2 will remain in force, with the exception that the lower bound for α is relaxed to $1 + C^{-1}$ instead of $2 + C^{-1}$. The corresponding class of all f's when a, α and β are fixed will be denoted by $\mathcal{F}'(\tau, C)$. We have to assume that this class is sufficiently rich: there exists $C' < C$ such that $\mathcal{F}'(\tau, C')$ is nonempty.

Theorem 3.1. *Define δ_n as in (2.8) where $t + 1 = \tau$. Then for all $\tau > 0$*

$$\lim_{\lambda \to 0} \liminf_{n \to \infty} \inf_{\hat{g}(0)} \sup_{f \in \mathcal{F}'(\tau, C)} P\{|\hat{g}(0) - g(0)| \geq \lambda \delta_n\} > 0$$

where the infimum is taken over all estimators $\hat{g}(0)$ at sample size n .

Introduce notation $A = 1/(\alpha + 1)$, $B = A\gamma$ where $\alpha = \alpha(0)$, $\gamma = \gamma(0)$ and define a rate exponent ρ by $\delta_n = n^{-\rho}$. In this notation,

$$\rho = \tau / (\tau D^{-1} + 1), \quad \text{where} \quad D = \frac{A + B}{2B + 1}.$$

To understand that lower bound result, consider the problem of endpoint estimation on the real line: suppose we have i. i. d. observations Y_i, $i = 1, \ldots, n$ with density ℓ, where for some $a, C > 0$,, $\alpha > 1$, $\beta > \alpha$ and some θ

$$\ell = \bar{\ell}(\theta - y), \ \bar{\ell}(y) = a y_+^\alpha + b(y) y_+^\beta, \ \ |b(y)| \leq C. \tag{3.1}$$

Remark 3.1: For this problem of endpoint estimation it is known that n^{-D} is an attainable rate (Hall (1982b), Csörgő and Mason (1989)), and we will see below that it is optimal. This problem with a nonparametric nuisance term $b(y) y_+^\beta$ in (3.1) is of functional estimation type, with a rate n^{-D} similar to those occurring in smoothing.

The nuisance term defines "indeterminate sharpness" in our terminology. In the two dimensional case with a support curve g which is Hölder smooth, we have an additional smoothing problem. The above optimal rate $n^{-\tau/(\tau D^{-1}+1)} = n^{-\rho}$ results from the superposition of these two nonparametric problems. Accordingly remark 2.2 describes this rate as the result of a balancing problem which involves three terms (cp. (2.7)), i. e. two bias terms and one variance term.

Remark 3.2: The general form of the rate exponent $\rho = \tau/(\tau/D + 1)$ is well known in edge estimation, see e. g. Korostelev and Tsybakov (1993a), (1993b). The most prominent case there has been the case $D = 1$ which corresponds to a one dimensional endpoint estimation problem of a uniform density, where in (3.1) $\alpha = 0$. For such a sharp support curve, even an asymptotic minimax constant has been found; see Korostelev, Simar and Tsybakov (1992). To make the connection, we discuss two limiting cases in the endpoint problem (3.1):

i) $\beta \to \infty$, *i. e.* $B \to \infty$ *where* $D \to 1/2$. In this case the term $b(y)y_+^\beta$, $\beta > \alpha$ becomes neglibible near 0, and an appropriate limiting problem for (3.1) is defined by

$$\bar{\ell}(y) = ay_+^\alpha \quad \text{for} \quad |y| \leq \kappa$$

for some small κ, i. e. a parametric problem. That would mean "determinate sharpness". The value of α is critical here: for $0 \leq \alpha < 1$ the parametric problem is nonregular, the rate n^{-A} of the largest order statistic is optimal, and this is better than n^{-D}. The previous superposition heuristic explains the rate exponent $\rho = \tau/(\tau A^{-1}+1)$ for the corresponding support curve problem, e.g. for the uniform density on a domain (Korostelev and Tsybakov (1993a)). For $\alpha > 1$ the endpoint problem turns regular and a parametric rate $n^{-D} = n^{-1/2}$ obtains. In theorem 3.1 that corresponds to the limiting case $\gamma \to \infty$. Thus for the support curve when $\alpha > 1$, $B \to \infty$ we get a smoothing problem similar to those of local averaging type, and the optimal rate exponent is of the well known form $\rho = \tau/(\tau D^{-1} + 1)$ $= \tau/(2\tau + 1)$.

ii) *$\beta \downarrow \alpha$, i. e. $B \to 0$ where $D \to A$.* An appropriate limiting problem for (3.1) is one where $\bar{\ell}$ is restricted by

$$\bar{\ell}(y) \geq a y_+^\alpha \quad \text{for} \quad |y| \leq \kappa$$

for some small κ. Here $n^{-1/(\alpha+1)} = n^{-A}$ is the optimal rate for any $\alpha \geq 0$, and it is again attained by the largest order statistic. (We abbreviate here; this reasoning can be justified by the results of Härdle, Park and Tsybakov (1993) or by our results for the endpoint problem below). We may conclude that the corresponding support curve problem should have rate exponent $\rho = \tau/(\tau A^{-1} + 1)$ for any $\alpha \geq 0$. Indeed this is the result of Härdle, Park and Tsybakov (1993) who impose a condition on the density f of (X, Y) similar to

$$f(x, y) \geq a\{g(x) - y\}_+^\alpha \quad \text{for} \quad 0 \leq \{g(x) - y\} \leq \kappa,\ x \in \mathcal{J}$$

for some small κ. In our terminology, this again is a case of "determinate sharpness" with rate governed solely by α.

Remark 3.3: We have seen in section 2 that the rate δ_n is attainable when $1 \leq \tau \leq 1 + \gamma/(\gamma + 2)$ and the function α satisfies $\alpha \geq 2 + C^{-1}$. The limitation to that narrower range in comparison with theorem 3.1 is due to the specific form of our estimator, which is comparatively simple given the complex situation.

Remark 3.4: For estimating an unknown exponent (or tail rate) α, Hall and Welsh (1984) established a best possible rate; attainability is shown e.g. by Csörgő, Deheuvels and Mason (1985). The tail rate functional is treated again by Donoho and Liu (1991) from the modulus of continuity viewpoint. We will apply that methodology to the endpoint functional and to the support curve problem.

Remark 3.5: Both estimation problems (endpoint and tail rate) are closely related to statistical issues in extreme value theory; in particular the nonparametric term $b(y)y_+^\beta$ in (3.1) ("indeterminate sharpness") constitutes a neighborhood of a generalized Pareto distribution (see Falk, Hüsler, Reiss (1994), chap. 2.2, Marohn (1991), Janssen and Marohn (1994) and the literature cited therein).

To derive Theorem 3.1 we shall follow Donoho and Liu (1991) and consider the value $g(0)$ as a functional on the set of densities f. It is then sufficient to estimate its Hellinger modulus of continuity, i. e. to exhibit a sequence of pairs f_0, $f_1 \in \mathcal{F}'(\tau, C)$ such that for the corresponding support curves g_0, g_1 we have

$$H(f_0, f_1) \preceq n^{-1/2} \quad \text{and} \quad |g_0(0) - g_1(0)| \succeq n^{-\rho} \tag{3.2}$$

where $H(\cdot,\cdot)$ is Hellinger distance. In the sequel the notation $n_1 \preceq n_2$ for two sequences means that $n_1 = O(n_2)$, $n_1 \succeq n_2$ means that $n_2 = O(n_1)$, and $n_1 \asymp n_2$ means that both $n_1 \preceq n_2$ and $n_1 \succeq n_2$. We shall use notation κ (or K) for positive constants, small or large respectively. The constant C is held fixed at its value in the class $\mathcal{F}'(\tau, C)$.

Consider again the endpoint problem (3.1) where $\alpha > 1$ and call $\mathcal{F}_0(C)$ the class of densities ℓ in (3.1) when θ varies in R. We will exhibit a sequence of pairs $\ell_0, \ell_1 \in \mathcal{F}_0(C)$ such that for the corresponding endpoints θ_0, θ_1

$$H(\ell_0, \ell_1) \preceq n^{-1/2} \quad \text{and} \quad |\theta_0 - \theta_1| \succeq n^{-D}. \tag{3.3}$$

Indeed this will follow from lemma 3.2 below by putting $\theta \asymp n^{-D}$. For proving (3.3), we will construct for two given functional values 0 and θ a pair of densities in $\mathcal{F}_0(C)$ which are at a minimal Hellinger distance. Consider a function

$$\ell_0(y) = ay^\alpha \quad \text{for} \quad 0 \le y \le \kappa.$$

Assume κ is small enough so that $\ell_0(y)$ can be continued to a density outside $[0, \kappa]$. For any $\theta > 0$ define

$$\begin{aligned} \ell_1(y,\theta) &= a(y-\theta)_+^\alpha + C(y-\theta)_+^\beta \quad \text{for} \quad 0 \le y \le y_0(\theta) \\ &= ay^\alpha \quad \text{for} \quad y_0(\theta) < y \le \kappa \end{aligned}$$

where the "cutoff point" $y_0(\theta)$ is selected such that

$$\int_0^\kappa \ell_1(y,\theta)dy = \int_0^\kappa \ell_0(y)dy. \qquad (3.4).$$

Provided that is possible, put $\ell_1(y,\theta) = \ell_0(y)$ for $y > \kappa$. In view of (3.4) $\ell_1(\cdot,\theta)$ then is also density. The next technical lemma makes this precise.

Lemma 3.1. *For sufficiently small $\theta > 0$, unique solutions $y = \tilde{y}(\theta)$* and $y = y_0(\theta)$ of

$$a(y-\theta)^\alpha + C(y-\theta)^\beta = ay^\alpha,\ \theta \le y \le \kappa$$

and of (3.4) respectively, exist and satisfy

$$\tilde{y}(\theta) \sim K_1 \theta^{A/(A+B)}, \quad y_0(\theta) \sim K_2 \theta^{A/(A+B)} \quad \text{as} \quad \theta \to 0, \quad \textit{where}$$
$$K_1 = ((A^{-1}-1)aC^{-1})^{A/(A+B)}, \quad K_2 = ((B+1)A^{-1}aC^{-1})^{A/(A+B)}.$$

Proof. Consider the function of y

$$a(y-\theta)^\alpha - ay^\alpha + C(y-\theta)^\beta.$$

For $y \to \theta$ it becomes negative, while at $y = \kappa$ it is positive for sufficiently small θ. Hence a solution exists for sufficiently small θ. Note that $\alpha = 1/A - 1$, $\beta = (B+1)/A - 1$, so any solution $\tilde{y}$ solves

$$a(y-\theta)^{1/A-1} - ay^{1/A-1} + C(y-\theta)^{(B+1)/A-1} = 0.$$

Put $\tilde{y} = \tilde{u}\theta$; then $\tilde{u} > 1$ since $\tilde{y} > \theta$, and we obtain

$$a((\tilde{u}-1)\theta)^{1/A-1} - a(\tilde{u}\theta)^{1/A-1} + C((\tilde{u}-1)\theta)^{(B+1)/A-1} = 0 \qquad (3.5)$$

or

$$1 - (1-1/\tilde{u})^{1/A-1} = Ca^{-1}(1-1/\tilde{u})^{1/A-1}(\tilde{u}-1)^{B/A}\theta^{B/A}. \qquad (3.6)$$

Suppose that $\tilde{u}$ stays bounded as $\theta \to 0$; then $1-(1-1/\tilde{u})^{1/A-1}$ is bounded away from 0 while the right hand side tends to 0, a contradiction. Hence $\tilde{u} \to \infty$. To prove uniqueness, consider the sign of the derivative of (3.5) at $\tilde{u}$. This derivative divided by $\theta^{1/A-1}\tilde{u}^{1/A-2}$ is

$$a(A^{-1}-1)((1-1/\tilde{u})^{1/A-2}-1) + C((B+1)A^{-1}-1)(1-1/\tilde{u})^{1/A-2}(\tilde{u}-1)^{B/A}\theta^{B/A}.$$

Since $\tilde{u} \to \infty$, and $B > 0$, the above tends to ∞ and the sign is eventually positive. Hence the solution $\tilde{y} = \tilde{u}\theta$ is unique for sufficiently small θ. We now expand the lhs in (3.6) and obtain

$$(A^{-1} - 1)\tilde{u}^{-1} \sim Ca^{-1}\tilde{u}^{B/A}\theta^{B/A}$$

which yields $\tilde{u} \sim ((A^{-1} - 1)aC^{-1})^{A/(A+B)}\theta^{-B/(A+B)}$ and the asymptotics of $\tilde{y}$ as claimed.

For $y_0(\theta)$, it suffices to consider (3.4) with an integration domain $[0, y_0(\theta)]$, or equivalently

$$aA(y - \theta)^{1/A} - aAy^{1/A} + (B + 1)^{-1}AC(y - \theta)^{(B+1)/A} = 0.$$

The argument is now analogous to the previous one, where only the constants and the exponents are changed.

Now we are ready for the basic estimate of the Hellinger modulus in the end-point problem.

Lemma 3.2. *As* $\theta \to 0$,

$$H^2(\ell_0, \ell_1(\cdot, \theta)) \le K_3\theta^{1/D}$$

where $\quad K_3 = aA + K_4 + K_5$,

$$K_4 = (A^{-1} - 1)^2(A^{-1} - 2)^{-1}a(2K_1)^{1/A-2},\ K_5 = K_6^2 Aa(2K_2)^{1/A},$$

$$K_6 = Ca^{-1}(2K_2)^{B/A}.$$

Proof. Define

$$z(\theta) = \theta^{A/D}.$$

Consider first the integral from 0 to z. Note that $D = \frac{A+B}{2B+1} < A + B$; hence $z = o(\tilde{y})$ and in this domain we have $\ell_1(y, \theta) < \ell_0(y)$. Consequently

$$\int_0^z \{\ell_0^{1/2} - \ell_1^{1/2}(\cdot, \theta)\}^2 \le \int_0^z \ell_0 = aAz^{1/A} = aA\theta^{1/D}. \tag{3.7}$$

Consider the domain $[z, \tilde{y}]$. Since $A/D = \frac{2AB+A}{A+B} < 1$ in view of $A < 1/2$, we have $y/\theta \to \infty$ uniformly in this domain. Define for $y \in [z, \tilde{y}]$

$$T = \ell_1(y,\theta)/\ell_0(y) =$$
$$= \left\{(y-\theta)^{1/A-1} + Ca^{-1}(y-\theta)^{(B+1)/A-1}\right\} y^{1-1/A}.$$

Putting $y = u\theta$ we obtain

$$T = (1-1/u)^{1/A-1} + Ca^{-1}(1-1/u)^{1/A-1}(u-1)^{B/A}\theta^{B/A}. \tag{3.8}$$

By the definition of $\tilde{y}$ we have $T \le 1$ here; since the second term on the rhs of (3.8) is positive, we have

$$|1-T| \le 1 - (1-1/u)^{1/A-1}. \tag{3.9}$$

Since $y \ge z$, we know that $1/u = o(1)$ uniformly over $y \ge z$. We may hence expand the rhs in (3.9) and obtain

$$|1-T| \le u^{-1}(A^{-1}-1)\sup_{u\ge 1}(1-1/u)^{1/A-2} \le (A^{-1}-1)\theta/y.$$

Here we used again that $1/A - 2 > 0$. Evaluating now the integral over this domain, we get

$$\int_z^{\tilde{y}} \{\ell_0^{1/2} - \ell_1^{1/2}(\cdot,\theta)\}^2 = \int_z^{\tilde{y}} \ell_0(1-T^{1/2})^2$$
$$\le (A^{-1}-1)^2 \int_z^{\tilde{y}} \ell_0(y)(\theta/y)^2 dy = \theta^2(A^{-1}-1)^2 a \int_z^{\tilde{y}} y^{1/A-3} dy$$
$$\le (A^{-1}-1)^2(A^{-1}-2)^{-1} a\theta^2 \tilde{y}^{1/A-2}$$

(note that $A < 1/2$ entails integrability here). Using lemma 3.1 we obtain

$$\int_z^{\tilde{y}} \{\ell_0^{1/2} - \ell_1^{1/2}(\cdot,\theta)\}^2 \le K_4\theta^2\theta^{(1-2A)/(A+B)} = K_4\theta^{1/D} \tag{3.10}$$

where

$$K_4 = (A^{-1}-1)^2(A^{-1}-2)^{-1} a (2K_1)^{1/A-2}.$$

The third integral over $[\tilde{y}, y_0]$ will be evaluated as follows. Defining T as in (3.7), we get from the definition of $\tilde{y}$ that $T \ge 1$. Then, since the first term on the rhs in (3.8) is < 1,

$$|1-T| \le Ca^{-1}u^{B/A}\theta^{B/A}.$$

For this we get

$$|1-T| \le Ca^{-1}(y/\theta)^{B/A}\theta^{B/A} \le Ca^{-1}y_0^{B/A} \le Ca^{-1}(2K_2)^{B/A}\theta^{B/(A+B)}.$$

Putting $K_6 = Ca^{-1}(2K_2)^{B/A}$, we obtain

$$\int_{\tilde{y}}^{y_0} \{\ell_0^{1/2} - \ell_1^{1/2}(\cdot,\theta)\}^2 = \int_{\tilde{y}}^{y_0} \ell_0(1-T^{1/2})^2$$
$$\le \theta^{2B/(A+B)}K_6^2 \int_{\tilde{y}}^{y_0} \ell_0 \le K_6^2 Aa\theta^{2B/(A+B)}y_0^{1/A} \le K_5\theta^{1/D},$$

where $\quad K_5 = K_6^2 Aa(2K_2)^{1/A}$.

The lemma follows from this result, (3.7) and (3.8).

In the two dimensional support curve problem, let f_0 be an element of the class $\mathcal{F}'(\tau, C')$ for a $C' < C$ and let g_0 be the corresponding support curve in the Hölder class $\Lambda^\tau(C')$. Suppose that

$$f_0(x,y) = a(x)\{g_0(x)-y\}_+^{\alpha(x)} + b(x,y)\{g_0(x)-y\}_+^{\beta(x)} \quad \text{for} \quad x \in \mathcal{J} \qquad (3.11)$$

where $|b(x,y)| \le C'$.

Lemma 3.3. *The term $b(x,y)$ in (3.11) can be modified such that for some small κ*

$$b(x,y) = 0 \quad \text{for} \quad 0 \le g_0(x) - y \le \kappa \quad \text{and} \quad |x| \le \kappa, \qquad (3.12)$$
$$|b(x,y)| \le C \quad \text{for} \quad x \in \mathcal{J}, \qquad (3.13)$$

and the resulting left-hand side in (3.11) is a density in $\mathcal{F}'(\tau, C)$.

Proof. First fix x and start with a one dimensional construction. Suppose that a function is of form

$$f_0(y) = ay^\alpha + b(y)y^\beta,\ y \ge 0$$

where $|b(y)| \le C'$. Define for $y \ge 0$

$$\bar{f}_0(y) = f_0(y) - b(y)y^\beta \chi_{[0,\kappa]}(y) + \lambda(C-C')y^\beta \chi_{(\kappa,K]}(y) \qquad (3.14)$$

where $K = (\frac{C}{C-C'})^{A/(B+1)}\kappa$ and $\lambda \in [-1,1]$ is chosen such that

$$\int \bar{f}_0(y)dy = \int f_0(y)dy. \tag{3.15}$$

Such a choice of λ is possible, since

$$\left| \int b(y) y^\beta \chi_{[0,\kappa]}(y) dy \right| \le C' \int_0^\kappa y^\beta dy = C' A (B+1)^{-1} \kappa^{(B+1)/A},$$

whereas

$$\int \lambda (C - C') y^\beta \chi_{(\kappa,K]}(y) dy = \lambda (C - C') \int_\kappa^K y^\beta dy = \lambda C' A (B+1)^{-1} \kappa^{(B+1)/A}.$$

Furthermore, it can be seen that $\bar{f}_0(y)$ has a representation

$$\bar{f}_0(y) = a y^\alpha + \bar{b}(y) y^\beta, \; y \ge 0 \tag{3.16}$$

where $|\bar{b}(y)| \le C$. Indeed, $\bar{b}(y) = 0$ on $[0,\kappa]$, and on $(\kappa, K]$ we have

$$|\bar{b}(y)| = |b(y) + \lambda(C - C')| \le C' + |\lambda|(C - C') \le C.$$

Then (3.16) implies that $\bar{f}_0$ is positive for sufficiently small κ. Thus, if f_0 is a density with $|b(y)| \le C'$ then $\bar{f}_0$ is a density with $|\bar{b}(y)| \le C$.

Consider now the representation (3.11) of f_0. For fixed x with $|x| \le \kappa$ apply the modification according to (3.14) with an argument $g_0(x) - y$ in place of y. Call this modified function $\bar{f}_0(x,y)$. Then (3.12) holds and (3.15) implies for each $x \in \mathcal{J}$

$$\int \bar{f}_0(x,y)dy = \int f_0(x,y)dy\,. \tag{3.17}$$

Integrating over $x \in \mathcal{J}$ we see that $\bar{f}_0(x,y)$ integrates to one, and since it is nonnegative it is a density. Then (3.17) implies that the marginal density of X is the same as that for f_0. Moreover, $\bar{f}_0(x,y)$ has a representation (3.11) in which $|b(x,y)| \le C$ (as a consequence of (3.16)). Hence $\bar{f}_0(x,y)$ is an element of the class $\mathcal{F}'(\tau, C)$, and the lemma is proved.

We assume now that the density f_0 fulfills (3.11) - (3.13); thus it is in $\mathcal{F}'(\tau, C)$ but the support curve g_0 is in the Hölder class $\Lambda^\tau(C')$. To construct the alternative

f_1, let φ be an infinitely differentiable function with support in $[-1, 1]$ such that $0 \le \varphi(x) \le 1$ and $\varphi(0) = 1$. Let $\kappa > 0$ and define a function

$$\theta(x) = \kappa\, m^{-\tau} \varphi(mx),\ x \in \mathcal{J}$$

where $m > 1$. Define a perturbed support curve g_1 by

$$g_1(x) = g_0(x) - \theta(x),\ x \in \mathcal{J}.$$

This function is in $\Lambda^\tau(C)$ for sufficiently large m if κ is chosen sufficiently small. We shall let m be dependent upon n in the sequel. Specifically, we put

$$m = n^{1/(\tau D^{-1}+1)}. \tag{3.18}$$

Lemma 3.4. *There is a density $f_1 \in \mathcal{F}'(\tau, C)$ which has support curve g_1 such that $H^2(f_1, f_0) \preceq n^{-1}$.*

Proof. Indicate the dependence of ℓ_0 and ℓ_1 on θ, a, α, β, C by $\ell_0(y; a, \alpha, \beta)$ and $\ell_1(y; \theta, a, \alpha, \beta, C)$. Relations (3.11) and (3.12) imply that f_0 can be represented

$$f_0(x, y) = \ell_0(g_0(x) - y); a(x), \alpha(x), \beta(x)) \quad \text{for } 0 \le g_0(x) - y \le \kappa \quad \text{and } |x| \le \kappa.$$

Accordingly define

$$f_1(x, y) = \ell_1(g_0(x) - y; \theta(x), a(x), \alpha(x), \beta(x), C) \text{ for } 0 \le g_0(x) - y \le \kappa \text{ and } |x| \le \kappa$$

and put $f_1 = f_0$ outside that domain. It follows from (3.4) that for each $x \in \mathcal{J}$

$$\int f_1(x, y)dy = \int f_0(x, y)dy$$

so that f_1 is a density which has the same marginal X-density as f_0. By construction of ℓ_1 the density f_1 fulfills

$$f_1(x, y) = a(x)\{g_1(x) - y\}_+^{\alpha(x)} + b(x, y)\{g_1(x) - y\}_+^{\beta(x)} \quad \text{for} \quad x \in \mathcal{J}$$

where $|b(x, y)| \le C$. We conclude that $f_1 \in \mathcal{F}'(\tau, C)$.

To estimate the Hellinger distance of f_1 and f_0, we argue from lemma 3.2 and observe that the constants there now depend on x. At this point we need an extension of lemma 3.2 with uniformity in a, α, β over the range $C^{-1} \leq a \leq C$, $1 + C^{-1} \leq \alpha \leq C$, $\alpha + C^{-1} \leq \beta \leq C$. Such a uniform version can easily be established, on the basis of a uniform version of lemma 3.1. With obvious notation, we conclude that $K_3(x)$ is uniformly bounded, while $1/D(x)$ fulfills a Lipschitz condition:

$$\left|D^{-1}(x_1) - D^{-1}(x_2)\right| \leq K|x_1 - x_2|^{1/C}. \tag{3.19}$$

We obtain

$$\begin{aligned} H^2(f_1, f_0) &= \int\int \{f_1^{1/2}(x,y) - f_0^{1/2}(x,y)\}^2 dydx \\ &\leq \int K_3(x)\theta(x)^{1/D(x)}dx \leq K \int \left\{m^{-\tau}\varphi(mx)\right\}^{1/D(x)} dx \end{aligned}$$

$$= K \int_{-\kappa/m}^{\kappa/m} \left\{m^{-\tau}\varphi(mx)\right\}^{1/D(0)} \exp\left[\{D^{-1}(x) - D^{-1}(0)\} \log\{m^{-\tau}\varphi(mx)\}\right] dx.$$

Now (3.19) implies that $|D^{-1}(x) - D^{-1}(0)| \leq K m^{-1/C}$ so that the term in $\exp(\ldots)$ tends to 0 uniformly in $x \in [-\kappa/m, -\kappa/m]$. Hence

$$\begin{aligned} H^2(f_1, f_0) &\leq K \int_{-\kappa/m}^{\kappa/m} \left\{m^{-\tau}\varphi(mx)\right\}^{1/D(0)} dx \\ &\asymp m^{-\tau/D(0)-1} \int \varphi(x)^{1/D(0)}dx \asymp n^{-1} \end{aligned}$$

in view of our selection (3.18) of m, which completes the proof.

The respective values of the target functional on f_1 and f_0 are $g_0(0)$ and $g_0(0) - \kappa\, m^{-\tau}\varphi(0)$, so that their distance is of order $m^{-\tau} = n^{-\tau/(\tau/D+1)}$. In view of lemma 3.4 this establishes (3.2).

4. PROOF OF THEOREM 2.1.

Observe that for $\eta = \alpha$ or β,

$$\int_{g(0)-u}^{\infty} \{g(x) - y\}_+{}^{\eta(x)}\, dy = \{\eta(x) + 1\}^{-1} \{g(x) - g(0) + u\}_+^{\eta(x)+1}. \tag{4.1}$$

If the function ζ is differentiable and ζ' satisfies a Lipschitz condition with exponent t in a neighbourhood of the origin, then

$$u^{\zeta(x)} = u^{\zeta(0)} \left\{1 + x\,\zeta'(0)\,\log u + O\big(x^2\,|\log u|^2 + |x|^{t+1}\,|\log u|\big)\right\}, \tag{4.2}$$

uniformly in pairs (x,u) such that $|x\,\log u|$ is bounded. Put $\zeta = \eta + 1$, let η satisfy the conditions imposed on α in the theorem, and let $\lambda = \lambda(h)$ denote a sequence of positive numbers diverging to infinity arbitrarily slowly. Since g' enjoys a Lipschitz condition with exponent t, we have uniformly in $u \in (\lambda\,h, 1)$ and $|x| \le h$,

$$\begin{aligned}
\{g(x) - g(0) + u\}^{\zeta(x)} &= \left\{u + x\,g'(0) + O\big(|x|^{t+1}\big)\right\}^{\zeta(x)} \\
&= u^{\zeta(x)}\left[1 + u^{-1}\,x\,\zeta(x)\,g'(0)\right. \\
&\qquad \left. + \tfrac{1}{2}\,u^{-2}\,x^2\zeta(x)\,\{\zeta(x) - 1\}\,g'(0)^2 + O\big(u^{-1}\,h^{t+1} + u^{-3}\,h^3\big)\right] \\
&= u^{\zeta(0)}\left[1 + u^{-1}\,x\,\zeta(0)\,g'(0) + x\,\zeta'(0)\,\log u\right. \\
&\qquad \left. + \tfrac{1}{2}\,u^{-2}\,x^2\,\zeta(0)\,\{\zeta(0) - 1\}\,g'(0)^2 + O\big(u^{-1}\,h^{t+1} + u^{-3}\,h^3\big)\right]. \qquad (4.3)
\end{aligned}$$

Therefore, combining (4.1)–(4.3),

$$\begin{aligned}
(2h)^{-1}\int_{-h}^{h} dx \int_{g(0)-u}^{\infty} a(x)\,\{g(x) - y\}_+{}^{\eta(x)}\,dy \\
= \zeta(0)^{-1}\,a(0)\,u^{\zeta(0)}\left[1 + \tfrac{1}{6}u^{-2}\,h^2\,\zeta(0)\,\{\zeta(0) - 1\}\,g'(0)^2\right. \\
\left. + O\big(u^{-1}\,h^{t+1} + u^{-3}\,h^3\big)\right]. \qquad (4.4)
\end{aligned}$$

Similarly, if η satisfies the conditions imposed on β in the theorem then

$$\begin{aligned}
(2h)^{-1}\int_{-h}^{h} dx \int_{g(0)-u}^{\infty} b(x)\,\{g(x) - y\}_+{}^{\eta(x)}\,dy \\
= \zeta(0)^{-1}\,b(0)\,u^{\zeta(0)}\left[1 + O\big\{(h/u)^{\delta}\big\}\right], \qquad (4.5)
\end{aligned}$$

where $\delta > 0$ depends on the exponents of Hölder continuity of b and β. Both (4.4) and (4.5) hold uniformly in $u \in (h^{1-\epsilon}, 1)$. Furthermore, $P(|X| \le h) = 2h\left\{e(0) + O\big(h^{t+1}\big)\right\}$. Combining this result with (4.4) and (4.5) we deduce that if U has the distribution of $g(0) - Y$ given that $|X| \le h$ then, uniformly in the same range of values of u,

$$\begin{aligned}
G(u) &\equiv P(U \le u) \\
&= \int_{-h}^{h} dx \int_{g(0)-u}^{\infty} f(x,y)\,dy \Big/ P(|X| \le h) \\
&= e(0)^{-1}\Big(\{\alpha(0)+1\}^{-1}\,a(0)\,u^{\alpha(0)+1} \\
&\qquad \times \big[1 + \tfrac{1}{6}\,u^{-2}\,h^2\,\alpha(0)\,\{\alpha(0)+1\}\,g'(0)^2\big] + \{\beta(0)+1\}^{-1}\,b(0)\,u^{\beta(0)+1}\Big) \\
&\qquad + O\big\{u^{\alpha(0)+1}\big(u^{-1}\,h^{t+1} + u^{-3}\,h^3\big) + u^{\beta(0)+1-\delta}\,h^\delta\big\} \\
&= a_1\,u^{\alpha(0)+1}\,\big\{1 + a_2\,u^{-2}\,h^2 + a_3\,u^{\gamma(0)} + O\big(u^{-1}\,h^{t+1} + u^{-3}\,h^3 + u^{\gamma(0)-\delta}\,h^\delta\big)\big\}\,,
\end{aligned}$$

where $\gamma = \beta - \alpha$, $a_1 = e(0)^{-1}\,\{\alpha(0)+1\}^{-1}\,a(0)$, $a_2 = \frac{1}{6}\,\alpha(0)\,\{\alpha(0)+1\}\,g'(0)^2$, $a_3 = b(0)\,\{\alpha(0)+1\}/[a(0)\,\{\beta(0)+1\}]$.

Inverting this expansion we deduce that

$$\begin{aligned}
G^{-1}(v) = b_1\,v^{1/\{\alpha(0)+1\}}\,\big\{1 &- b_2\,v^{-2/\{\alpha(0)+1\}}\,h^2 - b_3\,v^{\gamma(0)/\{\alpha(0)+1\}} \\
&+ O\big(v^{-1/\{\alpha(0)+1\}}\,h^{t+1} + v^{-3/\{\alpha(0)+1\}}\,h^3 \\
&+ v^{\gamma(0)/\{\alpha(0)+1\}-\delta}\,h^\delta\big)\big\}\,,
\end{aligned} \tag{4.6}$$

where

$$\begin{aligned}
b_1 &= [e(0)\,\{\alpha(0)+1\}/a(0)]^{1/\{\alpha(0)+1\}}\,, \\
b_2 &= \tfrac{1}{6}\,\alpha(0)\,[a(0)/e(0)\,\{\alpha(0)+1\}]^{2/\{\alpha(0)+1\}}\,g'(0)^2\,, \\
b_3 &= a(0)^{-\{\beta(0)+1\}/\{\alpha(0)+1\}}\,b(0)\,[\{\alpha(0)+1\}\,e(0)]^{\gamma(0)/\{\alpha(0)+1\}}\,\{\beta(0)+1\}^{-1}\,,
\end{aligned}$$

uniformly in $v \in (\lambda\,h^{\alpha(0)+1}, \frac{1}{2})$.

Since $g(0)$ is a location parameter, we may assume without loss of generality that $g(0) = 0$. In the work below we condition on the value of N, denoting the number of original data pairs (X_i, Y_i) in the interval of width $2h$ centred on the abscissa value $x = 0$. Let $U_1, U_2, \ldots, U_N$ be independent and identically distributed random variables with the distribution of U, and let $U_{(1)} \le U_{(2)} \le \ldots \le U_{(N)}$ denote the corresponding order statistics. In this notation, the sequence $\{\xi_i(\theta), 1 \le i \le N\}$

has the same distribution as $\{(U_{(r)}-U_{(i)})/(U_{(i)}-\theta),\ 1\le i\le N\}$. Without loss of generality, $\xi_i(\theta)=(U_{(r)}-U_{(i)})/(U_{(i)}-\theta)$.

Let $Z_1,\ldots,Z_N$ denote independent random variables with a common exponential distribution, and define

$$S_i=\sum_{j=1}^{i} Z_j/(N-j+1), \quad T_i=i^{-1}\sum_{j=1}^{i}(Z_j-1).$$

Noting Rényi's representation for order statistics we see that we may write

$$U_{(i)}=G^{-1}\{1-\exp(-S_i)\}, \quad 1\le i\le N\,. \tag{4.7}$$

For any real number w, $S_i=-\log(1-i\,N^{-1})+(i/N)\{T_i+O_p(i^{1/2}\,N^{-1})\}$ and

$$\{1-\exp(-S_i)\}^w=(i/N)^w\left\{1+w\,T_i+O_p\big(i^{-1}+i^{1/2}\,N^{-1}\big)\right\} \tag{4.8}$$

uniformly in $1\le i\le r$.

In the remainder of our proof we treat separately the cases $\alpha(0)>2$, $\alpha(0)=2$ and $1<\alpha(0)<2$. Recall that $A=\{\alpha(0)+1\}^{-1}$.

Case I: $\alpha(0)>2$. Given a positive sequence $\delta(n)\to 0$, let $i_1\ge 1$ denote the smallest positive integer such that $(nh/i_1)^A\,h\le\delta(n)$. The assumption $\epsilon(n)\equiv n\,h^{\alpha(0)+2}/r\to 0$, in that part of the theorem dealing with the case $\alpha(0)>2$, implies that

$$(N/r)^A\,h=O\big\{\epsilon(n)^A\big\}. \tag{4.9}$$

By (4.6)–(4.8) we have, uniformly in $i_1\le i\le r$,

$$\begin{aligned} b_1^{-1}\,U_{(i)}=(i/N)^A\,\big(1&+A\,T_i\\ &-\{1+o_p(1)\}\big\{b_2\,(N/i)^{2A}\,h^2+b_3\,(i/N)^{A\gamma(0)}\big\}\\ &+O_p\big[(N/i)^A\,h^{t+1}+i^{-1}+i^{1/2}\,N^{-1}\\ &+\big\{(N/i)^{2A}\,h^2+(i/N)^{A\gamma(0)}\big\}\,i^{-1/2}\big]\big). \end{aligned} \tag{4.10}$$

Given a random variable $\tilde{\theta}$ satisfying $N^A\,\tilde{\theta} \to 0$ in probability, define $\tilde{\theta}_i = (N/i)^A\,\tilde{\theta}$. Put

$$W_1 = r^{-1} \sum_{j=1}^{r} (Z_j - 1) \left\{ 1 - (1-A)\, r^A \sum_{i=j}^{r} i^{-(A+1)} \right\},$$

$$W_2 = r^{-1} \sum_{j=1}^{r} (Z_j - 1) \left(1 - \sum_{i=j}^{r} i^{-1} \right), \quad W_3 = (1-A)\, W_1 - W_2,$$

$$d_{11} = (1-2A)^{-1}\, b_1^{-1}, \quad d_{12} = 2\,(1-3A)^{-1}\, b_2,$$

$$d_{13} = -\gamma(0)\,[1 + A(0)\,\{\gamma(0) - 1\}]^{-1}\, b_3,$$

$$d_{21} = (1-A)^{-1}\, b_1^{-1}, \quad d_{22} = 2(1-2A)^{-1}\, b_2,$$

$$d_{23} = -\gamma(0)\,\{A\,\gamma(0) + 1\}^{-1}\, b_3, \quad d_{31} = A^2\,\{(1-A)\,(1-2A)\}^{-1}\, b_1^{-1}\,,$$

$$d_{32} = 4A^2\,(1-2A)^{-1}\,(1-3A)^{-1}\, b_2,$$

$$d_{33} = \gamma(0)^2\, A^2 \left(\{1 + A\,\gamma(0)\}\,[1 + A\,\{\gamma(0) - 1\}] \right)^{-1} b_3.$$

(Note that, since $\alpha(0) > 2$, $3A < 1$. Also, $d_{3i} = (1-A)\, d_{1i} - d_{2i}$.) In this notation we may prove successively from (4.10) that the following results hold, the first two uniformly in $i_1 \le i \le r$:

$$\begin{aligned}
1 + \xi_i(\tilde{\theta}) &= (U_{(r)} - \tilde{\theta})/(U_{(i)} - \tilde{\theta}) \\
&= (r/i)^A \big[1 + A\,(T_r - T_i) + b_1^{-1} \left\{ (N/i)^A - (N/r)^A \right\} \tilde{\theta} \\
&\quad + b_2 \left\{ (N/i)^{2A} - (N/r)^{2A} \right\} h^2 + b_3 \left\{ (i/N)^{A\gamma(0)} - (r/N)^{A\gamma(0)} \right\} \\
&\quad + O_p\big[(N/i)^A\, h^{t+1} + i^{-1} + i^{1/2}\, N^{-1} \\
&\quad + \left\{ (N/i)^{2A}\, h^2 + (i/N)^{A\gamma(0)} \right\} i^{-1/2} \big] \\
&\quad + o_p\left\{ i^{-1/2} + |\tilde{\theta}_i| + (N/i)^{2A}\, h^2 + (r/N)^{A\gamma(0)} \right\} \big]\,,
\end{aligned} \tag{4.11}$$

$$\begin{aligned}
&\log\{1 + \xi_i(\tilde{\theta})\} \\
&\quad = A\,(\log r - \log i) + A\,(T_r - T_i) + b_1^{-1} \left\{ (N/i)^A - (N/r)^A \right\} \tilde{\theta} \\
&\quad\quad + b_2 \left\{ (N/i)^{2A} - (N/r)^{2A} \right\} h^2 + b_3 \left\{ (i/N)^{A\gamma(0)} - (r/N)^{A\gamma(0)} \right\} \\
&\quad\quad + O_p\big[(N/i)^A\, h^{t+1} + i^{-1} + i^{1/2}\, N^{-1} \\
&\quad\quad + \left\{ (N/i)^{2A}\, h^2 + (i/N)^{A\gamma(0)} \right\} i^{-1/2} \big] \\
&\quad\quad + o_p\left\{ i^{-1/2} + |\tilde{\theta}_i| + (N/i)^{2A}\, h^2 + (r/N)^{A\gamma(0)} \right\} \big]\,,
\end{aligned} \tag{4.12}$$

$$
\begin{aligned}
r^{-1}\,(1-A)\,A^{-1}\sum_{i=1}^{r-1}\xi_i(\tilde\theta)
&= 1 + W_1 + d_{11}\,(N/r)^A\,\tilde\theta + d_{12}\,(N/r)^{2A}\,h^2 + d_{13}\,(r/N)^{A\,\gamma(0)}\\
&\quad + O_p\big\{(N/r)^A\,h^{t+1} + r^{1/2}\,N^{-1} + (i_1/r)^{1-A}\big\}\\
&\quad + o_p\big\{r^{-1/2} + (N/r)^A\,|\tilde\theta| + (N/r)^{2A}\,h^2 + (r/N)^{A\gamma(0)}\big\}\,, \qquad (4.13)
\end{aligned}
$$

$$
\begin{aligned}
r^{-1}\,A^{-1}\sum_{i=1}^{r-1}\log\{1+\xi_i(\tilde\theta)\}
&= 1 + W_2 + d_{21}\,(N/r)^A\,\tilde\theta + d_{22}\,(N/r)^{2A}\,h^2 + d_{23}\,(r/N)^{A\,\gamma(0)}\\
&\quad + O_p\big\{(N/r)^A\,h^{t+1} + r^{1/2}\,N^{-1} + r^{-1}\,\log r\big\}\\
&\quad + o_p\big\{r^{-1/2} + (N/r)^A\,|\tilde\theta| + (N/r)^{2A}\,h^2 + (r/N)^{A\gamma(0)}\big\}\,. \qquad (4.14)
\end{aligned}
$$

[The terms of orders $(i_1/r)^{1-A}$ and $r^{-1}\,\log r$ on the right-hand sides of (4.13) and (4.15), respectively, derive from extending the sums on the left-hand sides from $i_1 \le i \le r$ (which is their natural range, given the values of i for which (4.11) and (4.12) have been established) to $1 \le i \le r$. For example, in the case of (4.13) observe that $|\xi_i(\tilde\theta)| = O_p\{(r/i)^A\}$ uniformly in $1 \le i \le i_1$. Hence, the contribution to the left-hand side of (4.13) from such i's is of the same order as the sum of $r^{-1}\,(r/i)^A$ over those i's. That is, it is of order $(i_1/r)^{1-A}$.]

Therefore,

$$
\begin{aligned}
A\,r\left(\left[\sum_{i=1}^{r-1}\log\{1+\xi_i(\theta)\}\right]^{-1} - \left\{\sum_{i=1}^{r-1}\xi_i(\theta)\right\}^{-1} - r^{-1}\right)
&= W_3 + d_{31}\,(N/r)^A\,\tilde\theta + d_{32}\,(N/r)^{2A}\,h^2 + d_{33}\,(r/N)^{A\,\gamma(0)}\\
&\quad + O_p\big\{(N/r)^A\,h^{t+1} + r^{1/2}\,N^{-1} + (i_1/r)^{1-A}\big\}\\
&\quad + o_p\big\{r^{-1/2} + (N/r)^{2A}\,h^2 + (r/N)^{A\gamma(0)} + (N/r)^A\,|\tilde\theta|\big\}\,. \qquad (4.15)
\end{aligned}
$$

It follows from (4.15) that if $\tilde\theta$ is a solution of equation (2.4) then

$$
\begin{aligned}
-\tilde\theta &= d_{31}^{-1}\,(r/N)^A\,W_3 + d_{31}^{-1}\,d_{32}\,(N/r)^A\,h^2\\
&\quad + d_{31}^{-1}\,d_{33}\,(r/N)^{A\{\gamma(0)+1\}}\\
&\quad + O_p\big[h^{t+1} + (r/N)^A\,\big\{r^{1/2}\,N^{-1} + (i_1/r)^{1-A}\big\}\big]\\
&\quad + o_p\big\{(r/N)^A\,r^{-1/2} + |\tilde\theta| + (N/r)^A\,h^2 + (r/N)^{A\{\gamma(0)+1\}}\big\}\,. \qquad (4.16)
\end{aligned}
$$

Next we show that the term $\tau \equiv O_p[(r/N)^A \{r^{1/2} N^{-1} + (i_1/r)^{1-A}\}]$, on the right-hand side of (4.16), may be dropped. Since $r/N \to 0$ then $(r/N)^A r^{1/2} N^{-1} = o\{(r/N)^A r^{-1/2}\}$, and this term is addressed by the $o_p(\ldots)$ contribution to the right-hand side of (4.16). By definition of i_1, $(i_1/N)^A = O\{h\,\delta(n)^{-1}\}$, and so

$$(i_1/r)^{1-A} = O\big\{(N/r)^A\, h\,\delta(n)^{-1}\big\}^{(1-A)/A}. \tag{4.17}$$

In view of (4.9) we may choose $\delta(n)$ to converge to zero so slowly that the right-hand side of (4.17) equals $o\{(N/r)^{2A} h^2\}$, which is again subsumed into the $o_p(\ldots)$ contribution to the right-hand side of (4.16).

Standard methods may be used to prove that W_3 is asymptotically Normally distributed with zero mean and variance $A^2\{r(1-2A)\}^{-1}$. Therefore, defining $\sigma = d_{31}^{-1} A (1-2A)^{-1/2}$, $c_1 = -d_{32}/d_{31}$ and $c_2 = -d_{33}/d_{31}$, we see that from (4.16) (dropping the term corresponding to τ) that

$$\begin{aligned}\tilde\theta &= (r/N)^A r^{-1/2} \sigma W_4 + (N/r)^A h^2 c_1 + (r/N)^{A\{\gamma(0)+1\}} c_2 \\ &\quad + O_p\big(h^{t+1}\big) + o_p\big\{(r/N)^A r^{-1/2} + (N/r)^A h^2 + (r/N)^{A\{\gamma(0)+1\}}\big\},\end{aligned} \tag{4.18}$$

where W_4 is asymptotically Normal N(0,1). This is equivalent to the claimed expansion in Theorem 2.1. Arguing as in Hall (1982, pp. 566–567) the expansions above may be retraced to show that with probability tending to 1, a solution to (2.4) exists; and that the largest solution $\tilde\theta$ of (2.4) satisfies $N^A \tilde\theta \to 0$ in probability. These remarks also apply to the next two cases.

Case II: $\alpha(0) = 2$. Let i_1 be as in Case I, and as before, let $\tilde\theta$ denote a random variable equal to $o_p(N^{-A})$. Once again, (4.11) and (4.12) hold uniformly in $i_1 \le i \le r$, and (4.14) is true. In place of (4.13),

$$\begin{aligned} & r^{-1}(1-A) A^{-1} \sum_{i=1}^{r-1} \xi_i(\tilde\theta) \\ &\quad = 1 + W_1 + d_{11} (N/r)^A \tilde\theta + (1-A) A^{-1} b_2 (N/r)^{2A} h^2 \log r \\ &\qquad + d_{13} (r/N)^{A\gamma(0)} + O_p\big\{(N/r)^A h^{t+1} + r^{1/2} N^{-1} + (i_1/r)^{1-A}\big\} \\ &\qquad + o_p\big\{r^{-1/2} + (N/r)^A |\tilde\theta| + (N/r)^{2A} h^2 \log n + (r/N)^{A\gamma(0)}\big\}.\end{aligned}$$

Therefore, (4.15) holds as before but with the term $d_{32}\,(N/r)^{2A}\,h^2$ replaced by $A^{-1}\,(1-A)^2\,b_2\,(N/r)^{2A}\,h^2\,\log r$. The analogous change should be made to the right-hand side of (4.16), giving:

$$\begin{aligned}
-\tilde{\theta} = d_{31}^{-1}\,(r/N)^A\,W_3 + d_{31}^{-1}\,A^{-1}\,(1-A)^2\,b_2\,(N/r)^A\,h^2\,\log r \\
+ d_{31}^{-1}\,d_{33}\,(r/N)^{A\{\gamma(0)+1\}} \\
+ O_p\big[h^{t+1} + (r/N)^A\,\big\{r^{1/2}\,N^{-1} + (i_1/r)^{1-A}\big\}\big] \\
+ o_p\big\{(r/N)^A\,r^{-1/2} + |\tilde{\theta}| + (N/r)^A\,h^2\,\log n + (r/N)^{A\{\gamma(0)+1\}}\big\}\,.
\end{aligned}$$

In view of (4.17), and provided that $\delta(n)$ converges to zero so slowly that

$$\delta(n)\,(\log n)^{1/2} \to \infty,$$

the term $O_p[(r/N)^A\,\{r^{1/2}\,N^{-1} + (i_1/r)^{1-A}\}]$ on the right-hand side may be subsumed into the $o_p(\ldots)$ term. Therefore, in place of (4.18),

$$\begin{aligned}
\tilde{\theta} = (r/N)^A\,r^{-1/2}\,\sigma\,W_4 + (N/r)^A\,h^2\,\log r\,c_3 + (r/N)^{A\{\gamma(0)+1\}}\,c_2 \\
+ O_p\big(h^{t+1}\big) + o_p\big\{(r/N)^A\,r^{-1/2} + (N/r)^A\,h^2\,\log r + (r/N)^{A\{\gamma(0)+1\}}\big\}\,,
\end{aligned}$$

where $c_3 \equiv -A^{-1}\,(1-A)^2\,b_2/d_{31}$. This is equivalent to the claimed expansion in Theorem 2.1.

Case III: $1 < \alpha(0) < 2$. Here it is necessary to develop a refined version of formula (4.11). Our starting point is a more concise form of (4.8) in the special case $w = 1$, which follows via the discussion immediately preceding that result:

$$1 - \exp(-S_i) = (i/N)\,(1+T_i)\,\big\{1 + O_p\big(i^{1/2}\,N^{-1}\big)\big\}\,,$$

Hence,

$$\{1 - \exp(-S_i)\}^w = (i/N)^w\,\big\{1 + T_i^{(w)} + O_p\big(i^{1/2}\,N^{-1}\big)\big\}\,, \tag{4.19}$$

where $T_i^{(w)} \equiv (1+T_i)^w - 1 = w\,T_i + O_p(i^{-1})$. Using (4.19) in place of (4.8) we obtain, instead of (4.10), and uniformly in $i_1 \le i \le r$,

$$\begin{aligned}
b_1^{-1}\,U_{(i)} = (i/N)^A\,\big[1 + A\,T_i - \{1 + o_p(1)\}\,b_3\,(i/N)^{A\gamma(0)} \\
- \big\{\big(1 + T_i^{(A)}\big)\,\big(1 + T_i^{(-2A)}\big) + o_p(1)\big\}\,b_2(N/i)^{2A}\,h^2 \\
+ O_p\big\{(N/i)^A\,h^{t+1} + i^{-1} + i^{1/2}\,N^{-1} + (i/N)^{A\gamma(0)}\,i^{-1/2}\big\}\big];
\end{aligned}$$

and in place of (4.11) and (4.13),

$$
\begin{aligned}
1+\xi_i(\tilde\theta) &= (U_{(r)}-\tilde\theta)/(U_{(i)}-\tilde\theta)\\
&= (r/i)^A \big[1 + A\,(T_r - T_i) + b_1^{-1}\,\big\{(N/i)^A - (N/r)^A\big\}\,\tilde\theta\\
&\quad + b_2\,\big(1+T_i^{(A)}\big)\,\big(1+T_i^{(-2A)}\big)\,(N/i)^{2A}\,h^2\\
&\quad + b_3\,\big\{(i/N)^{A\gamma(0)} - (r/N)^{A\gamma(0)}\big\} + O_p\big\{(N/r)^{2A}\,h^2\\
&\quad + (N/i)^A\,h^{t+1} + i^{-1} + i^{1/2}\,N^{-1} + (i/N)^{A\gamma(0)}\,i^{-1/2}\big\}\\
&\quad + o_p\big\{i^{-1/2} + |\tilde\theta_i| + (N/i)^{2A}\,h^2 + (r/N)^{A\gamma(0)}\big\}\big]\,,\\
r^{-1}\,&(1-A)\,A^{-1}\sum_{i=1}^{r-1}\xi_i(\tilde\theta)\\
&= 1 + W_1 + d_{11}\,(N/r)^A\,\tilde\theta + d_{13}\,(r/N)^{A\gamma(0)}\\
&\quad + b_2\,A^{-1}\,(1-A)\,r^{A-1}\,N^{2A}\,h^2\sum_{i=1}^{r-1}\big(1+T_i^{(A)}\big)\,\big(1+T_i^{(-2A)}\big)\,i^{-3A}\\
&\quad + O_p\big\{(N/r)^{2A}\,h^2 + (N/r)^A\,h^{t+1} + r^{A-1} + r^{1/2}\,N^{-1}\big\}\\
&\quad + o_p\big\{r^{-1/2} + (N/r)^A\,|\tilde\theta| + r^{A-1}\,N^{2A}\,h^2 + (r/N)^{A\gamma(0)}\big\}\,.
\end{aligned}
$$

In view of the assumption $n\,h^{\alpha(0)+2}\to\infty$, made in that part of the theorem addressing the case $1<\alpha(0)<2$, the term $O_p(r^{A-1})$ is of smaller order than $r^{A-1}\,N^{2A}\,h^2$, and so may be incorporated into the remainder $o_p(r^{A-1}\,N^{2A}\,h^2)$. Similarly, the $O_p(r^{1/2}\,N^{-1})$ term is subsumed by the remainder $o_p(r^{-1/2})$. Results (4.12) and (4.14) hold as before. Therefore, instead of (4.18),

$$
\begin{aligned}
\tilde\theta &= (r/N)^A\,r^{-1/2}\,\sigma\,W_4 + r^{2A-1}\,N^A\,h^2\,W_5 + (r/N)^{A\{\gamma(0)+1\}}\,c_2\\
&\quad + O_p\big(h^{t+1}\big) + o_p\big\{(r/N)^A\,r^{-1/2} + r^{2A-1}\,N^A\,h^2\\
&\quad + (r/N)^{A\{\gamma(0)+1\}}\big\}\,,
\end{aligned}
\tag{4.20}
$$

where

$$
W_5 \equiv c_3\sum_{i-1}^{\infty}\big(1+T_i^{(A)}\big)\,\big(1+T_i^{(-2A)}\big)\,i^{-3A},
$$

and c_3 is defined as in the previous case. Result (4.20) is equivalent to the claimed expansion in Theorem 2.1.

REFERENCES

AKAHIRA, M. AND TAKEUCHI, K. (1979). Discretized likelihood methods — asymptotic properties of discretized likelihood estimators (DLE's). *Ann. Inst. Statist. Math.* **31**, 39–56.

COOKE, P.J. (1979). Statistical inference for bounds of random variables. *Biometrika* **66**, 367–374.

COOKE, P.J. (1980). Optimal linear estimation of bounds of random variables. *Biometrika* **67**, 257–258.

CSÖRGO, S. AND MASON, D. (1989). Simple estimators of the endpoint of a distribution. In: *Extreme Value Theory* (Proceedings Oberwolfach 1987), J. Hüsler, R.-D. Reiss, Eds., Lecture Notes in Statistics, 51, Springer-Verlag, New York.

CSÖRGO, S., DEHEUVELS, P. AND MASON, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13**, 1050–1077.

DEVROYE, L. AND WISE, G.L. (1980). Detection of abnormal behaviour via nonparametric estimation of the support. *SIAM J. Appl. Math.* **38**, 80–488.

DONOHO, D.L. AND LIU, R.C. (1991). Geometrizing rates of convergence, II. *Ann. Statist.* **19**, 633–667.

FALK, M., HÜSLER, J. AND REISS, R.-D. (1994). *Laws of Small Numbers: Extremes and Rare Events.* DMV Seminar 23, Birkhäuser-Verlag, Basel.

HÄRDLE, W., PARK, B.U. AND TSYBAKOV, A.B. (1993). Estimation of non-sharp support boundaries. Discussion paper 9317, Institute for Statistics and Econometrics, Humboldt University Berlin, submitted to *J. Multivar. Anal.*

HALL, P. (1982a). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44**, 137–142.

HALL, P. (1982b). On estimating the endpoint of a distribution. *Ann. Statist.* **10**, 556–568.

HALL, P. AND WELSH, A.H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12**, 1079–1084.

HILL, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.

JANSSEN, A. AND MAROHN, F. (1994). On statistical information of extreme order statistics, local extreme value alternatives, and Poisson point processes. *J. Multivar. Anal.* **48**, 1–41.

KOROSTELEV, A.P. AND TSYBAKOV, A.B. (1992). Efficient support of a probability density and estimation of support functionals. Discussion paper 9229, Institut de Statistique, Louvain-la-Neuve.

KOROSTELEV, A.P., SIMAR, L. AND TSYBAKOV, A.B. (1992). Efficient estimation of monotone boundaries. Discussion paper 9206, Institut de Statistique, Louvain-la-Neuve.

KOROSTELEV, A.P. AND TSYBAKOV, A.B. (1993a). Estimation of support of a probability density and of its functionals. *Problems of Information Transmission* **29**, 3–18.

KOROSTELEV, A.P. AND TSYBAKOV, A.B. (1993b). *Minimax Theory of Image Reconstruction.* Lecture Notes in Statistics 82.

MAMMEN, E. AND TSYBAKOV, A.B. (1992). Asymptotical minimax results in image analysis for sets with smooth boundaries. To appear, *Ann. Statist.*

MAROHN, F. (1991). Global sufficiency of extreme order statistics in location models of Weibull type. *Probab. Theory Related Fields* **88**, 261–268.

POLFEDT, T. (1970). The order of the minimum variance in a non-regular case. *Ann. Math. Statist.* **41**, 667–672.

SMITH, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174–1207.

TSYBAKOV, A.B. (1994). Nonparametric estimation of density level sets. Rapport Technique, 94/01, Statistique, Université Paris 6.

WOODROOFE, M. (1972). Maximum likelihood estimation of a translation parameter of a truncated distribution II. *Ann. Statist.* **3**, 474–488.