L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

Measuring the performance of boundary-estimation methods

Peter Hall

Australian National University, Canberra, Australia and CSIRO Mathematical and Information Sciences, Sydney, Australia

Marc Raimondo

Australian National University, Canberra, Australia

Abstract: The problem of local linear approximation to a curved boundary using gridded data is closely connected to both curve estimation methods in statistics and rational approximation in number theory. The problem is ill-posed, in the sense that orders of approximation at arbitrarily close points can be very different. This may be interpreted as a consequence of the problem's number-theoretic aspects, since irrational numbers with arbitrarily slowly convergent rational approximations are distributed in dense sets. On the other hand, by measuring performance in a "statistically average" way which excludes most of the pathologies, we may deduce useful results about optimal orders of approximation. In this respect, among others, statistical approaches to the problem are important. For example, measures of performance based on the L^1 norm are more appropriate than those founded on L^p norms for $p > 1$. The paper will describe these viewpoints, and outline the way in which they may be combined to produce a cohesive theory of curve estimation from gridded data. We shall start with the relatively simple case of approximation to a simple linear boundary, where data are observed without noise, and progress through an analysis of the number-theoretic connections, concluding with results in the context of stochastic or curved boundaries observed with noise.

Key words: Curve estimation, edge, gradient, grid, integral metric, irrational number, nonparametric, rational number, slope, vertex.

AMS subject classification: Primary 60G35; secondary 62G20.

1 Defining a sraight-line boundary

Imagine placing a straight line across a square lattice in the plane, thereby dividing the plane into two parts. Assuming that the line is not vertical, colour black those lattice vertices above the line and white the vertices below, with a third colour (red, say) for any vertices that lie on the line. Now remove the line, and attempt to reconstruct it from the pattern of vertex colours. This is a theoretical idealisation of a range of practical boundary estimation problems, where a curve representing the boundary between two areas of different colour is to be estimated from pixel data.

Even a brief consideration of this problem shows that its solution depends critically on the nature of the slope of the line. For example, if the slope is rational and if the line passes through some vertex, then the line necessarily passes through an infinite number of vertices. In this case, if we were able to observe the vertex colour pattern in a large enough region of the plane, we would see that there are at least two red vertices there, and from them we could trivially deduce the equation of the line. Then, we would know the line exactly.

On the other hand, if the line has rational slope but does not pass through any vertex, it cannot be determined exactly even if we know the colour of every vertex in the plane. This is perhaps most easily seen if the line, $\mathcal{L}$ say, is parallel to one of the axes of the square lattice. In that case there exists an infinite strip in the plane, with its sides parallel to the line and its width equal to the edge width of the lattice, such that any straight line contained wholly within the strip produces exactly the same vertex colour pattern as $\mathcal{L}$.

A similar situation arises for any line with rational gradient, where the intercept is chosen so that the line does not pass through any vertex. In such cases, while the gradient may be determined exactly from vertex colour data within the whole plane, the intercept will remain unknown beyond the fact that it lies within a certain nondegenerate interval — except when the line passes through a vertex. So, in the case of a line with rational slope we know either everything or nothing: either we can compute the line exactly from a finite amount of vertex colour data (when the line passes through a vertex) or we cannot compute it exactly even from an infinite amount of data (if it does not pass through any vertex).

The situation is quite different if the line has irrational slope. There, if the colour pattern is observed within an increasingly large region $\mathcal{R}$, say an $n \times n$ section of the lattice centred roughly on the line, then an approximation to the line may be constructed using only the colour pattern within $\mathcal{R}$. As $\mathcal{R}$ expands, the accuracy with which the line may be approximated increases. More explicitly, we may compute an approximation $\widehat{\mathcal{L}} = \widehat{\mathcal{L}}(\mathcal{P})$ to $\mathcal{L}$, using only the vertex colour pattern $\mathcal{P}$ within $\mathcal{R}$, such that the Hausdorff distance between $\widehat{\mathcal{L}} \cap \mathcal{R}$ and $\mathcal{L} \cap \mathcal{R}$ converges to zero as $\mathcal{R}$ increases.

In the case of irrational slope the rate of convergence of a good approximation $\widehat{\mathcal{L}}$ depends intimately on the nature of the irrational slope. It depends hardly at all on whether $\mathcal{L}$ intersects a lattice vertex; this influences only the constant multiple of the optimal rate of convergence of $\widehat{\mathcal{L}}$ to $\mathcal{L}$, not the rate itself. Thus, the problem of approximating straight-line boundaries is starkly ill-posed, since nearby slopes can produce very different convergence rates along infinite subsequences.

In Section 2 we shall treat examples of classes of irrational numbers, which capture the spirit of the boundary approximation problem and its solution. Section 3 will employ the examples to motivate development of more general boundary approximation problems, and will discuss ways in which the problems might be tackled. Section 4 will briefly survey the number-theoretic background to the methods. Later sections will develop theories for curved boundary estimation using local linear methods, borrowing ideas that are now well understood in more traditional statistical settings. For the latter, the reader is referred to Wand and Jones (1995, Chapter 5) and Fan and Gijbels (1996).

In Sections 1–5 we shall always assume that the lattice is fixed; without loss of generality it has its vertices at integer pairs (i, j) in the Cartesian plane, so that its edge width (the width of the side of the smallest square face of the lattice) is 1. In later sections we shall sometimes consider lattices of increasing fineness, so as to model the physical problem of approximating a curved boundary on a fine pixel grid. Technical details behind our arguments may be found in Hall and Raimondo (1996a,b).

While we shall concentrate on the case of a square lattice, for definiteness, the results that we shall describe are valid for any regular lattice that has the property that it contains a square lattice and is contained within the union of a finite number of square lattices. Thus, our results are available for lattices whose faces are hexagons or triangles. Lattices of the latter type are used in practice in J.P. Serra's image analyser. When considering an "$n \times n$" portion of a general lattice we interpret n as the square root of the number of vertices within a finite, square subset of the lattice.

2 Classes of irrational numbers

The irrational numbers with which many of us are most familiar are the so-called "quadratic irrationals", defined as the set of real numbers that may be expressed as solutions of quadratic equations with rational coefficients (or, without loss of generality, integer coefficients). These are a subset of the class of so-called *periodic irrationals*, and also of the larger class of *badly approximable* irrationals, which we shall discuss in Section 5. Straight-line boundaries with slope coming from one of these classes have special properties with respect to the boundary approximation problem. Indeed, in such cases the optimal rate of convergence (in the sense of the Hausdorff metric) of approximations based on vertex colours within an $n \times n$ subset of the lattice, is asymptotic to a constant multiple of n^{-1}.

The set of *algebraic irrationals* is larger than the class of quadratic irrationals, and is defined as the set of all real numbers that may be expressed as solutions to polynomial equations with rational coefficients. However, the most accurate available estimate of the rate of convergence in the linear boundary approximation problem for boundaries with slope equal to an algebraic irrational, is only the upper bound of $O(n^{-1+\epsilon})$ for all $\epsilon > 0$. Not even logarithmic refinements, such as $O(n^{-1} \log n)$, are available. The upper bound $O(n^{-1+\epsilon})$ is a corollary of deep number-theoretic work of Roth (1955), who determined the exact exponent in the Thue-Siegel inequality and for which work he was awarded the Fields medal in 1958.

Roth's result is virtually equivalent to the upper bound $O(n^{-1+\epsilon})$, for all $\epsilon > 0$, in our approximation problem. If we could improve on that rate then we could refine Roth's Theorem, as it is known. And of course, even if we could refine Roth's result, we would still have only scratched the surface as far as solving our problem goes, since the great majority of irrational numbers are not algebraic. Indeed, since the number of rationals is countable then the number of polynomial equations of degree p with rational coefficients is countable. Therefore, the number of solutions of such equations, for arbitrary p, is countable. Hence, the number of algebraic irrationals is countable, whereas the number of irrational numbers is uncountably infinite. By focusing only on algebraic irrationals we would be missing the great majority of irrational numbers.

It might be thought that because the algebraic irrationals are dense in the real line, they provide a good guide to the sort of behaviour that will be experienced when the slope of the line is a non-algebraic irrational. While this is true from some viewpoints, the argument has limitations. Indeed, irrationals that are not algebraic, and produce particularly pathological convergence rates in our line approximation problem, also comprise a dense subset of the real line. To elucidate this point we mention that if $\alpha_1, \alpha_2, \ldots$ and $\beta_1, \beta_2, \ldots$ are any two sequences of positive numbers converging to zero, then there exists a dense set of irrational numbers such that, whenever the slope of the linear boundary is in this set, the optimal rate of convergence in our approximation problem on an $n \times n$ grid is bounded above by α_n along one subsequence of values of n, and bounded below by β_n along another subsequence.

This is perhaps not a major issue if we confine attention to exactly linear boundaries — we may simply exclude such pathological irrational numbers from contention as possible gradients. However, in the problem of local linear estimation of a curved boundary the range of values of the gradient is an interval, and so includes representatives from any set which is dense in the real line. This fact, and the properties of irrational numbers noted above, make it clear that one must take care when defining boundary-estimation problems, to avoid becoming side-tracked by relatively unimportant cases.

3 Defining boundary-estimation problems

We need to pose boundary-estimation problems in such a way that we can deduce relatively simple principles behind rate-of-convergence properties. For that, we need some way of averaging over all possible choices of irrational gradients, so that the central issues in the problem will not be lost in through consideration of pathological special cases. There are at least two ways of doing this.

First, we might allow the slope of the boundary to be a random variable, and devote our discussion to its "average" properties. This is feasible for either straight or curved boundaries. If the boundary is linear then we may apply a random rotation to it, and more generally we may regard the boundary as a realization of a random curve whose equation is represented by $y = G(x)$, where G is a random, smooth function. Alternatively, we may choose to treat the boundary as fixed and curved, but estimate it at a randomly chosen point. Under such models we do not need to be too prescriptive about the type of averaging, since the more radical of the pathological cases described in the previous section arise only for sets of irrational numbers having measure zero. Therefore, if the random boundary, or the random point at which we estimate a fixed, curved boundary, is distributed in the continuum, then, by confining attention to almost sure properties we avoid all but reasonably regular cases. We shall outline this approach in Section 6.

Alternatively, in the case of a curved boundary we may average approximations in some way, for example by considering them in an integral metric. It turns out that the L^1 metric is more appropriate for this purpose than an L^p metric for $p > 1$, since it is relatively resistant to large deviations in the approximation error. (In view of the properties described in Section 2, it comes as no surprise to learn that the approximation error can change dramatically as we move from one point to another along a curved boundary.)

While the integral metric approach is attractive, not least because it is well established in the context of nonparametric curve estimation, it does require care. For example, if the boundary is linear then we are still faced with the ill-posed problem of the effect of rational-versus-irrational slope. The remedy is to avoid linear boundaries altogether. Now, one way of characterising a nonlinear boundary is to insist that its second derivative never vanish. Therefore, in Section 7 we shall study the L^1 performance of local linear approximations to twice-differentiable boundaries that do not have any points of inflexion.

4 Rational approximation by continued-fraction expansion

In order to appreciate the methods and results for general boundary-estimation problems it is necessary to understand the main elements of the theory of rational approximation by continued fractions. We shall survey them here, referring the reader to Leveque (1956, Chapter 9) and Khintchine (1963) for more detailed discussion. Section 5 will make the connections to boundary estimation explicit.

A non-integer real number u may be uniquely expressed as a continued fraction,

$$u = [a_0; a_1, a_2, \ldots] = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ldots}}},$$

where a_0 is an integer and $a_1, a_2, \ldots$ are strictly positive integers, called the *partial denominators* of u. The continued fraction expansion terminates if and only if u is rational. Up to the termination point (in the case of rational u), or for all n (if u is irrational), the *convergents* of u are defined to be the numbers

$$\frac{p_0}{q_0} = a_0, \quad \frac{p_1}{q_1} = a_0 + \frac{1}{a_1}, \quad \frac{p_2}{q_2} = a_0 + \frac{1}{a_1 + \frac{1}{a_2}}, \quad \ldots,$$

where p_n and q_n are relatively prime integers. The q_n's are strictly positive and form a strictly increasing sequence. By definition, p_n/q_n converges to u as $n \to \infty$. The sequence of odd-indexed convergents is decreasing, and the sequence of even-indexed convergents increases.

If u is irrational then the convergents provide a sequence of rational approximations to u, often referred to as "continued fraction approximations". The approximations are optimal in the sense that

$$\inf_{p,\, 1 \le q \le q_n} |u - (p/q)| = |u - (p_n/q_n)|. \tag{1}$$

They also satisfy

$$\{q_n(q_n + q_{n+1})\}^{-1} < |u - (p_n/q_n)| < (q_n q_{n+1})^{-1} \tag{2}$$

and

$$\text{if } p \text{ and } q \text{ are relatively prime, and } |u - (p/q)| < (2q^2)^{-1}, \text{ then } p/q \text{ is a convergent of } u \tag{3}$$

The quality of approximations by continued fractions is determined mostly by properties of large elements of the sequence $\{a_n\}$, or equivalently by large values of q_{n+1}/q_n, since it may be shown that $q_{n+1}/q_n \asymp a_n$ (meaning that the ratio of the left- and right-hand sides is bounded away from zero and infinity).

5 Relationship between convergents and rates of approximation to linear boundaries

The importance of continued fraction expansions to the problem of approximating linear boundaries, as defined in Section 1, is that the optimal rate of approximation (in the Hausdorff metric) to a line $\mathcal{L}$ with irrational gradient u, using vertex colour data observed on an $n \times n$ grid, is essentially equivalent to n times the optimal rate at which we can approximate u by a rational number p/q with q not exceeding n. In view of properties (1)–(3) the latter rate is the order of

$$\{q_{k(n)}(u)\, q_{k(n+1)}(u)\}^{-1},$$

where $k(n) = k(n, u)$ denotes the smallest k such that $q_k(u) \le n$. Call these results (R). A relationship between rational approximations and lines on square lattices is also expressed by Klein diagrams; see for example Klein (1907).

To illustrate the importance of the connection between convergents and boundary approximation we consider a simple example. The real number u is said to be *badly approximable* (or BA, for short) if $\sup_n a_n(u) < \infty$. The set of all BA numbers in the interval $[0, 1]$ has cardinality equal to that of the continuum (see e.g. Schmidt 1980, p. 23), but is of measure zero (e.g. Khintchine 1963, p. 69). All quadratic irrationals are BA, since for them the sequence $\{a_n\}$ is eventually periodic. However, not all algebraic irrationals are BA. In view of the asymptotic equivalence of the sequences a_n and q_{n+1}/q_n, and the fact that q_n is increasing, u is BA if and only if $(q_{k(n)}q_{k(n+1)})^{-1}$ is bounded between two constant multiples of n^{-2}.

From this result and (R) we see that an irrational number u is BA if and only if the optimal rate (in the Hausdorff metric) at which a line with gradient u may be approximated from vertex colour data in an $n \times n$ section of the lattice, is n^{-1}. As a corollary, the optimal rate is n^{-1} when the gradient of $\mathcal{L}$ is a quadratic irrational.

6 A stochastic number-theoretic view

Khintchine (1963), describing and developing work dating from the 1930's (see e.g. Khintchine, 1935; Lévy, 1937), gave a concise account of rates of rational approximation to irrational numbers when the latter are chosen randomly with respect to Lebesgue measure. In view of the equivalence between problems of rational approximation and boundary approximation noted in Section 5, we may apply Khintchine's results to our line estimation problem.

To pose that problem in a stochastic setting we assume that the linear boundary is placed into the plane according to a random mechanism. For our purposes the mechanism may be defined very generally; we need only ask that the distribution of slope, conditional on the line's intercept with any given axis, be continuous. This reflects the fact that, when the line has irrational gradient — which it will enjoy with probability 1 if the gradient has a continuous distribution — it is immaterial from the viewpoint of rates of approximation whether the line passes through a vertex.

It is known, for example from Theorem 30 of Khintchine (1963), that if $\psi(n) = n^{-1}L(n)$ for a positive, slowly varying function L then, for almost all real numbers u (with respect to Lebesgue measure), $\psi(n)\, q_{n+1}(u)/q_n(u) = O(1)$ if and only if

$$\sum_{n=1}^{\infty} \psi(n) < \infty\,; \tag{4}$$

and from Lévy (1937, p. 320) that $n^{-1} \log q_n(u) \to \pi^2/(12 \log 2)$ as $n \to \infty$. Hence, for almost all u, $\psi\{\log q_n(u)\}\, q_{n+1}(u)/q_n(u) = O(1)$ if and only if (4) holds.

This result, and the relationship between rational approximation and linear boundary approximation discussed in Section 5, may be used to show that if the boundary is stochastic in the sense defined in the previous section, then with probability one the optimal rate of approximation to the boundary, in the Hausdorff metric restricted to a region containing an $n \times n$ grid and using data from that grid, equals

$$O\{n^{-1}(\log n)\, L(\log n)^{-1}\} \tag{5}$$

if and only if (4) holds. Similarly, it may be proved that with probability one (4) is equivalent to asking that the optimal rate of approximation be no better than

$$O\{(n \log n)^{-1} L(\log n)\} \tag{6}$$

along any subsequence.

The bound at (5) implies that if the line is placed into the plane at random according to the regime suggested above, then with probability 1 it can be approximated at rate

$$O\{n^{-1}(\log n)(\log\log n)(\log\log\log n)^{1+\epsilon}\}, \tag{7}$$

for all $\epsilon > 0$, from vertex colour data within an $n \times n$ region of the square lattice; and the bound at (6) shows that the optimal convergence rate is no better than

$$O\{n^{-1}(\log n)^{-1}(\log\log n)^{-1}(\log\log\log n)^{-1-\epsilon}\},$$

along some subsequence. Moreover, these results are false if ϵ is replaced by 0.

7 Approximations to curved boundaries

The results derived in Section 6 may be readily extended to the case of local linear approximations to smooth curves on a square lattice. There it is convenient to introduce the concept of a grid of increasing fineness, so as to develop a theory for curve estimation using increasing amounts of information. Rather than assume that the lattice has fixed edge width we suppose it has edge width n^{-1}. For example, we might suppose that its vertices are at points $(n^{-1}i, n^{-1}j)$, where i, j range over the set of all integers.

Replace $\mathcal{L}$ by a smooth curve $\mathcal{C}$, for example given by the equation $y = g(x)$. As before, colour black the vertices above $\mathcal{C}$ and white the vertices below, and consider constructing a local linear approximation to g at x by employing the colours of all vertices that lie within the strip $\mathcal{S} = \mathcal{S}(x)$ defined by $\{(t, y) : x - h < t < x + h \text{ and } -\infty < y < \infty\}$. Here, h plays the role of bandwidth in more traditional curve estimation problems, and the asymptotics involve $h = h(n)$ converging to zero as $n \to \infty$, in such a manner that $nh \to \infty$.

We may define the local linear approximant, $\bar{g}(x)$, at x to be any straight-line segment that agrees with the vertex colour pattern within $\mathcal{S}(x)$; or any segment that has least number of disagreements, if no segment agrees completely. There are two sources of error in this approximation. First, there is a degree of bias, or systematic error, due to the fact that the part of $\mathcal{C}$ that lies within $\mathcal{S}$ is not exactly a straight line. As in more familiar, second-order nonparametric curve estimation problems, the bias is $O(h^2)$ as $h \to 0$ if g has two bounded derivatives. Secondly, there is approximation error arising from the fact that our only information about g is in the form of vertex colours. If the problem of estimating $g(x)$ has a random component, for example if x is taken to be a random variable with an absolutely continuous distribution, then the results developed in Section 6 for the case of a random line may be applied directly to the setting of approximating a random curve by a line segment within $\mathcal{S}$.

In particular, formula (7) may be used to bound the second type of approximation error, provided we replace n by nh and allow nh to increase without limit. Then, assuming that $\log(nh)$ increases like $\log n$, which will certainly be the case for optimal choice of h, we see that the second type of error is bounded above by

$$O\{(n^2h)^{-1}(\log n)^{1+\epsilon}\}\,. \tag{8}$$

Optimising the over-all convergence rate involves balancing systematic and non-systematic sources of error; that is, choosing h so that the bias term, of order h^2, is of the same size as the quantity at (8). This means taking h to be of size $\{n^{-2}(\log n)^{1+\epsilon}\}^{1/3}$, which gives a convergence rate of $O[\{n^{-2}(\log n)^{1+\epsilon}\}^{2/3}]$. The rate $n^{-4/3}$, multiplied by a positive power of $(\log n)^{-1}$, may be shown to be a minimax lower bound in this problem. In related work, Korostelev and Tsybakov (1993) have shown that the rate $n^{-4/3}$ is minimax optimal in the case of certain random grids. Thus, the local linear approximation $\bar{g}$ is within at most a logarithmic factor of achieving the optimal rate.

8 An L_1 view of boundary approximation

In the account of boundary approximation just above, we incorporated an element of randomness in order to remove the ill-posed nature of the problem. Without that randomness, the pointwise properties of rates of convergence defy simple description. Alternatively, we may address global rates of convergence in an L^p metric. We know from the work in earlier sections that, while there are many pathological cases where convergence rates are arbitrarily poor (along subsequences), and while such cases arise at points forming a dense set, they have measure zero. Hence, we are entitled to expect that they will not loom excessively large in an L^p measure of performance. Since the case $p = 1$ puts least emphasis on very large errors then it is potentially the most useful.

Let $\mathcal{C}$ have equation $y = g(x)$, and suppose we observe the vertex colour pattern at all vertices (in^{-1}, jn^{-1}) for integers i, j with $0 \le i \le n$ and

$-\infty < j < \infty$. (Thus, we adopt the "increasingly fine grid" model suggested in Section 7.) Construct the local linear approximation proposed in Section 7, so that for each x in a compact interval $\mathcal{I}$ (which we take without loss of generality to be $[0,1]$) we have an approximation $\bar{g}(x)$ to $g(x)$ using vertex colour data within the strip $\mathcal{S}(x)$. Employing property (2) of convergents we may prove that if g has two bounded derivatives then, for absolute constants A_1, A_2 and A_3,

$$|\bar{g}(x) - g(x)| < A_1 h \left\{q_{N(u)}(u)\, q_{N(u)+1}(u)\right\}^{-1} + B\, h^2 , \tag{9}$$

for all $x \in \mathcal{I}$, $0 < h < \frac{1}{2}$ and $nh > A_2$, where $u = g'(x)$, $N = N(u)$ is the largest integer such that $q_N(u) \le A_3 nh$, and $B = \sup_{\mathcal{I}} |g''|$. Here, $q_n(u)$ is the denominator of the n'th convergent, $p_n(u)/q_n(u)$; see Section 4 for a definition.

The second term on the right-hand side of (9) derives from bias, or equivalently from the systematic error induced by approximating a nonlinear curve by a short but linear segment. The first term results from the limited information available about g, in the form of vertex colours. That term can be arbitrarily large, owing to the sort of pathology noted at the end of Section 2. However, provided g'' is bounded away from zero the integral average of the first term is generally reasonable in size. In fact, it may be shown that if $h = h(n) \to 0$ in such a manner that $(nh)^{-2}(\log n)^2 \to 0$, then

$$\int_{\mathcal{I}} |\bar{g}(x) - g(x)|\, dx = O\{(n^2 h)^{-1}(\log n)^2 + h^2\} , \tag{10}$$

uniformly in functions g for which, for some $C > 1$,

$$C^{-1} \le \inf_{x \in \mathcal{I}} |g''(x)| \le \sup_{x \in \mathcal{I}} |g''(x)| \le C . \tag{11}$$

The lower bound in (11) ensures that g is not too much like a straight line.

Choosing h of size $(n^{-1} \log n)^{2/3}$ in (10) we obtain a rate of approximation, in the L^1 metric, of $O\{(n^{-1} \log n)^{4/3}\}$ uniformly in functions satisfying (11). Again, this convergence rate is close to the optimum of $n^{-4/3}$; see Section 7.

In principle, similar results may be derived for rates of approximation in L^p metrics, where $p > 1$. However, those rates are inferior to the L^1 rate by a polynomial order of magnitude. The reason is that, for $p > 1$, the L^p metric gives greater weight to larger values of the error $|\bar{g}(x) - g(x)|$.

To better appreciate the nature of this problem, observe from (9) that we have the bound

$$|\bar{g}(x) - g(x)| < A_1 A_3^{-2} h\, (nh)^{-2} \{q_{N(u)+1}(u)/q_{N(u)}(u)\} + B\, h^2 ,$$

which is potentially the key to deriving formulae such as (10). However, the ratio $Q_n(u) = q_{N(u)+1}(u)/q_{N(u)}(u)$ is very unstable. Bear in mind that, when finding the integral average of the right-hand side, we are in effect taking U to be a random variable with the Uniform distribution on $\mathcal{I}$, and (in the case of the L^p metric) asking that $E\{Q_n^p(U)\}$ be bounded. Now, it may be shown that the process $\{Q_n(U),\ n \geq 1\}$, is Markovian, and that (while the process is itself not stationary) it has a stationary limit distribution. Therefore, $Q_n(U) = O_p(1)$ as $n \to \infty$. However, the stationary distribution does not have any finite moments, and in fact $E\{Q_n^p(U)\} = \infty$ for all n and all $p \geq 1$. The term $(\log n)^2$ on the right-hand side of (10) is the result of taking a more subtle approach to this problem, necessary even in the case $p = 1$.

9 Estimating boundaries observed with noise

The noiseless model introduced in Section 7 may be written in the form

$$Y(i/n, j/n) = I\{j/n \leq g(i/n)\},$$

where $I(\cdot)$ is an indicator function, $Y(i/n, j/n)$ denotes the colour of the vertex at $(i/n, j/n)$ (white is represented by 1 and black by 0), and the equation $y = g(x)$ represents the boundary $\mathcal{C}$. In practice, due to a combination of systematic and stochastic errors, the colour of each vertex may be more appropriately represented by a number between $-\infty$ and ∞. In particular, we may write

$$Y(i/n, j/n) = f(i/n, j/n) + \epsilon_{ij},$$

where $f(\cdot,\cdot)$ is a function with a fault-type discontinuity along the curve $y = g(x)$, and the independent and identically distributed stochastic errors ϵ_{ij} have zero mean.

It will be assumed that f admits the representation

$$f(x, y) = f_1(x, y) + f_2(x, y)\, I\{y \leq g(x)\},$$

where f_1 and f_2 each have two uniformly bounded derivatives of all types, and f_2 is bounded away from zero. We suppose that g and its first two derivatives are bounded on $\mathcal{I}$. Several different analogues of the local linear estimators in Section 7 are possible; examples include versions based on least squares and on wavelets. We consider here only the former. It amounts to first computing a preliminary approximation, $\tilde{g}$, and then refining it using local linear smoothing within a window. We shall consider a particularly simple preliminary estimator, based on kernel methods, as follows.

Suppose we wish to estimate g at $x \in \mathcal{I}$. Write i_n for the integer nearest to nx, let K be a nonnegative, compactly supported, continuously differentiable function, let h_1 equal a constant multiple of $n^{-2/3}$, and put

$$T(j) = (nh_1^2)^{-1} \sum_k K'\{(j-k)/(nh_1)\} Y(i_n/n, j/n),$$

which is a statistical approximation to the first derivative of $f(i_n/n, \cdot)$ at j/n. Let $\hat{\jmath}$ denote a value which produces a global maximum of $|T|$ in the range $C_1 n \le j \le C_2 n$. Our preliminary estimator of $g(x)$ is $\tilde{g}(x) = \hat{\jmath}/n$.

Next we define an improved estimator. Let $\mathcal{W}$ be a square window of side length $h = h(n)$, with its centre at $(i_n/n, \hat{\jmath}/n)$ and, for the sake of definiteness, its axes aligned with those of the grid. Temporarily make the assumption that within $\mathcal{W}$, f assumes a constant value on either side of a line $\mathcal{L}$. We fit $\mathcal{L}$ by least squares in the class $\mathcal{M}(C, \mathcal{W})$ of all lines $\mathcal{L}$ that divide $\mathcal{W}$ into two sets of vertices of which the larger has no more than C times the number in the smaller (where $C > 1$ is arbitrary but fixed). Specifically, let $\mathcal{I}_1$ [respectively, $\mathcal{I}_2$] denote the set of vertex coordinates $w = (i/n, j/n)$ in $\mathcal{W}$ that lie above [below] $\mathcal{L}$, let $\sum^{(i)}$ denote the sum of $Y(w)$ over all $w \in \mathcal{I}_i$, let $\bar{Y}_i$ be the corresponding mean, and put

$$S(\mathcal{L}) = \sum_{i=1}^{2} \sum\nolimits^{(i)} \{Y(w) - \bar{Y}_i\}^2 .$$

Write $\widehat{\mathcal{L}}$ for a line that minimizes $S(\mathcal{L})$ among all straight lines in $\mathcal{M}(C, \mathcal{W})$ that do not pass through any vertices. (The minimum is of course not uniquely attained, and any measurable approach to breaking ties is allowed.) Write $\hat{g}(x)$ for the ordinate of the point on $\widehat{\mathcal{L}}$ with abscissa x.

Provided the distribution of the errors ϵ_{ij} has sufficiently light tails, it may be proved that $\hat{g}$ has properties similar to those ascribed to the local linear estimator $\bar{g}$ in the no-noise case in Sections 7 and 8. For example, let us assume that the moment generating function of the error distribution exists and is finite in a neighbourhood of the origin. If g'' is bounded and X is a continuous random variable (stochastically independent of the errors ϵ_{ij}), then it may be shown that, for suitable choice of h, with probability one $\hat{g}(X)$ converges to $g(X)$ at rate $O\{n^{-2}(\log n)^{1+\epsilon}\}^{2/3}$ (compare Section 7); and if $|g''|$ is bounded away from both zero and infinity then, again for appropriate h, $\hat{g}$ converges to g in L^1 at rate $O\{(n^{-1} \log n)^{4/3}\}$.

References

[1] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* London: Chapman and Hall.

[2] Hall, P. and Raimondo, M. (1996a). Approximating a line thrown at random onto a grid. *Ann. Appl. Probab.* To appear.

[3] Hall, P. and Raimondo, M. (1996b). On global performance of approximations to smooth curves using gridded data. Unpublished Manuscript.

[4] Khintchine, A.Ya. (1935). Zur metrischen Kettenbruch–Theorie. *Compositio Mathematica* **3**, 275–285.

[5] Khintchine, A.Ya. (1963). *Continued Fractions.* Groningen: Noordhoff.

[6] Klein, F. (1907). *Ausgewählte Kapital der Zahlentheorie.* Teubner.

[7] Korostelev, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction.* Springer Lecture Notes in Statistics Vol. 82. Berlin: Springer-Verlag.

[8] Leveque, W.J. (1956). *Topics in Number Theory.* Volume I. Addison–Wesley, Reading, Mass.

[9] Lévy, P. (1937). *Théorie de l'Addition des Variables Aléatories.* Paris: Gauthier–Villars.

[10] Roth, K.F. (1955). Rational approximations to algebraic numbers. *Mathematika* **2**, 1–20. (Corrigendum *ibid* p. 168.)

[11] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* London: Chapman and Hall.