Boundary Estimation

E. CARLSTEIN and C. KRISHNAMOORTHY*

A data set consists of independent observations taken at the nodes of a grid. An unknown boundary partitions the grid into two regions. All the observations coming from a particular region share a common distribution, but the distributions are different for the two different regions. These two distributions are entirely unknown and need not differ in their means, medians, or any other measure of "level." The grid is of arbitrary dimension, and its mesh is rectangular. Our objective is to estimate the boundary without making any distributional assumptions. We propose a class of estimators and obtain strong consistency for them (including rates of convergence and a bound on the error probability). The boundary estimate is selected from an appropriate collection of candidate boundaries, which must be specified by the user. The candidate boundaries as well as the true boundary must satisfy certain intuitively natural regularity assumptions, including a "smoothness" condition. The boundary estimation problem has applications in diverse fields, including quality control, epidemiology, forestry, marine science, meteorology, and geology. Our method provides (as special cases) estimators for the change point problem, the epidemic change model, templates, linear bisection of the plane, and Lipschitz boundaries. Each of these examples is explicitly analyzed. A simulation study provides numerical evidence that the boundary estimators work well; in this simulation, the two distributions actually share the same mean, median, variance, and skewness. Finally, as an illustration, a boundary estimate is calculated on a data grid of cancer mortality rates in the United States.

KEY WORDS: Change point; Cramér-von Mises; Empirical cumulative distribution function; Epidemic-change; Grid; Kolmogorov–Smirnov; Lipschitz; Partition; Template.

1. INTRODUCTION

1.1 The Statistical Problem

We observe a collection of independent r.v.s $\{X_i\}$, indexed by nodes i of a finite d-dimensional grid within the d-dimensional unit cube $\mathcal{U}_d := [0, 1]^d$. The unknown boundary Θ is simply a $(d-1)$-dimensional surface that partitions $\mathcal{U}_d$ into two regions, $\bar{\Theta}$ and $\underline{\Theta}$. All observations X_i made at nodes $i \in \bar{\Theta}$ are from distribution F, whereas all observations X_i made at nodes $i \in \underline{\Theta}$ are from distribution G. The objective is to estimate the unknown boundary Θ, using the observed data $\{X_i\}$. Figure 1 illustrates the setup in the case $d = 2$.

1.2 Distributional Assumptions

The distributions F and G are entirely unknown. We will not assume any knowledge of the functional forms or parametric families of F and G. No regularity conditions (e.g., continuity, discreteness) will be imposed on F and G. No prior information is needed regarding *how* F and G differ (e.g., they need not differ in their means, medians, or other measure of "level"). The only distributional assumption is that $F \neq G$. This nonparametric approach is needed when the user has insufficient prior knowledge of the underlying distributions or wants a robust corroborator for results from a parametric analysis.

1.3 Grid Assumptions

The grid is generated by divisions along each coordinate axis in $\mathcal{U}_d$ (see Fig. 1). Along the jth axis $(1 \le j \le d)$ are n_j divisions spaced equally at $1/n_j, 2/n_j, \ldots, n_j/n_j$. Observations are made at the resulting grid nodes $i := (i_1/n_1, i_2/n_2, \ldots, i_d/n_d) \in \mathcal{U}_d$, where $i_j \in \{1, 2, \ldots, n_j\}$. Thus the grid mesh may be rectangular rather than strictly square. A rectangular mesh allows for different sampling designs along the different dimensions; this in turn may reflect differing sampling costs in the different dimensions (see the examples). The collection of all nodes i is denoted by I, and the total number of observations is $|I| := \prod_{j=1}^{d} n_j$. In any set $A \subseteq \mathcal{U}_d$ the number of observations (i.e., grid nodes) is $|A| := \#\{i \in A\}$.

1.4 Boundaries

The notion of a boundary in $\mathcal{U}_d$ is formulated in a set theoretic way: the unknown boundary Θ is identified with the corresponding partition $(\bar{\Theta}, \underline{\Theta})$ of $\mathcal{U}_d$. This general formulation is free of the dimension d and allows enough flexibility to treat a wide variety of specific situations (see Sec. 2).

The sample-based estimate of Θ will be selected from a finite collection $\mathcal{T}$ of *candidate boundaries*, with generic element T. Again each candidate T is identified with its corresponding partition $(\bar{T}, \underline{T})$ of $\mathcal{U}_d$. The total number of candidates considered is $|\mathcal{T}| := \#\{T \in \mathcal{T}\}$.

The collection $\mathcal{T}$ must be specified explicitly by the user in accordance with certain *regularity conditions:* $\mathcal{T}$ must be rich enough to contain candidate boundaries close to the true Θ; the cardinality $|\mathcal{T}|$ nevertheless must be controlled in terms of the sample size $|I|$; and the candidate boundaries as well as the true boundary must be sufficiently "smooth." These regularity conditions are intuitively natural and technically manageable (see Secs. 4.4, 5.1, and 5.2), but they do require that the user have some prior knowledge about the form of Θ (see Sec. 2).

To generate a rich collection $\mathcal{T}$ with controlled cardinality, we consider candidate boundaries that are "anchored" to the grid nodes in I (see Sec. 2). It is natural to so anchor the Ts, because in practice one cannot hope to get better resolution from an estimated boundary than whatever degree of resolution is available from the data nodes I.

* E. Carlstein is Associate Professor, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599. C. Krishnamoorthy is Post-Doctoral Research Fellow, Division of Research Computing, Glaxo, Inc., Research Triangle Park, NC 27709. The research was supported by National Science Foundation Grants DMS-8701201 and DMS-8902973, and by a UNC Junior Faculty Development Award. The authors thank two referees, the editor, and an associate editor for their constructive comments.

Journal of the American Statistical Association
June 1992, Vol. 87, No. 418, Theory and Methods

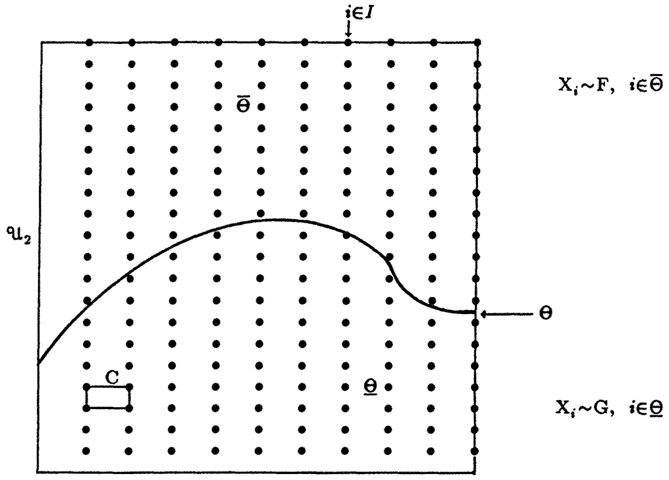

Figure 1. Grid With Boundary, the Two-Dimensional Case: $\mathcal{U}_2$ With $n_1 = 10$, $n_2 = 20$, and $|I| = 200$.

Because of our regularity restrictions, we can handle individual examples in which the set $\bar{\Theta}$ is nonconvex (see Examples 2.b, 2.c) or disconnected (see Example 1.b); compare Ripley and Rasson (1977), who assumed convexity and compactness but made no other regularity restrictions.

Even though the X_is are independent, the "spatial" indexing of the observations nevertheless is crucial in our formulation of boundary estimation. The physical layout of the d-dimensional grid is precisely what leads us to consider particular types of boundaries; for example, change point boundaries (Example 1.a) in the case $d = 1$ and linear bisection boundaries (Example 2.a) in the case $d = 2$. If we throw away the spatial indexing of the observations, then we are faced with a clustering problem.

1.5 Outline of the Article

The ideas discussed in the previous sections are illustrated by specific examples in Section 2. In Section 3 we describe the proposed boundary estimator and compare it to the related (but less general) estimators contained in the literature. Theoretical properties of our estimator are presented in Section 4, with the proofs of these deferred to the Appendix. In Section 5 we explicitly apply our method—and the theoretical results—to the examples from Section 2, and in Section 6 we present numerical applications (a simulation study and a United States cancer mortality data example).

2. EXAMPLES

2.1 The One-Dimensional Case

In the case $d = 1$ it is natural to think of $\mathcal{U}_d$ as a "time" axis; then I indexes observations at equally spaced intervals of time.

Example 1a. The Change Point Problem. The boundary Θ is simply an arbitrary real number $\theta \in (0, 1)$, inducing the partition $\bar{\Theta} := [0, \theta)$, $\underline{\Theta} := [\theta, 1]$. Estimation of the change point Θ has been studied extensively in the literature (see Shaban [1980] for an annotated bibliography); most of this other work assumes either parametric knowledge of F and G or that F and G differ in a known way (e.g., by a shift in "level"). In the quality-control setting, Θ demarcates a change from an "in control" production process to an "out of control" production process. Because we make no assumptions about G, our method will identify the onset of *any* type of disorder in the distribution of the output.

The candidate boundaries $T \in \mathcal{T}$ are essentially all the times at which observations were made; that is, all $t \in \{2/n_1, 3/n_1, \ldots, (n_1 - 1)/n_1\}$, with $\bar{T} := [0, t)$ and $\underline{T} := [t, 1]$. As suggested in Section 1.4, these candidates are anchored in I, yielding a collection $\mathcal{T}$ that is appropriately rich but with a cardinality of the same order as the sample size.

Example 1b. The Epidemic Change Model. The boundary Θ consists of two points $\{\theta_1, \theta_2\}$ with $0 < \theta_1 < \theta_2 < 1$, inducing the partition $\bar{\Theta} := [0, \theta_1) \cup [\theta_2, 1]$, $\underline{\Theta} := [\theta_1, \theta_2)$. In this model θ_1 represents a change to the epidemic distribution G, and θ_2 represents a return to the pre-epidemic distribution F. Again, previous work on the epidemic change model requires parametric knowledge of F and G and/or assumes a shift in "level"; see Siegmund (1986) and Bhattacharya and Brockwell (1976). Note that $\bar{\Theta}$ is not a connected set in this example.

The candidate boundaries $T \in \mathcal{T}$ are all pairs $\{t_1, t_2\}$ anchored in I, with $t_1 < t_2$, $t_1 \in \{1/n_1, 2/n_1, \ldots, (n_1 - 2)/n_1\}$, $t_2 \in \{2/n_1, 3/n_1, \ldots, (n_1 - 1)/n_1\}$, $\bar{T} := [0, t_1) \cup [t_2, 1]$, and $\underline{T} := [t_1, t_2)$.

2.2 The Two-Dimensional Case

In the case $d = 2$ it is natural to think of $\mathcal{U}_d$ as a geographic area; then I indexes observations that are regularly spaced in the east-west direction and also in the north-south direction.

Consider the following application from forestry: The observations $\{X_i : i \in I\}$ represent heights of trees, where F is the distribution for a healthy stand and G is the distribution for a diseased stand. If a disease kills very young and very old trees, then F and G may share the same "level" but may differ in terms of "dispersion." The point here is that our estimator will identify the boundary between "healthy" and "diseased" without any prior knowledge regarding the effect of the disease on the distribution of heights. If the disease spreads radially through the population, then the boundaries considered should be circular or elliptical templates (see Example 2.b).

Marine scientists often rely on voyages of commercial vessels to obtain data. On a trans-Atlantic voyage it is relatively inexpensive to record observations at a large number (n_1) of closely-spaced intervals in the east–west direction. But it is very expensive to extend the grid in the north–south direction (i.e., to increase n_2), because this would entail a whole new trans-Atlantic voyage. So it is important in practice to allow for different sampling designs in the different dimensions. Our method does allow the n_js to differ, and our theoretical analysis (see Sec. 5) shows how the n_js affect the rate of convergence for our estimator.

In image analysis, the data set may arise from a pixellated binary classification of the true but unknown shape $\bar{\Theta}$ against the background $\underline{\Theta}$. In this formulation, F and G are two

Bernoulli distributions with distinct parameter values (e.g., with $p_F \approx 1$ and $p_G \approx 0$).

Example 2a. Linear Bisection. The boundary Θ is an arbitrary straight line segment connecting endpoints on two distinct edges of $\mathcal{U}_2$; this boundary induces a partition ($\bar{\Theta}$, $\underline{\Theta}$) in an obvious way. The candidate boundaries $T \in \mathcal{T}$ are all straight line segments connecting endpoints on two distinct edges of $\mathcal{U}_2$, but the endpoints must be anchored to the grid. An endpoint on an east–west edge must have coordinate i_1/n_1, $i_1 \in \{2, 3, \ldots, n_1 - 1\}$, whereas an endpoint on a north–south edge must have coordinate i_2/n_2, $i_2 \in \{2, 3, \ldots, n_2 - 1\}$.

Example 2b. Templates. A *template* is a boundary in $\mathcal{U}_2$ that can be perturbed via a finite number of "parameters," which may allow for translation, rotation, elongation, and so forth. Circles, ellipses, and polygons can be handled as templates.

Consider, for example, an arbitrary rectangular template Θ with edges parallel to the edges of $\mathcal{U}_2$ and with vertices all in $(0, 1)^2$. The interior of the rectangle is $\underline{\Theta}$, and the remainder of $\mathcal{U}_2$ is $\bar{\Theta}$. Note that $\bar{\Theta}$ is not a convex set. The candidate boundaries $T \in \mathcal{T}$ are all rectangles with edges parallel to the edges of $\mathcal{U}_2$ and with vertices anchored to grid nodes. The east-west coordinate of a vertex must be in $\{1/n_1, 2/n_1, \ldots, (n_1 - 1)/n_1\}$, and the north-south coordinate must be in $\{1/n_2, 2/n_2, \ldots, (n_2 - 1)/n_2\}$. The region $\underline{T}$ is the interior of the rectangle together with its edges.

Example 2c. Lipschitz Boundaries. To define a *Lipschitz boundary,* we identify the lower edge of $\mathcal{U}_2$ as the z-axis and the left edge of $\mathcal{U}_2$ as the y-axis. A Lipschitz boundary Θ is the curve in $\mathcal{U}_2$ corresponding to a function $y_\Theta(\cdot)$: $[0, 1] \mapsto (0, 1)$ satisfying $|y_\Theta(z) - y_\Theta(z')| \le c_0|z - z'|\ \forall z, z' \in [0, 1]$. The constant c_0 simply controls the slope of the boundary. The points $(y, z) \in \mathcal{U}_2$ with $y > y_\Theta(z)$ comprise the region $\bar{\Theta}$; the remainder of $\mathcal{U}_2$ is $\underline{\Theta}$.

The class of Lipschitz boundaries is extremely rich. Note that these boundaries are not readily expressible as templates and that $\bar{\Theta}$ and $\underline{\Theta}$ need not be convex sets.

The candidate boundaries $T \in \mathcal{T}$ correspond to piecewise linear functions $y_T(\cdot)$: $[0, 1] \mapsto (0, 1)$, which are anchored to the grid in the following way: For each $z \in \{0/n_1, 1/n_1, \ldots, n_1/n_1\}$ the associated value of $y_T(z)$ is in $\{1/n_2, 2/n_2, \ldots, (n_2 - 1)/n_2\}$; at intermediate values of z the function $y_T(z)$ is defined by linear interpolation. We restrict $\mathcal{T}$ to those boundaries T for which $|y_T(z) - y_T(z - (1/n_1))| \le c/n_1\ \forall z \in \{1/n_1, 2/n_1, \ldots, n_1/n_1\}$, where $c := 3c_0 + 1$. This Lipschitz-type restriction controls the slope of the candidate boundaries. For this example, it is convenient to have $n_2 \ge n_1$. The region $\bar{T}$ is defined analogously to $\bar{\Theta}$.

2.3 The Three-Dimensional Case

Boundary estimation in the case $d = 3$ has natural applications to meteorology and geology. Planar bisection of $\mathcal{U}_3$ and templates in $\mathcal{U}_3$ can be handled analogously to Examples 2.a and 2.b. In geology the cost of extending the grid in the "depth" direction may again necessitate a sampling design with differing n_js.

3. THE BOUNDARY ESTIMATOR

3.1 The Basic Idea

Our main statistical tool for selecting an estimate $\hat{\Theta}$ from $\mathcal{T}$ is the *empirical cumulative distribution function* (ecdf). For a candidate boundary $T \in \mathcal{T}$, compute the ecdf $\bar{h}_T(x) := \sum_{i \in \bar{T}} \mathbf{I}\{X_i \le x\}/|\bar{T}|$, which treats all observations from region $\bar{T}$ as if they were identically distributed; similarly compute $\underline{h}_T(x) := \sum_{i \in \underline{T}} \mathbf{I}\{X_i \le x\}/|\underline{T}|$, which treats all observations from region $\underline{T}$ as if they were identically distributed. The former ecdf is actually a sample estimate of the unknown mixture distribution $[|\bar{T} \cap \bar{\Theta}|F(x) + |\bar{T} \cap \underline{\Theta}|G(x)]/|\bar{T}|$, whereas the latter ecdf analogously estimates $[|\underline{T} \cap \bar{\Theta}|F(x) + |\underline{T} \cap \underline{\Theta}|G(x)]/|\underline{T}|$. Therefore, the difference between the two ecdfs can be approximated as $|\bar{h}_T(x) - \underline{h}_T(x)| \approx |[|\bar{T} \cap \bar{\Theta}|/|\bar{T}|] - [|\underline{T} \cap \bar{\Theta}|/|\underline{T}|]| \cdot |F(x) - G(x)|$. Note that this last expression can never exceed $|F(x) - G(x)|$ (because $(\bar{T} \cap \bar{\Theta}) \subseteq \bar{T}$ and $(\underline{T} \cap \bar{\Theta}) \subseteq \underline{T}$); moreover, its maximizing value $|F(x) - G(x)|$ is attained precisely when $T = \Theta$. This suggests a natural approach for estimating Θ: Choose as the estimator the candidate boundary that maximizes $|\bar{h}_T(x) - \underline{h}_T(x)|$ over all $T \in \mathcal{T}$. This basic idea will now be refined and generalized.

3.2 Definition of the Boundary Estimator

Rather than restricting attention to the difference $|\bar{h}_T(\cdot) - \underline{h}_T(\cdot)|$ at a single specified x value, we instead consider the differences $d_i^T := |\bar{h}_T(X_i) - \underline{h}_T(X_i)|$ for each $i \in I$. This allows the data to lead us toward the informative x values without prior knowledge of F and G.

Now combine these differences d_i^T using a general norming function $S(d_1^T, d_2^T, \ldots, d_{|I|}^T)$. The norm $S(\cdot)$ must satisfy certain simple conditions (see Sec. 4), but special cases include the *Kolmogorov–Smirnov* norm $S_{\mathrm{KS}}(d_1, d_2, \ldots, d_N) := \sup_{1 \le i \le N}\{d_i\}$, the *Cramér–von Mises* norm $S_{\mathrm{CV}}(d_1, d_2, \ldots, d_N) := (\sum_{1 \le i \le N} d_i^2/N)^{1/2}$, and the *arithmetic-mean* norm $S_{\mathrm{am}}(d_1, d_2, \ldots, d_N) := \sum_{1 \le i \le N} d_i/N$.

Finally we must standardize to account for the inherent instability in the ecdf. Suppose $\bar{h}_{T_0}(\cdot)$ (say) is based on a very small amount of data (i.e., $|\bar{T}_0|$ is very small relative to $|I|$); then $d_i^{T_0}$ is unstable. Thus $d_i^{T_0}$ may be "large" (as compared to other d_i^Ts) merely due to random variability. We should downweight this particular candidate T_0 in our search through $\mathcal{T}$ for a maximizer. This downweighting is accomplished by the multiplicative factor $|\bar{T}_0|/|I|$. There are other intuitively reasonable downweightings (e.g., $(|\bar{T}_0|/|I|)^{1/k}$), but they seem to make the theoretical analysis in Appendix A.2 intractable.

The boundary estimator $\hat{\Theta}$ is defined as the candidate boundary in $\mathcal{T}$ that maximizes the criterion function $D(T) := (|\bar{T}|/|I|)(|\underline{T}|/|I|) \cdot S(d_1^T, d_2^T, \ldots, d_{|I|}^T)$ over all $T \in \mathcal{T}$. Formally, $\hat{\Theta} := \operatorname{argmax}_{T \in \mathcal{T}} D(T)$. Observe that $\hat{\Theta}$ is calculated solely from the data at hand. Theoretical properties of $\hat{\Theta}$ are presented in Section 4.

3.3 Related Methods

Our proposed approach is related to several other methods that recently have been studied in the literature. Each of these other methods suffers from at least one of the following limitations, however:

(a) Restrictions on F and G must be made; for example, F and G are both assumed to be Gaussian or are both assumed to be discrete with finite supports.
(b) Only the one-dimensional single-change point problem is considered (our Example 1.a).
(c) The norm $S(\cdot)$ must be specifically of the form $S_{\text{KS}}(\cdot)$ or $S_{\text{CV}}(\cdot)$.

For example, Deshayes and Picard (1981) proposed a hypothesis testing procedure in the one-dimensional single-change point scenario, based specifically on $S_{\text{KS}}(\cdot)$; note that they tested the null hypothesis of "no change" versus the alternative "some change occurs" and did not test or estimate the specific location of the boundary. Picard (1985) also tested for the existence of a single change point, under the assumption that F and G were both Gaussian with common mean; she used $S_{\text{KS}}(\cdot)$ and $S_{\text{CV}}(\cdot)$. Darkhovskii and Brodskii (1980) considered the one-dimensional single-change point problem under the assumptions that F and G were both discrete with finite supports and that Θ was in a known interval bounded away from the endpoints of $(0, 1)$; they used $S_{\text{CV}}(\cdot)$. Darkhovskii (1984, 1986) dealt with the case $d = 1$, again assuming that F and G were both discrete with finite supports and again using $S_{\text{CV}}(\cdot)$. Darkhovskii (1985) also considered the case $d = 1$, but used $S_{\text{KS}}(\cdot)$. Brodskii and Darkhovskii (1986) again assumed that F and G were both discrete with finite supports and only considered the norm $S_{\text{CV}}(\cdot)$. Asatryan and Safaryan (1986) studied the one-dimensional single-change point problem; they assumed that Θ was in a known interval bounded away from the endpoints of $(0, 1)$ and that F and G were both continuous. Csörgő and Horváth (1987) dealt with the one-dimensional single-change point problem assuming that F and G were both continuous; they used $S_{\text{KS}}(\cdot)$. Carlstein (1988) considered only the one-dimensional single-change point problem.

Our proposed method completely avoids limitations of types (a), (b), and (c). No restrictions (e.g., parametric conditions or regularity conditions) on F or G are needed, because we rely exclusively on the ecdf. Our set-theoretic formulation of boundaries is natural for index grids of any dimension and allows for a wide variety of examples (see Sec. 2). We abstract the important properties of the norm $S(\cdot)$ and then allow *any* choice of norm with those properties; $S_{\text{KS}}(\cdot)$ and $S_{\text{CV}}(\cdot)$ are just two special cases handled within our unified formulation.

A referee has pointed out additional related work by Gagalowicz (1981), Gagalowicz, Graffigne, and Picard (1988), and Therrien (1983).

4. THEORETICAL PROPERTIES OF THE BOUNDARY ESTIMATOR

4.1 Measuring the "Distance" Between Boundaries

To assess the performance of our boundary estimator $\hat{\Theta}$, we first must quantify the notion of "distance" between two boundaries (say, T and Θ). Our "distance" measure is $\partial(T, \Theta) := \min\{\lambda(\bar{T} \circ \bar{\Theta}), \lambda(\underline{T} \circ \bar{\Theta})\}$, where $\circ$ denotes set-theoretic symmetric difference—that is, $(A \circ B) := (A \cap B^c) \cup (A^c \cap B)$—and $\lambda(\cdot)$ is Lebesgue measure over $\mathcal{U}_d$. Intuitively, $\lambda(\bar{T} \circ \bar{\Theta})$ represents the "area" misclassified by $\bar{T}$ as an estimator of $\bar{\Theta}$. For a given boundary T the a priori labelling of the two induced regions as $(\bar{T}, \underline{T})$ rather than as $(\underline{T}, \bar{T})$ can be arbitrary (see, for example, Example 2.a). Therefore, a candidate boundary T is considered "close" to Θ if *either* $\bar{T}$ or $\underline{T}$ is nearly the same region of $\mathcal{U}_d$ as $\bar{\Theta}$.

The function $\partial(\cdot, \cdot)$ has the following desirable properties of a "distance."

Proposition 1. The function $\partial(\cdot, \cdot)$ is a *pseudometric;* that is, it satisfies:

(1.a) [Nonnegativity] $\partial(T, \Theta) \geq 0$.
(1.b) [Identity] $\partial(\Theta, \Theta) = 0$.
(1.c) [Symmetry] $\partial(T, \Theta) = \partial(\Theta, T)$.
(1.d) [Triangle inequality] $\partial(T, \Theta) \leq \partial(T, T') + \partial(T', \Theta)$.

Properties 1.a, 1.b, and 1.c are obvious from the definition; a proof of property 1.d is presented in Appendix A.1.

When we discuss consistency and probability of error for $\hat{\Theta}$ as an estimator of Θ, it will always be in the sense of ∂-distance.

4.2 The Norm $S(\cdot)$

There are some constraints on the choice of $S(\cdot)$. The following conditions are intuitively reasonable and enable us to simultaneously handle a whole class of boundary estimators $\hat{\Theta}$.

Definition. A function $S(\cdot): \mathbf{R}_+^N \mapsto \mathbf{R}_+^1$ is a *mean-dominant norm* if it satisfies:

(D.a) [Symmetry] $S(\cdot)$ is symmetric in its N arguments.
(D.b) [Homogeneity] $S(\alpha d_1, \alpha d_2, \ldots, \alpha d_N) = \alpha S(d_1, d_2, \ldots, d_N)$ whenever $\alpha \geq 0$.
(D.c) [Triangle inequality] $S(d_1 + d_1', d_2 + d_2', \ldots, d_N + d_N') \leq S(d_1, d_2, \ldots, d_N) + S(d_1', d_2', \ldots, d_N')$.
(D.d) [Identity] $S(1, 1, \ldots, 1) = 1$.
(D.e) [Monotonicity] $S(d_1, d_2, \ldots, d_N) \leq S(d_1', d_2', \ldots, d_N')$ whenever $d_i \leq d_i'$ $\forall i$.
(D.f) [Mean dominance] $S(d_1, d_2, \ldots, d_N) \geq \sum_{1 \leq i \leq N} \times d_i / N$.

It is straightforward to check that:

Proposition 2. The functions $S_{\text{KS}}(\cdot)$, $S_{\text{CV}}(\cdot)$, and $S_{\text{am}}(\cdot)$ are mean-dominant norms.

4.3 Asymptotic Results

To study the asymptotic properties of our method, we will let the number of grid-nodes increase: $|I| \rightarrow \infty$. Because the candidate boundaries and the estimator depend on the particular grid, we henceforth equip $\mathcal{T}$ and $\hat{\Theta}$ with explicit subscripts: $\mathcal{T}_I$ and $\hat{\Theta}_I$. Similarly, the number of observations in $A \subseteq \mathcal{U}_d$ is now $|A|_I$.

We assume that $F \neq G$; $\hat{\Theta}_I$ is based on a mean-dominant norm; the boundaries satisfy *regularity conditions* R.1–R.4 (described in Sec. 4.4). The main theoretical results are:

Theorem 1. (Strong consistency):

$$|I|^{\delta} \cdot \partial(\Theta, \hat{\Theta}_I) \to 0 \quad \text{as } |I| \to \infty, \text{ with probability 1.}$$

Theorem 2. (Bound on error probability):

$$\mathbf{P}\{\partial(\Theta, \hat{\Theta}_I) > \varepsilon\} \le c_1 \cdot |\mathcal{T}_I| \cdot \exp\{-c_2 \cdot \varepsilon^2 \cdot |I|\} \quad \text{for } |I| \text{ sufficiently large,}$$

where $c_1 > 0$ and $c_2 > 0$ are constants.

Proofs of these results are presented in Appendix A.2. The rate of convergence obtained in Theorem 1 depends on $\delta \ge 0$. Constraints on δ will follow from the regularity conditions discussed below. In Section 5 we see the actual rates that can be obtained in particular applications. Theorem 2 says that the probability of error decreases exponentially as a function of sample size, but that this effect is counterbalanced by the number of candidate boundaries considered. The precise nature of this tradeoff between $|I|$ and $|\mathcal{T}_I|$ is discussed in Section 4.4.

4.4 Regularity Conditions on Boundaries

Although Theorems 1 and 2 require no distributional assumptions, they do assume certain set theoretic regularity conditions on the boundaries. It will be seen that these regularity conditions are intuitively natural and that they are simple to check in specific applications (see Sec. 5).

Regularity Condition R.1 (Nontrivial partitions). For each $T \in \mathcal{T}_I$, $0 < \lambda(\bar{T}) < 1$ and $0 < |\bar{T}|_I/|I| < 1$. Also, $0 < \lambda(\bar{\Theta}) < 1$.

Condition R.1 prohibits consideration of trivial partitions. It is permissible to consider candidate boundaries with $\lambda(\bar{T})$ arbitrarily small as $|I| \to \infty$; for instance, $\bar{T} = [0, 2/n_1)$ in Example 1.a. Such candidates generally do not influence $\hat{\Theta}_I$, due to the downweighting factor in $D(T)$.

Regularity Condition R.2 (Richness of $\mathcal{T}_I$). For each I, $\exists T_I \in \mathcal{T}_I$ such that the sequence $\{T_I\}$ satisfies $|I|^{\delta} \cdot \partial(\Theta, T_I) \to 0$ as $|I| \to \infty$.

Condition R.2 requires $\mathcal{T}_I$ to contain some candidate boundary T_I that approximates the true Θ (at a rate corresponding to that desired in Theorem 1). If no such ideal candidate were available, we could not possibly hope to statistically select an estimator $\hat{\Theta}_I$ from $\mathcal{T}_I$ in such a way that the convergence of Theorem 1 holds. It is easy to satisfy R.2 when the candidate boundaries are anchored to the grid in a natural way (see Secs. 2 and 5).

Regularity Condition R.3 (Cardinality of $\mathcal{T}_I$). For each $\gamma > 0$, $|\mathcal{T}_I| \cdot \exp\{-\gamma \cdot |I|^{1-2\delta}\} \to 0$ as $|I| \to \infty$.

Condition R.3 quantifies the balance between $|I|$ and $|\mathcal{T}_I|$. Basically, the number of candidate boundaries must be substantially smaller than an exponential of the sample size. This constraint still allows for extremely rich collections $\mathcal{T}_I$ (see Sec. 5). In light of R.3 we can only obtain rates of convergence in Theorem 1 with $\delta < \frac{1}{2}$. Note that:

Proposition 3. Condition R.3 is satisfied for all $\delta \in [0, \frac{1}{2})$ whenever $|\mathcal{T}_I|$ is of the order $|I|^{\nu}$, $\nu > 0$.

To discuss the final regularity condition we need the notion of *cells* in $\mathcal{U}_d$. Recall that the grid (described in Sec. 1.3) induces a rectangular partition of $\mathcal{U}_d$. A cell is simply one of these d-dimensional rectangular regions, including its edges and vertices (see Fig. 1). Thus there are $|I|$ cells in $\mathcal{U}_d$, and they are not strictly disjoint. A generic cell is denoted C, and the collection of all cells in $\mathcal{U}_d$ is denoted $\mathcal{C}_I$. For an arbitrary set $A \subseteq \mathcal{U}_d$ the collection of perimeter cells of A is defined as $\mathcal{P}_I(A) := \{C \in \mathcal{C}_I: C \cap A \ne \phi \text{ and } C \cap A^c \ne \phi\}$. So $\mathcal{P}_I(A)$ consists of those cells that intersect with both A and A^c; that is, cells on the perimeter of A.

Regularity Condition R.4 (Smoothness of perimeter). Denote $\mathfrak{F}_I := \{\bar{\Theta}, \bar{T}: T \in \mathcal{T}_I\}$. We require $|I|^{\delta} \cdot \sup_{A \in \mathfrak{F}_I} \lambda(\mathcal{P}_I(A)) \to 0$ as $|I| \to \infty$.

Condition R.4 guarantees smoothness of the boundaries relative to the grid: as the grid mesh becomes finer, the cell-wise approximation to the boundary (i.e., $\mathcal{P}_I(\cdot)$) must shrink in area. This prohibits boundaries that wander through too many cells in $\mathcal{U}_d$ (see Example 1.c in Sec. 5).

It is trivial to directly check R.4 in the one-dimensional case (see Sec. 5.1). In the two-dimensional case we actually can reduce R.4 to a simple calculation of the lengths of boundaries. Consider a boundary in $\mathcal{U}_2$ that is expressible as a *rectifiable curve* $r(t)$; that is, $r(\cdot)$ is a continuous function from $[a, b] \subseteq \mathbf{R}^1$ into $\mathcal{U}_2$, with coordinates $r(t) = (r_1(t), r_2(t))$, satisfying $L(r) := \sup_{a=t_0<t_1<\cdots<t_k=b} \sum_{1\le i\le k} \|r(t_i) - r(t_{i-1})\| < \infty$. The quantity $L(r)$ is just the length of the boundary. The following relationship holds between the perimeter cells and the length.

Theorem 3. Let $d = 2$. If the set $A \in \mathfrak{F}_I$ corresponds to a boundary expressible as a rectifiable curve $r(\cdot)$, then $\lambda(\mathcal{P}_I(A)) \le 18 \cdot (L(r) + 1)/\min\{n_1, n_2\}$.

Proof of this result is presented in Appendix A.3. Now it is easy to check R.4 by approximating the lengths of boundaries (see Sec. 5.2).

5. EXAMPLES REVISITED

In this section we explicitly apply the theoretical results of Section 4 to the examples from Section 2. In each case a class of strongly consistent boundary estimators is obtained. Note that R.1 is satisfied (by construction) in each of Examples 1.a, 1.b, 2.a, 2.b, and 2.c; this regularity condition will not be discussed further.

5.1 The One-Dimensional Case

Example 1a. The Change Point Problem. In R.2 take T_I to be the point $\lfloor \theta n_1 \rfloor / n_1$, where $\lfloor z \rfloor$ denotes the largest integer less than z. Because $\partial(\Theta, T_I) \le 1/n_1$ and $|I| = n_1$, R.2 holds for any $\delta < \frac{1}{2}$. For R.3 we observe that $|\mathcal{T}_I| < |I|$; hence Proposition 3 applies. For each $A \in \mathfrak{F}_I$ there is only one perimeter cell. Therefore, $\lambda(\mathcal{P}_I(A)) = 1/n_1$ and R.4 is satisfied for any $\delta < \frac{1}{2}$. This establishes Theorems 1 and 2 for the class of change point estimators $\hat{\Theta}_I$; in the rate factor we can allow any $\delta < \frac{1}{2}$.

Example 1b. The Epidemic Change Model. Take T_I to be the pair $\{(\lfloor \theta_1 n_1 \rfloor + 1)/n_1, \lfloor \theta_2 n_1 \rfloor / n_1\}$, so that $\underline{T}_I$ is an "inner approximation" to $\underline{\Theta}$. Again $\partial(\Theta, T_I)$ is of order 1/

n_1 and $|I| = n_1$, so R.2 holds for any $\delta < \frac{1}{2}$. Next we can apply Proposition 3, because $|\mathcal{T}_I| < |I|^2$. For $A \in \mathfrak{F}_I$ there are now two perimeter cells, yielding $\lambda(\mathcal{P}_I(A)) = 2/n_1$ and satisfying R.4 for any $\delta < \frac{1}{2}$. Thus Theorems 1 and 2 hold: We have convergence of our estimators for the epidemic change model, with the rate factor allowing any $\delta < \frac{1}{2}$.

Example 1c. Rationals Versus Irrationals. We have heavily emphasized the set theoretic nature of our approach—in particular, we have exploited the partition $(\bar{\Theta}, \underline{\Theta})$ of $\mathcal{U}_d$. The sets $\bar{\Theta} := \{$*rational* numbers in $[0, 1]\}$ and $\underline{\Theta} := \{$*irrational* numbers in $[0, 1]\}$ constitute a perfectly legitimate partition of $\mathcal{U}_1$. Yet we would be surprised if the proposed method applied in this situation. Indeed an obvious problem arises with the "smoothness" of the perimeter (R.4): *every* cell is in $\mathcal{P}_I(\bar{\Theta})$, so $\lambda(\mathcal{P}_I(\bar{\Theta})) \equiv 1$ and R.4 is violated.

5.2 The Two-Dimensional Case

Example 2a. Linear Bisection. To check R.2 consider the following special case: Θ connects an endpoint on the lower edge of $\mathcal{U}_2$ to an endpoint on the left edge of $\mathcal{U}_2$; the coordinate on the lower edge is $\theta_1 \in (0, 1]$, and the coordinate on the left edge is $\theta_2 \in (0, 1]$. Take T_I to be an analogously oriented segment, with lower edge coordinate $\lfloor \theta_1 n_1 \rfloor / n_1$ and left edge coordinate $\lfloor \theta_2 n_2 \rfloor / n_2$. Then $\partial(\Theta, T_I) = \frac{1}{2}(\theta_1\theta_2 - \lfloor \theta_1 n_1 \rfloor \lfloor \theta_2 n_2 \rfloor / n_2 n_1) \le 1/\min\{n_1, n_2\}$, and hence R.2 is satisfied whenever

$$(\max\{n_1, n_2\})^{\delta}(\min\{n_1, n_2\})^{\delta-1} \to 0. \qquad (\star)$$

Other configurations of Θ similarly yield $(\star)$ as a sufficient condition for R.2. Because $|\mathcal{T}_I| < |I|^2$, we can handle R.3 via Proposition 3. Now observe that $L(r) \le \sqrt{2}$ for any $A \in \mathfrak{F}_I$, so by using Theorem 3 we find that R.4 is satisfied whenever $(\star)$ holds. Thus Theorems 1 and 2 apply to our estimators of a linear bisecting boundary, provided $(\star)$ holds.

Condition $(\star)$ forces the grid design to asymptotically become finer in both dimensions. Consider in particular $\min\{n_1, n_2\} = |I|^{\alpha}$ and $\max\{n_1, n_2\} = |I|^{1-\alpha}$, where $0 < \alpha \le \frac{1}{2}$. Then $(\star)$ holds for any $\delta \in [0, \alpha)$. We obtain the best rate of convergence $|I|^{\delta}$ when we can afford the symmetric grid design; that is, $\alpha = \frac{1}{2}$.

Example 2b. Templates. Consider the rectangular template Θ discussed in Section 2.2. Take T_I to be the candidate boundary corresponding to the largest $\underline{T} \subseteq \underline{\Theta}$. Then $\partial(\Theta, T_I) \le \lambda(\underline{\Theta}) - \lambda(\underline{T}_I) \le 4/\min\{n_1, n_2\}$, so that R.2 is satisfied whenever $(\star)$ holds. Because $|\mathcal{T}_I| < |I|^2$, we again can use Proposition 3 to deal with R.3. Note that $L(r) \le 4$ for all $A \in \mathfrak{F}_I$, so that Theorem 3 reduces R.4 to condition $(\star)$. Theorems 1 and 2 apply to our estimators of rectangular templates, provided $(\star)$ holds. The analysis of the case $\min\{n_1, n_2\} = |I|^{\alpha}$ and $\max\{n_1, n_2\} = |I|^{1-\alpha}$, $0 < \alpha \le \frac{1}{2}$ is exactly as presented in Example 2.a.

Example 2c. Lipschitz Boundaries. For R.2 consider T_I corresponding to the piecewise linear function $y_{T_I}(\cdot)$ defined at each $z \in \{0/n_1, 1/n_1, \ldots, n_1/n_1\}$ by $y_{T_I}(z) := \max\{i_2/n_2 \colon i_2 \in \{1, 2, \ldots, n_2 - 1\}$ and $i_2/n_2 \le y_\Theta(z')$ $\forall z' \in [z - (1/n_1), z + (1/n_1)]\}$. Note that $T_I \in \mathcal{T}_I$, because for each $z \in \{1/n_1, 2/n_1, \ldots, n_1/n_1\}$ we have:

$$\left| y_{T_I}(z) - y_{T_I}\left(z - \frac{1}{n_1}\right)\right| \le \sup_{z' \in [z-(2/n_1), z+(1/n_1)]} y_\Theta(z')$$
$$- \inf_{z' \in [z-(2/n_1), z+(1/n_1)]} y_\Theta(z') + 1/n_2 \le c_0 \cdot 3/n_1 + 1/n_1.$$

Because $y_{T_I}(\cdot)$ is dominated by $y_\Theta(\cdot)$, we can write (for $|I|$ sufficiently large)

$$\partial(\Theta, T_I) = \sum_{z \in \{1/n_1, 2/n_1, \ldots, n_1/n_1\}} \int_{z-1/n_1}^{z} (y_\Theta(z') - y_{T_I}(z'))\, dz',$$

with each integrand bounded above by the right side of the preceding inequality. Therefore R.2 is satisfied whenever

$$(n_2)^{\delta}(n_1)^{\delta-1} \to 0. \qquad (\dagger)$$

For R.3 note that $|\mathcal{T}_I| \approx (n_2)^{n_1}$, so Proposition 3 does not directly apply. In this situation R.3 reduces to

$$\gamma(n_1 n_2)^{1-2\delta} - n_1 \ln(n_2) \to \infty \quad \forall \gamma > 0. \qquad (\dagger\dagger)$$

Last, we use Theorem 3 to handle R.4. Each $A \in \mathfrak{F}_I$ has $L(r) \le \sqrt{c^2 + 1}$, because for $\bar{\Theta}$ we find

$$\sum_{1 \le i \le k} \|r(t_i) - r(t_{i-1})\|$$
$$= \sum_{1 \le i \le k} \sqrt{|y_\Theta(t_i) - y_\Theta(t_{i-1})|^2 + |t_i - t_{i-1}|^2} \le \sqrt{c^2 + 1}$$

whenever $0 = t_0 < t_1 < \cdots < t_k = 1$, and for $\bar{T}$ we find

$$L(r) = \sum_{z \in \{1/n_1, 2/n_1, \ldots, n_1/n_1\}} \times \sqrt{\left| y_T(z) - y_T\left(z - \frac{1}{n_1}\right)\right|^2 + \left(\frac{1}{n_1}\right)^2}$$
$$\le \sqrt{c^2 + 1}.$$

Thus R.4 reduces to $(\dagger)$. Theorems 1 and 2 apply to our estimators of Lipschitz boundaries, provided $(\dagger)$ and $(\dagger\dagger)$ are satisfied.

Consider the case $n_1 = |I|^{\alpha}$ and $n_2 = |I|^{1-\alpha}$, where $0 < \alpha \le \frac{1}{2}$. Conditions $(\dagger)$ and $(\dagger\dagger)$ will both be satisfied for any $\delta \in [0, \min\{\alpha, (1 - \alpha)/2\})$. Note that the best rate of convergence $|I|^{\delta}$ is obtained from the nonsymmetric grid design with $\alpha = \frac{1}{3}$.

6. NUMERICAL APPLICATIONS

6.1 Simulation Study

Here we illustrate the finite sample behavior of $\hat{\Theta}_I$ in a two-dimensional situation where no other boundary estimator is appropriate: F and G are unknown to the statistician and are from different parametric families, but they are both continuous with common mean, median, variance, and skewness. Specifically, F is the $N(0, 1)$ distribution and G has density $g(x) = .697128 \cdot x^2 \cdot \mathbf{I}\{|x| < 1.291\}$. The boundaries are as in Example 2.a; that is, linear bisection, with Θ connecting the points $(.85, .00)$ and $(.25, 1.00)$. Table 1 presents the results of the simulation, showing that the estimators perform well even for these moderately sized data grids. All three estimators improve markedly in response to

Table 1. Simulation Study

Grid size	*Norm S(·)*	$E\{\partial(\Theta, \hat{\Theta}_I)\}$	*(Standard error)*[a]
15 × 15	Arithmetic-mean	.0950	(.0038)
	Cramér–von Mises	.0956	(.0033)
	Kolmogorov–Smirnov	.1011	(.0025)
25 × 25	Arithmetic-mean	.0156	(.0007)
	Cramér–von Mises	.0209	(.0007)
	Kolmogorov–Smirnov	.0509	(.0014)
35 × 35	Arithmetic-mean	.0101	(.0002)
	Cramér–von Mises	.0114	(.0002)
	Kolmogorov–Smirnov	.0304	(.0008)

[a] Each "expected distance" is an empirical estimate based on 1,000 realizations of $\hat{\Theta}_I$; the associated standard error of this estimate is given in parentheses.

larger sample size, with $S_{\mathrm{am}}(\cdot)$ emerging as the best choice of norm and $S_{\mathrm{KS}}(\cdot)$ emerging as the worst choice of norm.

6.2 United States Cancer Mortality Data

Maps are a powerful tool for epidemiological investigation; perhaps the most celebrated example is John Snow's (1855) study of the London cholera epidemic. United States cancer mortality rate maps were compiled by Riggan et al. (1987), for use in "developing and examining hypotheses about the influence of various environmental factors" (page xi) and for investigating possible associations of cancer with "unusual demographic, environmental, industrial characteristics, or employment patterns" (page xii). In particular, for cancer of the trachea, bronchus, and lung (including pleura and other respiratory sites), White males exhibited a pronounced geographic mortality pattern during 1970–1979 (see Fig. 2). The heavy concentration in the southeastern region of the data grid (a region roughly equivalent to what historically and culturally is known as "the South") might be related to the following factors: prevalence of cigarette smoking, especially as associated with the region's cash crop, tobacco; access to preventive, diagnostic, and therapeutic health care,

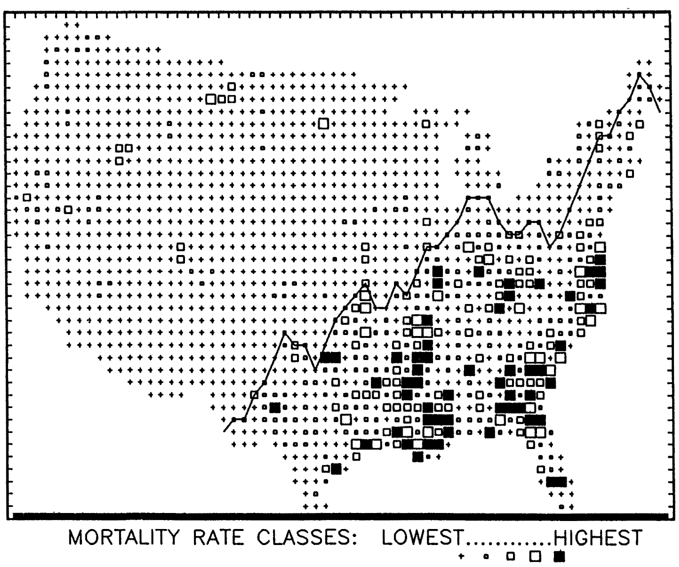

Figure 2. Cancer Mortality of White Males (1970–1979). Cancer of the trachea, bronchus, and lung. Solid line indicates the estimated Lipschitz boundary.

especially as a function of economic and educational conditions; and employment in the textile and furniture manufacturing industries, involving extensive exposure to airborne fibers and dusts. To explicitly delineate this region with minimal constraints on its shape, we estimated a Lipschitz boundary as shown in Figure 2. Besides encompassing the South, the estimated boundary interestingly also includes the heavily industrialized and polluted Eastern seaboard.

For this example the algorithm used the arithmetic-mean norm and employed a restricted search among candidates with Lipschitz constant equal to 2; the FORTRAN code is available from the authors. The value of the Lipschitz constant was chosen by visual comparison of several possibilities: Smaller values seemed to stifle the natural features of the data, whereas larger values allowed distracting jaggedness in the boundary.

APPENDIX: PROOFS

A.1 Proof of Property 1.d in Proposition 1

Expand $\partial(\Theta, T)$ in terms of Θ, T, and T':

$$\begin{aligned}\partial(\Theta, T) = \min\{&\lambda(\bar{T} \cap \underline{\Theta} \cap \bar{T}') + \lambda(\bar{T} \cap \underline{\Theta} \cap \underline{T}') \\ &+ \lambda(\underline{T} \cap \bar{\Theta} \cap \bar{T}') + \lambda(\underline{T} \cap \bar{\Theta} \cap \underline{T}'), \\ &\lambda(\underline{T} \cap \underline{\Theta} \cap \bar{T}') + \lambda(\underline{T} \cap \underline{\Theta} \cap \underline{T}') \\ &+ \lambda(\bar{T} \cap \bar{\Theta} \cap \bar{T}') + \lambda(\bar{T} \cap \bar{\Theta} \cap \underline{T}')\} \\ =: \min\{&U, V\}.\end{aligned}$$

Making the analogous expansions of $\partial(\Theta, T')$ and $\partial(T', T)$, in terms of Θ, T, and T' we see that $\partial(\Theta, T') + \partial(T', T)$ equals one of four possible expressions; for example:

$$\begin{aligned}W := {}&\lambda(\bar{T}' \cap \underline{\Theta} \cap \bar{T}) + \lambda(\bar{T}' \cap \underline{\Theta} \cap \underline{T}) + \lambda(\underline{T}' \cap \bar{\Theta} \cap \bar{T}) \\ &+ \lambda(\underline{T}' \cap \bar{\Theta} \cap \underline{T}) + \lambda(\bar{T} \cap \underline{T}' \cap \bar{\Theta}) + \lambda(\bar{T} \cap \underline{T}' \cap \underline{\Theta}) \\ &+ \lambda(\underline{T} \cap \bar{T}' \cap \bar{\Theta}) + \lambda(\underline{T} \cap \bar{T}' \cap \underline{\Theta}).\end{aligned}$$

Because the summands in U are a subset of the summands in W, we have: $\partial(\Theta, T) = \min\{U, V\} \le U \le W$. Similar inequalities hold for the other three possible expressions of $\partial(\Theta, T') + \partial(T', T)$.

A.2 Proof of Theorems 1 and 2

The main task is to establish:

$$\mathbf{P}\{|I|^{\delta} \cdot \partial(\Theta, \hat{\Theta}_I) > \varepsilon\} \le c_1 \cdot |\mathcal{T}_I| \cdot \exp\{-c_2 \cdot \varepsilon^2 \cdot |I|^{1-2\delta}\}$$
$$\text{for } |I| \text{ sufficiently large.} \quad (1)$$

Then Theorem 2 is the case $\delta = 0$. Theorem 1 follows from the Borel-Cantelli lemma because, by R.3, $\sum_{|I|} |\mathcal{T}_I| \cdot \exp \times \{-c_2 \cdot \varepsilon^2 \cdot |I|^{1-2\delta}\} \approx \sum_{|I|} \exp\{-\frac{1}{2} \cdot c_2 \cdot \varepsilon^2 \cdot |I|^{1-2\delta}\} < \infty$. We shall obtain Equation (1) via a series of lemmas.

Lemma 1. Define $\mathfrak{F}_I^* := \{\bar{\Theta}, \underline{\Theta}; \bar{T}, \underline{T}, \bar{T} \cap \bar{\Theta}, \underline{T} \cap \bar{\Theta}, \bar{T} \cap \underline{\Theta}, \underline{T} \cap \underline{\Theta}: T \in \mathcal{T}_I\}$. We have:

$$|I|^{\delta} \cdot \sup_{A \in \mathfrak{F}_I^*} |\lambda(A) - |A|_I / |I|| \to 0 \quad \text{as } |I| \to \infty.$$

Proof. Consider $A \in \mathfrak{F}_I^*$. Denote $\overline{\mathcal{P}}_I(A) := \{C \in \mathcal{C}_I : C \subseteq A\} \subseteq A$ and $\underline{\mathcal{P}}_I(A) := \{C \in \mathcal{C}_I : C \subseteq A^c\} \subseteq A^c$, so that $\{\overline{\mathcal{P}}_I(A), \mathcal{P}_I(A), \underline{\mathcal{P}}_I(A)\}$ form a partition of $\mathcal{C}_I$. There is a one-to-one correspondence between $\mathcal{C}_I$ and I, where each cell C is associated with a particular grid node i, which is a vertex of C. Now $\#\{C \in \overline{\mathcal{P}}_I(A)\}/|I| = \lambda(\overline{\mathcal{P}}_I(A)) \le \lambda(A) = 1 - \lambda(A^c) \le 1 - \lambda(\underline{\mathcal{P}}_I(A)) = [\#\{C \in \overline{\mathcal{P}}_I(A)\} + \#\{C \in \mathcal{P}_I(A)\}]/|I|$, and $\#\{C \in \overline{\mathcal{P}}_I(A)\} \le |\overline{\mathcal{P}}_I(A)|_I \le |A|_I = |I| - |A^c|_I \le |I| - |\underline{\mathcal{P}}_I(A)|_I \le |I| - \#\{C \in \underline{\mathcal{P}}_I(A)\} = \#\{C \in \overline{\mathcal{P}}_I(A)\} + \#\{C \in \mathcal{P}_I(A)\}$, so that $|\lambda(A) - |A|_I/|I|| \le |\lambda(A) - \#\{C \in \overline{\mathcal{P}}_I(A)\}/|I|| + |\#\{C \in \overline{\mathcal{P}}_I(A)\}/|I| - |A|_I/|I|| \le 2\cdot\#\{C \in \mathcal{P}_I(A)\}/|I| = 2\lambda(\mathcal{P}_I(A))$.

Thus sets A of the form $\bar\Theta$ and $\bar T$ are handled immediately by R.4. Sets A of the form $\underline\Theta$ and $\underline T$ also are handled by R.4, because $\mathcal{P}_I(A^c) = \mathcal{P}_I(A)$. For the remaining sets A in $\mathfrak{F}_I^*$, note that $\mathcal{P}_I(A_1 \cap A_2) \subseteq \mathcal{P}_I(A_1) \cup \mathcal{P}_I(A_2)$, so R.4 again applies.

We need some notation making the dependence on I explicit. The data now will be denoted $\{X_i^I : i \in I\}$, and analogously to Section 3 we denote $\bar h_T^I(x) := \sum_{j\in\bar T} \mathbf{I}\{X_j^I \le x\}/|\bar T|_I$, $\underline{h}_T^I(x) := \sum_{j\in\underline T} \mathbf{I}\{X_j^I \le x\}/|\underline T|_I$, $d_{Ii}^T := |\bar h_T^I(X_i^I) - \underline{h}_T^I(X_i^I)|$, $D_I(T) := (|\bar T|_I/|I|)(|\underline T|_I/|I|)S_{|I|}(d_{Ii}^T : i \in I)$, where $S_{|I|}(\cdot)$ is a mean-dominant norm with $|I|$ arguments. We also define $\bar\eta_T(x) := [\lambda(\bar T \cap \bar\Theta)F(x) + \lambda(\bar T \cap \underline\Theta)G(x)]/\lambda(\bar T)$, $\underline\eta_T(x) := [\lambda(\underline T \cap \bar\Theta)F(x) + \lambda(\underline T \cap \underline\Theta)G(x)]/\lambda(\underline T)$, $\delta_{Ii}^T := |\bar\eta_T(X_i^I) - \underline\eta_T(X_i^I)|$, $\Delta_I(T) := \lambda(\bar T)\lambda(\underline T)S_{|I|}(\delta_{Ii}^T : i \in I)$.

Lemma 2. For $|I|$ sufficiently large,

$$\mathbf{P}\{|I|^\delta \cdot \sup_{T\in\mathcal{T}_I}|D_I(T) - \Delta_I(T)| > \varepsilon\} \le K\cdot|\mathcal{T}_I|\cdot\exp\{-k\cdot\varepsilon^2\cdot|I|^{1-2\delta}\}.$$

Proof. Define $\tilde D_I(T) := \lambda(\bar T)\lambda(\underline T)S_{|I|}(d_{Ii}^T : i \in I)$ and note that $|\tilde D_I(T) - D_I(T)| = |(\lambda(\bar T) - |\bar T|_I/|I|)\lambda(\underline T) + (\lambda(\underline T) - |\underline T|_I/|I|)|\bar T|_I/|I||\,|S_{|I|}(d_{Ii}^T : i \in I)$. The last factor is at most 1, by properties (D.e) and (D.d) of $S_{|I|}$. Thus by Lemma 1 it will suffice for us to prove that the inequality of Lemma 2 holds with $\tilde D_I$ in place of D_I.

Denote $\bar H_T^{Ii} := |\bar h_T^I(X_i^I) - \bar\eta_T(X_i^I)|$, $\underline{H}_T^{Ii} := |\underline h_T^I(X_i^I) - \underline\eta_T(X_i^I)|$, and $e_{Ii}^T := \bar H_T^{Ii} + \underline H_T^{Ii}$, so that $d_{Ii}^T \le e_{Ii}^T + \delta_{Ii}^T$ and $\delta_{Ii}^T \le e_{Ii}^T + d_{Ii}^T$. Then $\tilde D_I(T) - \Delta_I(T) \le \lambda(\bar T)\lambda(\underline T)S_{|I|}(e_{Ii}^T : i \in I)$, by virtue of properties (D.e) and (D.c) of $S_{|I|}$. The same bound applies to $\Delta_I(T) - \tilde D_I(T)$, yielding $|\tilde D_I(T) - \Delta_I(T)| \le \lambda(\bar T)S_{|I|}(\bar H_T^{Ii} : i \in I) + \lambda(\underline T)S_{|I|}(\underline H_T^{Ii} : i \in I)$, by property (D.c).

Now observe that

$$\begin{aligned}\bar H_T^{Ii} \le{}& \left|\frac{\sum_{j\in\bar T\cap\bar\Theta}\mathbf{I}\{X_j^I \le X_i^I\}}{|\bar T\cap\bar\Theta|_I} - F(X_i^I)\right|\frac{|\bar T\cap\bar\Theta|_I}{|\bar T|_I}\\ &+ F(X_i^I)\left|\frac{|\bar T\cap\bar\Theta|_I}{|\bar T|_I} - \frac{\lambda(\bar T\cap\bar\Theta)}{\lambda(\bar T)}\right| + \left|\frac{\sum_{j\in\bar T\cap\underline\Theta}\mathbf{I}\{X_j^I \le X_i^I\}}{|\bar T\cap\underline\Theta|_I}\right.\\ &\left. - G(X_i^I)\right|\frac{|\bar T\cap\underline\Theta|_I}{|\bar T|_I} + G(X_i^I)\left|\frac{|\bar T\cap\underline\Theta|_I}{|\bar T|_I} - \frac{\lambda(\bar T\cap\underline\Theta)}{\lambda(\bar T)}\right|.\end{aligned}$$

The first modulus on the right side is bounded by

$$\bar F_T^I := \sup_{x\in\mathbf{R}}\left|\frac{\sum_{j\in\bar T\cap\bar\Theta}\mathbf{I}\{X_j^I \le x\}}{|\bar T\cap\bar\Theta|_I} - F(x)\right|,$$

and the factor $F(X_i^I)$ in the second summand on the right side is bounded by 1. The third and fourth summands on the right side are bounded similarly. Substituting these bounds into the right side, we obtain $\bar H_T^{Ii} \le \bar H_T^I$ (say), where $\bar H_T^I$ does *not* depend on i. Use an analogous argument to obtain $\underline H_T^{Ii} \le \underline H_T^I$. Therefore, applying properties (D.e), (D.b), and (D.d), we find that

$$|\tilde D_I(T) - \Delta_I(T)| \le \lambda(\bar T)\bar H_T^I + \lambda(\underline T)\underline H_T^I. \tag{2}$$

The right side of Equation (2) comprises four analogous summands. We shall deal explicitly with only the first of these; that is

$$\left|\left(\lambda(\bar T) - \frac{|\bar T|_I}{|I|}\right)\frac{|\bar T\cap\bar\Theta|_I}{|\bar T|_I} + \frac{|\bar T\cap\bar\Theta|_I}{|I|}\right|\bar F_T^I + \left|\left(\lambda(\bar T) - \frac{|\bar T|_I}{|I|}\right)\frac{|\bar T\cap\bar\Theta|_I}{|\bar T|_I} + \left(\frac{|\bar T\cap\bar\Theta|_I}{|I|} - \lambda(\bar T\cap\bar\Theta)\right)\right|.$$

For the second modulus, note that Lemma 1 applies—uniformly in $T \in \mathcal{T}_I$. Also Lemma 1 applies to the first term inside the first modulus. It now suffices to consider

$$\mathbf{P}\{\sup_{T\in\mathcal{T}_I}(|\bar T\cap\bar\Theta|_I/|I|)\bar F_T^I > \varepsilon|I|^{-\delta}\} \le \sum_{T\in\mathcal{T}_I}\mathbf{P}\{\bar F_T^I > \varepsilon|I|^{1-\delta}/|\bar T\cap\bar\Theta|_I\}.$$

Each probability in this summation is bounded by $K_1 \times \exp\{-k_1\varepsilon^2|I|^{1-2\delta}\}$; see Dvoretzky, Kiefer, and Wolfowitz (1956), specifically their Lemma 2 and the discussion following their Theorem 3.

Lemma 3. We can write $\Delta_I(T) = \rho(T)\cdot S_{|I|}(\delta_{Ii}^\Theta : i \in I)$, where $\rho(T) := |\lambda(\bar T\cap\bar\Theta)\lambda(\underline T) - \lambda(\underline T\cap\bar\Theta)\lambda(\bar T)|$.

Proof. Observe that $\delta_{Ii}^T = \rho(T)\delta_{Ii}^\Theta/\lambda(\bar T)\lambda(\underline T)$. Now apply property (D.b) of $S_{|I|}$.

Lemma 4. For every $T \in \mathcal{T}_I$, we have $\Delta_I(T) \le \Delta_I(\Theta)$.

Proof. By Lemma 3, it suffices to show $\rho(T) \le \rho(\Theta)$. In the definition of $\rho(T)$, consider the expression within the modulus. If this expression is positive, then $\rho(T)$ equals $\lambda(\bar T\cap\bar\Theta)\lambda(\underline T\cap\underline\Theta) - \lambda(\underline T\cap\bar\Theta)\lambda(\bar T\cap\underline\Theta) \le \lambda(\bar\Theta)\lambda(\underline\Theta)$. The negative case is handled similarly.

Lemma 5. For $\gamma > 0$, $\partial(\Theta, T) < \gamma \Rightarrow \rho(\Theta) - \rho(T) < \gamma$ and $\partial(\Theta, T) > \gamma \Rightarrow \rho(\Theta) - \rho(T) > k'\cdot\gamma > 0$.

Proof. Note that

$$\begin{aligned}\rho(\Theta) - \rho(T) = \min\{&\lambda(\underline\Theta\cap\bar T)\lambda(\bar\Theta) + \lambda(\bar\Theta\cap\underline T)\lambda(\underline\Theta),\\ &\lambda(\bar\Theta\cap\bar T)\lambda(\underline\Theta) + \lambda(\underline\Theta\cap\underline T)\lambda(\bar\Theta)\}.\end{aligned} \tag{3}$$

Comparing this expression to the definition of $\partial(\Theta, T)$, the first implication is clear. For the second implication, we have by hypothesis that $\max\{\lambda(\underline\Theta\cap\bar T), \lambda(\bar\Theta\cap\underline T)\} > \gamma/2$ and that $\max\{\lambda(\bar\Theta\cap\bar T), \lambda(\underline\Theta\cap\underline T)\} > \gamma/2$. Equation (3) then yields $\rho(\Theta) - \rho(T) > \min\{\lambda(\bar\Theta), \lambda(\underline\Theta)\}\gamma/2$.

Lemma 6. For $|I|$ sufficiently large,

$$\mathbf{P}\{|I|^\delta\cdot|\Delta_I(\hat\Theta_I) - \Delta_I(\Theta)| > \varepsilon\} \le \tilde K\cdot|\mathcal{T}_I|\cdot\exp\{-\tilde k\cdot\varepsilon^2\cdot|I|^{1-2\delta}\}.$$

Proof. Let $T_I^\Theta \in \mathcal{T}_I$ be the maximizer of $\Delta_I(\cdot)$ over $\mathcal{T}_I$. Then by Lemma 4 we have $\Delta_I(\Theta) \ge \Delta_I(T_I^\Theta) \ge \Delta_I(T)\ \forall T \in \mathcal{T}_I$, and by definition we have $D_I(\hat\Theta_I) \ge D_I(T)\ \forall T \in \mathcal{T}_I$. Now

$$\begin{aligned}|\Delta_I(\hat\Theta_I) - \Delta_I(\Theta)| \le{}& |\Delta_I(\hat\Theta_I) - D_I(\hat\Theta_I)| + |D_I(\hat\Theta_I)\\ &- \Delta_I(T_I^\Theta)| + |\Delta_I(T_I^\Theta) - \Delta_I(\Theta)|.\end{aligned} \tag{4}$$

The second modulus on the right side is bounded by $\sup_{T\in\mathcal{T}_I}|D_I(T) - \Delta_I(T)|$, because either $D_I(\hat\Theta_I) \ge \Delta_I(T_I^\Theta) \ge \Delta_I(\hat\Theta_I)$ or $\Delta_I(T_I^\Theta) \ge D_I(\hat\Theta_I) \ge D_I(T_I^\Theta)$. The same bound applies to the first modulus on the right side of Equation (4).

Using Lemmas 3 and 4 and properties (D.e) and (D.d) of $S_{|I|}$, the third modulus on the right side is bounded by $\Delta_I(\Theta) - \Delta_I(T_I) = [\rho(\Theta) - \rho(T_I)]S_{|I|}(\delta_{Ii}^\Theta : i \in I) \le \rho(\Theta) - \rho(T_I)$. For $|I|$ sufficiently large, R.2 ensures that $\partial(\Theta, T_I) < \frac{1}{2}\varepsilon|I|^{-\delta}$. Thus by Lemma 5 the third modulus on the right side of Equation (4) is *deterministically* bounded by $\frac{1}{2}\varepsilon|I|^{-\delta}$. Combining this with the bounds from the previous paragraph, we see that applying Lemma 2 completes this proof.

Lemma 7. Denote $\mu_F := \int_{-\infty}^{\infty} |F(x) - G(x)|\ dF(x)$, $\mu_G := \int_{-\infty}^{\infty} |F(x) - G(x)|\ dG(x)$, and $\mu := \mu_F\lambda(\bar{\Theta}) + \mu_G\lambda(\underline{\Theta})$. Then $\mu > 0$.

Proof. By assumption we have $\Lambda := \{x \in \mathbf{R}: |F(x) - G(x)| > 0\} \neq \phi$. It suffices to show that either $\int_\Lambda dF(x) > 0$ or $\int_\Lambda dG(x) > 0$. The case where Λ contains a discontinuity point of F or G is trivial, so we now will presume that F and G are continuous at each $x \in \Lambda$.

Select $x_0 \in \Lambda$ with (say) $F(x_0) > G(x_0)$. Then $\sigma := \{y \in (-\infty, x_0): F(x) > G(x)\ \forall x \in (y, x_0]\}$ is nonempty, by continuity. Denote $y_0 := \inf\{y \in \sigma\}$. If $y_0 = -\infty$, then $(-\infty, x_0] \subseteq \Lambda$ and thus $\int_\Lambda dF(x) \geq F(x_0) > G(x_0) \geq 0$. If $y_0 > -\infty$, then $F(y_0) \leq G(y_0)$ and also $(y_0, x_0] \subseteq \Lambda$, yielding $\int_\Lambda dF(x) \geq F(x_0) - F(y_0) \geq F(x_0) - G(y_0) > G(x_0) - G(y_0) \geq 0$.

Lemma 8. Equation (1) holds.

Proof. Applying Lemmas 3, 4, and 5 and property (D.f) of $S_{|I|}$, we have:

$$\partial(\Theta, T) > \gamma \Rightarrow |\Delta_I(\Theta) - \Delta_I(T)| = [\rho(\Theta) - \rho(T)]S_{|I|}(\delta_{Ii}^{\Theta}: i \in I) > k' \cdot \gamma \cdot \delta_I^{\Theta},$$

where $\delta_I^{\Theta} := \sum_{i\in I} \delta_{Ii}^{\Theta}/|I|$. Thus, $\mathbf{P}\{\partial(\Theta, \hat{\Theta}_I) > \varepsilon|I|^{-\delta}\} \leq \mathbf{P}\{|\Delta_I(\Theta) - \Delta_I(\hat{\Theta}_I)| > k'\varepsilon|I|^{-\delta}\delta_I^{\Theta}\} \leq \mathbf{P}\{|I|^{\delta}|\Delta_I(\Theta) - \Delta_I(\hat{\Theta}_I)| > k'\varepsilon\omega\} + \mathbf{P}\{\delta_I^{\Theta} < \omega\}$, where $\omega := \mu/2 > 0$ (by Lemma 7). The first probability on the right side is handled immediately by Lemma 6; the second probability on the right side is bounded by $\mathbf{P}\{|\delta_I^{\Theta} - \mu| > \omega\}$. Denote $\bar{\delta}_I^{\Theta} := \sum_{i\in\bar{\Theta}} \delta_{Ii}^{\Theta}/|\bar{\Theta}|_I$ and $\underline{\delta}_I^{\Theta} := \sum_{i\in\underline{\Theta}} \delta_{Ii}^{\Theta}/|\underline{\Theta}|_I$, so that $\delta_I^{\Theta} = \bar{\delta}_I^{\Theta}(|\bar{\Theta}|_I/|I|) + \underline{\delta}_I^{\Theta}(|\underline{\Theta}|_I/|I|)$. Note that

$$|\delta_I^{\Theta} - \mu| \leq \bar{\delta}_I^{\Theta}\left|\frac{|\bar{\Theta}|_I}{|I|} - \lambda(\bar{\Theta})\right| + \lambda(\bar{\Theta})|\bar{\delta}_I^{\Theta} - \mu_F| + \underline{\delta}_I^{\Theta}\left|\frac{|\underline{\Theta}|_I}{|I|} - \lambda(\underline{\Theta})\right| + \lambda(\underline{\Theta})|\underline{\delta}_I^{\Theta} - \mu_G|.$$

The first and third summands on the right side are handled using Lemma 1. Consider the second summand on the right side; a similar argument holds for the fourth summand. By Equation (2.3) of Hoeffding (1963), we have $\mathbf{P}\{|\bar{\delta}_I^{\Theta} - \mu_F| > \omega/4\} \leq 2 \cdot \exp\{-\tilde{c}|\bar{\Theta}|_I\}$. Lemma 1 ensures that $|\bar{\Theta}|_I$ eventually exceeds $|I|\lambda(\bar{\Theta})/2$ (say); therefore, the bound from Hoeffding (1963) can be absorbed into the earlier bound from Lemma 6, for $\varepsilon > 0$ sufficiently small.

A.3 Proof of Theorem 3

For any set $U \subseteq \mathcal{U}_2$, let $\tilde{U}$ denote the *closure* of U, let $\mathcal{B}(U) := \tilde{U} \cap \widetilde{U^c}$ *denote the boundary* of U, and let $\mathcal{J}(U)$ denote the *interior* of U. For the set A we are assuming: $\mathcal{B}(A) = \{r(t) \in \mathcal{U}_2: t \in [a, b]\}$, with $\mathcal{J}(A) \neq \phi$ and $\mathcal{J}(A^c) \neq \phi$.

Define $K := \lceil 2 \cdot L(r) \cdot \max\{n_1, n_2\}\rceil$. Consider intervals $[a_j, b_j]$, $1 \leq j \leq K$, such that $a = a_1 < b_1 = a_2 < b_2 = \cdots = a_K < b_K = b$. Let $r_j^*(t)$ be the curve $r(t)$ restricted to $t \in [a_j, b_j]$, and let $L_j := L(r_j^*)$. In particular choose the intervals so that $L_j = L(r)/K\ \forall j$. Denote $R_j := \{r_j^*(t) \in \mathcal{U}_2: t \in [a_j, b_j]\}$. Let $C_j \in \mathcal{C}_I$ be a cell containing the point $r_j^*(a_j)$, and let $\mathfrak{C}_j$ be the collection of neighboring cells $C \in \mathcal{C}_I$ that share a common edge or vertex with C_j (including C_j itself). Because $L_j \leq \frac{1}{2}(1/\max\{n_1, n_2\})$, the points in R_j form a subset of the points in $\mathcal{J}(\mathfrak{C}_j)$. Thus we have $\#\{C \in \mathcal{C}_I: C \in \mathcal{P}_I(A)\} \leq \#\{C \in \mathcal{C}_I: C \cap \mathcal{B}(A) \neq \phi\} = \#\{C \in \mathcal{C}_I: C \cap (\cup_{1\leq j\leq K} R_j) \neq \phi\} = \sum_{C\in\mathcal{C}_I} \mathbf{I}\{\cup_{1\leq j\leq K} (R_j \cap C) \neq \phi\} \leq \sum_{1\leq j\leq K} \sum_{C\in\mathcal{C}_I} \mathbf{I}\{R_j \cap C \neq \phi\} \leq \sum_{1\leq j\leq K} \sum_{C\in\mathcal{C}_I} \mathbf{I}\{C \in \mathfrak{C}_j\} \leq K \cdot 9$. Finally, recall that $\lambda(\mathcal{P}_I(A)) = \#\{C \in \mathcal{C}_I: C \in \mathcal{P}_I(A)\}/n_1n_2$.

[*Received April 1989. Revised July 1991.*]

REFERENCES

Asatryan, D., and Safaryan, I. (1986), "Nonparametric Methods for Detecting Changes in the Properties of Random Sequences," in *Detection of Changes in Random Processes,* ed. L. Telksnys, New York: Optimization Software, pp. 1–13.

Bhattacharya, P. K., and Brockwell, P. J. (1976), "The Minimum of an Additive Process With Applications to Signal Estimation and Storage Theory," *Zeitschrift fuer Wahrscheinlichkeitstheorie und Verwandte Gebiete,* 37, 51–75.

Brodskii, B. E., and Darkhovskii, B. S. (1986), "The A Posteriori Method of Detecting the Disruption of a Random Field," in *Detection of Changes in Random Processes,* ed. L. Telksnys, New York: Optimization Software, pp. 32–38.

Carlstein, E. (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics,* 16, 188–197.

Csörgő, M., and Horváth, L. (1987), "Nonparametric Tests for the Changepoint Problem," *Journal of Statistical Planning and Inference,* 17, 1–9.

Darkhovskii, B. S. (1984), "On Two Estimation Problems for Times of Change of the Probabilistic Characteristics of a Random Sequence," *Theory of Probability and Its Applications,* 29, 478–487.

——— (1985), "A Nonparametric Method of Estimating Intervals of Homogeneity for a Random Sequence," *Theory of Probability and Its Applications,* 30, 845–849.

——— (1986), "A General Method for Estimating the Instant of Change in the Probabilistic Characteristics of a Random Sequence," in *Detection of Changes in Random Processes,* ed. L. Telksnys, New York: Optimization Software, pp. 47–52.

Darkhovskii, B. S., and Brodskii, B. E. (1980), "A Posteriori Detection of the 'Disorder' Time of a Random Sequence," *Theory of Probability and Its Applications,* 25, 624–628.

Deshayes, J., and Picard, D. (1981), "Convergence de Processus à Double Indice: Application aux Tests de Rupture Dans un Modèle," *Comptes Rendus des Séances de L'Académie des Sciences,* 292, 449–452.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956), "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator," *The Annals of Mathematical Statistics,* 27, 642–669.

Gagalowicz, A. (1981), "A New Method for Texture Fields Synthesis: Some Applications to the Study of Human Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* PAMI-3, 520–533.

Gagalowicz, A., Graffigne, C., and Picard, D. (1988), "Texture Boundary Positioning," in *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics,* New York: Institute of Electrical and Electronics Engineers, pp. 16–19.

Hoeffding, W. (1963), "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association,* 58, 13–30.

Picard, D. (1985), "Testing and Estimating Change-Points in Time Series," *Advances in Applied Probability,* 17, 841–867.

Riggan, W. B., Creason, J. P., Nelson, W. C., Manton, K. G., Woodbury, M. A., Stallard, E., Pellom, A. C., and Beaubier, J. (1987), *U.S. Cancer Mortality Rates and Trends, 1950–1979,* (Vol. IV: Maps), U.S. Environmental Protection Agency, Washington, D.C.: U.S. Government Printing Office.

Ripley, B. D., and Rasson, J.-P. (1977), "Finding the Edge of a Poisson Forest," *Journal of Applied Probability,* 14, 483–491.

Shaban, S. A. (1980), "Change Point Problem and Two-Phase Regression: An Annotated Bibliography," *International Statistical Review,* 48, 83–93.

Siegmund, D. (1986), "Boundary Crossing Probabilities and Statistical Applications," *The Annals of Statistics,* 14, 361–404.

Snow, J. (1855), *On the Mode of Communication of Cholera* (2nd ed.), London: John Churchill.

Therrien, C. W. (1983), "An Estimation-Theoretic Approach to Terrain Image Segmentation," *Computer Vision, Graphics, and Image Processing,* 22, 313–326.