THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY OF NORTH CAROLINA

PRACTICAL CONSIDERATIONS IN BOUNDARY ESTIMATION: ROBUSTNESS, EFFICIENT COMPUTATION, AND BOOTSTRAPPING

by

C. Krishnamoorthy

and

E. Carlstein

Sept. 1992

Mimeo Series #2083

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

PRACTICAL CONSIDERATIONS IN BOUNDARY ESTIMATION: ROBUSTNESS, EFFICIENT COMPUTATION, AND BOOTSTRAPPING

Charu Krishnamoorthy
Research Computing, 2-3224
Glaxo Research Institute
Research Triangle Park, North Carolina 27709

Ed Carlstein*
Department of Statistics
University of North Carolina at Chapel Hill
CB# 3260, Phillips Hall
Chapel Hill, North Carolina 27599-3260

Abstract

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 the grid is of arbitrary dimension with rectangular mesh. In this scenario, the boundary can be consistently estimated.

The boundary estimate is selected from an appropriate collection $\mathcal{T}$ of *candidate boundaries* which must be specified by the user. The candidate boundaries must satisfy certain regularity assumptions, including a "richness" condition. In practice, one may be faced with a $\mathcal{T}$ that is not sufficiently "rich". How robust is the estimator in such a situation? This question is addressed via a.s. results comparing the asymptotic error of the estimator with the smallest possible error in $\mathcal{T}$.

Because the boundary estimator requires a search (through $\mathcal{T}$) and calculation over the whole data grid, efficient computation is a practical necessity. An algorithm is presented which reduces the computational burden by an order of magnitude relative to the naive approach.

A graphical bootstrap procedure is proposed for studying the variability of the boundary estimator. Simulations of this procedure indicate that it does accurately reflect the true sampling behavior of the boundary estimator.

*Research supported by National Science Foundation Grants DMS-8902973 & DMS-9205881.

1. INTRODUCTION

1.1 The Boundary Estimation Problem. We observe a collection of independent r.v.s $\{X_i^I\}$, indexed by nodes i of a finite d-dimensional grid I within the d-dimensional unit cube $\mathcal{U}_d:=[0, 1]^d$. The unknown *boundary* Θ is a (d−1)-dimensional surface that partitions $\mathcal{U}_d$ into two regions, $\overline{\Theta}$ and $\underline{\Theta}$. All observations X_i^I made at nodes $i \in \overline{\Theta}$ are from distribution F, while all observations X_i^I made at nodes $i \in \underline{\Theta}$ are from distribution G. The distributions F and G are entirely unknown; the only distributional assumption is that $F \neq G$. The objective is to estimate the unknown boundary Θ, using the observed data $\{X_i^I: i \in I\}$.

The grid is generated by divisions along each coordinate axis in $\mathcal{U}_d$. Along the j^{th} axis $(1 \leq j \leq d)$, there are n_j divisions which are equally spaced 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 $i_j \in \{1, 2, \ldots, n_j\}$. 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 $|A|_I := \#\{i \in A\}$.

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

1.2 The Boundary Estimator. Our tool for selecting an estimate $\hat{\Theta}_I$ from $\mathcal{T}_I$ is the *empirical cumulative distribution function (e.c.d.f.)*. For a candidate boundary $T \in \mathcal{T}_I$, compute the e.c.d.f.s $\bar{h}_T^I(x):=\sum_{i \in \overline{T}} \mathbb{I}\{X_i^I \leq x\}/|\overline{T}|_I$ and $\underline{h}_T^I(x):=\sum_{i \in \underline{T}} \mathbb{I}\{X_i^I \leq x\}/|\underline{T}|_I$, and consider the differences $d_{Ii}^T:=|\bar{h}_T^I(X_i^I)-\underline{h}_T^I(X_i^I)|$ for each $i \in I$. Now combine these differences d_{Ii}^T using a "mean-dominant norm" $S_{|I|}(d_{I1}^T, d_{I2}^T, \ldots, d_{I|I|}^T)$ [see Carlstein (1988)], i.e., a function $S_{|I|}(\cdot)$: $\mathbf{R}_+^{|I|} \mapsto \mathbf{R}_+^1$ satisfying the following definition:

(D.a) [Symmetry] $S_{|I|}(\cdot)$ is symmetric in its $|I|$ arguments;

(D.b) [Homogeneity] $S_{|I|}(\alpha d_1, \alpha d_2, \ldots, \alpha d_{|I|})=\alpha S_{|I|}(d_1, d_2, \ldots, d_{|I|})$ whenever $\alpha \geq 0$;

(D.c) [Triangle Inequal.] $S_{|I|}(d_1+d'_1, d_2+d'_2, \ldots, d_{|I|}+d'_{|I|}) \leq S_{|I|}(d_1, d_2, \ldots, d_{|I|})+S_{|I|}(d'_1, d'_2, \ldots, d'_{|I|})$;

(D.d) [Identity] $S_{|I|}(1, 1, \ldots,1)=1$;

(D.e) [Monotonicity] $S_{|I|}(d_1, d_2, \ldots, d_{|I|}) \leq S_{|I|}(d'_1, d'_2, \ldots, d'_{|I|})$ whenever $d_i \leq d'_i \; \forall i$;

(D.f) [Mean Dominance] $S_{|I|}(d_1, d_2, \ldots, d_{|I|}) \geq \sum_{1 \leq i \leq |I|} d_i / |I|$.

Finally, we standardize $S_{|I|}(\cdot)$ by a multiplicative factor to account for the inherent instability in the e.c.d.f. The boundary estimator $\hat{\Theta}_I$ is defined as the candidate boundary in $\mathfrak{T}_I$ which maximizes the criterion function $D_I(T):=(|\overline{T}|_I / |I|)(|\underline{T}|_I / |I|) \cdot S_{|I|}(d^T_{I1}, d^T_{I2}, \ldots, d^T_{I|I|})$ over all $T \in \mathfrak{T}_I$. Formally, $\hat{\Theta}_I := \text{argmax}_{T \in \mathfrak{T}_I} D_I(T)$.

The intuitive motivation for this estimator is given by Carlstein and Krishnamoorthy (1992) [referred to as C&K from here on]. A literature review comparing $\hat{\Theta}_I$ to other related methods is provided in C&K, as is an extensive discussion of examples in the d=1 and d=2 cases (including the change-point problem, the epidemic-change model, linear bisection of the plane, templates, and Lipschitz boundaries).

1.3 Performance of the Boundary Estimator. In order to assess the performance of $\hat{\Theta}_I$, we must quantify the notion of "distance" between two boundaries (say, T and Θ). Our "distance" measure is the pseudometric $\partial(T,\Theta):=\min\{\lambda(\overline{T} \circ \overline{\Theta}), \lambda(\underline{T} \circ \overline{\Theta})\}$, where $\circ$ denotes set-theoretic symmetric difference, i.e., $(A \circ B):=(A \cap B^c) \cup (A^c \cap B)$, and $\lambda(\cdot)$ is Lebesgue measure over $\mathfrak{U}_d$.

Consider the following set-theoretic regularity conditions on the boundaries:

REGULARITY CONDITION (R.1): [Non-trivial Partitions]

For each $T \in \mathfrak{T}_I$, $0 < \lambda(\overline{T}) < 1$ and $0 < |\overline{T}|_I / |I| < 1$. Also, $0 < \lambda(\overline{\Theta}) < 1$.

REGULARITY CONDITION (R.2): [Richness of $\mathfrak{T}_I$]

For each I, $\exists \, T_I \in \mathfrak{T}_I$ such that the sequence $\{T_I\}$ satisfies: $\partial(\Theta, T_I) \to 0$ as $|I| \to \infty$.

REGULARITY CONDITION (R.3): [Cardinality of $\mathfrak{T}_I$]

For each $\gamma > 0$, $|\mathfrak{T}_I| \cdot \exp\{-\gamma \cdot |I|\} \to 0$ as $|I| \to \infty$.

REGULARITY CONDITION (R.4): [Smoothness of Perimeter]

Denote $\mathfrak{F}_I := \{\bar{\Theta}, \bar{\mathrm{T}}:\ \mathrm{T} \in \mathcal{T}_I\}$ and $\mathcal{P}_I(\mathrm{A}) := \{\mathrm{C} \in \mathcal{C}_I:\ \mathrm{C} \cap \mathrm{A} \neq \phi$ and $\mathrm{C} \cap \mathrm{A}^c \neq \phi\}$, where $\mathcal{C}_I$ is the collection of $|I|$ "rectangular" cells C induced by the grid partition of $\mathcal{U}_d$, and where $\mathrm{A} \subseteq \mathcal{U}_d$. We require $\sup_{A \in \mathfrak{F}_I} \lambda(\mathcal{P}_I(\mathrm{A})) \to 0$ as $|I| \to \infty$.

In C&K, these regularity conditions are seen to be intuitively natural and are explicitly checked for several examples. Moreover, R.1 – R.4 are shown to imply strong consistency of the boundary estimator $\hat{\Theta}_I$, i.e., $\partial(\Theta, \hat{\Theta}_I) \overset{a.s.}{\to} 0$ as $|I| \to \infty$.

1.4 Some Practical Considerations in Boundary Estimation.

Robustness: Condition R.2 requires $\mathcal{T}_I$ to contain *some* candidate boundary T_I that "gets close" to the true Θ. 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 consistency holds. Yet, in practice, the collection $\mathcal{T}_I$ may be misspecified by the user and may be incompatible with the true unknown boundary Θ. Even though R.2 is violated, one still hopes that $\hat{\Theta}_I$ will contain some useful information about Θ. In Section 2, this robustness issue is quantified by comparing the error of $\hat{\Theta}_I$ with the minimal possible error in $\mathcal{T}_I$; theoretical results are given in terms of a.s. asymptotic behavior and in terms of probability of error.

Efficient Computation: Condition R.3 restricts the number of candidate boundaries relative to the sample size, but still allows for extremely large $|\mathcal{T}_I|$; if the user is concerned about violating R.2, then caution should be exercised in attempting to reduce $|\mathcal{T}_I|$. And, for higher dimensional data-sets ($d \geq 2$), $|I|$ tends to be large. Therefore, since the estimator requires calculations over all $i \in I$ and over all $\mathrm{T} \in \mathcal{T}_I$, it behooves us to develop efficient computational schemes. In Section 3, an algorithm is presented which improves dramatically on the naive approach.

Bootstrapping: Although $\hat{\Theta}_I$ is a consistent estimator of Θ (in the sense described above), it is difficult to characterize the sampling variability of the boundary estimator – especially in cases where the boundary is not readily expressible as a parametric function. Therefore, in Section 4, we suggest a graphical bootstrap procedure for visually assessing the sampling variability of $\hat{\Theta}_I$; this procedure was simulated, yielding encouraging results.

2. ROBUSTNESS

2.1 Motivation. Knowledge about the form of the boundary Θ (e.g., rectangular template, Lipschitz function, etc.) has been assumed in the formulation of $\hat{\Theta}_I$. This information is used to construct "rich enough" collections $\mathcal{T}_I$ (see R.2), so that boundaries which are "close" to the true boundary Θ are available as possible estimates. Thus, if the user has knowledge that Θ defines a circle, then $\mathcal{T}_I$ is constructed to include only circles. If further information is available on the *location* of Θ (say the center of the circle is known to be in the left half of $\mathcal{U}_2$), then a *smaller* set $\mathcal{T}_I$ of circles can be constructed which will still satisfy R.2. Smaller $|\mathcal{T}_I|$ means easier computability (see Section 3) and sharper bounds on the error probability (see Theorem 2 of C&K). Therefore there are practical incentives for small $\mathcal{T}_I$. It is, however, possible for $\mathcal{T}_I$ to be *too* small if it violates R.2. This may happen due to overzealousness in reducing the computational burden, or due to incorrect prior information. Thus, a question of practical interest is: What is the price paid if the collections $\mathcal{T}_I$ do not satisfy R.2 ?

We shall explore the case of imperfect knowledge about the *form* of Θ. For example, when $d=1$, if we do not know how many change-points (from F to G and vice-versa) there are, we can incorporate our uncertainty into the model by including in $\mathcal{T}_I$: cases of 1 change-point, 2 change-points (epidemic-change), ... , M change-points. By Proposition 3 of C&K, we will still satisfy R.3, the condition limiting the cardinality of $\mathcal{T}_I$. However, we do need to know M, an upper bound on the true number of change-points, in order to satisfy R.2. Another illustration to keep in mind is the case where the true Θ is a circle, but $\mathcal{T}_I$ contains only squares.

Formally, let $\Theta'_I := \arg\min_{T\in\mathcal{T}_I} \partial(\Theta,T)$ be the element in $\mathcal{T}_I$ closest to Θ. Then $\partial(\Theta'_I,\Theta)$ is the minimal error that we can expect to make in estimating Θ; if the error of our estimator, $\partial(\hat{\Theta}_I,\Theta)$, is of the same order as $\partial(\Theta'_I,\Theta)$ we shall be reasonably satisfied, i.e., $\hat{\Theta}_I$ is "robust". We will study the asymptotic behavior of $\partial(\hat{\Theta}_I,\Theta)$ and compare it to the asymptotic behavior of $\partial(\Theta'_I,\Theta)$. Notice that Θ'_I is not necessarily *unique*: There might well be a class of candidates Θ'_I that are closest in terms of the pseudometric ∂, but each of them might yield a different value of $\partial(\hat{\Theta}_I,\Theta'_I)$. Because of this possible ambiguity the study of $\partial(\hat{\Theta}_I,\Theta'_I)$ is not appropriate here.

2.2 Asymptotic Target Error. Assume that R.1 holds. Given Θ and the candidate family $\{\mathfrak{T}_I : \forall I\}$, define the *asymptotic target error*, η, as follows:

$$\eta := \overline{\lim_{|I|\to\infty}} \min_{T\in\mathfrak{T}_I} \partial(\Theta,\mathrm{T}) = \overline{\lim_{|I|\to\infty}} \partial(\Theta,\Theta_I').$$

We want to study the robustness of $\hat{\Theta}_I$ with respect to violation of R.2. The quantity η is the natural parameter for measuring the severity of the failure of R.2, because:

Proposition 1: **R.2 $\Leftrightarrow$ η=0.**

This also justifies the use of $\overline{\lim}$ rather than $\underline{\lim}$ in η's definition, since the latter could yield zero even if R.2 fails and hence does not adequately quantify the departure from our conditions. We will assess the robustness of $\hat{\Theta}_I$ by comparing $\partial(\hat{\Theta}_I,\Theta)$ to the target η.

Another quantity which arises naturally in the robustness analysis is:

$$\rho(\mathrm{T}) := |\,\lambda(\overline{\mathrm{T}}\cap\overline{\Theta})\lambda(\underline{\mathrm{T}})-\lambda(\underline{\mathrm{T}}\cap\overline{\Theta})\lambda(\overline{\mathrm{T}})\,|\,.$$

The difference $\rho(\Theta)-\rho(\mathrm{T})$ behaves similarly to $\partial(\Theta,\mathrm{T})$, as formalized by:

Proposition 2: **For $\gamma > 0$,**

$$\partial(\Theta,\mathrm{T}) < \gamma \Rightarrow \rho(\Theta)-\rho(\mathrm{T}) < \gamma\,,$$

$$\partial(\Theta,\mathrm{T}) > \gamma \;\Rightarrow\; \rho(\Theta)-\rho(\mathrm{T}) > \sigma\cdot\gamma > 0\,,$$

where σ:=min$\{\lambda(\overline{\Theta}),\lambda(\underline{\Theta})\}$. It follows that:

$$\rho(\Theta)-\rho(\mathrm{T}) \le \partial(\Theta,\mathrm{T}) \le [\rho(\Theta)-\rho(\mathrm{T})]/\sigma.$$

When the partition defined by Θ is "balanced", i.e., $\sigma=\frac{1}{2}$, we find $\partial(\Theta,\mathrm{T})=2[\rho(\Theta)-\rho(\mathrm{T})]$.

[Proofs of results are given in Section 2.4 below.] When R.2 does hold, then

$$\overline{\lim_{|I|\to\infty}} \min_{T\in\mathfrak{T}_I} \{\rho(\Theta)-\rho(\mathrm{T})\} = 0.$$

However, unlike $\partial(\Theta,\mathrm{T})$, the quantity $\rho(\Theta)-\rho(\mathrm{T})$ is *not* a pseudometric; in fact, it is not even symmetric in its arguments. Nevertheless, to analyze the robustness of $\hat{\Theta}_I$, it is useful to define the analog of η for $\rho(\Theta)-\rho(\mathrm{T})$:

$$\eta^* := \overline{\lim_{|I|\to\infty}} \min_{T\in\mathfrak{T}_I} \{\rho(\Theta)-\rho(\mathrm{T})\} = \overline{\lim_{|I|\to\infty}} [\rho(\Theta)-\rho(\Theta_I'')],$$

where $\Theta_I'' := \arg\min_{T\in\mathfrak{T}_I}\{\rho(\Theta)-\rho(\mathrm{T})\}$. The relationship between η and η^* is given by:

Proposition 3: **$0 \le \eta^* \le \eta \le \eta^*/\sigma$.**

Notice that R.2 $\Leftrightarrow$ $\eta=0$ $\Leftrightarrow$ $\eta^*=0$; also, when the partition defined by Θ is balanced, the third inequality becomes an equality.

2.3 **Robustness Results.** Assume now that: $F \neq G$; $\hat{\Theta}_I$ is based on a mean-dominant norm; regularity conditions R.1, R.3, R.4 hold. The main result provides an a.s. bound on $\hat{\Theta}_I$'s error that is linear in the target error.

Theorem 1: **[Robustness]**

$$\eta \leq \overline{\lim_{|I|\to\infty}} \partial(\Theta,\hat{\Theta}_I) \overset{a.s.}{\leq} \eta^*/\sigma\mu ,$$

where $\mu := \mu_F\lambda(\overline{\Theta})+\mu_G\lambda(\underline{\Theta})$ **and** $\mu_F := \int_{-\infty}^{\infty} |F(x)-G(x)|\,dF(x)$, $\mu_G := \int_{-\infty}^{\infty} |F(x)-G(x)|\,dG(x)$.

When the partition defined by Θ is balanced,

$$\eta \leq \overline{\lim_{|I|\to\infty}} \partial(\Theta,\hat{\Theta}_I) \overset{a.s.}{\leq} \eta/\mu.$$

Notice that $\sigma\mu > 0$ (by R.1 and Lemma 7 of C&K). The upper bounds in Theorem 1 show that the robustness of $\hat{\Theta}_I$ is favorably influenced by two intuitively natural factors: (i) *balance* between the amount of data in $\overline{\Theta}$ versus $\underline{\Theta}$, as measured by σ; (ii) *disparity* between the two distributions F and G, as measured by μ.

Example: Let F be the distribution with point mass at the origin, and let G be the *Uniform*[0,1] distribution. Then direct calculations yield $\mu_F+\mu_G=\frac{3}{2}$. Therefore, in the case of a balanced partition $(\overline{\Theta},\underline{\Theta})$, we obtain

$$\eta \leq \overline{\lim_{|I|\to\infty}} \partial(\Theta,\hat{\Theta}_I) \overset{a.s.}{\leq} \tfrac{4}{3}\eta \,.$$

Finally, we bound the probability of $\hat{\Theta}_I$ behaving in a non-robust way (as compared to the r.h.s. of Theorem 1).

Theorem 2: **[Bound on Error Probability]**

For any $0 < \alpha < 1$ and $0 < \varepsilon \leq \varepsilon(\alpha)$,

$$P\{\partial(\Theta,\hat{\Theta}_I) > (\eta^* + \varepsilon)/\sigma\mu(1-\alpha)\} \leq \widetilde{K} \cdot |\mathfrak{T}_I| \cdot \exp\{-K \cdot \varepsilon^2 \cdot |I|\}$$

for $|I|$ sufficiently large, where K and $\widetilde{K}$ are positive constants.

The probability of non-robustness decreases exponentially as a function of sample size. However, this effect is counterbalanced by the number of candidate boundaries considered (i.e., for fixed target error, it is easier to grossly mislead $\hat{\Theta}_I$ when the collection $\mathfrak{T}_I$ is larger). The bound in Theorem 2 explains why the joint growth of $|I|$ and $|\mathfrak{T}_I|$ must be controlled via R.3. Lastly, notice that the bound does not depend on the actual target error, i.e., it is unaffected by the *severity* of the R.2 violation.

2.4 Proofs.

Proof of Proposition 1: Immediate from the definitions. □

Proof of Proposition 2: Note that

$$\rho(\Theta)-\rho(\mathrm{T}) = \min\{\lambda(\underline{\Theta}\cap\overline{\mathrm{T}})\lambda(\overline{\Theta})+\lambda(\overline{\Theta}\cap\underline{\mathrm{T}})\lambda(\underline{\Theta}),\ \ \lambda(\overline{\Theta}\cap\overline{\mathrm{T}})\lambda(\underline{\Theta})+\lambda(\underline{\Theta}\cap\underline{\mathrm{T}})\lambda(\overline{\Theta})\}.$$

Comparing this expression to the definition of $\partial(\Theta,\mathrm{T})$, the first implication is clear. For the second implication, write $\partial(\Theta,\mathrm{T})=\min\{x+x',\ y+y'\}$ and $\rho(\Theta)-\rho(\mathrm{T})=\min\{\Gamma_\Theta(x,x'),\ \Gamma_\Theta(y,y')\}$, where $\Gamma_\Theta(z,z'):=\lambda(\overline{\Theta})z+\lambda(\underline{\Theta})z'$. Then,

$$\partial(\Theta,\mathrm{T}) > \gamma\ \Rightarrow\ x+x' > \gamma \text{ and } y+y' > \gamma\ \Rightarrow\ \sigma(x+x') > \sigma\gamma \text{ and } \sigma(y+y') > \sigma\gamma$$

$$\Rightarrow\ \Gamma_\Theta(x,x') > \sigma\gamma \text{ and } \Gamma_\Theta(y,y') > \sigma\gamma\ \Rightarrow\ \rho(\Theta)-\rho(\mathrm{T}) > \sigma\gamma.$$

The final two assertions of Proposition 2 are now immediate. □

Proof of Proposition 3: Since Θ_I', $\Theta_I'' \in \mathfrak{T}_I$, we have by definition and by Proposition 2 that

$$0 \le \rho(\Theta)-\rho(\Theta_I'') \le \rho(\Theta)-\rho(\Theta_I') \le \partial(\Theta,\Theta_I') \le \partial(\Theta,\Theta_I'') \le [\rho(\Theta)-\rho(\Theta_I'')]/\sigma.$$

Now take $\overline{\lim}_{|I|\to\infty}$ throughout. □

Proof of Theorem 1: For $0<\alpha<1$ and $0<\varepsilon\le\varepsilon(\alpha)$, we have by Theorem 2 and R.3 that

$$\sum_{|I|} \mathbf{P}\{\partial(\Theta,\hat{\Theta}_I) > (\eta^*+\varepsilon)/\sigma\mu(1-\alpha)\} < \infty\ ;$$

hence by the Borel-Cantelli Lemma $\mathbf{P}\{\partial(\Theta,\hat{\Theta}_I) > (\eta^*+\varepsilon)/\sigma\mu(1-\alpha)$ infinitely often $[\,|I|\,]\,\} = 0$. For $\alpha_k=\frac{1}{k}$ and $\varepsilon_k=\min\{\frac{1}{k},\varepsilon(\alpha_k)\}$, k>1, $\exists$ a null set A_k s.t. $\forall\omega\notin A_k$ we have $\partial(\Theta,\hat{\Theta}_I) \le (\eta^*+\varepsilon_k)/\sigma\mu(1-\alpha_k)$ when $|I| \ge \mathrm{i}(\mathrm{k},\omega)$. Thus, for $\omega \notin \bigcup_{k>1} A_k$, $\overline{\lim}_{|I|\to\infty}\partial(\Theta,\hat{\Theta}_I) \le \eta^*/\sigma\mu$. □

Proof of Theorem 2: A series of Lemmas will be presented, leading to Theorem 2.

Lemma 1: Define $\mathfrak{T}_I^*:=\{\overline{\Theta},\ \underline{\Theta};\ \overline{\mathrm{T}},\ \underline{\mathrm{T}},\ \overline{\mathrm{T}}\cap\overline{\Theta},\ \underline{\mathrm{T}}\cap\overline{\Theta},\ \overline{\mathrm{T}}\cap\underline{\Theta},\ \underline{\mathrm{T}}\cap\underline{\Theta}\colon\ \mathrm{T}\in\mathfrak{T}_I\}$. Then:

$$\sup_{A\in\mathfrak{T}^*_I} \left|\lambda(A) - |A|_I/|I|\right| \to 0 \text{ as } |I|\to\infty.$$

Proof: Follows from R.4, using the same argument as in the proof of C&K's Lemma 1. □

Define the following notation:

$$\bar{\eta}_T(x):=[\lambda(\bar{T}\cap\bar{\Theta})F(x)+\lambda(\bar{T}\cap\underline{\Theta})G(x)]/\lambda(\bar{T}),\ \ \underline{\eta}_T(x):=[\lambda(\underline{T}\cap\bar{\Theta})F(x)+\lambda(\underline{T}\cap\underline{\Theta})G(x)]/\lambda(\underline{T}),$$

$$\delta^T_{Ii} := |\bar{\eta}_T(X^I_i)-\underline{\eta}_T(X^I_i)|,\ \ \Delta_I(T):=\lambda(\bar{T})\lambda(\underline{T})S_{|I|}(\delta^T_{Ii}: i\in I).$$

Lemma 2: For $|I| \geq N_1(\varepsilon)$,

$$\mathbf{P}\{\sup_{T\in\mathfrak{T}_I}|D_I(T)-\Delta_I(T)| > \varepsilon\} \leq K_1|\mathfrak{T}_I|\exp\{-K_2\cdot\varepsilon^2\cdot|I|\}.$$

Proof: Follows from Lemma 1 and Dvoretzky, Kiefer & Wolfowitz (1956), using the same argument as in the proof of C&K's Lemma 2 (with δ=0). □

Lemma 3: We can write $\Delta_I(T)=\rho(T)\cdot S_{|I|}(\delta^{\Theta}_{Ii}: i\in I)$.

Proof: Exactly as in the proof of C&K's Lemma 3. □

Lemma 4: For every $T\in\mathfrak{T}_I$, we have $\Delta_I(T)\leq\Delta_I(\Theta)$.

Proof: Follows from Lemma 3, exactly as in the proof of C&K's Lemma 4. □

Lemma 5: For $|I| \geq N_2(\varepsilon)$,

$$\mathbf{P}\{|\Delta_I(\hat{\Theta}_I)-\Delta_I(\Theta)| > \eta^*+\varepsilon\} \leq K_3|\mathfrak{T}_I|\exp\{-K_4\cdot\varepsilon^2\cdot|I|\}.$$

Proof: By definition $\Theta''_I\in\mathfrak{T}_I$ is the maximizer of $\rho(\cdot)$, and hence, by Lemma 3, it is the maximizer of $\Delta_I(\cdot)$, over $\mathfrak{T}_I$. Then by Lemma 4 we have $\Delta_I(\Theta)\geq\Delta_I(\Theta''_I)\geq\Delta_I(T)\ \forall T\in\mathfrak{T}_I$, and by definition we have $D_I(\hat{\Theta}_I)\geq D_I(T)\ \forall T\in\mathfrak{T}_I$. Now,

$$(\dagger)\qquad |\Delta_I(\hat{\Theta}_I)-\Delta_I(\Theta)| \leq |\Delta_I(\hat{\Theta}_I)-D_I(\hat{\Theta}_I)|+|D_I(\hat{\Theta}_I)-\Delta_I(\Theta''_I)|+|\Delta_I(\Theta''_I)-\Delta_I(\Theta)|.$$

The second modulus on the r.h.s. is bounded by $\sup_{T\in\mathfrak{T}_I}|D_I(T)-\Delta_I(T)|$, because either $D_I(\hat{\Theta}_I)\geq\Delta_I(\Theta''_I)\geq\Delta_I(\hat{\Theta}_I)$ or $\Delta_I(\Theta''_I)\geq D_I(\hat{\Theta}_I)\geq D_I(\Theta''_I)$. The same bound applies to the first modulus on the r.h.s. of (†). Using Lemmas 4 and 3, and properties (D.e) & (D.d) of $S_{|I|}$, the third modulus on the r.h.s. is $\Delta_I(\Theta)-\Delta_I(\Theta''_I) = [\rho(\Theta)-\rho(\Theta''_I)]S_{|I|}(\delta^{\Theta}_{Ii}: i\in I)\leq\rho(\Theta)-\rho(\Theta''_I)$. Therefore,

$$|\Delta_I(\hat{\Theta}_I)-\Delta_I(\Theta)| \leq 2\cdot\sup_{T\in\mathfrak{T}_I}|D_I(T)-\Delta_I(T)|+\rho(\Theta)-\rho(\Theta''_I).$$

For $|I|$ sufficiently large, by the definition of η^*, we have the *deterministic* bound $\rho(\Theta)-\rho(\Theta''_I)<\eta^*+\varepsilon/3$. Now combine this with the probabilistic bound from Lemma 2. □

To prove Theorem 2, begin by applying Proposition 2, Lemma 4, Lemma 3, and property (D.f) of $S_{/I/}$, obtaining:

$$\partial(\Theta,T) > \gamma \;\Rightarrow\; |\Delta_I(\Theta)-\Delta_I(T)| = [\rho(\Theta)-\rho(T)]S_{/I/}(\delta^{\Theta}_{Ii}: i\in I) > \sigma\gamma\delta^{\Theta}_I,$$

where $\delta^{\Theta}_I := \sum_{i\epsilon I}\delta^{\Theta}_{Ii}/|I|$. Thus,

$$\mathbf{P}\{\partial(\Theta,\hat{\Theta}_I) > (\eta^*+\varepsilon)/\sigma\mu(1-\alpha)\} \;\le\; \mathbf{P}\{|\Delta_I(\Theta)-\Delta_I(\hat{\Theta}_I)| > (\eta^*+\varepsilon)\delta^{\Theta}_I/\mu(1-\alpha)\}$$
$$\le\; \mathbf{P}\{|\Delta_I(\Theta)-\Delta_I(\hat{\Theta}_I)| > \eta^*+\varepsilon\} \;+\; \mathbf{P}\{\delta^{\Theta}_I < \mu(1-\alpha)\}.$$

The first probability on the r.h.s. is immediately handled by Lemma 5. The second probability on the r.h.s. is bounded by $\mathbf{P}\{|\delta^{\Theta}_I-\mu| > \mu\alpha\}$. Denote $\bar{\delta}^{\Theta}_I := \sum_{i\epsilon\bar{\Theta}}\delta^{\Theta}_{Ii}/|\bar{\Theta}|_I$ and $\underline{\delta}^{\Theta}_I := \sum_{i\epsilon\underline{\Theta}}\delta^{\Theta}_{Ii}/|\underline{\Theta}|_I$, so that $\delta^{\Theta}_I = \bar{\delta}^{\Theta}_I(|\bar{\Theta}|_I/|I|) + \underline{\delta}^{\Theta}_I(|\underline{\Theta}|_I/|I|)$. Notice that:

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

The first and third summands on the r.h.s. are handled using Lemma 1. Consider the second summand on the r.h.s. (a similar argument holds for the fourth). By equation (2.3) of Hoeffding (1963), we have $\mathbf{P}\{|\bar{\delta}^{\Theta}_I-\mu_F| > \mu\alpha/4\} \;\le\; 2\cdot\exp\{-c_\alpha|\bar{\Theta}|_I\}$. Lemma 1 ensures that $|\bar{\Theta}|_I$ eventually exceeds $|I|\,\lambda(\bar{\Theta})/2$ (say); therefore the bound from Hoeffding (1963) can be combined with the earlier bound from Lemma 5, for $\varepsilon > 0$ sufficiently small (given α). □

3. EFFICIENT COMPUTATION

To calculate the boundary estimator $\hat{\Theta}_I$ based on data $\{X^I_i: i\in I\}$, we need to calculate the criterion function $D_I(T)$ for each candidate $T\in\mathcal{T}_I$. The number of simple computations C_I required to calculate $\hat{\Theta}_I$ is essentially

$$C_I \approx |\mathcal{T}_I| \times \#\{\text{computations for calculating } D_I(T) \text{ on a single } T\}.$$

The number of computations for calculating a single $D_I(T)$ seems on first thought to be of the order $|I|^2$, since we have to calculate $|I|$ distinct d^T_{Ii}s, and each d^T_{Ii} requires us to find the rank of X^I_i $(i\in I)$ relative to the subsets $\{X^I_j: j\in\bar{T}\}$ and $\{X^I_j: j\in\underline{T}\}$. This would yield

$$C^0_I \approx |\mathcal{T}_I| \times |I|^2.$$

Algorithm: Notice that the calculation of the d^T_{Ii}s involves ranking the X^I_is *within each individual candidate region,* $\bar{T}$ *and* $\underline{T}$. For fixed T, the $\bar{h}^I_T(X^I_i)$s $(1\le i\le |I|)$ can be calculated from an ordered subsequence of the ordered sequence $\{X^I_{(r)}: 1\le r\le |I|\}$; similarly for the $\underline{h}^I_T(X^I_i)$s. Thus

$\hat{\Theta}_I$ is a function only of $\left\{R(i), \mathbb{I}\{i \in \overline{T}\}:\ i \in I,\ T \in \mathcal{T}_I\right\}$, where $R(i)$ is the rank of X_i^I in the ordered sequence. The suggested computational algorithm exploits this structure.

First sort the X_i^Is so that $X_{(1)}^I \leq X_{(2)}^I \leq \cdots \leq X_{(r)}^I \leq \cdots \leq X_{(|I|-1)}^I \leq X_{(|I|)}^I$. Since $S_{|I|}(\cdot)$ is symmetric [property (D.a)], its arguments d_{Ii}^T may be entered according to rank order (r) rather than grid location i. The rank of $X_{(r)}^I$ is r, and its original grid index is stored in INDEX(r). Now $\hat{\Theta}_I$ is a function only of $\left\{\mathbb{I}\{\text{INDEX}(r) \in \overline{T}\}:\ 1 \leq r \leq |I|,\ T \in \mathcal{T}_I\right\}$. Sorting of the X_i^Is and calculation of the vector INDEX(r) together take at most order $|I|^2$ computations.

The collection of $\bar{h}_T^I(X_{(r)}^I)$s (and $\underline{h}_T^I(X_{(r)}^I)$s) is now calculated recursively, as follows. The quantity $|\overline{T}|_I \bar{h}_T^I(X_{(1)}^I)$ is 1 if the the smallest observation $X_{(1)}^I$ comes from $\overline{T}$, and 0 if it does not. Given $|\overline{T}|_I \bar{h}_T^I(X_{(r)}^I)$, the quantity $|\overline{T}|_I \bar{h}_T^I(X_{(r+1)}^I)$ is one more than $|\overline{T}|_I \bar{h}_T^I(X_{(r)}^I)$ if $X_{(r+1)}^I$ comes from $\overline{T}$, and the same as $|\overline{T}|_I \bar{h}_T^I(X_{(r)}^I)$ if it does not. Formally,

$$|\overline{T}|_I \bar{h}_T^I(X_{(1)}^I) = \mathbb{I}\{\text{INDEX}(1) \in \overline{T}\},$$

$$|\overline{T}|_I \bar{h}_T^I(X_{(r+1)}^I) = |\overline{T}|_I \bar{h}_T^I(X_{(r)}^I) + \mathbb{I}\{\text{INDEX}(r+1) \in \overline{T}\}, \text{ for } r = 1, 2, \ldots, |I| - 1.$$

Also, $|\underline{T}|_I \underline{h}_T^I(X_{(r)}^I) = r - |\overline{T}|_I \bar{h}_T^I(X_{(r)}^I)$. Although the symbols $|\underline{T}|_I$, $|\overline{T}|_I$, and $X_{(r)}^I$ are used in the explanation above, the recursive *calculations* (r.h.s.) depend only on $\mathbb{I}\{\text{INDEX}(r) \in \overline{T}\}$. The numerical values of $|\overline{T}|_I$ and $|\underline{T}|_I$ are obtained "for free", since $|\overline{T}|_I = |\overline{T}|_I \bar{h}_T^I(X_{(|I|)}^I)$ and $|\underline{T}|_I = |I| - |\overline{T}|_I$. Thus, for fixed $T \in \mathcal{T}_I$, calculation of $D_I(T)$ requires us to go through $\left\{\mathbb{I}\{\text{INDEX}(r) \in \overline{T}\}:\ 1 \leq r \leq |I|\right\}$ exactly once.

Combining both phases of the algorithm, we have

$$C_I^* \approx |I|^2 + |\mathcal{T}_I| \times |I|,$$

which is an order of magnitude less than the initial naive approach, i.e.,

$$C_I^*/C_I^0 \approx |\mathcal{T}_I|^{-1} + |I|^{-1}.$$

Examples: In most applications, $|\mathcal{T}_I|$ is a function of $|I|$. Table 1 gives the approximate number of computations C_I^* for several specific examples. [See C&K for precise descriptions of $\mathcal{T}_I$ in each case.]

TABLE 1: Number of Simple Computations.

Boundary Type Θ,T	*Dimension* d	*# of Candidates* $\vert\mathcal{T}_I\vert$	*# of Computations* C_I^*
Change-Point	1	$\vert I\vert$	$\vert I\vert^2$
Epidemic-Change	1	$\vert I\vert^2$	$\vert I\vert^3$
Template (Oriented Rectangle)	2	$\vert I\vert^2$	$\vert I\vert^3$

4. BOOTSTRAPPING

The boundary estimator $\hat{\Theta}_I$ is applicable in a broad range of situations: there are no restrictions on the distributions F and G; the class of boundaries (Θ, $\mathcal{T}_I$) is quite general. In keeping with this omnibus spirit, it is appropriate to use Efron's (1979) bootstrap for assessing the sampling variability of $\hat{\Theta}_I$. Dümbgen (1991) has studied bootstrap procedures for a related class of estimators in the 1-dimensional single change-point case. We now propose a graphical bootstrap procedure that is particularly useful in the planar (d=2) case.

The true underlying probability structure is determined by the unknown triplet (F, G; Θ). Having estimated Θ with $\hat{\Theta}_I$, the e.c.d.f.s constructed from the two regions induced by $\hat{\Theta}_I$ provide estimates of F and G:

$$\hat{F}_I(x) := \sum_{i\in\hat{\overline{\Theta}}_I} \mathbb{I}\{X_i^I \le x\}/|\hat{\overline{\Theta}}_I|_I ,$$

$$\hat{G}_I(x) := \sum_{i\in\hat{\underline{\Theta}}_I} \mathbb{I}\{X_i^I \le x\}/|\hat{\underline{\Theta}}_I|_I .$$

We now have the estimated probability structure ($\hat{F}_I$, $\hat{G}_I$; $\hat{\Theta}_I$), from which bootstrap samples can be drawn. Generate $\{X_i^*: i\in\hat{\overline{\Theta}}_I\}$ as i.i.d. observations from $\hat{F}_I$, and $\{X_i^*: i\in\hat{\underline{\Theta}}_I\}$ as i.i.d. observations from $\hat{G}_I$. From this bootstrap sample $\{X_i^*: i\in I\}$, calculate $\hat{\Theta}_I^*$ via the same formula as $\hat{\Theta}_I$. By repeating this process, say, B times, we obtain bootstrap realizations $\{\hat{\Theta}_{I1}^*, \hat{\Theta}_{I2}^*, \ldots, \hat{\Theta}_{IB}^*\}$ from ($\hat{F}_I$, $\hat{G}_I$; $\hat{\Theta}_I$). These bootstrap realizations $\hat{\Theta}_{Ik}^*$ are used to model the sampling variability of $\hat{\Theta}_I$.

Each $\hat{\Theta}^*_{Ik}$ determines a partition of the grid I. If each of these $\hat{\Theta}^*_{Ik}$s were a number, we could look at the variance, percentiles, and histogram of these numbers in order to assess the variability of $\hat{\Theta}_I$. But in some situations (e.g., Lipschitz curves [see C&K]), the boundaries are not readily expressible in terms of a fixed finite set of parameters. Moreover, in the 2-dimensional case it is natural to want a direct visual representation of $\hat{\Theta}_I$'s variability in $\mathfrak{U}_2$. For this purpose we introduce the *(1 − α)100% bootstrap indifference zone*, $Z^*_{(1-\alpha)}$, which is defined as follows:

$$p^*_i := \sum_{k=1}^{B} I\{i \in \hat{\bar{\Theta}}^*_{Ik}\}/B \ , \quad i \in I \ ;$$

$$Z^*_{(1-\alpha)} := \{i \in I\colon \alpha/2 \le p^*_i \le 1 - \alpha/2\} \ ;$$

i.e., $Z^*_{(1-\alpha)}$ excludes those points $i \in I$ which either belong to more than $(1-\alpha/2)100\%$ of the $\hat{\bar{\Theta}}^*_{Ik}$s or belong to more than $(1-\alpha/2)100\%$ of the $\hat{\underline{\Theta}}^*_{Ik}$s.

The zone $Z^*_{(1-\alpha)}$ is conditional on the estimated probability structure $(\hat{F}_I, \hat{G}_I; \hat{\Theta}_I)$. A large or wide $Z^*_{(1-\alpha)}$ typically indicates high spatial variability amongst the $\hat{\Theta}^*_{Ik}$s, and similarly a non-existent or narrow $Z^*_{(1-\alpha)}$ means that the $\hat{\Theta}^*_{Ik}$s have little spatial variability. Furthermore, if the boundary $\hat{\Theta}_I$ lies completely inside $Z^*_{(1-\alpha)}$, this suggests that the spatial "bias" of $\hat{\Theta}^*_I$ (as an "estimator" of $\hat{\Theta}_I$) is small. Finally, by analogy, we pass these conclusions about the bootstrap distribution of $\hat{\Theta}^*_I$ (variability and bias) back to the sampling distribution of $\hat{\Theta}_I$.

In practice, the entire bootstrap procedure described above is implemented on a single set of data $\{X^I_i\colon i \in I\}$.

Simulations: We simulated the entire bootstrap procedure 100 times for the case of a linear bisecting boundary [see C&K] in a 2-dimensional 15×15 grid; the true boundary Θ connects the points $(0.67, 0.00)$ and $(0.40, 1.00)$ in $\mathfrak{U}_2$. For each of these 100 simulations, we obtained first $\hat{\Theta}_I$, then a picture of $Z^*_{(.90)}$, and the proportion (q^*) of grid points falling inside $Z^*_{(.90)}$; the bootstrap procedure used B=1000. Figures 1.a, 1.b, 1.c show the results for three of the 100 simulations; these were selected to represent the performance of the bootstrap under "good", "average", and "poor" estimates $\hat{\Theta}_I$. Each figure illustrates $\hat{\Theta}_I$, $Z^*_{(.90)}$, and q^*.

As a basis of comparison for judging the bootstrap, we generated 1000 realizations of $\hat{\Theta}_I$ from the true (F, G; Θ). The true sampling variability of $\hat{\Theta}_I$ is thus measured by the analogous quantities

$Z_{(.90)}$ and q, which are shown in Figure 2 (along with Θ).

Notice that the bootstrapped zones $Z^*_{(.90)}$ in Figure 1 are reasonably comparable in shape and size (q^*) to the true $Z_{(.90)}$ and q in Figure 2. Also note that each $Z^*_{(.90)}$ in Figure 1 contains its $\hat{\Theta}_I$, as does $Z_{(.90)}$ contain Θ in Figure 2. So it seems that the proposed graphical bootstrap procedure is useful for visually assessing the sampling behavior of $\hat{\Theta}_I$.

<u>FIGURE 1:</u> Simulated Bootstrap Indifference Zones.

$Z^*_{(.90)}$ *is indicated by:* . . . . . $\hat{\Theta}_I$ *is indicated by:* ______

<u>Figure 1.a</u>
(q^*=.16)

```
1 1 1 1 . . . 0 0 0 0 0 0 0 0
1 1 1 1 . . . 0 0 0 0 0 0 0 0
1 1 1 1 1 . . 0 0 0 0 0 0 0 0
1 1 1 1 1 . . 0 0 0 0 0 0 0 0
1 1 1 1 1 .   0 0 0 0 0 0 0 0
1 1 1 1 1 1 . . 0 0 0 0 0 0 0
1 1 1 1 1 1 . . 0 0 0 0 0 0 0
1 1 1 1 1 1 .   0 0 0 0 0 0 0
1 1 1 1 1 1 . . . 0 0 0 0 0 0
1 1 1 1 1 1 1 . . 0 0 0 0 0 0
1 1 1 1 1 1 1 . . 0 0 0 0 0 0
1 1 1 1 1 1 1 .   0 0 0 0 0 0
1 1 1 1 1 1 1 . . . 0 0 0 0 0
1 1 1 1 1 1 1 . . . 0 0 0 0 0
1 1 1 1 1 1 1 . .   0 0 0 0 0
```

<u>Figure 1.b</u>
(q^*=.28)

```
1 1 1 1 . . . . . . 0 0 0 0 0
1 1 1 1 . . . . . 0 0 0 0 0 0
1 1 1 1 . . . . . 0 0 0 0 0 0
1 1 1 1 1 . . . . 0 0 0 0 0 0
1 1 1 1 1 . . . . 0 0 0 0 0 0
1 1 1 1 1 . . . . 0 0 0 0 0 0
1 1 1 1 1 1 . . . 0 0 0 0 0 0
1 1 1 1 1 1 .   . 0 0 0 0 0 0
1 1 1 1 1 1 . . . 0 0 0 0 0 0
1 1 1 1 1 1 . . . 0 0 0 0 0 0
1 1 1 1 1 1 . . . 0 0 0 0 0 0
1 1 1 1 1 . . . . 0 0 0 0 0 0
1 1 1 1 1 . . . . . 0 0 0 0 0
1 1 1 1 1 . . . . . 0 0 0 0 0
1 1 1 1 1 . . .   . 0 0 0 0 0
```

Figure 1.c
(q^*=.22)

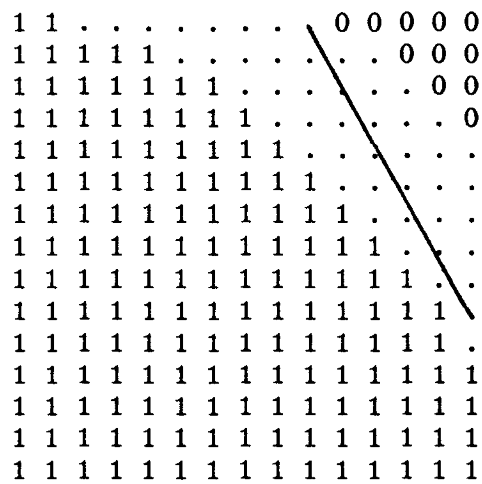
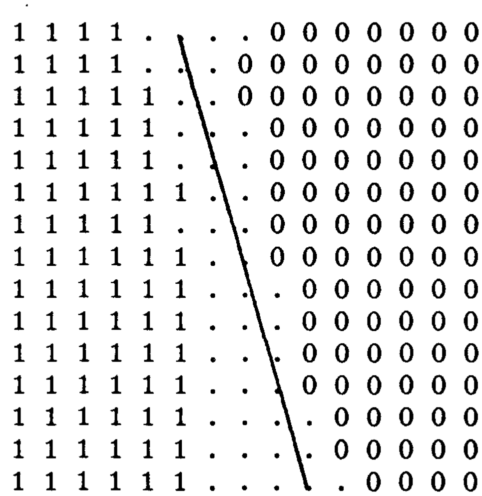


FIGURE 2: Simulated Behavior of $\hat{\Theta}_I$.

$Z_{(.90)}$ *is indicated by:* . . . . . Θ *is indicated by:* ______

(q=.21)



REFERENCES

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

Carlstein, E. and Krishnamoorthy, C. (1992), "Boundary Estimation," *Journal of the American Statistical Association*, 87, 430-438.

Dümbgen, L. (1991), "The Asymptotic Behavior of Some Nonparametric Change-Point Estimators," *Annals of Statistics*, 19, 1471-1495.

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

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.

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