The Annals of Statistics
1994, Vol. 22, No. 2, 824–834

LINEAR ESTIMATORS IN CHANGE POINT PROBLEMS[1]

BY J. A. HARTIGAN

Yale University

Observations X_i are uncorrelated with means $\theta_i, i = 1, \dots, n$, and variances 1. The linear estimators $\widehat{\theta} = T\mathbf{X}$, for some $n \times n$ matrix T, are widely used in smoothing problems, where it is assumed that neighbouring parameter values are similar. The smoothness assumption is violated in change point problems, where neighbouring parameter values are equal, except at some unspecified change points where there are jumps of unknown size from one parameter value to the next. In the case of a single change point in one dimension, for any linear estimator, the expected sum of squared errors between estimates and parameters is of order $\sqrt{n}$ for some choice of parameters, compared to order 1 for the least squares estimate. We show similar results for *adaptive* shift estimators, in which the linear estimator uses a kernel estimated from the data. Finally, for a change point problem in two dimensions, the expected sum of squared errors is of order $n^{3/4}$.

1. Introduction. The vector $\mathbf{X}$ of n observations has expectation $\boldsymbol{\theta}$. The parameter vector is estimated by a linear function of the observations, $\widehat{\boldsymbol{\theta}} = T\mathbf{X}$, using some $n \times n$ matrix T. Such estimators are simple to compute and analyse, and have been studied in an immense literature: see Buja, Hastie and Tibshirani (1989) for a review of linear smoothers, Cleveland (1979) for smoothing in nonparametric regression, Craven and Wahba (1979) and Silverman (1985) for spline smoothing. In smoothing problems, it is usually assumed that neighbouring parameter values are close to each other; for example, Rice and Rosenblatt (1983) assume that the parameter values are taken at the points $0, 1/n, \dots, 1$ from a continuously differentiable function on the unit interval. Another type of justification is possible in a Bayesian framework: if it is assumed that $\mathbf{X}$ and $\boldsymbol{\theta}$ are joint normal, then the posterior mean of $\boldsymbol{\theta}$ given X is linear in X; such models are considered for splines by Wahba (1978) and for image restoration of smoothly varying pictures by Besag (1986).

In this paper, we evaluate the performance of linear estimators for some change point problems where the parameter values do not vary smoothly, and where we would expect the linear estimates to do poorly. Intuitively, if a parameter value near a discontinuity is estimated by a weighted sum of observations which includes observations from both sides of the discontinuity, then bias will be introduced by the observations on the other side of the discontinuity to the parameter value of interest. Thus the discontinuity is blurred in the estimated parameters.

Received July 1991; revised September 1993.
[1]Research partially supported by NSF Grant DMS-86-17919.
AMS 1991 *subject classifications.* Primary 62F10; secondary 62G07.
Key words and phrases. Linear estimators, change point problems.

We will assume that the errors $X_i - \theta_i$ are uncorrelated with means 0 and variances 1. The expected sum of squared errors for the estimator T at parameter value $\boldsymbol{\theta}$ is

$$M(T,\boldsymbol{\theta}) = \mathrm{tr}\left[T'T + (T-I)'(T-I)\boldsymbol{\theta}\boldsymbol{\theta}'\right].$$

For each Lebesgue measurable set S of parameter vectors, define the minimax risk $M(S) = \inf_T \sup_{\boldsymbol{\theta} \in S} M(T,\boldsymbol{\theta})$.

In the following, we will consider lower bounds for the minimax risk in three cases. In the first case, we consider a *circular* change point problem on an even number of data points $n = 2m$; the parameter values are equal to $\frac{1}{2}\Delta$ for a set of m contiguous indices, and equal to $-\frac{1}{2}\Delta$ on the remaining indices. The lower bound for the risk of linear estimators is of order $\Delta n^{1/2}$, although least squares in this case has risk of order 1. The same order bound applies to change point problems in which the first J parameter values are $\frac{1}{2}\Delta$, and the last $n-J$ parameter values are $-\frac{1}{2}\Delta$. If there are k change points, where $kn^{-1/2} \to 0$, I expect the lower bound to be order $k\Delta n^{1/2}$.

In the second case, the set of parameter values is again circular, but the linear estimator is assumed to be a *shift* estimator: $T_{ij} = f(|i-j|)$, where $f(x) = f(n-x)$. The lower bound remains of order $n^{1/2}$ even when the shift estimator is *adaptive*: the function f is chosen based on the data X and the parameter $\boldsymbol{\theta}$ to give the smallest sum of squared errors for each particular pair $X, \boldsymbol{\theta}$. (This is not a choice available in practice, but it provides a lower bound for all adaptive methods.)

In the third case, we examine linear estimators for image segmentation problems, in which the $n \times n$ torus is divided into four $\frac{1}{2}n \times \frac{1}{2}n$ blocks, in each of which the parameter values are equal. In this case, the minimax risk is at least of order $n^{3/2}$.

The above bounds are comparable to the results of Van Eeden (1985) on kernel density estimation when there are discontinuities in the true density, which give the asymptotic integrated mean squared error $O(n^{-1/2})$.

We have shown that linear estimators may perform relatively badly when discontinuities are present in the parameters. Some authors have considered modifications of linear estimators to allow for the presence of such discontinuities. McDonald and Owen (1986) identify change points using local least squares fits, and use linear smoothers within intervals between the change points. Müller (1992) considers classes of one-sided kernels and infers the presence of change points when there are large differences between the estimates based on right-sided and left-sided kernel estimates at the change point. He develops conditions under which the change points can be consistently estimated.

2. The change point problem. A one-dimensional change point problem has parameter values that are constant in *k blocks*

$$\theta_1 = \theta_2 = \cdots = \theta_{i_1}, \qquad \theta_{i_1+1} = \cdots = \theta_{i_2}, \ldots, \qquad \theta_{i_{k-1}} = \cdots = \theta_n.$$

We will consider initially the *circular* change point problem where the parameter values are of form

$$-\frac{1}{2}\Delta, -\frac{1}{2}\Delta, \dots, -\frac{1}{2}\Delta, \qquad \frac{1}{2}\Delta, \frac{1}{2}\Delta, \dots, \frac{1}{2}\Delta, \qquad -\frac{1}{2}\Delta, \dots, -\frac{1}{2}\Delta,$$

where $n = 2m$, and a block of m $\frac{1}{2}\Delta$'s is preceded and followed by a total of m $-\frac{1}{2}\Delta$'s. We can think of the n indices as being equally spaced on the circle, and the parameter values are equal within each of two contiguous blocks on the circle, and of opposite sign in the two blocks. This class of change point problems is susceptible to Fourier analysis. It will be convenient to extend the X and θ sequences by the "wrap-around" rule $X_{i+n} = X_i, \theta_{i+n} = \theta_i$.

For any distribution P over the set S, let $\Sigma = \int \boldsymbol{\theta}\boldsymbol{\theta}' dP(\boldsymbol{\theta}), T_P = \Sigma(I + \Sigma)^{-1}$. Then $T_P\mathbf{X}$ is a linear estimator of minimum Bayes risk corresponding to P, and that risk bounds the minimax risk from below:

$$\begin{aligned}
\operatorname{tr}\left[\Sigma(I+\Sigma)^{-1}\right] &= \int M(T_P, \boldsymbol{\theta})\, dP(\boldsymbol{\theta}) \\
&\leq \int M(T, \boldsymbol{\theta})\, dP(\boldsymbol{\theta}) \\
&\leq \sup_{\boldsymbol{\theta} \in S} M(T, \boldsymbol{\theta}), \\
\operatorname{tr}\left[\Sigma(I+\Sigma)^{-1}\right] &\leq \inf_T \sup_{\boldsymbol{\theta} \in S} M(T, \boldsymbol{\theta}).
\end{aligned}$$

Indeed, if Σ can be chosen so that the corresponding Bayes estimator $T_P X$ has constant risk $M(T_P, \boldsymbol{\theta})$ over $\boldsymbol{\theta}$ in S, then T_P is a minimax estimator. Our technique will be to seek distributions P for which the minimum Bayes risk is as large as possible; this will give us the largest possible lower bound for the minimax risk.

THEOREM 1. *Suppose that* $\mathbf{X}$ *has mean* $\boldsymbol{\theta}$ *and variance–covariance matrix* I. *Let* $n = 2m$ *be even. Let* S *consist of the change point parameters* $\boldsymbol{\theta}^J$, $1 \leq J \leq n$:

$$\theta_i{}^J = \begin{cases} +\frac{1}{2}\Delta, & \text{for } J < i \leq J + m, \\ -\frac{1}{2}\Delta, & \text{for } J + m < i \leq J + n. \end{cases}$$

For each linear estimator $\widehat{\boldsymbol{\theta}} = T\mathbf{X}$,

$$\sup_{\boldsymbol{\theta} \in S} M(T, \boldsymbol{\theta}) > \frac{\frac{1}{2}n}{\left(n/\Delta^2 + 1\right)^{1/2}} - 1.$$

PROOF. Each of the $\boldsymbol{\theta}^J$'s contains a block of m consecutive equal parameter values, between two blocks of equal parameter values of the opposite sign. Consider the prior distribution that selects each $\boldsymbol{\theta}^J$ with probability $1/n$.

Under this distribution, the mean value of $\boldsymbol{\theta}$ is zero and the variance–covariance matrix is

$$\sigma_{ij} = \begin{cases} \Delta^2\left[\dfrac{1}{4} - \dfrac{|i-j|}{n}\right], & \text{for } |i-j| \le m, \\ \Delta^2\left[\dfrac{1}{4} - \dfrac{n-|i-j|}{n}\right], & \text{for } |i-j| \ge m. \end{cases}$$

We have chosen the distribution of $\boldsymbol{\theta}$ to be stationary under shifts modulo n; the covariance matrix is a *circulant*, having eigenvectors

$$e^J\colon\ e_i^J = \exp\left[\sqrt{-1}w(i-1)J\right], \qquad 1 \le i, J \le n, \text{ where } w = 2\pi/n.$$

The corresponding eigenvalues are zero for J even, and $2\Delta^2/[n(1-\cos(Jw))]$ for J odd.

The Bayes estimator has the same mean squared error for each choice of parameter vector $\boldsymbol{\theta}^J$. The minimum Bayes and minimax risk is

$$B = \sum_{J \text{ odd}, J<n} \left[1 + \frac{n(1-\cos Jw)}{2\Delta^2}\right]^{-1}.$$

The function $f(x) = [1 + n(1 - \cos xw)/2\Delta^2]^{-1}$ is nonnegative, decreasing for $0 \le x \le m$, and satisfies $f(n-x) = f(x)$. Note also that $f(0) = 1$. Thus

$$\int_0^n f(x)\,dx - 2 \le 2 \sum_{J \text{ odd},\ J<n} f(J) < \int_0^n f(x)\,dx + 2.$$

From the integral $\int_0^n f(x)\,dx = n/(n/\Delta^2 + 1)^{1/2}$, $|B - \frac{1}{2}n/(n/\Delta^2+1)^{1/2}| \le 1$. For any particular T, $\sup_{\theta \in S} M(T, \boldsymbol{\theta}) \ge B$, so the desired inequality follows. This concludes the proof. □

While most of the eigenvalues are of order n^{-1}, about $\sqrt{n}$ of them are of order 1, and these force the mean squared error to be of order $\sqrt{n}$. The corresponding Bayes estimate $T\mathbf{X}$ behaves as follows: each row of T contains order $\sqrt{n}$ elements that are not negligible; away from the change points, θ_i is estimated with standard error of order $n^{-1/4}$, adding up to a mean squared error of order $nn^{-1/2} = \sqrt{n}$ over the $O(n)$ points. For $\sqrt{n}$ parameter values near the change points, the nonnegligible weights in T will produce estimates with standard errors of order 1, again adding up to a mean squared error of order $\sqrt{n}$.

The inequality for the circular change point problem may be applied to the problem where there is a single change point with a jump of size Δ; that is, S consists of the set of parameter vectors $\boldsymbol{\theta}^J, -\boldsymbol{\theta}^J$, $1 \le J \le n$, where $\boldsymbol{\theta}_i{}^J = -\frac{1}{2}\Delta$, $1 \le i \le J$, $\boldsymbol{\theta}_i{}^J = \frac{1}{2}\Delta$, $J < i \le n$.

Define S^* to consist of parameter vectors $\boldsymbol{\theta}^*, \boldsymbol{\theta} \in S$, where $\boldsymbol{\theta}^*$ is a parameter vector of length $2n$: $\theta_i^* = \theta_i$, $1 \le i \le n$, $\theta_i^* = -\theta_{i-n}$, $n < i \le 2n$. Note that S^* defines a circular change point problem as considered in Theorem 1.

For any $n \times n$ matrix T, define T^*, a $2n \times 2n$ matrix corresponding to T by

$$T^*_{ij} = \begin{cases} T_{ij}, & 1 \le i, j \le n, \\ T_{i-n, j-n}, & n < i, j \le 2n, \\ 0, & i \le n, j > n, \\ 0, & j \le n, i > n. \end{cases}$$

Using T^* is equivalent to applying T separately to the sequence $X_1, X_2, \dots, X_n$ and to the sequence $X_{n+1}, \dots, X_{2n}$. Then $M(T^*, \theta^*) = 2M(T, \theta)$, so that

$$\sup_{\theta \in S} M(T, \theta) \ge \frac{\frac{1}{2}n}{(1 + 2n/\Delta^2)^{1/2}} - 1.$$

3. Least squares estimates of change point parameters. Consider the change point model in which neighbouring parameter values are equal except for a single jump of size Δ. Define $S_j = \sum_{i=1}^{j} X_i$. The least squares estimate at a change point j is $\widehat{\theta}^j$:

$$\widehat{\theta}_i^j = \begin{cases} \dfrac{S_j}{j}, & 1 \le i \le j, \\ \dfrac{S_n - S_j}{n - j}, & j + 1 \le i \le n. \end{cases}$$

A least squares estimate for the model is an estimate $\widehat{\theta}^J$ corresponding to a change point J for which $S_j^2/j + (S_n - S_j)^2/(n - j)$ is maximal over $1 \le j \le n$.

If the true change point is A, so that

$$\theta_1 = \theta_2 = \cdots = \theta_A = \theta_{A+1} + \Delta = \cdots = \theta_n + \Delta,$$

error is measured by the sum of squared deviations

$$d(\widehat{\theta}^J, \theta) = \sum_{i=1}^{n} (\theta_i{}^J - \theta_i)^2.$$

We will bound the expected value of this quantity, which is denoted by $\mathbb{P}d(\widehat{\theta}^J, \theta)$.

THEOREM 2. *Suppose the errors are independent unit normal. Assume* $\Delta > 0$. *For sample sizes n and change points A such that* $\Delta^2 A(n - A)/(4n \log n) > 1 + \varepsilon$, *there exists a constant* $c(\varepsilon)$ *such that* $\mathbb{P}d(\widehat{\theta}^J, \theta) < c(\varepsilon)$.

PROOF. The condition requires that the change point be not too close to either 1 or n; if it is too close, the situation approximates the case $\Delta = 0$; in that case, the sum of squared deviations will be about $2 \log \log n$ [from the behaviour of the maximum of standardized partial sums of independent random variables, Darling and Erdös (1956)]; the correlations between the partial sums causes the maximum of n standardized partial sums to behave like the maximum of $\log n$ independent unit normals]. Thus if the expected sum of squared deviations is to be bounded, the minimum block size cannot be too small.

Let $\overline{X} = S_n/n, \overline{\theta} = \sum_{i=1}^{n}\theta_i/n$. A least squares estimate $\widehat{\theta}^J$ satisfies $\overline{\theta}^J = \overline{X}$. Thus

$$\begin{aligned}\mathbb{P}d\big(\widehat{\theta}^J, \theta\big) &= \mathbb{P}d\big(\widehat{\theta}^J - \overline{X}, \theta - \overline{\theta}\big) + n\mathbb{P}\big(\overline{X} - \overline{\theta}\big)^2\\ &= \mathbb{P}d\big(\widehat{\theta}^J - \overline{X}, \theta - \overline{\theta}\big) + 1\\ &= \mathbb{P}\Big[d\big(\widehat{\theta}^J, \theta - \overline{\theta}\big) \mid S_n = 0\Big] + 1.\end{aligned}$$

Thus it is sufficient to prove the theorem assuming $\overline{\theta} = 0$ and conditioning on $S_n = 0$. Also, since $\mathbb{P}[d(\widehat{\theta}^A, \theta) \mid S_n = 0] = 1$ and since $d(\widehat{\theta}^J, \theta) \le 2d(\widehat{\theta}^A, \theta) + 2d(\widehat{\theta}^A, \widehat{\theta}^J)$, it will be sufficient to prove that $\mathbb{P}d(\widehat{\theta}^A, \widehat{\theta}^J)$ is bounded under the condition given.

Define the random variables Z_j, $1 \le j \le n$:

$$\begin{aligned}&\text{for } j < A, && S_j = \frac{j}{A}S_A + \sqrt{\frac{j(A-j)}{A}}Z_j;\\ &\text{for } j > A, && S_j = \frac{n-j}{n-A}S_A + \sqrt{\frac{(n-j)(j-A)}{n-A}}Z_j;\\ &\text{for } j = A, && Z_j = 0.\end{aligned}$$

The random variable Z_j, $j \ne A$, is a unit normal random variable independent of S_A. Let $\sigma^2 = A(n-A)/n$. Then, conditional on $S_n = 0$, $S_A \sim N(\sigma^2\Delta, \sigma^2)$; let $Z = (S_A - \sigma^2\Delta)/\sigma$.

If $J < A$, the least squares property requires that

$$S_J^2 \ge \frac{J(n-J)}{A(n-A)}S_A^2,$$

which implies

$$|Z_J| \ge |S_A|\frac{\sqrt{(n-J)/(n-A)} - \sqrt{J/A}}{\sqrt{A-J}} \ge \frac{|S_A|}{\sigma^2\big[2/\sqrt{|A-J|} + 1/\sigma\big]},$$

which we will write as $|Z_J| \ge |S_A|B_J$. (The same bound holds when $J > A$.)

Using this bound on S_A, after some algebra,

$$d\big(\widehat{\theta}^J, \widehat{\theta}^A\big) = \frac{n(A-J)}{A(n-J)}\Big[Z_J^2 + \frac{n}{n-A}S_A^2\Big] \le 4Z_J^2.$$

Bounding the maximum of nonnegative variables by their sum,

$$\mathbb{P}\Big[d\big(\widehat{\theta}^J, \widehat{\theta}^A\big) \mid S_n = 0\Big] \le 4\sum_{j \ne A}\mathbb{P}Z_j^2\big\{|Z_j| \ge |S_A|B_j\big\}.$$

We now adjust for the variance of S_A; set $Y_j = (Z_j - \sigma B_j Z)/(1+\sigma^2 B_j^2)^{1/2}$ and use $\mathbb{P}Z_j^2\{Y_j > K > 0\} \le \mathbb{P}Z_j^2\{Z_j > K\}$:

$$\begin{aligned}\mathbb{P}Z_j^2\{|Z_j| \ge |S_A|B_j\} &\le 2\mathbb{P}Z_j^2\{Z_j \ge |S_A|B_j\}\\ &\le 2\mathbb{P}Z_j^2\{Z_j \ge S_A B_j\} = 2\mathbb{P}Z_j^2\left\{Y_j \ge \frac{\Delta\sigma^2 B_j}{\left(1+\sigma^2 B_j^2\right)^{1/2}}\right\}\\ &\le 2\mathbb{P}Z_j^2\{Z_j \ge K_j\} = 2\mathbb{P}\{Z_j \ge K_j\} + \frac{2K_j \exp\left(-\frac{1}{2}K_j^2\right)}{\sqrt{2\pi}}\\ &\le 2\left[\frac{K_j}{\sqrt{2\pi}} + 1\right]\exp\left(-\tfrac{1}{2}K_j^2\right),\end{aligned}$$

where $K_j = (1/\sqrt{2})\Delta/(2/\sqrt{|A-J|}) + 1/\sigma) \le \Delta\sigma^2 B_j[1+\sigma^2 B_j^2]^{-1/2}$.

The quantities K_j behave like $\Delta\sqrt{|A-j|}$ for $|A-j|$ small, and like $\Delta\sigma$ for $|A-j|$ large:

$$\frac{K_j}{\Delta} \ge \begin{cases} \dfrac{\sqrt{|A-j|}}{\sqrt{2}(2+C)}, & \text{for } |A-j| \le C^2\sigma^2,\\[2ex] \dfrac{\sigma}{\sqrt{2}(2/C+1)}, & \text{for } |A-j| \ge C^2\sigma^2.\end{cases}$$

Thus the sum over $j \ne A$ of the tail integrals is bounded by the sum over $|A-j|$ small, and $|A-j|$ large; the first of these is a power series, and the second consists of less than n constant terms. Set

$$\Delta_1 = \frac{\Delta}{2+C}, \qquad \Delta_2 = \frac{\Delta}{2/C+1};$$

$$\begin{aligned}\frac{\mathbb{P}\left[d(\widehat{\theta}^J, \widehat{\theta}^A)\,|S_n = 0\right]}{8} &\le \left[1 + \frac{\Delta_1}{\sqrt{4\pi}}\right]\left[1 - \exp\left(-\frac{\Delta_1^2}{4}\right)\right]^{-2}\\ &\quad + n\left[1 + \frac{\Delta_2\sigma}{\sqrt{4\pi}}\right]\exp\left(-\frac{\Delta_2^2\sigma^2}{4}\right).\end{aligned}$$

The condition in the theorem asserts that $\Delta^2\sigma^2/(4\log n) > 1 + \varepsilon$. Thus we can choose C depending on ε so that

$$n\left[1 + \frac{\Delta_2\sigma}{\sqrt{4\pi}}\right]\exp\left(-\frac{\Delta_2^2\sigma^2}{4}\right)$$

is bounded. The first term in the sum is bounded for each choice of C. Thus the theorem is proved. □

The theorem asserts that the expected sum of squared errors is $O(1)$ when the size of each block of constant parameter values exceeds $4\log n/\Delta^2$; although this result has been proved only for a single change point, I expect the same result to apply to many change points. I suspect that block sizes exceeding $C \log\log n$ might be sufficient to guarantee asymptotic boundedness of the expected sum of squared errors. With block sizes exceeding εn for some $\varepsilon > 0$, I would expect to be able to find a similar bound for error distributions constrained only to have finite fourth moments.

4. Adaptive choice of shift estimators.

THEOREM 3. *Let $n = 2m$ be an even integer. Let S consist of the change point parameters θ^J, $1 \le J \le n$:*

$$\theta_i^J = \begin{cases} +\frac{1}{2}\Delta, & \text{for } J < i \le J + m, \\ -\frac{1}{2}\Delta, & \text{for } J + m < i \le J + n. \end{cases}$$

Let $X_i - \theta_i$, $i = 1, \dots, n$, be independent unit normal variables. Let T be a shift estimator $T = f(|i-j|)$, with $f(|i-j|) = f(n - |i-j|)$, where f may be chosen based on the observations. Then

$$M(T, \theta^j) > \frac{1}{3}\min\left[n, \frac{\Delta\sqrt{n}}{\pi} - 4\right].$$

PROOF. Take the true parameter to be θ^n without loss of generality. The matrix T has the Fourier eigenvectors $\mathbf{e}^j$: $e_i^j = \exp(\sqrt{-1}w(i-1)j)$, $1 \le i, j \le n$, where $w = 2\pi/n$. The corresponding eigenvalues are $\lambda_j = \Sigma f(i)\cos(w(i-1)j)$. We will use the Fourier transforms of data and parameters:

$$\mathbf{X} - \theta^n = \sum \frac{Z_j \mathbf{e}^j}{\sqrt{n}},$$

$$\theta^n = \sum \frac{c_j \mathbf{e}^j}{\sqrt{n}},$$

where the Z_j are independent unit complex Gaussians for $1 \le j \le m$, and

$$c_j = \begin{cases} \dfrac{2\Delta/\sqrt{n}}{\exp(\sqrt{-1}\,jw) - 1}, & j \text{ odd}, \\ 0, & j \text{ even}. \end{cases}$$

The optimal adaptive T chooses the λ_i to minimize the sum of squared errors for each particular data vector $\mathbf{X}$ and parameter vectors θ:

$$\begin{aligned} \text{SSE} &= \sum_i \sum (T_{ij} X_j - \theta_i)^2 \\ &= \sum_i |\lambda_i Z_i + (\lambda_i - 1)c_i|^2. \end{aligned}$$

For i even, choose $\lambda_i = 0$, since $c_i = 0$; for i odd, minimize a quadratic equation in λ_i, to obtain

$$\mathrm{SSE} = \sum_{i\,\mathrm{odd}} \frac{|c_i|^2|Z_i|^2 - \left(\mathrm{Re}(c_i\overline{Z}_i)\right)^2}{|c_i|^2 + |Z_i|^2 - 2\,\mathrm{Re}\left(c_i\overline{Z}_i\right)}$$

Now set $U_i = \mathrm{Re}(c_i\overline{Z}_i)/|c_i|$ and $V_i = \mathrm{Im}(c_i\overline{Z}_i)/|c_i|$, to obtain

$$\mathrm{SSE} = \sum_{i\,\mathrm{odd}} \frac{|c_i|^2 V_i^2}{V_i^2 + \left(U_i - |c_i|\right)^2},$$

where the U_i and V_i are independent unit normals. The given SSE is the minimal sum of squares for this particular data set.

To evaluate the mean sum of squares, we first bound the individual terms:

$$\mathbb{P}\frac{c^2V^2}{V^2 + (U-c)^2} \geq \mathbb{P}V^2\left[1 - \frac{V^2 + (U-c)^2 - c^2}{c^2}\right] = 1 - \frac{4}{c^2}.$$

Now use $|c_i|^2 = 2\Delta^2/(n(1 - \cos iw))$, and set $\alpha = \Delta/(w\sqrt{n})$. We will assess the effect of the large $|c_i|$, which occur for i near 1 and for i near n.

For any adaptive estimate T and parameter vector $\boldsymbol{\theta}$, and for each odd $k \leq m$,

$$\begin{aligned} M(T,\boldsymbol{\theta}) &\geq \sum_{i\,\mathrm{odd},\, i<n} \left[1 - 2n(1-\cos iw)/\Delta^2\right]^+ \\ &\geq 2 \sum_{i\,\mathrm{odd},\, i\leq k} \left[1 - 2n(1-\cos iw)/\Delta^2\right] \\ &\geq 2 \sum_{i\,\mathrm{odd},\, i\leq k} \left[1 - ni^2w^2/\Delta^2\right] \\ &\geq (k+1)\left[1 - \tfrac{1}{3}(k+1)^2/\alpha^2\right]. \end{aligned}$$

The cubic $x - \frac{1}{3}x^3/\alpha^2$ has its maximum, over $x \geq 0$, at $x = \alpha$; if $\alpha \geq 2$, there is an even integer $k+1$ in the interval $[\alpha - 2, \alpha]$ such that

$$k + 1 - \tfrac{1}{3}(k+1)^3/\alpha^2 \geq \alpha - 2 - \tfrac{1}{3}(\alpha-2)^3/\alpha^3 \geq \tfrac{2}{3}(\alpha - 2).$$

We now need to consider two cases; in the first case, $\alpha \leq m + 1$, and then we set $k + 1 \in [\alpha - 2, \alpha] \leq m + 1$ so that

$$M(T,\boldsymbol{\theta}) \geq \tfrac{1}{3}\left[\Delta\sqrt{n}/\pi - 4\right].$$

In the second case, $\alpha > m + 1$, and then we set $k + 1 = m$, to obtain

$$M(T,\boldsymbol{\theta}) \geq m\left[1 - \tfrac{1}{3}m^2/\alpha^2\right] \geq \tfrac{1}{3}n.$$

Combining the two cases produces the inequality asserted in the theorem. □

5. Partitions of the square. Suppose that $\mathbf{X}$ has mean θ and variance–covariance matrix I. Suppose that the observations X_{ij} are indexed by the pixels i, j in the square, with $1 \le i \le n,\ 1 \le j \le n$.

THEOREM 4. *Let $n = 2m$ be even. Let S consist of the parameters θ^{IJ}, $1 \le I, J \le n$, where $\theta_{ij}^{IJ} = 2\theta_i^I\theta_j^J/\Delta$, for θ^I as defined in Theorem 1. Let T be an arbitrary linear estimator. Then*

$$\sup_{\{\theta \in S\}} M(T, \theta) \ge \frac{1}{6} n \left[\min\left(n, \frac{\Delta\sqrt{32n}}{\pi}\right) - 7\right].$$

PROOF. Each of the θ^{IJ}'s consists of four $m \times m$ blocks of values that are equal to either $\frac{1}{2}\Delta$ or $-\frac{1}{2}\Delta$, with the blocks perhaps wrapped across opposite edges of the square. Consider the distribution that selects each θ_{IJ} with probability $1/n^2$.

The mean value of θ is zero and the variance–covariance matrix is $\mathbb{P}\theta_{ij}\theta_{kl} = 4\Sigma_{ik}\Sigma_{jl}/\Delta^2$, where

$$\Sigma_{ij} = \begin{cases} \Delta^2\left[\dfrac{1}{4} - \dfrac{|i-j|}{n}\right], & \text{for } |i-j| \le m, \\[2ex] \Delta^2\left[\dfrac{1}{4} - \dfrac{n-|i-j|}{n}\right], & \text{for } |i-j| \ge m, \end{cases}$$

The covariance matrix is a direct product of circulants, having eigenvectors $\mathbf{e}^{IJ} = \mathbf{e}^I \times \mathbf{e}^J$, $\mathbf{e}^j$: $e_i^j = \exp(\sqrt{-1}w(i-1)j)$, $1 \le i, j \le n$, where $w = 2\pi/n$. The corresponding eigenvalues are zero for I or J even, and $16\Delta^2/[n^2(1-\cos Iw)(1-\cos Jw)]$ for I, J odd. These are the products of the corresponding eigenvalues in the one-dimensional case. The k-dimensional problem may be treated similarly as a product of k one-dimensional problems.

Let $\alpha\sqrt{8n}/\pi$. We will use the result established in Theorem 1, that

$$\sum_{J \text{ odd}, J<n} \left[1 + \frac{n(1-\cos Jw)}{2\Delta^2}\right]^{-1} \ge \frac{(1/2)n}{(n/\Delta^2+1)^{1/2}} - 1,$$

substituting $16\Delta^2/[n(1-\cos Iw)]$ for $2\Delta^2$.

The Bayes and minimax risk is

$$\begin{aligned} B &= \sum_{I,J \text{ odd}} \frac{16\Delta^2}{n^2(1-\cos Iw)(1-\cos Jw) + 16\Delta^2} \\ &\ge \sum_{I \text{ odd}} \left[\frac{(1/2)n}{[1 + n(1-\cos Iw)/8\Delta^2]^{1/2}} - 1\right] \\ &\ge \sum_{I \text{ odd}, I \le m} \left[n - 2 - \frac{n^2I^2w^2}{(32\Delta^2)}\right]^+ \\ &\ge \frac{(n-2)(k+1)}{2} - \frac{n(k+1)^3}{6\alpha^2} \qquad \text{for } k \le m. \end{aligned}$$

For $\alpha \leq m$, take $k+1$ as the even integer in $[\alpha - 2, \alpha]$, to obtain (as in the proof of Theorem 3),

$$B \geq \frac{1}{3}n(\alpha - 2) - k - 1 \geq \frac{1}{3}n\left(\alpha - \frac{7}{2}\right) = \frac{1}{6}n\left[\Delta\sqrt{32n}/\pi - 7\right].$$

For $\alpha > m$, take $k + 1 = m$, to obtain

$$B \geq \frac{1}{3}nm - m \geq \frac{1}{6}n[n - 7].$$

The two inequalities together imply the inequality asserted in the theorem. □

Thus, in two dimensions, the best linear estimator does relatively worse than the best linear estimator in one dimension. The order of the error is $N^{3/4}$, where $N = n^2$ is the total number of observations.

Least squares estimates will estimate the true parameter values with mean sum of squared errors of order 1; in fact the correct block is identified with probability 1 as $n \to \infty$, and the mean sum of squared errors approaches 4.

REFERENCES

BESAG, J. (1986). On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B* **48** 259–302.

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models (with discussion). *Ann. Statist.* **17** 453–555.

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403.

DARLING, D. A. and ERDÖS, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155.

MCDONALD, J. A. and OWEN, A. B. (1986). Smoothing with split linear fits. *Technometrics* **28** 195–208.

MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761.

RICE, J. and ROSENBLATT, M. (1983). Smoothing splines, regression derivatives and convolution. *Ann. Statist.* **11** 141–156.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B* **47** 1–52.

VAN EEDEN, C. (1985). Mean integrated squared error of kernel estimators when the density and its derivatives are not necessarily continuous. *Ann. Inst. Statist. Math.* **37** 461–472.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

DEPARTMENT OF STATISTICS
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06520