The Annals of Statistics
1982, Vol. 10, No. 1, 278–286

MONOTONE REGRESSION ESTIMATES FOR GROUPED OBSERVATIONS[1]

F. T. WRIGHT

University of Missouri – Rolla

The maximum likelihood estimator of a nondecreasing regression function with normally distributed errors has been considered in the literature. Its asymptotic distribution at a point is related to a solution of the heat equation, and its rate of convergence to the underlying regression function is of order $n^{-1/3}$. This estimator can be modified by grouping adjacent observations and then "isotonizing" the corresponding means. It is shown that the resulting estimator has an asymptotic normal distribution for certain group sizes and its rate of convergence is of order $n^{-2/5}$. The results of a simulation study for small sample sizes are presented and grouping procedures are discussed.

1. Introduction. Suppose that for each t in the bounded interval (a, b) there is a probability distribution $D(t)$ with mean $m(t)$ and that we wish to estimate the regression function $m(t)$ on (a, b). In many situations it would be reasonable to assume that $m(t)$ is nondecreasing, so estimates which are also nondecreasing would be considered. In this note we restrict attention to regression functions which are nondecreasing on (a, b). (It will be clear what changes should be made in the following if $m(t)$ is not nonincreasing.) Brunk (1955) studied the least squares estimates subject to this monotonicity restriction. If $D(t)$ is $\mathcal{N}(m(t), \sigma^2)$, then these estimates are also the maximum likelihood estimates. We consider modifications obtained by grouping adjacent observations, using the algorithm proposed by Brunk to obtain a nondecreasing estimate from the group means and then interpolating. The schemes considered for grouping and interpolating may be of interest in nonparametric estimation of regression functions.

Suppose that $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_k$ are the means of independent random samples taken at $t_1, t_2, \cdots, t_k$ with $k > 1$ and $a < t_1 < t_2 < \cdots < t_k < b$. Suppose that the sample at t_i is of size n_i, and set $n = \sum_{i=1}^{k} n_i$. There are several methods for computing the estimate proposed by Brunk. We now describe the minimum lower sets algorithm. (Barlow et al., 1972, give a detailed discussion of these estimators which includes several of the computation algorithms.) Choose $i(1)$ to be the largest positive integer i which minimizes the pooled sample mean $\sum_{j=1}^{i} n_j \bar{X}_j / \sum_{j=1}^{i} n_j$, and then choose $i(2)$ to be the largest integer $i > i(1)$ which minimizes

$$\textstyle\sum_{k=i(1)+1}^{i} n_j \bar{X}_j / \sum_{j=i(1)+1}^{i} n_j.$$

Continuing this process, we obtain $0 = i(0) < i(1) < i(2) < \cdots < i(h) = k$ and the estimates are

$$\bar{m}(t_i) = \textstyle\sum_{j=i(r-1)+1}^{i(r)} n_j \bar{X}_j / \sum_{j=i(r-1)+1}^{i(r)} n_j$$

for $i = i(r-1)+1, \cdots, i(r)$ and $r = 1, 2, \cdots, h$. The $\bar{m}(t_i)$, which are nondecreasing in i, are referred to as the isotonized estimates and any nondecreasing function which is defined on (a, b) and has value $\bar{m}(t_i)$ at t_i for $i = 1, 2, \cdots, k$ could be thought of as an estimate of $m(\cdot)$. Let $\bar{m}$ be such a function.

Received July 1980; revised May 1981.

[1] This research was supported in part by the National Science Foundation under Grant MCS-78-01514.

AMS 1970 *subject classifications*. Primary 62F10; secondary 62E20.

Key words and phrases. Isotone regression, grouped observations, interpolation, asymptotic distribution, and rates of convergence.

Let $a = a_0 < a_1 < \cdots < a_p = b$ be a partition of (a, b), let b_i denote the midpoint of (a_{i-1}, a_i) for $i = 1, 2, \cdots, p$, let $\bar{Y}_i$ be the pooled sample mean for those samples corresponding to t_j with $a_{i-1} < t_j \le a_i$, that is

$$\bar{Y}_i = \textstyle\sum_{(a_{i-1},a_i]} n_j \bar{X}_j / \sum_{(a_{i-1},a_i]} n_j$$

where $\sum_A$ denotes the summation over those js for which $t_j \in A$, and let $m_i = \sum_{(a_{i-1},a_i]} n_j$. We assume that the partition has been chosen so that $m_i > 0$ and we think of $\bar{Y}_i$ as associated with the midpoint b_i for $i = 1, 2, \cdots, p$. Applying the minimum lower sets algorithm to $\bar{Y}_1, \bar{Y}_2, \cdots, \bar{Y}_p$ with sample sizes $m_1, m_2, \cdots, m_p$, one obtains nondecreasing estimates $m^*(b_1), m^*(b_2), \cdots, m^*(b_p)$. In Section 2 we discuss an interpolation scheme which extends the estimate m^* to (a, b) and controls the large sample bias. For $t_0 \in (b_{j-1}, b_j)$ we set $m^*(t_0) = \alpha\, m^*(b_{j-1}) + (1 - \alpha)m^*(b_j)$ with α given by (5).

The m.l.e., $\bar{m}$, pools observations only if there is a reversal, that is if $\bar{X}_i > \bar{X}_{i+1}$. However, the estimate considered here, m^*, pools all observations which are taken in a given interval of the partition and then pools further if there are reversals among the resulting pooled sample means. Since m^* typically involves more pooling than $\bar{m}$, it has a smaller variance but a larger bias and these two effects must be balanced.

Assuming that $m'(t_0) > 0$, Brunk (1970) showed that the asymptotic distribution of $\bar{m}(t_0)$ has density which is related to a solution of the heat equation and that the rate of convergence of $\bar{m}(t_0)$ to $m(t_0)$ is of order $n^{-1/3}$. Parsons (1979), Leurgans (1979) and Wright (1981) have studied the asymptotic distribution of $\bar{m}(t_0)$ when $m'(t_0) = 0$. In the next section, we show that for properly chosen partitions $m^*(t_0)$ had an approximate normal distribution for large n and the rate of convergence is of order $n^{-2/5}$. So not only does $m^*(t_0)$ converge to $m(t_0)$ more rapidly than does $\bar{m}(t_0)$, it also has a large sample distribution that is easier to work with. Neither the density nor the percentage points have been obtained for the large sample distribution of $\bar{m}(t_0)$.

Section 3 contains the results of a Monte Carlo study of the mean square errors of $\bar{m}(\cdot)$ and $m^*(\cdot)$. Based on the asymptotic results of Section 2 and the Monte Carlo study, a recommendation is made concerning the number of intervals in a regular partition and the behavior of the resulting estimator is discussed.

2. Asymptotic results. If either estimator is to be consistent for a broad class of underlying regression functions, then the observation points must become dense in (a, b). So we allow k; t_j and $\bar{X}_j$ for $j = 1, 2, \cdots, k$; p; and a_i, b_i and $\bar{Y}_i$ for $i = 1, 2, \cdots, p$ to depend on n, but for notational convenience we will not show this dependence.

We consider the behavior of the regression estimates at a fixed point $t_0 \in (a, b)$ for large n. We assume that

$$m(t) = m(t_0) + (t - t_0)m'(t_0) + \tfrac{1}{2}(t - t_0)^2 m''(t_0)\{1 + o(1)\} \tag{1}$$

as $t \to t_0$ and $m'(t_0) > 0$. Let F_n be the empirical distribution function corresponding to the observation points, that is

$$F_n(t) = \textstyle\sum_{\{j \le k: t_j \le t\}} n_j / n,$$

and let F be a distribution function which has a density f in a neighborhood of t_0. Throughout this discussion, we consider a fixed, positive δ and assume that

$$\begin{gathered} \sup_t |F_n(t) - F(t)| = o(n^{-(1-\delta)/2}), \quad f(t_0) > 0 \quad \text{and} \\ f(t) = f(t_0) + (t - t_0)f'(t_0)(1 + o(1)) \quad \text{as } t \to t_0. \end{gathered} \tag{2}$$

If the observation points are the realization of an i.i.d. sequence of random variables whose common distribution F has density f which satisfies the conditions of (2), then (2) holds with probability one because of the law of the iterated logarithm for empirical distribution functions. Denoting the observations at t_j by $X_{j,\ell}$ for $\ell = 1, 2, \cdots, n_j$, we assume that the errors,

(3) $$\{X_{j,\ell} - m(t_j): \ell = 1, 2, \cdots, n_j, j = 1, 2, \cdots, k\},$$

are i.i.d. with common variance $\sigma^2 \in (0, \infty)$.

We now place restrictions on the sequence of partitions. For sufficiently large n, t_0 is in $(b_{j-1}, b_j]$ for some j, but of course, j depends on n. We assume that there is a positive constant C for which

(4) $$\max_{\ell=j-1,j} | n^{\delta}(a_\ell - a_{\ell-1}) - C| \to 0 \quad \text{as} \quad n \to \infty.$$

Because some of the results that follow may be of interest in nonparametric regression estimation, we first consider the effects of grouping and interpolation on the estimation of the regression function. Then we discuss the isotonization of the estimate. Define $m_0(t_0) = \alpha \bar{Y}_{j-1} + (1 - \alpha)\bar{Y}_j$ with

(5) $$\alpha = (a_j - a_{j-1})(b_j - t_0)/\{(a_j - a_{j-1})(b_j - t_0) + (a_{j-1} - a_{j-2})(t_0 - b_{j-1})m_j/m_{j-1}\}.$$

As we shall see, this choice of α causes the ratio of the large sample bias to the large sample standard deviation to converge to zero. It should be noted that if $a_j - a_{j-1} = a_{j-1} - a_{j-2}$ and $m_j = m_{j-1}$, then this α is the one that would be used in simple linear interpolation. The following result is proved in the Appendix.

LEMMA 1. *If the observation points, the observations, the partitions, and the regression function satisfy assumptions* (1, 2, 3 *and* 4) *with* $0 < \delta < 1/3$, *then*

$$\alpha E(\bar{Y}_{j-1}) + (1 - \alpha)E(\bar{Y}_j) - m(t_0) = 0(n^{-2\delta}) + o(n^{-(1-\delta)/2}).$$

In the proof of Lemma 2, it is shown that under the assumptions of Lemma 1, m_j and m_{j-1} are both of the form $Cf(t_0)n^{1-\delta}\{1 + o(1)\}$. Hence, $\{\alpha^2 m_{j-1}^{-1} + (1 - \alpha)^2 m_j^{-1}\}^{-1/2} \leq 0(n^{(1-\delta)/2})$ and so if assumptions (1, 2, 3 and 4) are satisfied and $1/5 < \delta < 1/3$,

$$[\sigma^2\{\alpha^2 m_{j-1}^{-1} + (1 - \alpha)^2 m_j^{-1}\}]^{-1/2}\{m_0(t_0) - m(t_0)\}$$

has a standard normal limiting distribution. Examining the proofs of Lemmas 1 and 2, we see that this result is still valid if $m'(t_0)$ is not positive.

If the underlying regression function, m, is nondecreasing, then one might find the estimate m_0 unacceptable since it may be decreasing over some subintervals of (a, b). So we turn our attention to the estimator $m^*(t_0) = \alpha m^*(b_{j-1}) + (1 - \alpha)m^*(b_j)$ with α given by (5). However, the following lemma, which is proved in the Appendix, shows that, asymptotically, $m^*(b_{j-1})$ and $m^*(b_j)$ behave essentially like $\bar{Y}_{j-1}$ and $\bar{Y}_j$, respectively.

LEMMA 2. *If the assumptions* (1, 2, 3 *and* 4) *are satisfied with* $0 < \delta < 1/3$, *then*

$$P\{m^*(b_{j-1}) \neq \bar{Y}_{j-1} \text{ or } m^*(b_j) \neq \bar{Y}_j\} \to 0 \quad \text{as} \quad n \to \infty.$$

It is interesting to note that Lemma 2 shows that for large n, $m^*(t_0)$ essentially pools all the observations from an interval of length $Cn^{-\delta}$ with $\delta < 1/3$. (Later, we will see that $\delta = 1/5$ is asymptotically optimal). However, Brunk (1970) has shown that for large n, $\bar{m}(t_0)$ pools observations from a smaller interval of length $Cn^{-1/3}$ and then only if it is necessary to obtain a nondecreasing estimate.

In establishing the asymptotic distribution of $m^*(t_0)$ we may redefine $m^*(b_{j-1})$ and $m^*(b_j)$ on a set A_n provided $P(A_n) \to 0$. So under assumptions (1, 2, 3 and 4) with $1/5 < \delta < 1/3$, $m^*(t_0)$ and $m_0(t_0)$ have the same large sample distributions. This gives the following result.

THEOREM 3. *If assumptions* (1, 2, 3 *and* 4) *are satisfied with* $1/5 < \delta < 1/3$, *then*

$$[\sigma^2\{\alpha^2 m_{j-1}^{-1} + (1 - \alpha)^2 m_j^{-1}\}]^{-1/2}\{m^*(t_0) - m(t_0)\}$$

has a standard normal limiting distribution.

This result is related to the asymptotic distribution theory developed for the basic estimator of the failure rate function in Barlow and Van Zwet (1971). However, the interpolation scheme used here allows us to consider arbitrary t_0 and not just the midpoints, b_j. As in their work, we could obtain a normal limit for $\delta \leq 1/5$, but the centering constants become much more complicated, involving derivatives of m at t_0. Such results would seem to be of limited practical significance in this setting. Furthermore, as in the work of Barlow and Van Zwet, the choice of δ that gives the smallest asymptotic mean square error is $\delta = 1/5$; this is the choice of δ for which the asymptotic variance and the asymptotic bias squared are of the same order. It should be noted that while $\bar{m}(t_0)$ converges to $m(t_0)$ at the rate of $n^{-1/3}$, the rate of m^* is $n^{-2/5}$ if $\delta = 1/5$. The determination of the constant C is discussed in the next section.

Under the assumption that m'' exists, Stone (1980) has shown that the rate $n^{-2/5}$ is the best possible for a nonparametric regression estimator. On the other hand, the large sample results for $\bar{m}(t_0)$ established by Brunk (1970) are based on the assumption that m' exists and in this case, Stone's Theorem shows that $n^{-1/3}$ is the best possible rate for a nonparametric regression estimator.

It is clear that the asymptotic distribution of the estimator m^* at a finite collection of distinct points could be obtained using the techniques employed here. In fact, these estimates would be asymptotically independent.

3. Small sample properties. Since the sampling distributions of $\bar{m}$ and m^* are quite complex, simulation techniques were employed to compare the two estimators for small sample sizes. (The results of earlier simulation studies of $\bar{m}$ and related estimators are contained in Cryer et al. (1972) and Wright (1978).) Regression functions $m(t) = 4t, 4\sqrt{t}, 4t^2, 2(2t-1)^5, 2(2t-1)^{1/5}$ and $2 \sin \pi(t - 1/2)$ were considered on the interval [0, 1]. With $k = 24$ and 50, equally spaced observation points, $t_i = (i-1)/(k-1)$ for $i = 1, 2, \cdots, k$, were chosen. With the regression function m fixed, independent, normally distributed random variables with mean $m(t_i)$ and standard deviation 1/2 were simulated for $i = 1, 2, \cdots, k$. Based on these k observations, the maximum likelihood estimator $\bar{m}$ was computed using the minimum lower sets algorithm. Next the observations' points were grouped into p groups with the number in each group as nearly constant as possible. Then for various choices of p, the modified estimate m^* was computed. Based on 1000 iterations, the mean squared error (m.s.e.) of both estimators at each observation point was approximated and then these m.s.e.'s at the points t_i were summed to approximate the total m.s.e. for both of the estimators. This procedure was carried out for the six regression functions and for $k = 24$ and 50.

The results of this Monte Carlo study are summarized in Tables 1 and 2. For each regression function and both choices of k the total m.s.e. of $\bar{m}(\cdot)$ was estimated six times. In these twelve cases the largest discrepancy among the six estimates was .06 and so this gives an indication of the size of the possible sampling errors in these estimated m.s.e.'s. It is interesting to note that for the first four functions given in both Tables 1 and 2, the total

TABLE 1
Total mean squared errors of $\bar{m}$ and m^* for $k = 24$ equispaced t values on [0, 1]

Function m(t)	Estimator $\bar{m}$	m^* p = 3	p = 4	p = 5	p = 6	p = 8	p = 12
$4t$	2.18	.84	1.06	1.20	1.34	1.60	1.87
$4\sqrt{t}$	2.09	1.63	1.58	1.53	1.64	1.75	2.02
$4t^2$	2.17	1.30	1.22	1.21	1.31	1.56	1.86
$2 \sin \pi(t - 1/2)$	.94	.48	.57	.65	.63	.75	.88
$2(2t-1)^5$	1.89	4.99	3.77	2.84	2.54	2.13	2.30
$2(2t-1)^{1/5}$	1.89	10.22	3.06	4.87	2.30	2.01	1.91

TABLE 2
Total mean squared errors of $\bar{m}$ and m^* for $k = 50$ equispaced t values on [0, 1]

Function m(t)	$\bar{m}$	m*, p = 4	p = 5	p = 6	p = 8	p = 10	p = 16
$4t$	2.86	.95	1.15	1.28	1.59	1.82	2.06
$4\sqrt{t}$	2.70	1.72	1.70	1.66	1.74	1.93	1.99
$4t^2$	2.82	1.20	1.24	1.26	1.47	1.76	1.98
$2 \sin \pi(t - ½)$	1.14	.54	.61	.61	.68	.77	.83
$2(2t - 1)^5$	2.41	5.82	4.09	3.06	2.06	2.07	1.67
$2(2t - 1)^{1/5}$	2.49	5.42	8.78	3.69	3.02	2.64	2.32

TABLE 3
Pointwise mean squared errors of $\bar{m}$ and m^* for $k = 24$ equispaced t values on [0, 1]

	m(t) = 4t		m(t) = 4t²		m(t) = 2(2t − 1)⁵		m(t) = 2(2t − 1)^{1/5}	
t	$\bar{m}$	m*, p = 3	$\bar{m}$	m*, p = 5	$\bar{m}$	m*, p = 8	$\bar{m}$	m*, p = 12
.000	.173	.083	.149	.122	.209	.230	.161	.165
.043	.103	.067	.074	.078	.170	.160	.077	.086
.087	.083	.053	.057	.051	.125	.187	.054	.048
.130	.080	.041	.052	.037	.098	.147	.047	.045
.174	.081	.032	.053	.031	.072	.081	.046	.035
.217	.084	.025	.053	.031	.056	.062	.047	.040
.261	.079	.020	.059	.037	.044	.046	.048	.036
.304	.079	.017	.060	.046	.038	.038	.051	.040
.348	.079	.017	.061	.036	.031	.031	.061	.044
.391	.082	.018	.067	.034	.030	.026	.074	.053
.435	.077	.022	.074	.039	.027	.027	.099	.071
.478	.083	.029	.082	.053	.027	.025	.173	.289
.522	.084	.029	.086	.052	.026	.025	.183	.284
.565	.086	.023	.086	.037	.028	.027	.106	.074
.609	.085	.019	.094	.030	.030	.026	.074	.057
.652	.085	.017	.096	.032	.034	.031	.064	.048
.696	.076	.017	.101	.045	.041	.039	.051	.044
.739	.078	.020	.103	.034	.049	.047	.047	.037
.783	.078	.025	.109	.030	.059	.062	.045	.041
.826	.087	.032	.114	.035	.073	.083	.047	.036
.870	.084	.041	.111	.047	.094	.154	.049	.045
.913	.080	.052	.121	.064	.134	.197	.054	.048
.957	.101	.066	.131	.088	.167	.164	.076	.084
1.000	.174	.081	.200	.119	.237	.216	.164	.157

m.s.e. for m^* was less than that of $\bar{m}$ for every choice of p considered. But, of course, the optimal choice of p depends on the function considered. For the remaining two functions, $2(2t-1)^5$ and $2(2t-1)^{1/5}$, the choice of p is especially critical since for small p, m^* behaves much worse than $\bar{m}$ in these cases. For the function $2(2t-1)^5$ and small p, it is at the points near 0 and 1 that m^* has larger m.s.e. than $\bar{m}$ and this is due to the bias of the estimator. For the function $2(2t-1)^{1/5}$ and such p, it is at the points near ½ that m^* behaves so poorly, and again this is due to bias. Clearly, a procedure for choosing p is needed.

To indicate how these estimators behave locally, the pointwise m.s.e.'s of $\bar{m}$ and m^* (with the optimal choices of p from Table 1) are given in Table 3 for $k = 24$ and several choices of m. As one can see, the estimator $\bar{m}$ typically does not perform as well at the extreme observation points. This was observed by Cryer et al. (1972) and was found to be due primarily to bias. This was generally true here for m^* also, but in some cases the effects were not as drastic.

We suppose that the intervals in the partition are all of the same length $Cn^{-\delta}\{1 + o(1)\}$. We have already seen that the large sample analysis suggests that δ should be chosen to be ⅕, provided (1, 2, 3 and 4) hold. So with this choice of δ, we consider choosing C so that m^* performs reasonably well over the entire interval (a, b). However, the optimal choice (in terms of m.s.e.) for C does depend on the point being considered. As a first step we consider the choice of C which minimizes the large sample m.s.e. of $m^*(t_0)$ assuming $t_0 = b_j$. Now with $\delta = ⅕$, $\alpha = 0$ and n large, the variance of $m^*(t_0)$ is approximately $n^{-4/5}\sigma^2/(Cf(t_0))$ and the bias squared is approximately

$$n^{-4/5}[(C^2/12)\{m'(t_0)f'(t_0)/f(t_0) + m''(t_0)/2\}]^2;$$

see the derivation of expression (13) in the Appendix. So the resulting expression for the asymptotic m.s.e. is minimized by

$$C = \left[\frac{36\sigma^2}{\{m'(t_0)f'(t_0)/f(t_0) + m''(t_0)/2\}^2 f(t_0)}\right]^{1/5}.$$

We assume that the distribution governing the observation points, t_i, is known so that $f(t_0)$ and $f'(t_0)$ are known. Using a preliminary estimate such as $\bar{m}$, one could estimate σ, m' and m'' and then use $C' = (6\hat{\sigma}/A)^{2/5}$ where $\hat{\sigma}$ is the estimate of σ and A is the average of the estimated value of

$$|m'(t)f'(t)/f(t) + m''(t)/2|\sqrt{f(t)}.$$

However, as we shall see, C' and $\delta = ⅕$ generally yield a smaller p than is optimal for the sample sizes and functions considered here. So we recommend $C^* = (6\hat{\sigma}/M)^{2/5}$ where M is the maximum estimated value rather than the average.

If the distribution F is uniform on [0, 1], as in the simulation study discussed here, then the formula for C becomes $C = (12\sigma/|m''(t_0)|)^{2/5}$ and one might use $C^* = (12\hat{\sigma}/M)^{2/5}$, where M is the largest estimated value of $|m''(t)|$. To give some idea of how C and hence C^* varies with the underlying regression function, the value of p was computed for each of the six functions considered here with $k = 24$ and 50. The true values of σ and $\max|m''(t)|$ were used except for the functions $m(t) = 4\sqrt{t}$ and $m(t) = 2(2t - 1)^{1/5}$. For these two functions, which do not have a finite second derivative,

$$\max_j|m(t_j) - 2m(t_{j-1}) + m(t_{j-2})|/(t_j - t_{j-1})^2$$

was used rather than $\max|m''(t)|$.

For the function $m(t) = 4t$, $m''(t) \equiv 0$ indicating a choice for p as small as possible, and $p = 2$ is optimal in this case. For this function and the smallest p's given in Tables 1 and 2 the pooling reduces the m.s.e. of $\bar{m}$ by more than 60%. For the function $m(t) = 4t^2$, $p = 2$ for both $k = 24$ and 50. However, it is unlikely that one would want to use so few groups but again for the smallest p's considered in Tables 1 and 2, the reductions in m.s.e. due to pooling are approximately 40% and 60%. For the function $m(t) = 2\sin\pi(t - ½)$, $p = 3(4)$ for $k = 24$ (50) and these are the optimal choices of p given in Tables 1 and 2. In these two cases, the m.s.e. of m^* is about ½ of that of $\bar{m}$. For the function $m(t) = 4\sqrt{t}$, $p = 8$ (16) for $k = 24$ (50) and the reductions in m.s.e. are 16% and 26%. While those were by no means the optimal choices of p, the resulting estimator still behaved significantly better than $\bar{m}$. For $m(t) = 2(2t - 1)^{1/5}$, $p = 15$ (29) for $k = 24$ (50). While these are larger values of p than those studied, we do note that for the largest values of p in Tables 1 and 2, m^* behaved as well or better than $\bar{m}$. For the function $m(t) = 2(2t - 1)^5$, $p = 7$ (8) for $k = 24$ (50). This case with $k = 24$ is the only one of those considered in which $\bar{m}$ has a significantly smaller m.s.e. than m^* with the choice of C under discussion. The m.s.e. of m^* would seem to be approximately 20% larger. However, with $k = 50$, the m.s.e. of m^* is about 15% smaller than that of $\bar{m}$. For this function and $k = 24$ (50), the value of p corresponding to the C based on the average of $|m''(t)|$, rather than $\max|m''(t)$, is 4 (5). For this p, m^* has a considerably larger m.s.e. and in fact, the m.s.e. of m^* is 100% (70%) larger than that of $\bar{m}$. For this reason we have chosen the C based on the maximum.

Studying the pointwise m.s.e. in the few cases above for which m^* did not dramatically outperform $\bar{m}$, it can be seen that the reason is bias. In regions in which m is "extremely" convex or concave, grouping and interpolating linearly introduces considerable bias. If the researcher anticipated such regions the partition could be refined there.

One of the referees suggested choosing the partition adaptively. This approach, which certainly seems to merit investigation, might provide a nice solution to the problem of choosing C and even provide nonregular partitions to reduce bias in those regions where m^* performs poorly.

APPENDIX

We give the proof of Lemma 2 before that of Lemma 1.

PROOF OF LEMMA 2. The proof of that $P\{m^*(b_{j-1}) \neq \bar{Y}_{j-1}\} \to 0$ and that $P\{m^*(b_j) \neq \bar{Y}_j\} \to 0$ are similar and we only give the latter. We first show $P\{m^*(b_j) \neq \bar{Y}_j\}$ is bounded above by the sum of

$$P\{\max_{\ell<j} A(\ell, j-1) > \bar{Y}_j\} \tag{6}$$

and

$$P\{\min_{\ell>j} A(j+1, \ell) < \bar{Y}_j\}, \tag{7}$$

where $A(\ell, \ell') = \sum_{\nu=\ell}^{\ell'} m_\nu \bar{Y}_\nu / \sum_{\nu=\ell}^{\ell'} m_\nu$. To see this, one uses the representation (cf. Barlow et al., 1972, page 19)

$$m^*(b_j) = \max_{1\le\ell\le j} \min_{j\le\ell'\le p} A(\ell, \ell')$$

and takes complements. If $\max_{\ell<j} A(\ell, j-1) \le \bar{Y}_j$ and $\min_{\ell>j} A(j+1, \ell) \ge \bar{Y}_j$, then for any $\ell \le j$, $A(\ell, \ell') \ge A(\ell, j)$ for each $\ell' \ge j$. Furthermore, for any $\ell \le j$, $A(\ell, j) \le \bar{Y}_j$ which implies that $m^*(b_j) = \bar{Y}_j$. The proofs that (6) and (7) converge to zero are alike and so we give the one for (6). Expression (6) is bounded above by the sum of

$$P\{\max_{\ell<j} A(\ell, j-1) > m(a_{j-1})\} \tag{8}$$

and

$$P\{m(a_{j-1}) > \bar{Y}_j\}. \tag{9}$$

Since m is nondecreasing, (8) is bounded above by

$$P\{\max_{\ell<j} A_c(\ell, j-1) > m(a_{j-1}) - E(\bar{Y}_{j-1})\} \tag{10}$$

where $A_c(\ell, \ell') = \sum_{\nu=\ell}^{\ell'} m_\nu\{\bar{Y}_\nu - E(\bar{Y}_\nu)\} / \sum_{\nu=\ell}^{\ell'} m_\nu$ and (9) is bounded above by

$$P\{\bar{Y}_j - E(\bar{Y}_j) < m(a_{j-1}) - E(\bar{Y}_j)\}. \tag{11}$$

Before obtaining estimates for $E(\bar{Y}_{j-1})$ and $E(\bar{Y}_j)$ we need to estimate m_{j-1} and m_j. Now

$$\begin{aligned} m_j = n\{F_n(a_j) - F_n(a_{j-1})\} &= n\{F(a_j) - F(a_{j-1})\} + o(n^{(1+\delta)/2}) \\ &= C f(t_0) n^{1-\delta}\{1 + o(1) + O(n^{-(1-3\delta)/2})\} = C f(t_0) n^{1-\delta}\{1 + o(1)\} \end{aligned}$$

since $\delta < 1/3$. Similarly, it can be shown that $m_{j-1} = C f(t_0) n^{1-\delta}\{1 + o(1)\}$.

Now we obtain estimates for $E(\bar{Y}_j) - m(a_{j-1})$ and $m(a_{j-1}) - E(\bar{Y}_{j-1})$. By assumption,

$$m(t) = m(t_0) + (t - t_0)m'(t_0) + \tfrac{1}{2}(t - t_0)^2 m''(t_0)\{1 + o(1)\}$$

and

$$f(t) = f(t_0) + (t - t_0)f'(t_0)\{1 + o(1)\}.$$

So

$$E(\bar{Y}_j) - m(t_0) = (n/m_j)\left[\int_{(a_{j-1}, a_j]} \{m(t) - m(t_0)\}\, dF(t) + \int_{(a_{j-1}, a_j]} \{m(t) - m(t_0)\} d\{F_n(t) - F(t)\}\right]$$

and, integrating by parts, the latter integral is bounded above by

$$O\{(a_j - a_{j-1}) \sup_t |F_n(t) - F(t)|\} = o(n^{-(1+\delta)/2}). \tag{12}$$

The first integral can be written as the sum of

$$f(t_0)m'(t_0)\{(a_j - t_0)^2 - (a_{j-1} - t_0)^2\}/2 = f(t_0)m'(t_0)(a_j - a_{j-1})(b_j - t_0)$$

and

$$\{f'(t_0)m'(t_0) + f(t_0)m''(t_0)/2\}3^{-1}\{(a_j - t_0)^3 + (t_0 - a_{j-1})^3\}\{1 + o(1)\}.$$

Hence, $E(\bar{Y}_j) - m(t_0)$ can be written

$$(n/m_j)f(t_0)m'(t_0)(a_j - a_{j-1})(b_j - t_0) + O(n^{-2\delta}) + o(n^{-(1-\delta)/2}), \tag{13}$$

and similarly

$$E(\bar{Y}_{j-1}) - m(t_0) = (n/m_{j-1})f(t_0)m'(t_0)(a_{j-1} - a_{j-2})(b_{j-1} - t_0) + O(n^{-2\delta}) + o(n^{-(1-\delta)/2}). \tag{14}$$

Since

$$m(t_0) - m(a_{j-1}) = (t_0 - a_{j-1})m'(t_0) + O((t_0 - a_{j-1})^2),$$

we have that

$$E(\bar{Y}_j) - m(a_{j-1}) = m'(t_0)(b_j - a_{j-1}) + o(n^{-\delta}) = 2^{-1}Cm'(t_0)n^{-\delta}\{1 + o(1)\}$$

and

$$m(a_{j-1}) - E(\bar{Y}_{j-1}) = m'(t_0)(a_{j-1} - b_{j-1}) + o(n^{-\delta}) = 2^{-1}Cm'(t_0)n^{-\delta}\{1 + o(1)\}.$$

Applying Chebychev's inequality, expression (11) is bounded by $O(n^{3\delta-1})$, which converges to zero provided $\delta < 1/3$. To treat expression (10), fix n, let $\bar{Y}_{j,i}$; $i = 1, 2, \cdots, m_j$ and $j = 1, 2, \cdots, p$ denote the observations which are averaged to obtain $\bar{Y}_j$, let

$$Z_i = Y_{j-1,i} - E(Y_{j-1,i}) \quad \text{for } i = 1, 2, \cdots, m_{j-1},$$

$$Z_{m_{j-1}+i} = Y_{j-2,i} - E(Y_{j-2,i}) \quad \text{for } i = 1, 2, \cdots, m_{j-2}, \cdots,$$

$$Z_{m_{j-1}+m_{j-2}+\cdots+m_2+i} = Y_{1,i} - E(Y_{1,i}) \quad \text{for } i = 1, 2, \cdots, m_1$$

and set $S_\ell = \sum_{\nu=1}^{\ell} Z_\nu$. So (10) is bounded above by

$$P[\max_{m_{j-1} \le \ell \le m_{j-1}+m_{j-2}+\cdots+m_1} \ell^{-1}S_\ell > 2^{-1}Cm'(t_0)n^{-\delta}\{1 + o(1)\}]$$

and applying the Hájek-Rényi inequality (cf. Bauer, 1972), we see that this probability is bounded by

$$O(n^{2\delta}(m_{j-1}^{-1} + \sum_{\ell=m_{j-1}+1}^{\infty} \ell^{-2})) = O(n^{-(1-3\delta)})$$

which converges to zero if $\delta < 1/3$.

PROOF OF LEMMA 1. Using the expressions given in (13) and (14), we see that

$$\alpha E(\bar{Y}_{j-1}) + (1-\alpha)E(\bar{Y}_j) - m(t_0) = O(n^{-2\delta}) + o(n^{-(1-\delta)/2}).$$

Acknowledgement. The author wishes to express his thanks to the referees and editors for their helpful suggestions and also to Mr. Wei-Kei Shiue for preparing the computer program for the Monte Carlo study.

REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions; The Theory and Application of Isotonic Regression.* Wiley, New York.

BARLOW, R. E. and VAN ZWET, W. R. (1971). Comparisons of several nonparametric estimators of the failure rate function. *Operations Research and Reliability* (Proceedings of the NATO Conference at Turin, 1969). Gordon and Breach, New York, 375–399.

BAUER, HEINZ (1972). *Probability Theory and Elementary Measure Theory.* Holt, Rinehart and Winston; New York.

BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.* **26** 607–616.

BRUNK, H. D. (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference.* Cambridge Univ. Press, 177–195.

CRYER, J. D., ROBERTSON, TIM, WRIGHT, F. T. and CASADY, ROBERT J. (1972). Monotone median regression. *Ann. Math. Statist.* **43** 1459–1469.

LEURGANS, SUE (1979). Asymptotic distribution of slope of greatest convex minorant estimators. Technical Report #1946. Univ. of Wisconsin-Madison, Mathematics Research Center.

PARSONS, VAN L. (1979). The asymptotic distribution of an isotonic regression estimator at a point. (unpublished manuscript).

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

WRIGHT, F. T. (1978). Estimating strictly increasing regression functions. *J. Amer. Statist. Assoc.* **73** 636–639.

WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MISSOURI-ROLLA
326 MATHEMATICS-COMPUTER SCIENCE BUILDING
ROLLA, MO 65401