Journal of the Korean Statistical Society (1998), 27: 4, pp 485–494

Adaptive Estimation of Monotone Functions[†]

Yung-Gyung Kang[1]

ABSTRACT

In the white noise model we construct an adaptive estimate for $f(0)$ for a decreasing function f. We also show that the maximum mean square error of this estimate attains the same rate as the minimax risk simultaneously over a range of Lipschitz classes of order less than or equal to one.

Keywords: White noise model; Adaptive estimation; Decreasing function; Minimax risk; Lipschitz classes

1. INTRODUCTION

The estimation of order-restricted functions is an area of nonparametric function estimation that much research has been done. Order constraints are somewhat different from smoothness conditions under which we usually deploy our theory of nonparametric function estimation. Order-restricted functions include unimodal densities, monotone regression functions, and monotone hazard rates. Many statisticians have looked at the problem of estimating such functions in both applied and theoretical perspectives and many methods have been proposed and evaluated. Most of those estimators are made by combining the pool-adjacent-violators (PAV) algorithm, so called isotonic regression, and their choices of smoothing techniques, such as spline smoothing. This line of research can be found in articles and books including Grenander(1956), Barlow, Bartholomew, Bremmer and Brunk(1972), Mammen(1991), Friedman and Tibshirani(1984), and Kelly and Rice(1990).

In terms of application it is not hard to find order-restricted functions in real life. In seismology it is assumed on the basis of considerable empirical evidence that the probability of earthquake aftershocks is decreasing as time goes by after the main shock. See Reasenberg and Jones(1989). In survival analysis

[†]This research was supported in part by the Research Institute of Basic Sciences at the Seoul National University.

[1]Department of Computer Science and Statistics, Hanshin University, Osan 447-791, Korea

a dose response curve is usually monotone as in Kelly and Rice(1990). In addition, if we browse through an introductory level Economics textbook we can easily find many monotone or at least unimodal functions. For instance, the law of downward-slopping demand is generally accepted and Engel's curves, which show how family expenditures for consumption change as incomes rise, are either increasing or unimodal. See Samuelson and Nordhaus(1985) and Hildenbrand and Hildenbrand(1986).

In spite of the potential usefulness of the estimation of monotone functions, many statisticians who work under the minimax paradigm have not paid much attention to this problem due to Kiefer(1982). From a theoretical point of view, one might wonder whether this order constraint in addition to smoothness conditions can change the minimax linear functional estimation problem. Kiefer(1982) answered that the minimax rate cannot be improved by adding an order constraint. However, in this article we shall show that the extra order constraint completely changes the adaptation problem. That is, for a fixed class of functions adding an order constraint does not improve the minimax rate but putting the extra order contraint makes it possible to construct an estimate that is minimax for more than one function class. Such an estimate is called "adaptive". More precisely, for a whole range of Lipschitz function classes when functions are decreasing, it is possible to construct a fully adaptive estimator as far as rates are concerned. Without knowing the order of the Lipschitz class, we can estimate $f(0)$ as good as when we know the order in the minimax sense; the estimate can adapt itself to the smoothness of the function to be estimated.

This result is somewhat surprising. Because Brown and Low(1996) have shown that we have to pay a penalty of at least the logarithm of the sample size when we try to adapt over any two symmetric function classes. By a symmetric function class we mean $\mathcal{F}$ such that $f \in \mathcal{F}$ implies $-f \in \mathcal{F}$. Therefore, without an order constraint Lipschitz function classes are impossible to adapt over.

In the next section we shall show the construction of our adaptive estimator and present its adaptability. This result will be proven in section 3.

2. MAIN RESULT

In recent years many statisticians have been interested in the nonparametric function estimation in Gaussian white noise rather than nonparametric regression or density estimation. Suppose we observe a stochastic process $X_n(t)$ from the

white noise model

$$X_n(dt) = f(t)dt + \frac{1}{\sqrt{n}}W(dt), \quad -1 \le t \le 1, \tag{2.1}$$

where W is Brownian motion starting at time $t = -1$ and we need to estimate the drift term $f(x)$ based on $X_n(t)$. Brown and Low(1996) and Nussbaum(1996) showed that "generally" estimating $f(x)$ from the white noise model is asymptotically equivalent to estimating the mean function in the nonparametric regression model or estimating a density from a random sample of size n. The literature on function estimation in white noise is vast by now and some samples are Ibragimov and Hasminskii(1984), Lepskii(1991), Lepskii(1992), Pinsker(1980), Low(1992), Low(1995), and Donoho and Liu(1991).

Now we assume that $f \in \mathcal{F}_D(\alpha, M)$ for $0 < \alpha \le 1$ and $M > 1$. Here we write $\mathcal{F}(\alpha, M)$ for the Lipschitz class such that

$$\mathcal{F}(\alpha, M) = \{f : [-1, 1] \to \mathbf{R}, |f(y) - f(x)| \le M|y - x|^\alpha, \forall x, y\}. \tag{2.2}$$

and $\mathcal{F}_D(\alpha, M) = \mathcal{F}(\alpha, M) \cap \{f : f(x) \ge f(y), \text{ if } x < y\}$. We want to estimate a linear functional of f , namely $f(0)$, based on X_n. We shall construct an estimator $\hat{f}_n$ that depends on neither α nor M and prove that the rate of its maximum risk coincides with that of the minimax risk for the Lipschitz class $\mathcal{F}(\alpha, M)$ that is $M^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}}$ for the mean square error loss. The existence of such an adaptive estimator is a new result.

The construction of an adaptive estimate for the value of a function at a given point when the function is monotone can be described in the following way. Let K be a kernel with support on $[-1, 1]$. For $j = 0, 1, 2, \ldots$ let $h_j = 2^{-j} n^{-\frac{1}{3}}$ and then we can define a sequence of kernel estimates for $f(0)$

$$\hat{f}_j = \int \frac{1}{h_j} K\left(\frac{t}{h_j}\right) X_n(dt). \tag{2.3}$$

Then, a proper bandwidth selection scheme will give us a choice of j based on the smoothness of f around zero. The bandwidth selection scheme is defined as follows. Let

$$\hat{f}_{Lj} = \frac{1}{h_j}[X_n(-h_j) - X_n(-h_{j-1})] \tag{2.4}$$

and

$$\hat{f}_{Rj} = \frac{1}{h_j}[X_n(h_{j-1}) - X_n(h_j)]. \tag{2.5}$$

Then for a fixed $\lambda > 0$ set

$$I_j = \mathbf{1}\left(\hat{f}_{Lj} - \hat{f}_{Rj} \leq \lambda 2^{\frac{j+1}{2}} n^{-\frac{1}{3}}\right) \prod_{i=0}^{j-1} \mathbf{1}\left(\hat{f}_{Lj} - \hat{f}_{Rj} > \lambda 2^{\frac{i+1}{2}} n^{-\frac{1}{3}}\right). \tag{2.6}$$

Note that $Var[\hat{f}_{Lj} - \hat{f}_{Rj}]$ is of the order $2^{j+1} n^{-\frac{2}{3}}$. In a word, we start with the longest bandwidth and choose j for which $\hat{f}_{Lj} - \hat{f}_{Rj}$ is smaller than a constant factor of its standard deviation for the first time. Since given $X_n(t)$ exactly one I_j is equal to 1, $\{I_j\}$ defines a bandwidth selection procedure. Now the adaptive procedure for estimating $f(0)$ can be written

$$\hat{f}_n = \sum_{j=0}^{\infty} \hat{f}_j I_j. \tag{2.7}$$

As mentioned before, Kiefer(1982) showed that there is a constant D such that

$$\inf_{\delta_n} \sup_{f \in \mathcal{F}_D(\alpha, M)} E\left[\delta_n - f(0)\right]^2 \geq D M^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}}. \tag{2.8}$$

Note that the minimax rate is the same rate as that of $\mathcal{F}(\alpha, M)$. Therefore, if we prove that the maximum risk of our estimator $\hat{f}_n$ is bounded from above by some constant factor of $M^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}}$ then our estimator is proven to be simultaneously minimax, in a word, adaptive. From the above construction we know that our estimator does not depend on the values of M and α but for any $\mathcal{F}_D(\alpha, M)$ its maximum risk is the same order as the rate of the minimax risk for a fixed $\mathcal{F}_D(\alpha, M)$. Therefore, the following theorem completes the adaptability of $\hat{f}_n$.

Theorem 2.1. *For the estimator $\hat{f}_n$ defined above with a positive constant λ, there is a constant C such that*

$$\sup_{f \in \mathcal{F}_D(\alpha, M)} E\left[\hat{f}_n - f(0)\right]^2 \leq C M^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}} \tag{2.9}$$

for all α and M where $0 < \alpha \leq 1$ and $M > 1$.

3. PROOF OF THE MAIN RESULT

Before proving the theorem, we shall present and prove the following lemmas first. Throughout the lemmas let $\tau_j = E[\hat{f}_{Lj} - \hat{f}_{Rj}]$ and suppose f is decreasing.

Lemma 3.1. *There is a constant C_1 not depending on f such that*

$$E\left[\hat{f}_j - f(0)\right]^2 \leq C_1 2^j n^{-\frac{2}{3}}. \tag{3.1}$$

Proof: Simple calculations show that

$$E\left[\hat{f}_j - f(0)\right]^2 I_j = E\left[\hat{f}_j - E\hat{f}_j\right]^2 I_j + \left[E\hat{f}_j - f(0)\right]^2 EI_j. \tag{3.2}$$

Note that $E\hat{f}_{Lj}$ is equal to the average of f over $[-h_{j-1}, -h_j]$ and $E\hat{f}_{Rj}$ is the average of f over $[h_j, h_{j-1}]$. Also, $E\hat{f}_j$ is a weighted average of f over $[-h_j, h_j]$. Therefore, for a decreasing function f

$$f(-h_j) - f(h_j) \leq \tau_j \tag{3.3}$$

and

$$f(h_j) \leq E\hat{f}_j \leq f(-h_j). \tag{3.4}$$

Thus,

$$|E\hat{f}_j - f(0)| \leq \tau_j. \tag{3.5}$$

On the other hand, $\hat{f}_j - E\hat{f}_j$ is a Normal random variable with mean 0 and variance $2^j n^{-\frac{2}{3}} \int K^2(t)dt$. So there is a constant D_1 such that

$$E\left[\hat{f}_j - E\hat{f}_j\right]^2 \leq D_1 2^j n^{-\frac{2}{3}}. \tag{3.6}$$

It is easy to see that the inequality in Lemma 3.1 holds when $\tau_j \leq 2\lambda 2^{\frac{j+1}{2}} n^{-\frac{1}{3}}$. We need to show it is true for $\tau_j \geq 2\lambda 2^{\frac{j+1}{2}} n^{-\frac{1}{3}}$. In this case,

$$\begin{aligned} EI_j &= E\mathbf{1}\left(\hat{f}_{Lj} - \hat{f}_{Rj} \leq \lambda 2^{\frac{j+1}{2}} n^{-\frac{1}{3}}\right) \prod_{i=0}^{j-1} \mathbf{1}\left(\hat{f}_{Li} - \hat{f}_{Ri} \geq \lambda 2^{\frac{i+1}{2}} n^{-\frac{1}{3}}\right) \\ &\leq E\mathbf{1}\left(\hat{f}_{Lj} - \hat{f}_{Rj} \leq \lambda 2^{\frac{j+1}{2}} n^{-\frac{1}{3}}\right). \end{aligned} \tag{3.7}$$

Since $\hat{f}_{Lj} - \hat{f}_{Rj}$ is a Normal random variable with mean τ_j and variance $2^{j+1} n^{-\frac{2}{3}}$, the following inequalities hold.

$$\begin{aligned} EI_j &\leq P\left(Z \leq \lambda - \frac{\tau_j n^{\frac{1}{3}}}{2^{\frac{j+1}{2}}}\right) \\ &\leq P\left(Z \leq -\frac{\tau_j n^{\frac{1}{3}}}{2^{\frac{j+3}{2}}}\right) \\ &\leq \exp\left(-\frac{\tau_j^2 n^{\frac{2}{3}}}{2^{j+4}}\right). \end{aligned} \tag{3.8}$$

Using calculus, we can show that

$$\begin{aligned} \tau_j^2 EI_j &\leq \tau_j^2 \exp\left(-\frac{\tau_j^2 n^{\frac{2}{3}}}{2^{j+4}}\right) \\ &= 2n^{-\frac{2}{3}} 2^{j+3} e^{-1} \\ &= 2^4 e^{-1} 2^j n^{-\frac{2}{3}}. \end{aligned} \tag{3.9}$$

Hence, for all τ_j we have proved that $E[\hat{f}_j - f(0)]^2 I_j$ is bounded by a constant factor of $2^j n^{-\frac{2}{3}}$. □

Lemma 3.2. *Suppose that f is decreasing and $\lambda > 0$. Suppose also that there is an integer J such that*

$$\tau_J = E\left[\hat{f}_{LJ} - \hat{f}_{RJ}\right] \leq c^* 2^{\frac{J}{2}} n^{-\frac{1}{3}}. \tag{3.10}$$

Then, the risk of the adaptive estimator $\hat{f}_n$ has an upper bound

$$E\left[\hat{f}_n - f(0)\right]^2 \leq C_2 2^J n^{-\frac{2}{3}} \tag{3.11}$$

where C_2 is a constant depending only on λ.

Proof: First we break the risk of $\hat{f}_n$ into two parts;

$$E\left[\hat{f}_n - f(0)\right]^2 = \sum_{j=0}^{J+k} E\left[\hat{f}_j - f(0)\right]^2 I_j + \sum_{j=J+k+1}^{\infty} E\left[\hat{f}_j - f(0)\right]^2 I_j. \tag{3.12}$$

Here the integer k depending on λ chosen later. An upper bound for the first term is easy to obtain by Lemma 3.1

$$\sum_{j=0}^{J+k} E\left[\hat{f}_j - f(0)\right]^2 I_j \leq \sum_{j=0}^{J+k} C_1 2^j n^{-\frac{2}{3}} \leq D_1 2^J n^{-\frac{2}{3}} \tag{3.13}$$

for some constant D_1.

For the analysis of the second term, we should use the independence among $\hat{f}_{Lj} - \hat{f}_{Rj}$'s. That is true due to the properties of the Brownian motion. For $j \geq J + k + 1$,

$$\begin{aligned} EI_j &\leq \prod_{i=J+k}^{j-1} P\left(\hat{f}_{Li} - \hat{f}_{Ri} > \lambda 2^{\frac{i+1}{2}} n^{-\frac{1}{3}}\right) \\ &\leq \left[P\left(Z > \lambda - \frac{c^*}{2^{\frac{k}{2}}}\right)\right]^{j-J-k} \\ &\leq \beta^{j-J-k}. \end{aligned} \tag{3.14}$$

In the above,

$$\beta = P\left(Z > \lambda - \frac{c^*}{2^{\frac{k}{2}}}\right) \tag{3.15}$$

and k is chosen to be $\lambda - \frac{c^*}{2^{\frac{k}{2}}} > 0$ so that $\beta < \frac{1}{2}$.

Now we have

$$\begin{aligned}
\sum_{j=J+k+1}^{\infty} E\left[\hat{f}_j - E\hat{f}_j\right]^2 I_j &= \sum_{j=J+k+1}^{\infty} E\left[\hat{f}_j - E\hat{f}_j\right]^2 EI_j \\
&\leq \sum_{j=J+k+1}^{\infty} D_2 2^j \beta^{j-J-k} n^{-\frac{2}{3}} \\
&= \sum_{j=J+k+1}^{\infty} D_3 2^J (2\beta)^{j-J} n^{-\frac{2}{3}} \\
&= D_4 2^J n^{-\frac{2}{3}}
\end{aligned} \tag{3.16}$$

for some positive constants D_2 and D_4. The first equality holds because of the independence between $\hat{f}_j$ and I_j. Recall their definitions. From the fact that the square of the bias of $\hat{f}_j$ is bounded by τ_j^2

$$\begin{aligned}
\sum_{j=J+k+1}^{\infty} E\left[E\hat{f}_j - f(0)\right]^2 &\leq \sum_{j=J+k+1}^{\infty} (c^*)^2 2^J n^{-\frac{2}{3}} EI_j \\
&\leq (c^*)^2 2^J n^{-\frac{2}{3}}.
\end{aligned} \tag{3.17}$$

The last inequality is true because

$$\sum_{j=J+k+1}^{\infty} EI_j \leq 1. \tag{3.18}$$

Hence, we have proved that the second term of the risk $\sum_{j=J+k+1}^{\infty} E\left[\hat{f}_j - f(0)\right]^2 I_j$ is also bounded from above by some constant factor of $2^J n^{-\frac{2}{3}}$. Now we ready to prove the main result. □

Proof: [Proof of the Theorem 2.1] First we have to show that for a function in $\mathcal{F}_D(\alpha, M)$, the assumption of Lemma 3.2 is satisfied, i.e., there is an integer J

such that $\tau_J \leq c^* 2^{\frac{J}{2}} n^{-\frac{1}{3}}$ for some constant c^*. If $f \in \mathcal{F}_D(\alpha, M)$, then

$$\begin{aligned} \tau_j &\leq E\left(\hat{f}_{Lj} - \hat{f}_{Rj}\right) \\ &\leq f\left(-2^{-(j-1)} n^{-\frac{1}{3}}\right) - f\left(2^{-(j-1)} n^{-\frac{1}{3}}\right) \\ &\leq M\left(2^{-(j-2)} n^{-\frac{1}{3}}\right)^{\alpha}. \end{aligned} \tag{3.19}$$

Now let J be the smallest integer such that

$$2^J n^{-\frac{2}{3}} \geq M^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}}. \tag{3.20}$$

Then if $j \geq J$,

$$\begin{aligned} 2^{j\alpha} &\geq M^{\frac{2\alpha}{2\alpha+1}} n^{\frac{2\alpha-2\alpha^2}{6\alpha+3}} \\ M 2^{-j\alpha} n^{-\frac{\alpha}{3}} &\leq M^{\frac{1}{2\alpha+1}} n^{-\frac{\alpha}{2\alpha+1}}. \end{aligned} \tag{3.21}$$

Combining all the above results, we can show that for $j \geq J$

$$\tau_j \leq 2^{2\alpha} 2^{\frac{J}{2}} n^{-\frac{1}{3}}. \tag{3.22}$$

Therefore the condition of Lemma 3.2 is satisfied with $c^* = 2^{2\alpha}$.

Since J is the smallest integer satisfying

$$2^J n^{-\frac{2}{3}} \geq M^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}}, \tag{3.23}$$

it follows that

$$2^J n^{-\frac{2}{3}} \leq 2M^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}}. \tag{3.24}$$

Using Lemma 3.2, we show that there is a constant C such that

$$\sup_{f \in \mathcal{F}_D(\alpha, M)} E\left[\hat{f}_n - f(0)\right]^2 \leq C M^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}}. \tag{3.25}$$

□

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical inference under order restrictions*, Wiley, New York.

Brown, L. and Low, M.(1996). "Asymptotic equivalence of nonparametric regression and white noise," *Annals of Statistics*, **24**, 2384-2398.

Brown, L. and Low, M. (1996). "A constrained risk inequality with applications to nonparametric functional estimation," *Annals of Statistics*, **24**, 2524-2535.

Donoho, D.L. and Liu, R.c. (1991). "Geometrizing rates of convergence, III," *Annals of Statistics*, **19**, 668-701.

Friedman, J. and Tibshirani, R. "(1984) The monotone smoothing of scatterplots," *Technometrics*, **26**, 243-250.

Grenander, U. (1956). "On the theory of mortality measurement, II," *Skand. Aktuarietidskr.* **39**, 125-153.

Hildenbrand, K. and Hildenbrand, W. (1986). "On the mean income effect: a data analysis of the U.K. family expenditure survey," In: *Contributions to Mathematical Econimics*, eds. W. Hildenbrand and A. Mas-Colell. New York: North Holland.

Ibragimov, I.A. and Hasminskii, R.Z. (1984). "On nonparametric estimation of the value of a linear functional in Gaussian White noise," *Theory of Probability and its Application*, **29**, 18-32.

Kelly, C. and Rice, J. (1990). "Monotone smoothing with application to dose-response curves and the assessment of synergism," *Biometrics*, **46**, 1071-1085.

Kiefer, J. (1982). "Optimum rates for non-parametric density and regression estimates, under order restrictions," *Statistics and probability : Essay in honor of C. R. Rao*, 419-428 (Kallianpur, G., Krishnaiah, P. R. and Ghosh, J. K. editors), North-Holland 1982.

Lepskii, O,V. (1991). "On a problem of adaptive estimation in Gaussian White noise," *Theory of Probability and its Application*, **35**, 454-466.

Lepskii, O.V. (1992). "On problems of adaptive estimation in White Gaussian noise," *Advances in Soviet Mathematics*, **12**, 87-106.

Low, M.G. (1992). "Renormalization and white noise approximation for nonparametric functional estimation problems," *Annals of Statistics*, **20**, 545-554.

Low, M.G. (1995). "Bias-variance tradeoffs in functional estimation problems," *Annals of Statistics*, **23**, 824-835.

Mammen, E. (1991). "Estimating a smooth monotone regression function," *Annals of Statistics*, **19**, 724-740.

Nussbaum, M. (1996). "Asymptotic equivalence of density estimation and Gaussian white noise," *Annals of Statistics*, **24**, 2399-2430.

Pinsker, M.S. (1980). "Optimal filtering of square integrable signals in Gaussian white noise," *Problems Inform. Transmission*, **16**, 52-68.

Reasenberg, P. A. and Jones, L. M.(1989). "Earthquake hazard after a main shock in southern california," *Science*, **243**, 1173-1176.

Samuelson, P. and Nordhaus, W. (1985). *Economics*, 12th ed., McGraw Hill.