Minimax optimal designs for nonparametric regression — a further optimality property of the uniform distribution

Stefanie Biedermann
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: Stefanie.Biedermann@ruhr-uni-bochum.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: Holger.Dette@ruhr-uni-bochum.de
FAX: +49 2 34 32 14 559

July 24, 2000

Abstract

In the common nonparametric regression model $y_i = g(t_i) + \sigma(t_i)\,\varepsilon_i,\ i = 1, \ldots, n$ with i.i.d. noise and nonrepeatable design points t_i we consider the problem of choosing an optimal design for the estimation of the regression function g. A minimax approach is adopted which searches for designs minimizing the maximum of the asymptotic integrated mean squared error, where the maximum is taken over an appropriately bounded class of functions (g, σ). The minimax designs are found explicitly, and for certain special cases the optimality of the uniform distribution can be established.

Keywords and phrases: Nonparametric regression, kernel estimation, locally optimal designs, minimax designs, mean squared error

AMS 1991 Subject Classifications: Primary 62K05; Secondary 62G05

1 Introduction

Consider the common nonparametric regression model

$$Y_{i,n} = g(t_{i,n}) + \sigma(t_{i,n})\varepsilon_{i,n} \qquad i = 1, \ldots, n, \tag{1.1}$$

where the $\varepsilon_{i,n}$ form a triangular array of rowwise independent identically distributed random variables with variance 1 and mean 0, g, σ are unknown smooth functions and $\{t_{i,n} \mid i = 1, \ldots, n\}$ is a fixed design in the interval $[0, 1]$. Much effort has been devoted to the problem of estimating the regression and variance function [see e.g. the recent monographs of Härdle (1990), Wand and Jones (1995), Fan and Gijbels (1996)], and many of the developed estimation methods are meanwhile standard methods in applied regression analysis.

Although it is well known (and intuitively clear) that the asymptotic properties of the various nonparametric estimators depend sensitively on the underlying designs, the problem of designing experiments in the nonparametric setup (1.1) has found much less attention in the literature. This is mainly due to the fact that – similar as in the case of nonlinear regression models – the optimality criteria usually depend on the unknown regression and variance function. Müller (1984) studied optimal designs for estimating derivatives of the regression function by minimizing the asymptotic integrated mean squared error of a kernel estimate. He showed that the optimality criteria are local in the sense of Chernoff (1953) and the asymptotically optimal designs depend on the unknown regression and variance function. Because of these difficulties Cheng, Hall and Titterington (1998) proposed a sequential approach for defining an optimal design measure in the context of local linear regression.
A further optimality criterion is considered in the present paper, which is different from the methods considered by the aforementioned authors and mainly motivated by two observations. On the one hand the designs proposed by Müller (1984) are not robust with respect to misspecifications of the regression and variance function, on the other hand there are many situations where sequential designs cannot be applied (e.g. experiments in agriculture). For these reasons, we propose a minimax approach which seeks for designs minimizing the maximum of the asymptotic integrated mean squared error, where the maximum is taken over certain Sobolev balls for the regression and variance function. Section 2 states some basic terminology from nonparametric kernel estimation, which is necessary for the formulation of the minimax optimality criterion. The main results can be found in Section 3 showing that the minimax optimal designs are equal to the weight function used in the definition of the integrated mean squared error. This highlights the particular role of the uniform distribution, which turns out to be the minimax optimal design with respect to the classical integrated mean squared error criterion.

2 Asymptotic representation of the integrated mean squared error of a nonparametric regression estimator

For the introduction of the optimality criterion we need an asymptotic representation of the integrated mean squared error of a nonparametric estimate of the regression function g, which is nowadays standard in mathematical statistics [see Härdle (1990), Wand and Jones (1995) or Fan and Gijbels (1996)]. We assume that the regression function is k-times continuously differentiable, i.e. $g \in C^k[0,1]$, and the variance function is Lipschitz continuous of order γ, i.e. $\sigma \in \mathrm{Lip}_\gamma[0,1]$ for some $\gamma \in (0,1]$. Following Gasser and Müller (1984), we consider

$$g_{n,v}(t) = \frac{1}{h^{v+1}(t)} \sum_{i=1}^{n} \int_{s_{i-1,n}}^{s_{i,n}} K_v\Big(\frac{t-x}{h(t)}\Big) dx \cdot Y_{i,n} \tag{2.1}$$

as an estimate of the v-th derivative $g^{(v)}$ of the regression function ($v = 0, 1, \ldots, k-1$). Here K_v is a Lipschitz continuous function with compact support, say $[-1,1]$, such that

$$\frac{(-1)^j}{j!} \int_{-1}^{1} K_v(x) x^j dx = \begin{cases} 0 & \text{if } 0 \le j < k; j \ne v \\ 1 & \text{if } j = v \\ B & \text{if } j = k \end{cases}, \tag{2.2}$$

where $B \ne 0$. The triangular array $\{s_{i,n} \mid i = 0, \ldots, n\}$ is defined by $s_{0,n} = 0, s_{n,n} = 1, s_{i,n} = (t_{i,n} + t_{i+1,n})/2 \quad (i = 1, \ldots, n-1)$, and the design points $t_{i,n}$ are supposed to satisfy a Sacks and Ylvisaker (1970) condition

$$\int_0^{t_{i-1,n}} f(t)dt = \frac{i-1}{n-1} \qquad i = 1, \ldots, n \tag{2.3}$$

with a positive design density f. The quantity $h(t)$ denotes the bandwidth which may depend on a specific point t (local bandwidth). The locally optimal bandwidth minimizes the asymptotic mean squared error at the point t and is given by

$$h^*(t) = \Big\{\frac{(2v+1)V\sigma^2(t)}{2(k-v)nB^2 f(t)(g^{(k)}(t))^2}\Big\}^{1/(2k+1)}, \tag{2.4}$$

where $V = \int_{-1}^1 K_v^2(x)dx$ [see Gasser and Müller (1984)], and it is assumed that $g^{(k)}(t) \neq 0$. Insertion of the optimal bandwidth in the weighted integrated mean squared error

$$E\int_0^1 \{\hat{g}_{n,v}(t) - g^{(v)}(t)\}^2 w(t)dt$$

yields [using an appropriate modification of the kernel at the boundary, see Müller (1984)]

$$n^{2(k-v)/(2k+1)} E\int_0^1 \{\hat{g}_{n,v}(t) - g^{(v)}(t)\}^2 w(t)dt = c \cdot \Psi_{g,\sigma}^{(v)}(f) + o(1)\,, \tag{2.5}$$

where w is a positive continuous weight function on the interval $[0,1]$, c is a constant independent of f and

$$\Psi_{g,\sigma}^{(v)}(f) = \int_0^1 \Big\{\frac{\sigma^2(t)w(t)}{f(t)}\Big\}^{2(k-v)/(2k+1)} \{(g^{(k)}(t))^2 w(t)\}^{(2v+1)/(2k+1)} dt. \tag{2.6}$$

3 Optimal designs minimizing the maximum integrated mean squared error

Müller (1984) determined the design density f which minimizes the criterion (2.6) for fixed g, σ^2, i.e.

$$f^*(t) = \frac{\varphi(t)}{\int_0^1 \varphi(x)dx} \tag{3.1}$$

where $\varphi(t) = [(\sigma^2(t))^{2(k-v)}(g^{(k)}(t))^{2(2v+1)}(w(t))^{2k+1}]^{1/(4k+1-2v)}$. The optimal design is local in the sense of Chernoff (1953) and might be not robust with respect to misspecification of the variance and regression function [see also Example 3.2]. For this reason we propose a minimax criterion for the determination of optimal designs and call a design density f^* minimax optimal for the estimation of the vth derivative of the regression function if it minimizes

$$\max\{\Psi_{g,\sigma}^{(v)}(f) \mid (g,\sigma) \in \mathcal{F}\}. \tag{3.2}$$

Here $\mathcal{F}$ is an appropriate class of functions given either by

$$\mathcal{F}_2 = \Big\{(g,\sigma) \in C^{(k)}[0,1] \times \text{ Lip}_\gamma[0,1] \quad \Big| \int_0^1 \sigma^2(t)dt \leq \varepsilon, \int_0^1 (g^{(k)}(t))^2 w(t)dt \leq \eta\Big\} \tag{3.3}$$

or by

$$\mathcal{F}_\infty = \Big\{(g,\sigma) \in C^{(k)}[0,1] \times \text{ Lip}_\gamma[0,1] \quad \Big| \sup_{t\in[0,1]} |\sigma^2(t)| \leq \varepsilon, \sup_{t\in[0,1]} |g^{(k)}(t)|^2 w(t) \leq \eta\Big\} \tag{3.4}$$

Theorem 3.1 *Let $\mathcal{F}$ be either defined by (3.3) or (3.4) then the design with density*

$$f^*(t) = \frac{w(t)}{\int_0^1 w(x)dx}$$

is optimal with respect to the minimax criterion (3.2).

Proof: We consider only the case $\mathcal{F} = \mathcal{F}_2$, the corresponding result for the sup-norm is proved by similar arguments. Without loss of generality it is assumed that $\int_0^1 w(t)dt = 1$, i.e. $f^* = w$. The proof is performed in two steps showing

$$(1) \qquad \sup\{\Psi_{g,\sigma}^{(v)}(f^*) \mid (g,\sigma) \in \mathcal{F}_2\} = [\varepsilon^{2(k-v)}\eta^{2v+1}]^{1/(2k+1)}$$

$$(2) \qquad \forall\, f\ \exists\,(g,\sigma) \in \mathcal{F}_2 : \Psi_{g,\sigma}^{(v)}(f) \geq [\varepsilon^{2(k-v)}\eta^{2v+1}]^{1/(2k+1)}$$

(1) The first part follows by a direct application of Hölder's inequality observing that with the notation $p = \frac{2k+1}{2(k-v)}$, $q = \frac{2k+1}{2v+1}$ the integrated mean squared error is given by

$$\begin{aligned} (3.5) \qquad \Psi_{g,\sigma}^{(v)}(w) &= \int_0^1 |\sigma^2(t)|^{1/p}|(g^{(k)}(t))^2 w(t)|^{1/q}dt \\ &\leq \Big\{\int_0^1 \sigma^2(t)dt\Big\}^{1/p}\Big\{\int_0^1 (g^{(k)}(t))^2 w(t)dt\Big\}^{1/q} \leq \varepsilon^{1/p}\eta^{1/q}, \end{aligned}$$

and that there is equality in (3.6) for the functions

$$(3.6) \qquad \sigma(t) \equiv \sqrt{\varepsilon}; \quad g(t) = \sqrt{\eta}\int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} \frac{ds}{\sqrt{w(s)}}$$

[note that $(g,\sigma) \in \mathcal{F}_2$].

(2) Let f denote an arbitrary positive density on the interval $[0,1]$ and let

$$(3.7) \qquad p' = 1 + \frac{1}{p} = \frac{4k+1-2v}{2k+1}; \quad q' = 1 + p = \frac{4k+1-2v}{2(k-v)}$$

(note that $1/p' + 1/q' = 1$). For the functions

$$(3.8) \qquad \sigma_*(t) = \sqrt{\varepsilon w(t)}; \quad g_*(t) = \sqrt{\eta}\frac{t^k}{k!}$$

it follows (observing $\int_0^1 w(t)dt = 1$) that $(g,\sigma) \in \mathcal{F}_2$ and

$$\begin{aligned} \Psi_{g_*,\sigma_*}^{(v)}(f) &= \varepsilon^{1/p}\eta^{1/q}\int_0^1 |\frac{w(t)}{f(t)}|^{1/p} w(t)dt = \varepsilon^{1/p}\eta^{1/q}\int_0^1 (w(t))^{p'}\frac{dt}{|f(t)|^{p'/q'}} \\ &= \varepsilon^{1/p}\eta^{1/q}\|\frac{w}{|f|^{1/q'}}\|_{p'}^{p'} \geq \varepsilon^{1/p}\eta^{1/q}\Big\{\frac{\|w\|_1}{\|f^{1/q'}\|_{q'}}\Big\}^{p'} = \varepsilon^{1/p}\eta^{1/q} \end{aligned}$$

where $\|s\|_p = (\int_0^1 |s(t)|^p dt)^{1/p}$ denotes the L^p-norm with respect to the Lebesgue measure and the inequality follows again from Hölder's inequality. This proves (2) and completes the proof of Theorem 3.1.

□

Example 3.2. Consider the case $k = 2, v = 0$ and $w(x) \equiv 1$. The locally optimal design (with respect to the asymptotic integrated MSE) is given by (3.1) and we are interested in how a misspecification of the regression or variance function affects the performance of this design. To this end we consider the following scenario

$$\begin{array}{lll} (1) & g(x) = e^x; & \sigma(x) = 1 \\ (2) & g(x) = e^x; & \sigma(x) = e^x \\ (3) & g(x) = e^x; & \sigma(x) = e^{2x} \\ (4) & g(x) = \sin x; & \sigma(x) = 1 \\ (5) & g(x) = \sin x; & \sigma(x) = e^x \\ (6) & g(x) = \sin x; & \sigma(x) = e^{2x}. \end{array}$$

For a misspecification we investigate four cases

$$\begin{array}{lll} (A) & g \text{ correct}; & \sigma \text{ constant} \\ (B) & g \text{ correct}; & \sigma^{-1} \text{ instead of } \sigma \\ (C) & g^2 \text{ instead of } g; & \sigma \text{ correct} \\ (D) & e^g \text{ instead of } g; & \sigma \text{ correct} \end{array}$$

and calculate the locally optimal design from Müller (1984). Note that in case (A), (B) the variance function is misspecified, while (C), (D) correspond to an incorrect assumption for the regression function. The corresponding results are listed in Table 3.1, which also contains a column for the minimax design which turns out to be the uniform design in this case. The table shows the asymptotic integrated MSE obtained by the particular design f^* in (3.1) for a misspecification relative to the asymptotic integrated MSE obtained by the design f^* for the correct regression and variance function.

	A	B	C	D	minimax
1	1	1	0.4650	0.4625	0.9970
2	0.9562	0.8374	0.4519	0.4494	0.9326
3	0.9573	0.8406	0.4404	0.4381	0.9066
4	1	1	0.7193	0.6676	0.9776
5	0.9600	0.8502	0.8289	0.7977	0.8997
6	0.8658	0.5643	0.9379	0.7894	0.7894

Table 3.1. *Asymptotic relative efficiency of the optimal design f^* in (3.1) for various misspecifications of the regression and variance function. The last column: minimax design obtained from Theorem 3.1.*

We observe that a misspecification of the regression and variance function has a serious impact on the performance of the resulting design. This dependency is even stronger, if the regression function is not adequately specified. The minimax design has a reasonable performance in all considered cases. For these reasons the locally optimal designs proposed by Müller (1984) should not be used in practice unless there is information about the variance and regression structure available. Otherwise the minimax designs proposed in this paper are recommended.

Acknowledgements: The authors would like to thank I. Gottschlich for typing most parts of this paper with considerable technical expertise. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, Reduction of complexity in multivariate data structures) is gratefully acknowledged.

References

Cheng, M.-H., Hall, P., Titterington, D.M. (1998). Optimal design for curve estimation by local linear smoothing. Bernoulli 4, 3-14.

Chernoff, H. (1953). Locally optimal design for estimating parameters. Ann. Math. Statist. 24, 586-602.

Fan, J., Gijbels, I. (1996). Local polynomial modelling and its applications. Chapman and Hall, London

Gasser, T., Müller, H.G. (1984). Estimating regression functions and their derivatives by kernel functions. Scand. J. of Statist. 11, 171-185.

Härdle, W. (1990). Applied nonparametric regression. Cambridge University Press, Boston.

Hall, P., Caroll, R.J. (1989). Variance function estimation in regression: the effect of estimating the mean. J. Roy. Statist. Soc. B 51, 3-14.

Hall, P., Kay, J.W., Titterington, D.M. (1990). Asymptotically optimal difference – based estimation of variance in nonparametric regression. Biometrika 77, 521-528.

Müller, H.G. (1984). Optimal designs for nonparametric kernel regression. Stat. & Probab. Letters 2, 285-290.

Müller, H.G., Stadtmüller, U. (1993). On variance function estimation with quadratic forms. J. Statist. Plan. Inf. 35, 213-231.

Rice, J. (1984). Bandwidth choice for nonparametric kernel regression. Ann. Statist. 12, 1215-1230.

Sacks, J., Ylvisaker, D. (1970). Designs for regression problems for correlated errors. Ann. Math. Statist. 41, 2057-2074.

Wand, M.P., Jones, M.C. (1995). Kernel smoothing. Chapman and Hall, London.