Z. Hlávka

On nonparametric estimators of location of maximum

Acta Universitatis Carolinae. Mathematica et Physica, Vol. 52 (2011), No. 1, 5--13

Persistent URL: http://dml.cz/dmlcz/143663

On Nonparametric Estimators of Location of Maximum

ZDENĚK HLÁVKA

Praha

Received April 30, 2010
Revised June 19, 2010

An estimator of the maximum of a regression function and its location is often of greater interest than an estimator of the regression curve itself. We review properties of nonparametric estimators of the location of maximum and investigate the influence of the density of design points on the asymptotic distribution of the estimator. Classical calculus of variations is used to find the optimal distribution of the design points for the nonparametric kernel estimator of the location of maximum.

1. Introduction

Let us consider the nonparametric regression model $Y_i = m(x_i)+\varepsilon_i$, where the fixed design points $0 = x_1 < x_2 < \cdots < x_n = 1$ are such that $F_X(x_i) - F_X(x_{i-1}) = 1/(n-1)$, for some distribution function $F_X(.)$. The symbol $f_X(.)$ denotes the corresponding probability density function if it exists.

Our investigations will be based on the Gasser-Müller (GM) kernel regression estimate [1]:

$$\hat{m}(x) = \frac{1}{b_n}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{b_n}\right) du\, Y_i, \tag{1}$$

where $s_{i-1} = \frac{1}{2}(x_i + x_{i-1})$, b_n is the bandwidth and $K(.)$ the kernel function.

The parameter of interest, θ, is the location of the maximum of the regression function $m(.)$, i.e., $\theta = \arg\max_{x\in(0,1)} m(x)$. The natural estimator of the location of maximum is the so-called empirical location of the maximum, i.e., $\hat{\theta}_n = \arg\max_{x\in(0,1)} \hat{m}(x)$, see [3].

MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín

The work was partially supported by GAČR 201/08/0486 and MSM 0021620839.

Key words and phrases. Kernel regression, location of maximum, optimal design

E-mail address: hlavka@karlin.mff.cuni.cz

In Section 2, we review the effect of the distribution of design points on the asymptotic distribution of the estimator. The optimal design of the experiment is derived in Section 3. A simulation study is carried out in Section 4 and Section 5 concludes.

2. Nonparametric estimator of location of maximum

In order to investigate the dependency of the asymptotic distribution of the empirical location of the maximum, $\hat{\theta}_n$, on the density of the design points, f_X, we need some assumptions on the regression function $m(.)$, the kernel function $K(.)$, and the bandwidth b_n:

(A1) Assume that $m(.)$ has unique maximum in $\theta \in (0, 1)$, $m(.)$ is four times continuously differentiable, and there exist $x_l < \theta < x_u$, $c > 0$, and $\rho \geq 1$ such that $m(.)$ is increasing on $\langle x_l, \theta\rangle$ and decreasing on $\langle \theta, x_u\rangle$ and $|m(t) - m(\theta)| > > c|t - \theta|^\rho$ for $t \in \langle x_l, x_u\rangle$.

(A2) The kernel function $K(.)$ is of order $(0, 2)$, cf. [1, 3], i.e., $B_0 = \int K(u)du = 1$, $B_1 = \int K(u)udu = 0$, and $B_2 = (1/2)\int K(u)u^2du$ is finite. Assume that the kernel $K(.)$ is two times continuously differentiable and that $K^{(2)}(.)$ is Lipschitz continuous.

(A3) Assume that $\liminf_{n\to\infty} nb_n^4 > 0$, $nb_n^5/\log n \to \infty$ and that for some $r > 2$ the conditions $E|\varepsilon_1|^r < \infty$ and $\liminf_{n\to\infty} b_n n^{1-2/r}/\log n > 0$ are satisfied.

Lemma 1 *Assume that assumptions (A1)–(A3) hold and that the design points are uniformly distributed. If* $nb_n^7 \to d^2 \geq 0$ *then*

$$(nb_n^3)^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathscr{D}} N\left(-\frac{dm^{(3)}(\theta)B_2}{m^{(2)}(\theta)}, \frac{\sigma^2 V'}{\{m^{(2)}(\theta)\}^2}\right),$$

where $V' = \int\{K^{(1)}(u)\}^2du$.

Proof. See Theorem 3.1(B) in [3] with $k = 2$ and $\nu = 0$. □

Thus, choosing b_n such that $nb_n^7 \to 0$, for example $b_n = n^{-1/6}$, and denoting by $u_{1-\alpha/2}$ the $1 - \alpha/2$ quantile of the standard Normal distribution, we obtain an approximate $1 - \alpha$ confidence interval for θ as

$$\left(\hat{\theta}_n - \frac{u_{1-\alpha/2}}{\sqrt{nb_n^3}}\frac{\sigma\sqrt{V'}}{m^{(2)}(\theta)}, \hat{\theta}_n + \frac{u_{1-\alpha/2}}{\sqrt{nb_n^3}}\frac{\sigma\sqrt{V'}}{m^{(2)}(\theta)}\right). \tag{2}$$

The simulated example in Figure 1 provides a simple illustration. For two bandwidths, the true regression curve $m(x) = \cos\{2\pi(x - 0.3)\}$ and its estimate, $\hat{m}(x)$, are denoted by the dotted and the solid line, respectively. Similarly, the dotted and solid vertical lines denote respectively the true location of maximum θ and the empirical location of maximum $\hat{\theta}_n$. The asymptotic 95% confidence intervals for the location of maximum are plotted as two dashed vertical lines in each plot. For both bandwidths

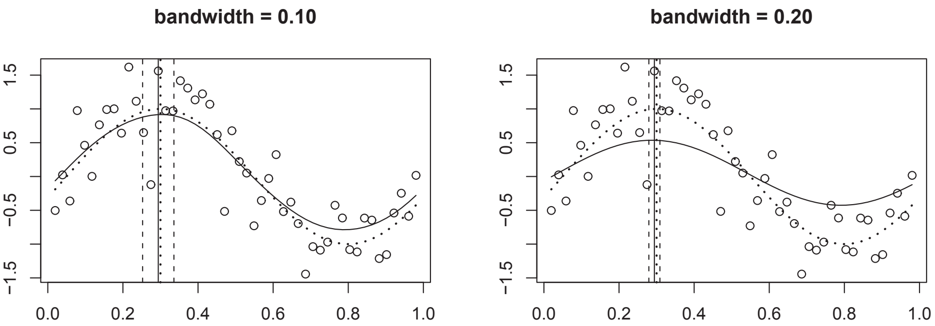

FIGURE 1. Two estimates of the location of the maximum with uniformly distributed design points and different bandwidths. The true regression line and the true location of maximum are denoted by dotted lines. Solid lines denote the GM kernel regression estimates and the corresponding empirical location of maxima, dashed lines mark 95% confidence intervals for the location of maximum

in Figure 1, the approximate 95% confidence intervals cover the true location of maximum $\theta = 0.3$. Notice that the shorter confidence interval on the right-hand side plot in Figure 1 covers the true value of the parameter θ although the regression function estimate $\hat{m}(x)$ is highly biased due to large bandwidth.

2.1 Non-uniformly distributed design points

Let us now investigate the influence of the distribution of design points on the asymptotic behavior of the empirical location of maximum.

First, we summarize and reformulate some already known results [2, 3] concerning the variance of the nonparametric GM kernel regression estimator, $\hat{m}(x)$, and the variance of its derivative, $\hat{m}^{(1)}(x)$.

Lemma 2 *Assume that assumptions (A1)–(A3) hold and that the design points are distributed according to a probability density $f_X(.)$ satisfying assumption (A4):*

(A4) *The density of the design points, $f_X(.)$, is strictly positive and there exists $\gamma \in (0, 1\rangle$ and $L > 0$ such that $|f(x_1) - f(x_2)| \leq L|x_1 - x_2|^{\gamma}$ for all $x_1, x_2 \in \langle 0, 1\rangle$.*

Then

$$Var\{\hat{m}(x)\} = \frac{\sigma^2}{nb_n}\left[\int_{-1}^{1} K^2(u)du\{f(x)\}^{-1} + o(1)\right].$$

If the first derivative of the kernel function, $K^{(1)}(.)$, is Lipschitz continuous then

$$Var\{\hat{m}^{(1)}(x)\} = \frac{\sigma^2}{nb_n^3}\left[\int_{-1}^{1} \{K^{(1)}(u)\}^2 du\{f(x)\}^{-1} + o(1)\right],$$

where $\hat{m}^{(1)}(x) = \frac{1}{b_n^2}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K^{(1)}\{(x-u)/b_n\}duY_i$.

Proof. See page 288 in [2] and pages 225–226 in [3]. □

The asymptotic distribution of the empirical location of maximum with non-uniformly distributed design points may be derived by proceeding similarly as in [3].

Theorem 1 *Assume that assumptions (A1)–(A3) hold and that the design points are distributed according to* $f_X(.)$ *satisfying assumption (A4). If* $nb_n^7 \to d^2 \geq 0$ *then*

$$(nb_n^3)^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathscr{D}} N\left(-\frac{dm^{(3)}(\theta)B_2}{m^{(2)}(\theta)}, \frac{\sigma^2 V'}{\{m^{(2)}(\theta)\}^2 f_X(\theta)}\right).$$

Proof. In the same way as in the proof of Theorem 3.1 in [3] it can be shown that, for any $0 < \delta < 1/2$, $\sup_{t\in\langle\delta,1-\delta\rangle}|\hat{m}^{(2)}(t) - m^{(2)}(t)| \to 0$ a.s. and by Lemma 2.3 in [3] it follows that $|\hat{m}^{(2)}(\hat{\theta}_n) - m^{(2)}(\theta)| \to 0$ a.s.

Notice that $\hat{m}^{(1)}(\hat{\theta}_n) = m^{(1)}(\theta) = 0$ by definition and $m^{(2)}(\theta) < 0$ by (A1). Next, from the Taylor expansion of $\hat{m}^{(1)}(\hat{\theta}_n)$ it follows that:

$$\begin{aligned}
\hat{\theta}_n - \theta &= \frac{\hat{m}^{(1)}(\hat{\theta}_n) - \hat{m}^{(1)}(\theta)}{\hat{m}^{(2)}(\theta^*)} \\
&= \frac{m^{(1)}(\theta) - \hat{m}^{(1)}(\theta)}{m^{(2)}(\theta)}\left\{\frac{\hat{m}^{(2)}(\theta^*) + m^{(2)}(\theta) - \hat{m}^{(2)}(\theta^*)}{\hat{m}^{(2)}(\theta^*)}\right\} \\
&= \frac{m^{(1)}(\theta) - \hat{m}^{(1)}(\theta)}{m^{(2)}(\theta)} + R_n,
\end{aligned}$$

for some $\theta^* \in (\theta, \hat{\theta}_n)$ with the remainder term:

$$R_n = \{m^{(1)}(\theta) - \hat{m}^{(1)}(\theta)\}\{m^{(2)}(\theta) - \hat{m}^{(2)}(\theta^*)\}/\{m^{(2)}(\theta)\hat{m}^{(2)}(\theta^*)\}.$$

Finally, the second part of Lemma 2 implies that

$$Var\{\hat{m}(\theta)^{(1)}\} = \frac{\sigma^2}{nb_n^3}\left[\int_{-1}^{1}\{K^{(1)}(u)\}^2 du\{f(t)\}^{-1} + o(1)\right]$$

and the result follows as in the proof of Theorem 3.1 in [3]. □

Theorem 1 describes the dependency of the variance of the empirical location of maximum, $\hat{\theta}_n$, on the distribution of the fixed design points $f_X(.)$. In the following Section 3, we address the naturally occurring practical question of finding the optimal design of the experiment and we propose a density of design points minimizing the variability of the empirical location of maximum.

3. Optimal distribution of design points

The problem of finding the optimal distribution of design points in nonparametric kernel regression has been previously addressed in [2] from the point of view of the integrated mean squared error (IMSE) of the GM kernel regression estimator $\hat{m}(x)$.

Choosing a probability measure H with a positive and continuous density $h(.)$ on $\langle 0, 1\rangle$ and considering:

$$IMSE = E\int\{\hat{m}(x) - m(x)\}dH(x) \approx \frac{1}{nb_n}\int K^2(u)du\int\frac{h(x)}{f(x)}dx,$$

the AIMSE (asymptotic IMSE) optimal density of the design points $f_X^*(x) = h(x)^{1/2}/\int h(u)^{1/2}du \propto h(x)^{1/2}$ has been derived in [2].

Unfortunately, the AIMSE optimal design is very difficult to interpret and almost impossible to apply in practice because the probability measure H lacks any clear interpretation. In the following Section 3, we overcome this obstacle by obtaining similar designs that minimize the variability of the empirical location of maximum.

3.1 Optimal distribution of design points for the estimation of the location of maximum

In practical problems concerning the estimation of the location of maximum, we may be able to gather a prior information concerning the location of the maximum. Such information might stem from a preliminary stage of the experiment or from past experience of other researchers. Let the symbol A denote a probability measure corresponding to the prior distribution of the location of the maximum and let us assume that A has a positive and continuous density $a(.)$ such that:

(A5) There exists $\delta > 0$ such that $a(x) > \delta$, for all $x \in \langle 0, 1\rangle$.

For simplicity, we assume throughout this section that σ^2 and $m^{(2)}(\theta)$ are constant. The following Theorem 2 suggests the optimal choice of the design points using the information contained in the prior density, $a(.)$, of the location of maximum θ.

Theorem 2 *Assume that (A1)–(A3) and (A5) hold,* $0 < \sigma^2 < \infty$, *and that* $m^{(2)}(\theta) = m_2$ *does not depend on the location of maximum* θ.

(1) *Assuming that the prior density* $a(.)$ *satisfies (A4), the density of design points* $f_V(x) \propto a^{1/2}(x)$ *minimizes the expectation of the asymptotic variance of the empirical location of maximum,* $\int Var(\hat{\theta}_n|\theta = u)a(u)du$, *with respect to the prior density* $a(.)$.

(2) *Assuming that* $a^{4/3}(.)$ *satisfies (A4), the density of design points* $f_L(x) \propto a^{2/3}(x)$ *minimizes the expected length of confidence intervals with respect to the prior density* $a(.)$.

Proof. We prove only the first part because the second part is very similar. At first, we recall that for a density of design points $f_X(.)$ satisfying (A4) it follows from Theorem 1 that $Var(\hat{\theta}_n|\theta = x) = cf_X^{-1}(x)$, where c is a constant depending on n, b_n, $K(.)$, σ^2, and m_2.

The minimization problem

$f_V = \arg\min_{f_X}\int_0^1 Var(\hat{\theta}_n|\theta = x)a(x)dx = \arg\min_{f_X}\int_0^1 f_X^{-1}(x)a(x)dx$ belongs to the classical calculus of variations. Denoting $F(x, y, y') = F(x, F_X, f_x) = f_X^{-1}(x)a(x)$, the neccessary condition for an extreme of $I(f_X) = I(y') = \int_0^1 F(x, y, y')dx$ is $F'_y - \frac{\mathrm{d}}{\mathrm{d}x} F'_{y'} =$ $= 0$, see e.g. [4, 6]. In our setup, $F'_y = 0$ and $F'_{y'}(x) = -f_X^{-2}(x)a(x)$ and the above condition thus implies that the optimal density of design points $f_V(.)$ has to satisfy $\frac{\mathrm{d}}{\mathrm{d}x}\{f_V^{-2}(x)a(x)\} = 0$, i.e., $f_V^{-2}(x)a(x) = \text{constant}$.

Next, let $f_V^\star(x) \propto a^{1/2}(x)$ denote the candidate solution and notice that it satisfies (A4) because $a^{1/2}(x_1) - a^{1/2}(x_2) < \{a^{1/2}(x_1) - a^{1/2}(x_2)\}\{a^{1/2}(x_1) + a^{1/2}(x_2)\}/\delta^{1/2} =$ $= \{a(x_1) - a(x_2)\}/\delta^{1/2}$ by (A5) and $a(.)$ satisfies (A4) by our assumptions.

It remains to verify that the candidate solution $f_V^\star(.)$ minimizes the expected variance. Considering another probability density functions f_Y and $f_Z = \alpha f_Y + (1 - \alpha) f_V^\star$ for $\alpha \in \langle 0, 1\rangle$ and defining $k = \{\int_0^1 a^{1/2}(u)du\}^{-1}$ and $Z(\alpha) = \int_0^1 f_Z^{-1}(x)a(x)dx =$ $= \int_0^1 [\alpha\{f_Y(x) - ka^{1/2}(x)\} + ka^{1/2}(x)]^{-1} a(x)dx$, it is easy to verify that $Z(\alpha)$ is continuously differentiable, $Z'(0) = 0$ and, if f_Y and $f_V^\star$ are not equal A-a.e., $Z^{(2)}(\alpha) > 0$, for $\alpha \in \langle 0, 1\rangle$. This implies that $Z'(\alpha) > 0$ for $\alpha \in (0, 1\rangle$ and, therefore, $I(f_Y) =$ $= \int f_Y^{-1}(x)a(x)dx = Z(1) > Z(0) = \int \{f_V^\star(x)\}^{-1} a(x)dx = I(f_V^*)$ and the assertion follows. □

4. Simulation study

The simulation study was implemented in the statistical computing environment R [5]. All simulation results are based on the Gasser–Müller (GM) kernel regression estimator using the quartic kernel and 2500 simulations.

In order to avoid boundary problems, we consider periodic regression functions on $\langle 0, 1\rangle$ and we calculate the GM estimator as if the design was circular, i.e., as if the design points were located on a circle.

Throughout the simulation study, the prior distribution of the location of maximum is $\theta \sim N(0.4, 0.01)$ restricted to the interval $\langle 0, 1\rangle$. More precisely, the prior density of the location of maximum is:

$$a(\theta) \propto \exp\{-5(\theta - 0.4)^2\} I(\theta \in \langle 0, 1\rangle),$$

where $I(.)$ denotes the indicator function.

The density of the design points in the simulation study will be controlled by a parameter r: for a fixed value of the parameter r, the density of the design points, $f_{X,r}(.)$, is proportional to the r-th power of the prior density $a(.)$, i.e., $f_{X,r}(x) \propto a^r(x)$. For example, the value $r = 0$ corresponds to uniformly distributed design points, $f_{X,0}(x) = I(x \in \langle 0, 1\rangle)$. The value $r = 1$ would mean that the density of design points is equal to the prior density of the location of maximum, $f_{X,1}(x) = a(x)$. Higher values of the parameter r mean that the design points are more concentrated in the neighborhood of the mode of the prior distribution $a(\theta)$.

In each step of the simulation, for a fixed sample size n, the standard deviation $\sigma \in \{0.1, 0.01\}$, the bandwidth $b_n \in \langle 0.02, 0.50 \rangle$, and the parameter controlling the density of design points $r \in \langle 0, 1.2 \rangle$, we:

(1) calculate the design points according to the density $f_{X,r}(x) \propto a^r(x)$,
(2) simulate the responses $Y_i = m(x_i) + \sigma\varepsilon_i$, where $m(.)$ is a chosen regression function and ε_i are iid $N(0, 1)$ (pseudo-)random variables,
(3) calculate the empirical location of maximum using the GM estimator with bandwidth b_n and function `optimise()` in R [5].

For each sample size n, standard deviation σ, the bandwidth b_n, and each value of the parameter r, we then calculate the Mean Squared Error (MSE) and the Mean Absolute Deviation (MAD) of the empirical location of maximum from 2500 simulations. In all tables, the MSE and MAD are presented only for the best value of the parameter r (denoted as r_{opt}) for each bandwidth. The bandwidths with the smallest MSE or MAD are denoted by $\star$. For example, in Table 1, for $n = 20$ observations and $\sigma = 0.1$, the MSE is minimized for the bandwidth $b_n = 0.15$ and the density of design points proportional to $a^{0.2}(x)$ ($r_{opt} = 0.20$).

In Table 1, we investigate the regression function:

$$m_1(x) = \cos\{2\pi(x - \theta)\},$$

where the location of maximum is equal to the parameter θ. In this situation, Theorem 1 implies that for bandwidths decreasing with the sample size at an appropriate rate, the empirical location of maximum $\hat{\theta}_n$ is an asymptotically unbiased estimator of θ since $m_1^{(3)}(\theta) = 0$. However, contrary to our expectations and also contrary to assumptions of Theorem 1, we observe in Table 1 that the optimal bandwidth actually increases with the sample size. This strange behavior could be caused by the symmetry of the regression function $m_1(.)$ in the neighborhood of the location of the maximum that favors larger bandwidths – this phenomenon is visible also in the right-hand side plot in Figure 1 where the oversmoothed estimator leads to a shorter confidence interval even if the corresponding nonparametric regression function estimator is seriously biased.

In order to investigate also a possibly biased estimator of the location of maximum, we use the regression function:

$$m_2(x) = \theta^{-2}\cos\{2\pi(x^2 - \theta^2)\}.$$

Note that the third derivative $m_2^{(3)}(\theta) \neq 0$ and that the second derivative $m_2^{(2)}(\theta) = -16\pi^2$ does not depend on the location of maximum θ.

The setup of the simulation study for the regression function $m_2(.)$ remains the same, the only difference is that the results are calculated on a modified grid of bandwidths from 0.01 to 0.10.

The simulation results for the regression function $m_2(.)$ are summarized in Table 2. In this case, we observe that the optimal bandwidth decreases with increasing sample size. For the simulated sample size $n = 2500$, the optimal values of the parameter r

TABLE 1. Results of 2500 simulations using the regression function $m_1(.)$: the powers r_{opt} of the prior density defining the distribution of design points with the smallest MSE and MAD for sample sizes $n = 20$ and 200, standard deviations $\sigma = 0.1$ and 0.01, and bandwidths $b_n \in (0.02, 0.50)$. The star $\star$ denotes the bandwidth with the smallest MSE or MAD for each combination of sample size and standard deviation

$n = 20$	$\sigma = 0.1$		$\sigma = 0.01$	
b_n	$r_{opt}(MAD)$	$r_{opt}(MSE)$	$r_{opt}(MAD)$	$r_{opt}(MSE)$
0.02	0.150 (3.1141)	0.150 (0.1619)	0.275 (1.5370)	0.250 (0.0377)
0.04	0.175 (2.7769)	0.175 (0.1244)	0.275 (1.3666)$\star$	0.250 (0.0243)
0.07	0.250 (2.4402)	0.175 (0.0985)	0.250 (1.3936)	0.200 (0.0240)$\star$
0.10	0.250 (1.9896)	0.175 (0.0684)	0.250 (1.4522)	0.250 (0.0244)
0.15	0.200 (1.7832)$\star$	0.200 (0.0479)$\star$	0.250 (1.6393)	0.200 (0.0313)
0.20	0.200 (1.9082)	0.100 (0.0484)	0.200 (1.8701)	0.100 (0.0413)
0.30	0.050 (2.3033)	0.050 (0.0582)	0.075 (2.3144)	0.050 (0.0554)
0.40	0.025 (2.4870)	0.025 (0.0651)	0.000 (2.4850)	0.000 (0.0620)
0.50	0.000 (2.4846)	0.000 (0.0648)	0.000 (2.4846)	0.000 (0.0619)
$n = 200$				
0.02	0.275 (1.6564)	0.275 (0.0425)	0.500 (0.4598)	0.500 (0.0034)
0.04	0.375 (1.3910)	0.250 (0.0309)	0.500 (0.2086)	0.375 (0.0007)
0.07	0.450 (0.8102)	0.400 (0.0113)	0.425 (0.1444)$\star$	0.325 (0.0003)
0.10	0.400 (0.5123)	0.400 (0.0044)	0.375 (0.1445)	0.225 (0.0003)$\star$
0.15	0.375 (0.3324)	0.225 (0.0018)	0.225 (0.1673)	0.175 (0.0003)
0.20	0.125 (0.2739)	0.125 (0.0012)	0.125 (0.1922)	0.125 (0.0004)
0.30	0.125 (0.2473)	0.075 (0.0009)	0.075 (0.2225)	0.075 (0.0005)
0.40	0.025 (0.2388)	0.025 (0.0008)	0.050 (0.2338)	0.050 (0.0006)
0.50	0.025 (0.2373)$\star$	0.025 (0.0008)$\star$	0.025 (0.2312)	0.025 (0.0005)

observed in the simulation study (0.8 and 0.575 for MAD and 0.5 and 0.4 for MSE) are already reasonably close to theoretical values (2/3 for MAD and 1/2 for MSE) provided by Theorem 2.

5. Conclusion

The density of design points derived in Theorem 2 increases the precision of the nonparametric kernel estimator of the location of maximum. Compared to existing literature [2], we obtain an easily interpretable and applicable result and we show that the AIMSE optimal design, $f_X^* \propto a^{1/2}(x)$, is also the MSE optimal design for the location of maximum. Moreover, in Theorem 2 we show that a density of design points proportional to $a^{2/3}(x)$ is optimal from a point of view of the expected length of confidence intervals. Finally, a small simulation study suggests that our theoretical findings are appropriate for larger sample sizes and that the optimal density of design points and the bandwidth are intrinsically related.

TABLE 2. Results of 2500 simulations using the regression function $m_2(.)$: the powers r_{opt} of the prior density defining the distribution of design points with the smallest MSE and MAD for sample sizes $n = 200$, 1000, and 2500, standard deviations $\sigma = 0.1$ and 0.01, and bandwidths $b_n \in (0.01, 0.10)$. Stars ($\star$) denote the optimal bandwidth for each combination of sample size and standard deviation

$n = 200$	$\sigma = 0.1$		$\sigma = 0.01$	
b_n	$r_{opt}(MAD)$	$r_{opt}(MSE)$	$r_{opt}(MAD)$	$r_{opt}(MSE)$
0.01	0.500 (1.0213)	0.500 (0.0178)	0.500 (0.3001)	0.375 (0.0014)
0.02	0.500 (0.8329)	0.475 (0.0110)	0.425 (0.1498)	0.425 (0.0004)
0.04	0.500 (0.4496)	0.375 (0.0034)	0.375 (0.0606)	0.350 (0.0001)
0.06	0.450 (0.2691)	0.400 (0.0012)	0.075 (0.0571)$\star$	0.075 (0.0001)$\star$
0.08	0.150 (0.2684)$\star$	0.150 (0.0012)$\star$	0.025 (0.1473)	0.025 (0.0004)
0.10	0.025 (0.3711)	0.025 (0.0021)	0.025 (0.3528)	0.025 (0.0016)
$n = 1000$				
0.01	0.650 (0.6551)	0.650 (0.0071)	0.775 (0.1513)	0.725 (0.0004)
0.02	0.650 (0.4885)	0.650 (0.0038)	0.475 (0.0566)	0.475 (0.0001)
0.04	0.725 (0.2093)	0.500 (0.0007)	0.000 (0.0514)$\star$	0.075 (0.0000)$\star$
0.06	0.325 (0.1996)$\star$	0.475 (0.0006)$\star$	0.000 (0.1627)	0.100 (0.0003)
0.08	0.025 (0.3280)	0.100 (0.0013)	0.025 (0.3261)	0.100 (0.0012)
0.10	0.025 (0.5346)	0.100 (0.0033)	0.100 (0.5349)	0.100 (0.0032)
$n = 2500$				
0.01	0.550 (0.5098)	0.550 (0.0042)	0.625 (0.0973)	0.475 (0.0002)
0.02	0.775 (0.3271)	0.500 (0.0018)	0.575 (0.0370)$\star$	0.400 (0.0000)$\star$
0.04	0.800 (0.1443)$\star$	0.500 (0.0003)$\star$	0.000 (0.0726)	0.000 (0.0001)
0.06	0.300 (0.1971)	0.625 (0.0005)	0.000 (0.1897)	0.050 (0.0004)
0.08	0.000 (0.3514)	0.625 (0.0014)	0.050 (0.3526)	0.050 (0.0014)
0.10	0.400 (0.5579)	0.625 (0.0034)	0.400 (0.5583)	0.625 (0.0034)

References

[1] GASSER, T., MÜLLER, H.-G.: *Estimating regression functions and their derivatives by the kernel method*. Scand. J. Stat. **11** (1984), 171–185.

[2] MÜLLER, H.-G.: *Optimal designs for nonparametric kernel regression*. Statist. Probab. Lett. **2** (1984), 285–290.

[3] MÜLLER, H.-G.: *Kernel estimators of zeros and of location and size of extrema of regression functions*. Scand. J. Stat. **12** (1985), 221–232.

[4] NOŽIČKA, F.: *Calculus of variations* [in Czech]. In Rektorys et al.: Přehled užité matematiky, 5th edition, SNTL, Praha, 1988, 841–858.

[5] R Development Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009.

[6] SMIRNOW, W. I.: Lehrgang der höheren Mathematik, Teil IV. VEB, Berlin, 1958.