

Nonparametric regression for threshold data

Ursula U. MÜLLER

Key words and phrases: binary data, kernel regression, optimal bandwidth, stochastic resonance.

Mathematics subject classification codes (1991): primary 62G07; secondary 62G20.

ABSTRACT

Consider a detector which records the times at which the endogenous variable of a nonparametric regression model exceeds a certain threshold. If the error distribution is known, the regression function can still be identified from these threshold data. The author constructs estimators for the regression function that are transformations of kernel estimators. She determines the bandwidth that minimizes the asymptotic mean average squared error. Her investigation was motivated by recent work on stochastic resonance in neuro-science and signal detection theory, where it was observed that detection of a subthreshold signal is enhanced by the addition of noise. The author compares her model with several others that have been proposed in the recent past.

RÉSUMÉ

Supposons qu'un appareil enregistre les temps auxquels la variable endogène d'un modèle de régression non paramétrique dépasse un certain seuil. Si la loi des erreurs est connue, la fonction de régression peut quand même être identifiée à partir de telles observations. L'auteure propose des estimateurs pour cette fonction de régression qu'elle obtient par transformation d'estimateurs à noyau. Elle détermine la taille de fenêtre qui minimise une erreur quadratique moyenne expérimentale asymptotique. Ses recherches ont été motivées par des travaux récents portant sur la résonance stochastique en théorie de la détection du signal et en neuro-science, domaines où l'on a observé que la présence d'un bruit facilite la détection d'un signal inférieur à un seuil donné. L'auteure compare son modèle à plusieurs autres qui ont été proposés ces dernières années.

1. INTRODUCTION

In a system with a threshold, a subthreshold signal may be detected if noise, either from the background or artificially generated, is added to the input. If the noise is too low, it does not help much. If it is too high, it drowns out the signal. It is plausible and has been observed both empirically and through simulations that there is an optimal level of noise. This property of the system is known as

stochastic resonance, although this name is not really appropriate unless the signal is periodic.

The term *stochastic resonance* was introduced by Benzi *et al.* (1981) in the context of a model describing the periodic recurrence of ice ages. Climate changes are modeled as transitions in a double-well potential system pushed by a signal, the orbital eccentricity, which causes small variations of the solar energy influx. Since the periodic forcing for switching from one state point to the other is very weak, it must be assisted by other factors such as short term climate fluctuations that are modeled as noise. However, if there is too much noise, the transitions become independent of the frequency of the periodic signal. Consequently there must be an optimal noise level, i.e., stochastic resonance.

Since then, stochastic resonance has been extended to a large variety of physical systems with simpler thresholds (for an overview, see Gammaitoni *et al.* 1998). It attracted particular attention in neuro-science these past few years. Various models for information processing in the nervous system were proposed that are explained by stochastic resonance. Here, theory turns away from bistable systems which do not describe neuronal dynamics well. Moreover, neuronal models with *aperiodic* input signal are often more appropriate than models limited to periodic inputs and have gained considerable attention (Collins *et al.* 1996). Perhaps the simplest model for neural dynamics regards a single neuron as a threshold crossing detector as follows: the cell is stimulated by an external input (signal); if its membrane voltage exceeds a fixed threshold, the cell fires and is reset. It can be assumed that the sensory system is optimized, so that the spike train of a firing cell contains significant information. The presence of noise, e.g., background noise from other neurons, leads to maximal performance, i.e., stochastic resonance (Wiesenfeld & Moss 1995). Stochastic resonance was studied not only theoretically but could also be exhibited in several experiments. For example, Douglass *et al.* (1993) demonstrated stochastic resonance in the firing rate of sinusoidally stimulated mechanoreceptor cells of crayfish.

Although a large amount of literature is available, there is little statistical work on this subject. For the models described, stochastic resonance is usually exhibited through simulations. Various measures of detectability are used. In the case of *periodic* signals, this is typically the signal-to-noise ratio (e.g., Wiesenfeld & Moss 1995). If the signal is *aperiodic*, usually a correlation measure is considered (e.g., Collins *et al.* 1995). A more familiar approach from a statistical point of view was considered by Stemmler (1996) in the case of a (nearly) constant signal. Instead of commonly adopted measures, which break down for constant signal, he uses Fisher information. Recently, Greenwood *et al.* (1999) have derived efficient estimators and analyzed stochastic resonance statistically for single and multiple thresholds.

As in Stemmler (1996) and Greenwood *et al.* (1999), the model considered here is embedded in a neuro-physiological context. We consider the simple neuron detector described above: if the incoming noisy signal reaches a certain threshold, the cell fires and is reset. The observations are the times when the threshold is exceeded. For the quantification of neuronal responses, it is standard practice to sum all spikes in a fixed period to obtain an estimation of the firing rate. Clearly there is additional information in the timing of the spikes, especially when the shape of the signal is unknown. This information will be used in our statistical approach based on kernel regression methods. We will derive a consistent estimator of a subthreshold signal from the exceedance times data. The problem is cast as nonparametric regression.

Consider the nonparametric regression problem $Y(t_i) = s(t_i) + \epsilon(t_i)$ with independent mean zero error variables $\epsilon(t_i)$. Let $a > 0$ be a threshold, and suppose that we do not observe the realizations $Y(t_i)$ but only the times at which the threshold a is exceeded. Then the observations (*threshold data*) are Bernoulli variables coded by 1 or 0 as follows:

$$X(t_i) = \mathbf{1}\{s(t_i) + \epsilon(t_i) > a\} = \mathbf{1}\{Y(t_i) > a\}; \quad (1)$$

this was suggested by McCulloch & Pitts (1943) as a model of a neuron.

We can always estimate the probabilities

$$E\{X(t_i)\} = p(t_i) = P\{X(t_i) = 1\} = P\{s(t_i) + \epsilon(t_i) > a\}, \quad (2)$$

say by kernel methods. To identify the signal $s(t)$, the distribution of the $\epsilon(t_i)$'s, say F_{t_i} , must be known and invertible. Then

$$s(t_i) = a - F_{t_i}^{-1}\{1 - p(t_i)\}. \quad (3)$$

The estimator of $s(t_i)$ will be taken to be the kernel estimator for $p(t_i)$ transformed as in (3).

If the noise is artificially generated, then it is reasonable to assume that the error distribution function is completely known. If the noise is background noise, we may at least know the form of the distribution. For example, it may be plausible to assume that the errors are normally distributed. However, we cannot identify the noise amplitude from the $X(t_i)$'s. One way of dealing with the problem is to get information about the noise from elsewhere, for example by using a second detector with a different threshold; see Greenwood *et al.* (1999) for a treatment of the constant signal case. A second solution is to note that even if the noise distribution is known up to a scale parameter, the signal can still be identified up to a one-parameter family of transformations. Hence most of the information is retained, unless the signal is constant, of course.

As in every approach involving kernel estimators, the choice of the bandwidth is of crucial importance. Our criterion for bandwidth selection will be the asymptotic mean average squared error of the estimator of s , which we derive. A formula for the asymptotically optimal bandwidth can then be written. Since the formula involves unknown quantities, we estimate the optimal bandwidth by plugging in estimators for them. An asymptotic approach is reasonable since a linearization of the mean squared error $E\{\hat{s}(t) - s(t)\}^2$ is necessary. This already involves asymptotics. Because of this approximation, it does not seem worthwhile to pursue a more sophisticated bandwidth selection technique. Our approach is similar to that of Ruppert *et al.* (1995), who derive an asymptotically optimal bandwidth for the classical setting, i.e., fully observed data. The main difference between this article and theirs is the nonlinear link occurring here, in particular in the mean squared error expression. Through the linearization of this expression, the calculation of the optimal bandwidth becomes similar to the standard case, and familiar results can be used.

This paper is organized as follows. In Section 2 we give the kernel estimator, basic notation and assumptions. Section 3 is the main section of this article. We derive the asymptotic expressions for the mean squared error, the mean average squared error, and the resulting optimal local and global optimal bandwidths. Some remarks and references concerning the suggested plug-in estimation will be given in Section 4. Section 5 concludes the article with an example and a comparison with existing techniques, emphasizing stochastic resonance.

2. KERNEL REGRESSION

Consider the model introduced in Section 1, Equations (1)–(3), and let the time points be equally spaced in $[0, 1]$, i.e., $t_i = i/n$, $i = 1, \dots, n$. We assume that the distribution functions of the noise F_{t_i} are continuous with mean zero. Often, the $\epsilon(t_i)$ will be identically distributed, and the distribution will be normal. We also require that s have two continuous derivatives and be bounded from below, viz.

$$s(t) \geq -c \text{ for every } t \in [0, 1] \quad \text{for some } c > 0.$$

In addition, F_t should be four times continuously differentiable, and there should exist $p_1, p_2 \in (0, 1)$ such that for all $0 \leq t \leq 1$,

$$1 - F_t(a + c) \geq p_1 \quad \text{and} \quad F_t(0) \geq p_2.$$

These conditions and the subthreshold assumption $s(t) < a$ imply that p has two continuous derivatives and is bounded away from 0 and 1, viz.

$$p(t) \in [p_1, 1 - p_2] \text{ for each } t, \tag{4}$$

a condition that is easily verified. To simplify the notation, we write $G_t(x) = 1 - F_t(a - x)$, so that $p(t) = G_t\{s(t)\}$ and $s(t) = G_t^{-1}\{p(t)\}$.

We treat the problem of estimating the probabilities $p(t)$ as a nonparametric regression problem and estimate $p(t)$ by a modified kernel estimator $\tilde{p}_h(t)$, where $h > 0$ denotes the bandwidth. We obtain an estimator for the signal by

$$\hat{s}_h(t) = G_t^{-1}\{\tilde{p}_h(t)\}$$

with

$$\tilde{p}_h(t) = \begin{cases} \hat{p}_h(t) & \text{if } \hat{p}_h(t) \in C = [p_1/2, 1 - p_2/2], \\ 1/2 & \text{otherwise.} \end{cases}$$

Here, $\hat{p}_h(t)$ is a classical kernel estimator. If the values of $\hat{p}_h(t)$ are near 0 and 1, $G_t^{-1}\{\hat{p}_h(t)\}$ is undefined. For simplicity, we set $\tilde{p}_h(t)$ equal to $1/2$ if $\hat{p}_h(t) \notin C$ (or to a suitable different constant). Because of (4), we have $p(t) \in C$. The kernel estimator used here is the Nadaraya-Watson estimator

$$\hat{p}_h(t) = \frac{\sum_{i=1}^n \frac{1}{h} K\left(\frac{t-t_i}{h}\right) X(t_i)}{\sum_{i=1}^n \frac{1}{h} K\left(\frac{t-t_i}{h}\right)}.$$

For the estimation at inner points $t \in [h, 1 - h]$ considered in this article, let $K : \mathbb{R} \rightarrow \mathbb{R}$ be some second order kernel function, i.e.,

$$\int K(u)du = 1, \quad \int uK(u)du = 0, \quad \int u^2K(u)du = c \neq 0.$$

We also need that the derivative K' be bounded and assume that the support of K is $[-1, 1]$. For $t \notin [h, 1 - h]$, a different approach with some boundary kernel K should be chosen (cf., e.g., Gasser *et al.* 1985).

Instead of the Nadaraya-Watson estimator, one could also consider a local linear kernel estimator (cf. Ruppert *et al.* 1995), which is known for its superior boundary behaviour but requires more work. The proofs given here can be adapted to this approach in a straightforward way. Later on, we take $h = h_n \rightarrow 0$ and $nh^3 \rightarrow \infty$ as $n \rightarrow \infty$.

3. ASYMPTOTIC ERROR AND OPTIMAL BANDWIDTH

In nonparametric regression theory, generally accepted measures of the quality of the estimation are the mean squared error $E\{\hat{s}_h(t) - s(t)\}^2$ and the mean average squared error $n^{-1} \sum_{t_i \in T} E\{\hat{s}_h(t_i) - s(t_i)\}^2$. In the latter case, summation is usually restricted to some interval $T = [c, d] \subset (0, 1)$ because of boundary effects. These quantities will also be studied here and taken as criteria for an asymptotically optimal local and an asymptotically optimal global bandwidth. Another popular global error measure would be the mean *integrated* squared error $\int_T E\{\hat{s}_h(t) - s(t)\}^2 dt$. The proofs given herein adapt to this case in a straightforward way.

The approach of this section will be to derive a Taylor approximation for the mean squared error (which immediately gives the approximation for the mean average squared error). The bandwidth h that minimizes the leading terms of the expansion will then be called optimal. The distinguishing characteristic of our model is that it involves the nonlinear transformation $\hat{s}_h(t) = G_t^{-1}\{\tilde{p}_h(t)\}$. An additional problem arising from the modification $\tilde{p}_h(t)$ of the kernel estimator $\hat{p}_h(t)$ will be seen to be negligible, because it can easily be verified that $\tilde{p}_h(t)$ coincides asymptotically with the common estimator $\hat{p}_h(t)$.

The following lemmas recall and extend well known auxiliary results from classical theory. They provide approximation formulas for the bias, variance and higher moments of $\hat{p}_h(t)$. The derivations are standard and can be found either in Eubank (1988) or Müller (1999).

In the following, let h be sufficiently small so that the time points t where estimation takes place satisfy $t \in [h, 1-h]$. The terminology and conditions introduced in Section 2 will be assumed throughout.

LEMMA 1. *Consider the asymptotics $n \rightarrow \infty$, $h = h_n \rightarrow 0$, and $nh^2 \rightarrow \infty$. For $\ell = 2, 3$, the approximation*

$$E[\hat{p}_h(t) - E\{\hat{p}_h(t)\}]^\ell = (nh)^{1-\ell} E\{X(t) - p(t)\}^\ell \int_{-1}^1 K^\ell(u) du + o\{(nh)^{1-\ell}\}$$

then holds uniformly in $t \in [h, 1-h]$, where $X(t)$ is a Bernoulli random variable with parameter $p(t)$. Furthermore,

$$E[\hat{p}_h(t) - E\{\hat{p}_h(t)\}]^4 = O\{(nh)^{-2}\} = o\{(nh)^{-1}\},$$

uniformly in $t \in [h, 1-h]$.

LEMMA 2. *For $n \rightarrow \infty$, $h = h_n \rightarrow 0$ and $nh^2 \rightarrow \infty$, the bias and the mean squared error of $\hat{p}_h(t)$ can be approximately uniformly in $t \in [h, 1-h]$ by*

$$E\{\hat{p}_h(t) - p(t)\} = h^2 p''(t) \mu_2(K) / 2 + o(h^2) + O\{(nh^2)^{-1}\}$$

and

$$\begin{aligned} E\{\hat{p}_h(t) - p(t)\}^2 &= \text{Var}\{\hat{p}_h(t)\} + [E\{\hat{p}_h(t)\} - p(t)]^2 \\ &= (nh)^{-1} p(t) \{1 - p(t)\} R(K) + h^4 p''(t)^2 \mu_2(K)^2 / 4 \\ &\quad + o\{(nh)^{-1}\} + o(h^4) + O\{(nh^2)^{-2}\} \end{aligned} \quad (5)$$

with $\mu_2(K) = \int_{-1}^1 u^2 K(u) du$ and $R(K) = \int_{-1}^1 K^2(u) du$. Furthermore,

$$E\{\hat{p}_h(t) - p(t)\}^3 = o\{(nh)^{-1}\} + O(h^6) + O\{(nh^2)^{-3}\}, \quad (6)$$

$$E\{\hat{p}_h(t) - p(t)\}^4 = o\{(nh)^{-1}\} + O(h^8) + O\{(nh^2)^{-4}\}, \quad (7)$$

uniformly in $t \in [h, 1 - h]$.

We now state our main theorem. It concerns the signal estimator \hat{s} , a function of the modification \tilde{p}_h of the estimator \hat{p}_h , whose properties are described in Lemmas 1 and 2. We give the asymptotic mean squared error (locally), the asymptotic mean average squared error (globally) and the respective optimal bandwidths.

THEOREM. *Consider the asymptotics $n \rightarrow \infty$, $h = h_n \rightarrow 0$, and $nh^3 \rightarrow \infty$. The mean squared error $MSE(h, t) = E\{\hat{s}_h(t) - s(t)\}^2$ exists and may be approximated as follows by the asymptotic mean squared error, viz.*

$$AMSE(h, t) = \frac{(nh)^{-1}p(t)\{1 - p(t)\}R(K) + h^4p''(t)^2\mu_2(K)^2/4}{G'_t\{s(t)\}^2} \quad (8)$$

where $R(K) = \int_{-1}^1 K^2(u)du$ and $\mu_2(K) = \int_{-1}^1 u^2K(u)du$. This approximation is valid up to an term of order $o\{(nh)^{-1} + h^4\}$ and uniform in $t \in [h, 1 - h]$,

For the mean average squared error $MASE(h) = n^{-1} \sum_{t_i \in T} E\{\hat{s}_h(t_i) - s(t_i)\}^2$, the approximation

$$AMASE(h) = \frac{R(K)}{n^2h} \sum_{t_i \in T} \frac{p(t_i)\{1 - p(t_i)\}}{G'_{t_i}\{s(t_i)\}^2} + \frac{h^4\mu_2(K)^2}{4n} \sum_{t_i \in T} \frac{p''(t_i)^2}{G'_{t_i}\{s(t_i)\}^2}.$$

holds up to a term of order $o\{(nh)^{-1} + h^4\}$. The asymptotically optimal local bandwidth is

$$h_{opt}(t) = \left[\frac{R(K)p(t)\{1 - p(t)\}}{n\mu_2(K)^2p''(t)^2} \right]^{1/5} \quad (9)$$

and the asymptotically optimal global bandwidth is

$$h_{opt} = \left[\frac{R(K) \sum_{t_i \in T} \frac{1}{G'_{t_i}\{s(t_i)\}^2} p(t_i)\{1 - p(t_i)\}}{n\mu_2(K)^2 \sum_{t_i \in T} \frac{1}{G'_{t_i}\{s(t_i)\}^2} p''(t_i)^2} \right]^{1/5}. \quad (10)$$

PROOF. Let $t \in [h, 1 - h]$ and $\hat{s}_h(t) = G_t^{-1}\{\tilde{p}_h(t)\}$ as in Section 2. Consider the loss function $H_t : (0, 1) \rightarrow \mathbb{R}$ defined by

$$H_t(x) = [G_t^{-1}(x) - G_t^{-1}\{p(t)\}]^2.$$

Applying the so-called delta method, we obtain a linearization of the mean squared error,

$$\begin{aligned} E\{\hat{s}_h(t) - s(t)\}^2 &= E[H_t\{\tilde{p}_h(t)\}] \\ &= \frac{1}{2}H_t^{(2)}\{p(t)\}E\{\tilde{p}_h(t) - p(t)\}^2 \\ &\quad + \frac{1}{3!}H_t^{(3)}\{p(t)\}E\{\tilde{p}_h(t) - p(t)\}^3 + E[R_t\{\tilde{p}_h(t)\}] \end{aligned} \quad (11)$$

with $H_t^{(2)}\{p(t)\}/2 = 1/G'_t\{s(t)\}^2$ and

$$R_t\{\tilde{p}_h(t)\} = \{\tilde{p}_h(t) - p(t)\}^4 \int_0^1 \frac{(1-z)^3}{3!} H_t^{(4)}[p(t) + z\{\tilde{p}_h(t) - p(t)\}] dz.$$

In order to establish the asserted approximation, we use the auxiliary results

$$E\{\tilde{p}_h(t) - p(t)\}^\ell - E\{\hat{p}_h(t) - p(t)\}^\ell = o\{(nh)^{-1}\} + o(h^4), \quad \ell \in \mathcal{N} \quad (12)$$

$$E[R_t\{\tilde{p}_h(t)\}] = o\{(nh)^{-1}\} + o(h^4), \quad (13)$$

which hold uniformly in $t \in [h, 1-h]$, as will be verified at the end of the proof. Both relations, combined with (6) and $nh^3 \rightarrow \infty$, imply that the two last terms of (11) have the order $o\{(nh)^{-1}\} + o(h^4)$. Similarly, the mean squared error of $\tilde{p}_h(t)$ in the first term can be replaced by the familiar approximation (5) for the mean squared error of $\hat{p}_h(t)$. This establishes (8), uniformly for $t \in [h, 1-h]$. Since $T \subset [h, 1-h]$ for sufficiently small h , the approximation formula $AMASE(h)$ for the mean average squared error $MASE(h) = n^{-1} \sum_{t_i \in T} MSE(h, t_i)$ is immediately derived from this result. The optimal local and global bandwidth given in (9) and (10) are obtained by simple calculus, differentiating $AMSE(h, t)$ and $AMASE(h)$ with respect to h . Hence only the auxiliary statements (12) and (13) remain to be shown.

For the proof of (12), it should first be noticed that $C = [p_1/2, 1-p_2/2] \subset (0, 1)$ was chosen such that $p(t) \in C$. By equation (4), we have $p(t) \in [p_1, p_2] \subset C$. Hence there exists some $\delta > 0$ such that $[p(t) - \delta, p(t) + \delta] \subset C$ for all $t \in [0, 1]$. This will be used in the following chain of equalities and inequalities, which holds for arbitrary nonnegative integer ℓ :

$$\begin{aligned} & |E\{\tilde{p}_h(t) - p(t)\}^\ell - E\{\hat{p}_h(t) - p(t)\}^\ell| \\ &= \left| \sum_{k=1}^{\ell} \binom{\ell}{k} E[\{\tilde{p}_h(t) - \hat{p}_h(t)\}^k \{\hat{p}_h(t) - p(t)\}^{\ell-k}] \right| \\ &\leq \sum_{k=1}^{\ell} \binom{\ell}{k} E[\mathbf{1}_{[0,1] \setminus C}\{\hat{p}_h(t)\} \cdot |\tilde{p}_h(t) - \hat{p}_h(t)|^k |\hat{p}_h(t) - p(t)|^{\ell-k}] \\ &\leq \sum_{k=1}^{\ell} \binom{\ell}{k} E[\mathbf{1}_{[0,1] \setminus C}\{\hat{p}_h(t)\}] \\ &= (2^\ell - 1)P\{\hat{p}_h(t) \notin C\} \\ &\leq (2^\ell - 1)P\{|\hat{p}_h(t) - p(t)| > \delta\} \\ &\leq (2^\ell - 1)E\{\hat{p}_h(t) - p(t)\}^4 / \delta^4. \end{aligned}$$

In the last step, Markov's inequality was applied. Relation (12) now follows from (7) and the fact that $nh^3 \rightarrow \infty$.

For the proof of (13), we will use the boundedness of $H_t^{(4)}$ on C , which results from the fact that F_t is four times continuously differentiable. Thus

$$\begin{aligned} & |E[R_t\{\tilde{p}_h(t)\}]| \\ &= \left| E \left[\{\tilde{p}_h(t) - p(t)\}^4 \cdot \int_0^1 \frac{(1-z)^3}{3!} H_t^{(4)}[p(t) + z\{\tilde{p}_h(t) - p(t)\}] dz \right] \right| \\ &\leq E\{\tilde{p}_h(t) - p(t)\}^4 \cdot \sup_{x \in C} |H_t^{(4)}(x)|. \end{aligned}$$

Using (12) with $\ell = 4$ and $E\{\hat{p}_h(t) - p(t)\}^4 = o((nh)^{-1}) + o(h^4)$, one then gets (13) immediately. Hence the proof is complete. \square

Let us briefly discuss the approximate mean squared error calculated in (8). Its numerator is the Taylor approximation of the mean squared error $E\{\hat{p}_h(t) - p(t)\}^2$ of the kernel estimator $\hat{p}_h(t)$. This well known formula yields the decomposition of $E\{\hat{p}_h(t) - p(t)\}^2$ into variance and squared bias of $\hat{p}_h(t)$ (Equation (5), Lemma 2). In particular, the characteristic variance-bias trade-off becomes evident.

With the optimal asymptotic bandwidth h_{opt} at hand, the minimal value of $AMASE$ may now be written as

$$AMASE(h_{opt}) = \frac{5}{4} \left[\frac{\mu_2(K)^2}{n^4} \sum_{t_i \in T} \frac{p''(t_i)^2}{G'_{t_i}\{s(t_i)\}^2} \right]^{\frac{1}{5}} \left[\frac{R(K)}{n} \sum_{t_i \in T} \frac{p(t_i)\{1 - p(t_i)\}}{G'_{t_i}\{s(t_i)\}^2} \right]^{\frac{4}{5}}.$$

This value depends on the squared second derivatives of $p(t) = G_t\{s(t)\}$. Smooth signals will, in general, lead to small values of $p''(t)^2$ and thus to small values of $\inf_{h>0} AMASE(h)$. Furthermore, this formula shows the influence of the kernel K , which appears only in the expression $\mu_2(K)^{2/5} R(K)^{4/5}$. The (second order) kernel with support $[-1, 1]$ which minimizes this term, and thus the asymptotic mean average squared error under the further constraint $K \geq 0$, is the Epanechnikov kernel

$$K^*(u) = 3(1 - u^2)\mathbf{1}_{[-1,1]}(u)/4.$$

The optimal kernel K^* is not much better than other kernels, for example the Gaussian kernel (cf. Wand & Jones 1995). What is really crucial is the correct choice of the bandwidth h .

4. DATA-DRIVEN BANDWIDTH SELECTION

In this section, we construct an optimal data-driven bandwidth. We assume that the kernel K is given. In particular, the kernel constants $R(K) = \int_{-1}^1 K^2(u) du$ and $\mu_2(K) = \int_{-1}^1 u^2 K(u) du$ are known.

Recall the asymptotically optimal local bandwidth in our theorem, viz.

$$h_{opt}(t) = n^{-1/5} c(p(t), p''(t)), \quad c(p(t), p''(t)) = \left[\frac{R(K)p(t)\{1 - p(t)\}}{\mu_2(K)^2 p''(t)^2} \right]^{1/5}.$$

The following arguments will also apply to the optimal *global* bandwidth h_{opt} . Both are of the form $n^{-1/5}c$, with c depending on the unknown probability function p and its second derivative, p'' . At first it should be mentioned that both optimal bandwidths, $h_{opt}(t)$ and h_{opt} , have the (optimal) convergence rate $n^{-1/5}$, and that this rate is maintained for any bandwidth h of the form $h = n^{-1/5}c$, where $c > 0$ is an arbitrary constant. However, the choice of c has a strong influence on the finite sample behaviour. Hence, in order to guarantee a good bandwidth approximation in a concrete application, p and p'' should be estimated reasonably well.

We estimate $h_{opt}(t)$ by a so-called plug-in strategy. This means that we estimate the unknown values $p(t)$ and $p''(t)$ with a preliminary estimator and plug them into the formula above. For the pilot estimator, a large variety of methods is available, since estimating p is a classical nonparametric regression problem with binary responses. One can, for example, apply cross-validation or the so-called blocking method introduced by Härdle & Marron (1995). In the latter, the design space is divided into blocks, and a polynomial of low degree is fitted to every block. Another approach would be to carry out some preliminary kernel smoothing. However, a new bandwidth selection problem then arises and, at some point, a pilot bandwidth has to be determined with a different technique.

A comprehensive overview of bandwidth selection techniques is given by Wand & Jones (1995). Several plug-in methods for the classical setting are discussed in Ruppert *et al.* (1995). Further discussions of bandwidth selection techniques can be found in the books of Eubank (1988) and Härdle (1990).

5. THE STOCHASTIC RESONANCE EFFECT

The model considered here was suggested by models for neuron firing triggered by a noisy signal. The literature in this field emphasizes the stochastic resonance effect, i.e., the existence of an optimal noise level for the detectability of the signal. This effect has mainly been shown empirically through simulations. In several papers (Collins *et al.* 1995, 1996; Henegan *et al.* 1996; Chialvo *et al.* 1997a,b) a box kernel is used to estimate the probability that a threshold is crossed. This kernel corresponds with our more sophisticated estimator $\hat{p}_h(t)$ but with a fixed bandwidth whose choice is justified by physical arguments. The estimator is compared with the signal using Pearson correlation or modifications thereof, mostly called “normalized power norm” or “cross-correlation coefficient,” namely

$$C = \frac{\sum_{i=1}^n \{s(t_i) - \bar{s}\} \{\hat{p}_h(t_i) - \overline{\hat{p}_h}\}}{[\sum_{i=1}^n \{s(t_i) - \bar{s}\}^2]^{1/2} [\sum_{i=1}^n \{\hat{p}_h(t_i) - \overline{\hat{p}_h}\}^2]^{1/2}} \in [-1, 1].$$

Here, \hat{p}_h stands for any estimator of the exceedance probabilities, \bar{s} and $\overline{\hat{p}_h}$ denote mean averages.

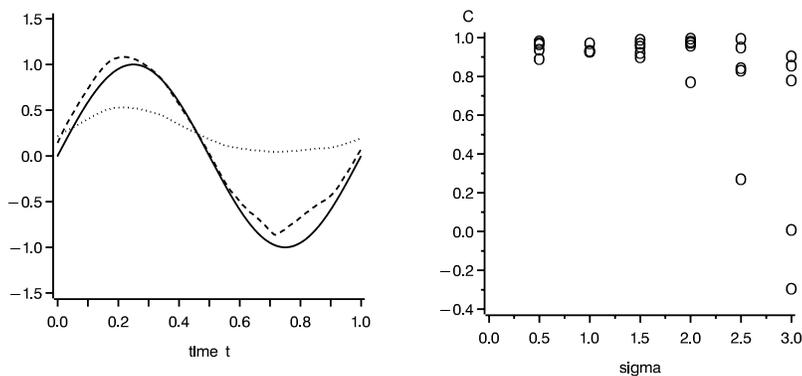


FIGURE 1. (a) Left panel: simulation of signal estimates \hat{s}_h (dashed line) for $s(t) = \sin(2\pi t)$ (solid line), with $n = 1000$, $a = 1$, $h = 0.16$ and i.i.d. $N(0, 1.08^2)$ noise; the dotted line shows the probability estimates \hat{p}_h . (b) Right panel: realizations of the Pearson correlation C (circles) between the sinusoid s from the left panel and its estimates \hat{s}_h , for $\sigma = 0.5, 1, \dots, 3$ and $n = 100$.

An illustration of our technique is given in Figure 1, whose left panel shows an estimate of a simple sinusoidal signal $s(t) = \sin(2\pi t)$. In this example, we chose $n = 1000$ time points, a threshold $a = 1$ and independent normal $N(0, 1.08^2)$ noise. The bandwidth $\hat{h}_{opt} = 0.16$ was estimated by plug-in methods, using a modification of the blocking method of Härdle & Marron (1995), which is relatively close to the theoretically optimal bandwidth $h_{opt} = 0.13$. Besides the estimated signal \hat{s}_h , this figure also shows the estimated probabilities \hat{p}_h that correspond to the box kernel mentioned above. Of course, these estimators are not consistent for

the signal. Nevertheless, both the Pearson correlation between s and \hat{p}_h and the correlation between s and our consistent estimator \hat{s}_h are close to 1 ($C = 0.977$ resp. $C = 0.995$). This is not surprising, however, since C is invariant with respect to linear transformations and \hat{p}_h still catches essential characteristics of the shape of s .

For $n = 100$, the right panel of Figure 1 shows five realizations of the Pearson correlation between s and \hat{s}_h , for each value of $\sigma = 0.5, 1, \dots, 3$. Papers in this field usually perform more elaborate simulation studies and get empirical estimates of mean and standard error of C . In particular, a concave curve of the estimated mean as a function of σ can then be drawn, producing stochastic resonance. That there is an optimal noise level already emerges in our picture, however: for $\sigma = 1$, the five values of C are all close to one; for smaller and larger σ , the correlation is not always high. For a comparison, we performed the same simulation but with $n = 1000$ time points, getting values C close to one throughout. This observation illustrates well a phenomenon first reported by Collins *et al.* (1995) in a different setting, namely “stochastic resonance without tuning”: since the variance of the signal estimator and hence the variance of C decreases with n , the correlation is high for a broad range of σ 's. Although there is stochastic resonance, i.e., an optimal level of noise, C cannot detect it if the time points are too densely spaced.

For threshold data in the nonparametric regression model considered here, a stochastic resonance effect analogous to that shown by simulation in the literature would imply that the asymptotic mean squared and the asymptotic mean average squared error are convex as functions of the standard deviation of the noise. We do not expect this behaviour for all signals or for all error distributions. For the example with the sinusoid from above, stochastic resonance is easily verified. In Figure 2, we plotted the asymptotic mean average squared error, with the optimal bandwidth, as a function of the noise level σ . Here, we used the “tuned” version, i.e., $\inf_h \{AMASE(\sigma, h)\}$ multiplied by the convergence rate $n^{4/5}$. Since the sums appearing in the formula approximate integrals, the same convex curve is produced for all n sufficiently large. In particular, a sharp optimal noise level, $\sigma = 1.08$, can be derived which we have already used for the simulation in Figure 1.

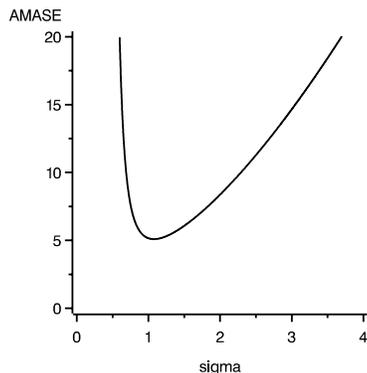


FIGURE 2. Plot of $n^{4/5} \inf_h AMASE(\sigma, h)$ for the sinusoid example, $s(t) = \sin(2\pi t)$, with n sufficiently large (here $n = 1000$).

In general, a proof is not straightforward, not even if we restrict attention to well-behaved unimodal distributions such as the normal $N(0, \sigma^2)$ distribution. The

function p of Section 2 is then expressed as $p(t) = \Phi[\{s(t) - a\}/\sigma]$, in terms of Φ , the distribution function of the standard normal $N(0, 1)$ with density ϕ . As for the asymptotic mean squared error, it is equal to

$$\frac{1}{nh} \frac{\sigma^2 \Phi \left\{ \frac{s(t)-a}{\sigma} \right\} \Phi \left\{ \frac{a-s(t)}{\sigma} \right\}}{\phi^2 \left\{ \frac{s(t)-a}{\sigma} \right\}} R(K) + \frac{h^4}{4} \left\{ \frac{a-s(t)}{\sigma^2} s'(t)^2 + s''(t) \right\}^2 \mu_2(K)^2.$$

The first term of the sum corresponds to the inverse Fisher information I_s^a in Greenwood *et al.* (1999) and shows the typical stochastic resonance behaviour: it tends to infinity when $\sigma^2 \rightarrow \infty$ or $\sigma^2 \rightarrow 0$ (cf. Greenwood *et al.* 1999). The second term, however, varies like $1/\sigma^4$. The behaviour of the sum requires further analysis.

The stochastic resonance behaviour of *AMSE* and further aspects will be investigated numerically and by simulations in a forthcoming paper of Müller & Ward (1999).

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada, and the Crisis Points Group of the Peter Wall Institute for Advanced Studies at the University of British Columbia. I thank P. E. Greenwood, N. E. Heckman, L. M. Ward and W. Wefelmeyer for several helpful discussions on the theory and applications of nonparametric regression and stochastic resonance.

REFERENCES

- Benzi, R., Sutera, A., & Vulpiani, A. (1981). The mechanism of stochastic resonance. *J. Phys. A: Math. Gen.*, 14, L453–457.
- Chialvo, D. R., Longtin, A., & Müller-Gerking, J. (1997a). Stochastic resonance in models of neuron ensembles. *Phys. Rev. E* (3), 55, 1798–1808.
- Chialvo, D. R., Longtin, A., & Müller-Gerking, J. (1997b). Stochastic resonance in models of neuron ensembles revisited. *Phys. Rev. E* (3), in press.
- Collins, J. J., Chow, C. C., Capela, A. C., & Imhoff, T. T. (1996). Aperiodic stochastic resonance. *Phys. Rev. E* (3), 54, 5575–5584.
- Collins, J. J., Chow, C. C., & Imhoff, T. T. (1995). Stochastic resonance without tuning. *Nature*, 376, 236–239.
- Douglas, J. K., Wilkens L., Pantazelou E., & Moss, F. (1993). Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, 365, 337–340.
- Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.
- Gammaitoni, L., Hänggi, P., Jung, P., & Marchesoni, F. (1998). Stochastic resonance. *Rev. Modern Phys.*, 70, 223–287.
- Gasser, T., Müller, H.-G., & Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Stat. Soc. Ser. B*, 47, 238–252.
- Greenwood, P. E., Ward, L. M., & Wefelmeyer, W. (1999). Statistical analysis of stochastic resonance in a simple setting. *Phys. Rev. E*, in press.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Härdle, W., & Marron, J. S. (1995). Fast and simple scatterplot smoothing. *Comput. Statist. Data Anal.*, 20, 1–17.

- Henegan, C., Chow, C. C., Collins, J. J., Imhoff, T. T., Lowen, S. B., & Teich, M. C. (1996). Information measures quantifying aperiodic stochastic resonance. *Phys. Rev. E* (3), 54, R2228–R2231.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5, 115–133.
- Müller, U. U. (1999). Nonparametric regression for threshold data. *Univ. Bremen, Mathematik–Arbeitspapiere A*, 52.
- Müller, U. U., & Ward, L. M. (1999). Stochastic resonance in a statistical model of a time-integrating detector. Unpublished manuscript.
- Ruppert, D., Sheather S. J., & Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, 90, 1257–1270.
- Stemmler, M. (1996). A single spike suffices: The simplest form of stochastic resonance in model neurons. *Network: Computations in Neural Systems*, 7, 687–716.
- Wand, M. P., & Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
- Wiesenfeld, K., & Moss, F. (1995). Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDS. *Nature*, 373, 33–36.

Received 30 July 1998

Accepted 3 September 1999

*Ursula U. Müller
 Institut für Statistik
 Fachbereich Mathematik/Informatik
 Universität Bremen
 28334 Bremen
 Germany*

e-mail: uschi@math.uni-bremen.de