Nonparametric regression for threshold data

Ursula U. Müller*
University of British Columbia and University of Bremen

Abstract

Consider a detector which records the times at which the realizations of a nonparametric regression model exceed a certain threshold. If the error distribution is known, the regression function can still be identified from these threshold data. We construct estimators for the regression function. They are transformations of kernel estimators. We determine the bandwidth which minimizes the asymptotic mean average squared error, and construct adaptive estimators for the optimal bandwidth by plug-in methods. Our work is motivated by recent work on stochastic resonance in neuro-science and signal detection theory. In this work it was observed empirically and by simulations that detection of a subthreshold signal is enhanced by the addition of noise, and that there is an optimal noise level. The present work seems to be the first effort to study theoretically how to make best use of this type of threshold data. We compare our model with several models in the literature.

Keywords: binary data, kernel regression, optimal bandwidth, stochastic resonance.

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. The glaciation sequence has an average periodicity of about 10^5 years. The only known comparable periodicity in earth dynamics is the modulation period of the orbital eccentricity. This causes small variations of the solar energy influx, i.e. a weak signal. Benzi et al. model climate changes as transitions in a double-well potential system pushed by the signal. The two minima in the system represent a largely ice-covered earth and the earth in its current state. Since the periodic forcing for switching from one state point to the other is very weak, the frequency of hopping across the barrier of the two wells must be assisted by other factors such as short term climate fluctuations which 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). Especially in neuro-

*work supported by NSERC, Canada, and the Crisis Points Group of the Peter Wall Institute for Advanced Studies at UBC.

science it attracted particular attention within the last years. Various models for information processing in the nervous system were proposed which are explained by stochastic resonance. Here, and in contrast to the original context, 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. Stochastic resonance now comes in quite naturally. Neurons typically have a threshold below which they do not respond. It can be assumed that the sensory system is optimized so that the spike train of a firing cell contains significant information about the signal. The presence of noise, e.g. background noise from other neurons, leads to maximal performance, i.e. stochastic resonance (Wiesenfeld and Moss, 1995). Stochastic resonance was studied not only theoretically but could also be exhibited in several experiments. For example, Douglass et al. (1993) studied the neural response of sinusoidally stimulated mechanoreceptor cells of crayfish. They applied a tone and some external noise to a neuron (an experimental change of the intrinsic noise is also possible but not straightforward) and, varying the noise level, could demonstrate stochastic resonance in the firing rate.

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 (see, for example, Wiesenfeld and 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 in the case of a (nearly) constant signal was considered e.g. by Stemmler (1996). Instead of commonly adopted measures as those above, 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.

Like the articles of 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 which gives an estimation of the firing rate. Clearly there is additional information in the timing of the spikes, especially then 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 are Bernoulli variables coded by 1 and 0 as follows:

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

which was suggested by Mc Culloch and Pitts (1943) as a model of a neuron.

We can always estimate the probabilities

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

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

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

The estimator of $s(t_i)$ will be taken to be the kernel estimator for $p(t_i)$ transformed as in (1).

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. It may, for example, be plausible in certain contexts to assume that the errors are normally distributed. We can, however, not identify the noise amplitude from the $X(t_i)$. There are two ways of dealing with the problem. One 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 constant signal); or, better still, by using *several* detectors with different thresholds. This will also improve the quality of the estimators of the signal. The second 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, of course, the signal is constant.

As in every approach using 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 and their article 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 utilized.

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 a threshold $a > 0$ and a signal $s : [0,1] \to \mathbf{R}$ which is subthreshold, i.e. $s(\cdot) < a$. Let $t_i = i/n$, $i = 1, \ldots, n$, be equally spaced time points in $[0,1]$. For $t \in [0,1]$ let F_t be a distribution function. Consider noise represented by independent random variables $\epsilon(t_i)$ with distribution functions F_{t_i}, $i = 1, \ldots, n$. The noisy signal is $s(t_i) + \epsilon(t_i)$, $i = 1, \ldots, n$. The *threshold data* are the exceedances of the noisy signal over the threshold. They are independent Bernoulli variables

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

with parameter

$$p(t_i) = 1 - F_{t_i}(a - s(t_i)).$$

We assume that the noise distributions are positive with mean zero and that the distribution functions are continuous. (Often the $\epsilon(t_i)$ will be identically distributed, and the distribution will be normal.) In this case, the signal can be identified from the threshold data (see (1)),

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

We also assume that

$$s \text{ has two continuous derivatives,} \tag{2}$$

$$s \text{ is bounded from below: } s(t) \geq -c \text{ for every } t \in [0,1] \quad (c > 0), \tag{3}$$

$$F_t \text{ is four times continuously differentiable,} \tag{4}$$

$$\text{there exist } p_1, p_2 \in (0,1) \text{ such that for all } t \in [0,1], \quad 1 - F_t(a + c) \geq p_1 \text{ and } F_t(0) \geq p_2. \tag{5}$$

These conditions imply

$$p \text{ has two continuous derivatives,} \tag{6}$$

$$p \text{ is bounded away from 0 and 1, i.e. } p(t) \in [p_1, 1 - p_2] \text{ for each } t. \tag{7}$$

Since $p(t) = 1 - F_t(a - s(t))$, relation (6) is an immediate consequence of (2) and (4). The upper bound $p(t) \leq 1 - p_2$ in (7) follows from the subthreshold assumption $s(t) < a$ and $F_t(0) \geq p_2$, see (5), which give $p(t) = 1 - F_t(a - s(t)) \leq 1 - F_t(0) \leq 1 - p_2$. The lower bound is obtained using $s(t) > -c$ (3), i.e. $a - s(t) \leq a + c$, which combined with (5) implies $p(t) = 1 - F_t(a - s(t)) \geq 1 - F_t(a + c) \geq p_1$.

To simplify the notation we write

$$G_t(x) = 1 - F_t(a - x).$$

Then $p(t) = G_t(s(t))$ 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

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

Here $\tilde{p}_h(t)$ is a classical kernel estimator $\hat{p}_h(t)$ if the values of $\hat{p}_h(t)$ lie in a compact subset C of $[0,1]$. If they are near 0 and 1 an estimator $G_t^{-1}(\hat{p}_h(t))$ is not defined. For simplicity set $\tilde{p}_h(t)$ equal to an arbitrary constant there. The kernel estimator which will be chosen here is the Nadaraya–Watson estimator

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

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

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

We also need that the derivative K' is 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 (see, for example, Gasser, Müller and Mammitzsch, 1985).

Instead of the Nadaraya–Watson estimator one could also consider a local linear kernel estimator (cf. Ruppert, Sheather and Wand, 1995), which is known for its superior boundary behavior but requires more work. The proofs carried out here can be adapted to this approach in a straightforward way.

For each t the estimator for $p(t)$ is now defined as follows:

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

Because of (7) we have $p(t) \in C$. The value $\tilde{p}_h(t) = 1/2$ for the case $\hat{p}_h(t) \notin C$ was chosen arbitrarily and could be replaced by any suitable constant $c \in (0, 1)$. This formal trick is, as already mentioned, only necessary in order to guarantee that $\hat{s}_h(t) = G_t^{-1}(\tilde{p}_h(t))$ is well-defined, and is without relevance for the asymptotic behavior. In the finite sample situation the extreme case $\hat{p}_h(t) = 0$ or 1 only occurs if $X(t_i) = 0$ for each i (resp. $X(t_i) = 1$ for each i). In this case no information about the signal can be obtained and the smoothing approach breaks down.

Later we take

$$h = h_n \to 0 \quad \text{and} \quad nh^3 \to \infty \quad \text{as} \quad n \to \infty.$$

3 Asymptotic Error and Optimal Bandwidth

In nonparametric regression theory, generally accepted measures of the goodness of the estimation are the mean squared error $E(\hat{s}_h(t) - s(t))^2$ and the mean average squared error $1/n \sum_{t_i \in T} E(\hat{s}_h(t_i) - s(t_i))^2$. In the latter case, because of boundary effects, summation is usually restricted to some interval $T = [c, d] \subset (0, 1)$. 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 besides the mean average squared error would be the mean *integrated* squared error $\int_T E(\hat{s}_h(t) - s(t))^2 dt$. The proofs of this article 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 which 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 $\hat{s}_h(t) = G_t^{-1}(\tilde{p}_h(t))$. The further problem arising through 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 $\hat{p}_h(t)$. Part of the derivation of the asymptotic mean squared error reduces to the known case not involving a nonlinear transformation.

The next two lemmas state well-known results from classical theory. In particular, we give approximation formulas of variance and bias of $\hat{p}_h(t)$ (compare Eubank, 1988, e.g.). In addition, also higher order moments will be analyzed. These results are necessary for the derivation of the asymptotic mean squared and the mean average squared error in the setting considered here. The asymptotic mean squared errors and the optimal local and global bandwidths will then be derived in Theorem 3.3. The proofs of Lemma 3.1 and 3.2 will be given in the Appendix.

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 3.1 *Consider the asymptotics $n \to \infty$, $h = h_n \to 0$ and $nh^2 \to \infty$. For the variance and the central third moments of $\hat{p}_h(t)$ we have the approximation*

$$E(\hat{p}_h(t) - E(\hat{p}_h(t)))^l = \frac{1}{(nh)^{l-1}} \cdot E(X(t) - p(t))^l \cdot \int_{-1}^{1} K^l(u)du + o(\frac{1}{(nh)^{l-1}}) \quad (l = 2, 3)$$

uniformly in $t \in [h, 1-h]$. Here $E(X(t) - p(t))^2 = Var(X(t)) = p(t)(1-p(t))$ and $E(X(t) - p(t))^3 = p(t) - 3p(t)^2 + 2p(t)^3$ are the second and third moments of the $B(1, p(t))$ distribution. For the central fourth moments one obtains the same order as for the third moments:

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

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

Lemma 3.2 *For $n \to \infty$, $h = h_n \to 0$ and $nh^2 \to \infty$ the bias and the mean squared error of $\hat{p}_h(t)$ are uniformly in $t \in [h, 1-h]$ approximated as follows:*

$$\begin{aligned} E(\hat{p}_h(t) - p(t)) &= h^2 \cdot \frac{1}{2} \cdot p''(t) \cdot \mu_2(K) + o(h^2) + O(\frac{1}{nh^2}), \quad (8) \\ E(\hat{p}_h(t) - p(t))^2 &= Var(\hat{p}_h(t)) + (E(\hat{p}_h(t)) - p(t))^2 \\ &= \frac{1}{nh} p(t)(1-p(t))R(K) + \frac{h^4}{4} p''(t)^2 \mu_2(K)^2 \\ &\quad + o(\frac{1}{nh}) + o(h^4) + O(\frac{1}{(nh^2)^2}) \quad (9) \end{aligned}$$

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

$$\begin{aligned} E(\hat{p}_h(t) - p(t))^3 &= o(\frac{1}{nh}) + O(h^6) + O(\frac{1}{(nh^2)^3}), \quad (10) \\ E(\hat{p}_h(t) - p(t))^4 &= o(\frac{1}{nh}) + O(h^8) + O(\frac{1}{(nh^2)^4}), \quad (11) \end{aligned}$$

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

We state our main theorem. It is about 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 3.1 and 3.2. We give the asymptotic mean squared error (locally), the asymptotic mean average squared error (globally) and the respective optimal bandwidths.

Theorem 3.3 *Consider the asymptotics $n \to \infty$, $h = h_n \to 0$ and $nh^3 \to \infty$. The mean squared error $MSE(h, t) = E(\hat{s}_h(t) - s(t))^2$ exists and is approximated by the asymptotic mean squared error $AMSE(h, t)$ up to an term of order $o(1/(nh) + h^4)$ as follows:*

$$AMSE(h, t) = \frac{1}{G_t'(s(t))^2} \cdot \Big(\frac{1}{nh} p(t)(1-p(t)) \cdot R(K) + \frac{h^4}{4} p''(t)^2 \cdot \mu_2(K)^2 \Big) \quad (12)$$

uniformly in $t \in [h, 1-h]$, where $R(K) = \int_{-1}^{1} K^2(u)du$ and $\mu_2(K) = \int_{-1}^{1} u^2 K(u)du$. For the mean average squared error $MASE(h) = 1/n \sum_{t_i \in T} E(\hat{s}_h(t_i) - s(t_i))^2$ one obtains up to an *term of order $o(1/(nh)+h^4)$ the approximation*

$$
\begin{aligned}
AMASE(h) &= \frac{1}{n^2 h} R(K) \sum_{t_i \in T} \frac{1}{G'_{t_i}(s(t_i))^2} \cdot p(t_i)(1-p(t_i)) \\
&\quad + \frac{h^4}{n} \cdot \frac{1}{4} \cdot \mu_2(K)^2 \sum_{t_i \in T} \frac{1}{G'_{t_i}(s(t_i))^2} \cdot p''(t_i)^2. \qquad (13)
\end{aligned}
$$

The asymptotically optimal local bandwidth is

$$
h_{opt}(t) = n^{-1/5} \cdot \Big(\frac{R(K) p(t)(1-p(t))}{\mu_2(K)^2 p''(t)^2} \Big)^{1/5} \qquad (14)
$$

and the asymptotically optimal global bandwidth is

$$
h_{opt} = n^{-1/5} \cdot \Big(\frac{R(K) \sum_{t_i \in T} \frac{1}{G'_{t_i}(s(t_i))^2} p(t_i)(1-p(t_i))}{\mu_2(K)^2 \sum_{t_i \in T} \frac{1}{G'_{t_i}(s(t_i))^2} p''(t_i)^2} \Big)^{1/5}. \qquad (15)
$$

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 $H_t : (0,1) \to \mathbf{R}$ defined by

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

Expand $H_t(\tilde{p}_h(t))$ in a Taylor series around $p(t)$ as follows:

$$
\begin{aligned}
H_t(\tilde{p}_h(t)) &= H_t(p(t)) + H'_t(p(t))(\tilde{p}_h(t) - p(t)) + \frac{1}{2} H_t^{(2)}(p(t))(\tilde{p}_h(t) - p(t))^2 \\
&\quad + \frac{1}{3!} H_t^{(3)}(p(t))(\tilde{p}_h(t) - p(t))^3 + R_t(\tilde{p}_h(t))
\end{aligned}
$$

with

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

Since the estimator $\hat{p}_h(t)$ is a weighted sum of Bernoulli variables, the moments $E(\hat{p}_h(t) - p(t))^k$ and thus $E(\tilde{p}_h(t) - p(t))^k$ exist for $k = 1, \ldots, 4$. Hence the expected value of $H_t(\tilde{p}_h(t))$, i.e. the mean squared error $E(H_t(\tilde{p}_h(t))) = E(G_t^{-1}(\tilde{p}_h(t)) - G_t^{-1}(p(t)))^2 = E(\hat{s}_h(t) - s(t))^2$, is

$$
\begin{aligned}
& E(H_t(\tilde{p}_h(t))) \\
&= H_t(p(t)) + H'_t(p(t)) E(\tilde{p}_h(t) - p(t)) + \frac{1}{2} H_t^{(2)}(p(t)) E(\tilde{p}_h(t) - p(t))^2 \\
&\quad + \frac{1}{3!} H_t^{(3)}(p(t)) E(\tilde{p}_h(t) - p(t))^3 + E(R_t(\tilde{p}_h(t))) \\
&= \frac{1}{2} H_t^{(2)}(p(t)) E(\tilde{p}_h(t) - p(t))^2 + \frac{1}{3!} H_t^{(3)}(p(t)) E(\tilde{p}_h(t) - p(t))^3 + E(R_t(\tilde{p}_h(t))). \qquad (16)
\end{aligned}
$$

Here we have used $H_t(p(t)) = 0$ and $H'_t(p(t)) = 0$.
In order to establish the asserted approximation, we use the auxiliary results, which hold uniformly in $t \in [h, 1-h]$,

$$
\begin{aligned}
E(\tilde{p}_h(t) - p(t))^l - E(\hat{p}_h(t) - p(t))^l &= o(\frac{1}{nh}) + o(h^4) \quad \text{for every } l \in \mathbf{N}, \qquad (17) \\
E(R_t(\tilde{p}_h(t))) &= o(\frac{1}{nh}) + o(h^4). \qquad (18)
\end{aligned}
$$

They will be verified at the end of the proof. Inserting (17) and (18) into (16), we get

$$E(H_t(\tilde{p}_h(t))) = \frac{1}{2}H_t^{(2)}(p(t))E(\hat{p}_h(t) - p(t))^2 + \frac{1}{3!}H_t^{(3)}(p(t))E(\hat{p}_h(t) - p(t))^3 + o(\frac{1}{nh}) + o(h^4).$$

Relation (10) and $nh^3 \to \infty$ give $E(\hat{p}_h(t) - p(t))^3 = o(\frac{1}{nh}) + o(h^4)$ uniformly in $t \in [h, 1-h]$. Hence

$$E(H_t(\tilde{p}_h(t))) = \frac{1}{2}H_t^{(2)}(p(t))E(\hat{p}_h(t) - p(t))^2 + o(\frac{1}{nh}) + o(h^4). \tag{19}$$

Inserting (9), and using $nh^3 \to \infty$, and $H_t^{(2)}(p(t)) = 2/(G_t'(s(t)))^2$ into (19), we obtain uniformly for $t \in [h, 1-h]$ the desired approximation (12) of $E(\hat{s}_h(t) - s(t))^2 = E(H_t(\tilde{p}_h(t)))$,

$$E(\hat{s}_h(t) - s(t))^2 = \frac{1}{G_t'(s(t))^2} \cdot \Big(\frac{1}{nh}p(t)(1-p(t)) \cdot R(K) + \frac{h^4}{4}p''(t)^2 \cdot \mu_2(K)^2\Big) + o(\frac{1}{nh}) + o(h^4).$$

Since $T \subset [h, 1-h]$ for sufficiently small h, the approximation formula AMASE(h) for the mean average squared error $MASE(h) = 1/n \sum_{t_i \in T} MSE(h, t_i)$ is immediately derived from this result. The optimal local and global bandwidth given in (14) and (15) are obtained by simple calculus, differentiating $AMSE(h,t)$ and $AMSE(h)$ with respect to h. Hence only the auxiliary statements (17) and (18) remain to be shown.

For the proof of (17) 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 (7) 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:

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

In the last step, Markov's inequality was applied. Relation (17) now follows from (11) and $nh^3 \to \infty$.

For the proof of (18) we will use the boundedness of $H_t^{(4)}$ on C which holds by assumption (4). Then,

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

$$\leq \quad E(\tilde{p}_h(t)-p(t))^4 \cdot E\Big(\int_0^1 \frac{(1-z)^3}{3!}|H_t^{(4)}(p(t)+z(\tilde{p}_h(t)-p(t)))|dz\Big)$$
$$\leq \quad E(\tilde{p}_h(t)-p(t))^4 \cdot \sup_{x\in C}|H_t^{(4)}(x)|.$$

Further using (17) for $l=4$ and $E(\hat{p}_h(t)-p(t))^4 = o((nh)^{-1})+o(h^4)$, this immediately establishes (18). Hence the proof is complete. □

Let us discuss the approximate mean squared error calculated in (12),

$$AMSE(h,t) \quad = \quad \frac{1}{G_t'(s(t))^2}\cdot\Big(\frac{1}{nh}p(t)(1-p(t))\cdot R(K)+\frac{h^4}{4}p''(t)^2\cdot\mu_2(K)^2\Big).$$

The expression in large parentheses is the Taylor approximation of the mean squared error $E(\hat{p}_h(t)-p(t))^2$ of the kernel estimator $\hat{p}_h(t)$. In the usual nonparametric regression setting with binary responses, this formula is well-known, giving the decomposition of $E(\hat{p}_h(t)-p(t))^2$ into variance and squared bias of $\hat{p}_h(t)$,

$$Var(\hat{p}_h(t)) \quad \doteq \quad \frac{1}{nh}p(t)(1-p(t))\cdot R(K),$$
$$(E(\hat{p}_h(t)-p(t)))^2 \quad \doteq \quad \frac{h^4}{4}p''(t)^2\cdot\mu_2(K)^2$$

(see Lemma 3.1 and Lemma 3.2, relation (8)). In particular, the characteristic variance–bias trade-off becomes evident: With decreasing h, the bias decreases and the variance increases. The aim is to find an optimal balance between both terms.

With the optimal asymptotic bandwidth h_{opt} at hand, the minimal value of $AMASE$ (see (13)), namely $AMASE(h_{opt})$, can now be derived:

$$\inf_{h>0} AMASE(h)$$
$$= \quad \frac{5}{4}n^{-\frac{4}{5}}\Big(\mu_2(K)^2\frac{1}{n}\sum_{t_i\in T}\frac{1}{G_{t_i}'(s(t_i))^2}p''(t_i)^2\Big)^{\frac{1}{5}}\Big(R(K)\frac{1}{n}\sum_{t_i\in T}\frac{1}{G_{t_i}'(s(t_i))^2}p(t_i)(1-p(t_i))\Big)^{\frac{4}{5}}.$$

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 $\inf_{h>0} AMASE(h)$. Further, this formula shows the influence of the kernel K, which appears only in the expression $\mu_2(K)^{2/5}\cdot R(K)^{4/5}$. The (second order) kernel which minimizes this term, and thus the asymptotic mean average squared error under the further constraint $K\geq 0$, is the Epanechnikov kernel K^*. This kernel is unique up to a scale parameter. If the scale parameter is chosen such that the kernel has support $[-1,1]$, the Epanechnikov kernel is

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

The optimal kernel K^* is not much better than other kernels, for example the Gaussian kernel (see Wand and 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 $\mu_2(K)=\int_{-1}^1 u^2K(u)\,du$ are known.

Recall the asymptotically optimal local bandwidth in Theorem 3.3,

$$h_{opt}(t) = n^{-1/5} \cdot c(p(t), p''(t)), \qquad c(p(t), p''(t)) = \Big(\frac{R(K)p(t)(1-p(t))}{\mu_2(K)^2 p''(t)^2} \Big)^{1/5}.$$

The following arguments will also apply to the optimal *global* bandwidth h_{opt} (Theorem 3.3). Both are of the form $n^{-1/5} \cdot c$, with the constant 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 $h = n^{-1/5} \cdot c$, where $c > 0$ is an arbitrary constant. The choice of c has, however, a strong influence on the finite sample behavior. 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. In order to find a pilot estimator for p, one can, for example, apply various quick and simple methods, beginning with certain "rules of thumb" up to a more recent approach, the so-called blocking method, introduced by Härdle and Marron (1995). There the design space, the unit interval in our setting, 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. Then, however, a new bandwidth selection problem arises. This could again be tackled with a third smoother, and so on, but at some point a pilot bandwidth has to be determined with a different technique, usually cross-validation, or the blocking method, or by fitting a parametric model.

A comprehensive overview of bandwidth selection techniques is given by Wand and Jones (1995), Chapter 3. They consider density estimation, but the methods carry over to nonparametric regression. In the latter setting several plug-in methods are discussed in Ruppert et al. (1995), where an asymptotic approach similar to this article is considered. 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 many different neuron models, using various measures of signal detection. In several papers (Collins et al. 1995 and 1996; Henegan et al. 1996; Chialvo et al. 1997a and 1997b) 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 bandwidth chosen fixed, which 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

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

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

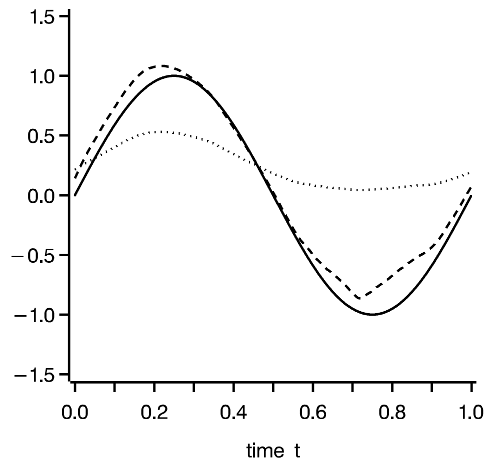

Figure 1: Simulation of signal estimates $\hat{s}_h$ (dashed line) for $s(t) = \sin(2\pi t)$ (solid line), with n=1000, a=1, h=0.16 and i.i.d. $N(0, 1.08^2)$ noise. The dotted line shows the estimates $\hat{p}_h$ for the probabilities.

An illustration of our technique is given in Figure 1 which 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 Härdle and Marron's (1995) blocking method, 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 $\hat{p}_h$ which correspond to the box kernel mentioned above. Of course, these estimators are not consistent for the signal. Nevertheless, the Pearson correlation between s and $\hat{p}_h$ as well as the correlation between s and our consistent estimator $\hat{s}_h$ both are close to 1 in this example ($C = 0.977$ resp. $C = 0.995$). This is, however, not surprising since C is invariant with respect to linear transformations and $\hat{p}_h$ still catches essential characteristics of the shape of s.

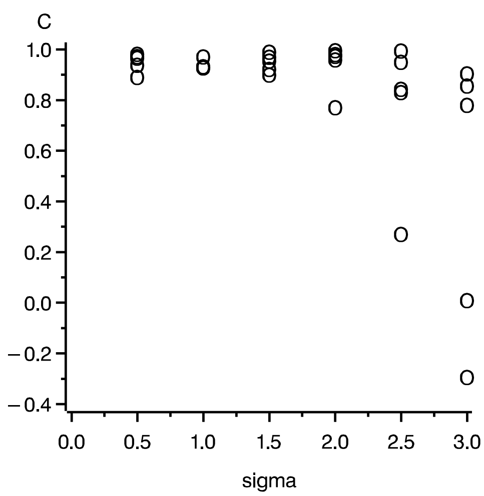

Figure 2: Realizations of the Pearson correlation C between the sinusoid s from Figure 1 and its estimates $\hat{s}_h$, for $\sigma = 0.5, 1, \ldots, 3$ and $n = 100$.

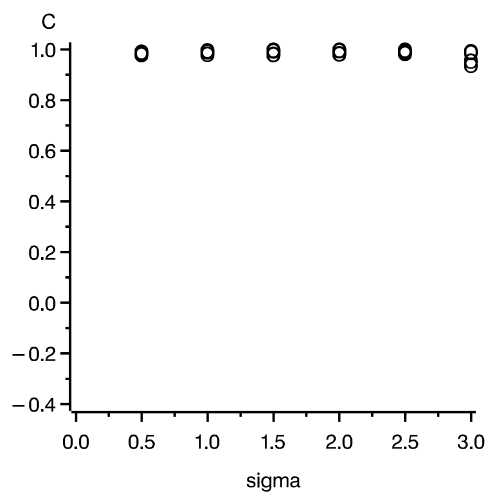

Figure 3: Simulation of the Pearson correlation C like Figure 2, but with $n = 1000$.

For $n = 100$, Figure 2 shows, for each value of $\sigma = 0.5, 1, \ldots, 3$, five realizations of the Pearson correlation between s and $\hat{s}_h$. 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: For $\sigma = 1$ the five values of C are all close to one; for smaller and larger σ, the correlation is not always high. In Figure 3 the same simulation was performed but with $n = 1000$ time points, getting values C close to one throughout. This picture, combined with Figure 2, illustrates well a phenomenon first observed by Collins et al. (1995) in a different setting, called "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 be 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 behavior for all signals or for all error distributions. For the example with the sinusoid from above stochastic resonance is easily verified. In Figure 4 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 with 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.

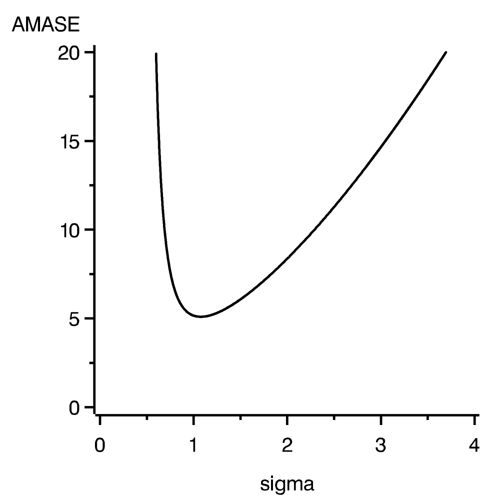

Figure 4: Plot of $n^{4/5} \cdot \inf_h AMASE(\sigma, h)$ for the sinusoid example, $s(t) = \sin(2\pi t)$, with n sufficiently large (here $n = 1000$).

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. Then the function p of Section 2 is $p(t) = \Phi\{(s(t) - a)/\sigma\}$ with Φ denoting the standard normal $N(0, 1)$ distribution function and the asymptotic mean squared error is

$$\begin{aligned} AMSE(h,t) &= \frac{1}{nh}\frac{\sigma^2\Phi(\frac{s(t)-a}{\sigma})\Phi(\frac{a-s(t)}{\sigma})}{\phi^2(\frac{s(t)-a}{\sigma})}R(K) \\ &+\frac{h^4}{4}\Big(\frac{a-s(t)}{\sigma^2}s'(t)^2+s''(t)\Big)^2\mu_2(K)^2. \end{aligned}$$

(ϕ is the $N(0,1)$ density). 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 behavior: It tends to infinity when $\sigma^2 \to \infty$ or $\sigma^2 \to 0$ (see Greenwood *et al.* 1999). The second term, however, varies like $1/\sigma^4$. The behavior of the sum requires further analysis.

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

6 Appendix

In the following, Lemmas 3.1 and 3.2 will be proved. Both proofs require an auxiliary result concerning the approximation of a kernel weighted sum through an integral, which will be given first.

Lemma 6.1 *Let the terminology from Section 2 be given. Consider some fixed* $m \in \mathbf{N}$ *and let* f *be a function on* $[0,1]$ *with bounded derivative. Then we have for the asymptotics* $n \to \infty$ *and* $h = h_n \to 0$

$$\frac{1}{n}\sum_{i=1}^{n} K^m(\frac{t_i-t}{h})f(t_i) = \int_0^1 K^m(\frac{s-t}{h})f(s)ds + O(\frac{1}{nh})$$

uniformly in $t \in [0,1]$.

Proof. We prove the existence of a constant $c > 0$ such that

$$\sup_{t\in[0,1]}\Big|\int_0^1 K^m(\frac{s-t}{h})f(s)ds - \frac{1}{n}\sum_{i=1}^{n} K^m(\frac{t_i-t}{h})f(t_i)\Big| \le \frac{c}{nh}. \tag{20}$$

Set $t_0 = 0$ and let $H(\cdot) = K^m((\cdot - t)/h)f(\cdot)$. In the following we will need the derivative of H,

$$H'(\cdot) = \frac{1}{h}mK^{m-1}\Big(\frac{\cdot - t}{h}\Big)\cdot K'\Big(\frac{\cdot - t}{h}\Big)\cdot f(\cdot) + K^m\Big(\frac{\cdot - t}{h}\Big)\cdot f'(\cdot),$$

especially

$$\sup_{x\in[0,1]}|H'(x)| \le \frac{c}{h}$$

for some constant c, which follows directly from the assumed boundedness of the derivatives K' and f'. Then we get (20) as follows:

$$
\begin{aligned}
& \Big| \int_0^1 K^m(\frac{s-t}{h}) f(s) ds - \frac{1}{n}\sum_{i=1}^n K^m(\frac{t_i-t}{h}) f(t_i) \Big| \\
= & \Big| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \Big(K^m(\frac{s-t}{h}) f(s) - K^m(\frac{t_i-t}{h}) f(t_i) \Big) ds \Big| \\
= & \Big| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \Big(H(s) - H(t_i) \Big) ds \Big| \\
\le & \sum_{i=1}^n \int_{t_{i-1}}^{t_i} |s-t_i| \cdot \sup_{x\in[0,1]} |H'(x)| ds \\
\le & \frac{1}{n} \cdot \sup_{x\in[0,1]} |H'(x)| \\
\le & \frac{c}{nh}.
\end{aligned}
$$

□

Proof of Lemma 3.1.
Let $t \in [h, 1-h]$. For arbitrary $l \in \mathbf{N}$,

$$
\begin{aligned}
E\Big(\hat{p}_h(t) - E(\hat{p}_h(t))\Big)^l &= E\left(\frac{\sum_{i=1}^n \frac{1}{h} K(\frac{t_i-t}{h}) \cdot X_i}{\sum_{i=1}^n \frac{1}{h} K(\frac{t_i-t}{h})} - \frac{\sum_{i=1}^n \frac{1}{h} K(\frac{t_i-t}{h}) \cdot E(X_i)}{\sum_{i=1}^n \frac{1}{h} K(\frac{t_i-t}{h})} \right)^l \\
&= E\left(\frac{\frac{1}{(nh)^l} (\sum_{i=1}^n K(\frac{t_i-t}{h}) \cdot (X_i - p(t_i)))^l}{(\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i-t}{h}))^l} \right).
\end{aligned}
$$

Consider the denominator and let h be such that $[-1,1] \subset [-t/h, (1-t)/h]$, which is possible since $t \in [h, 1-h]$. Application of Lemma 6.1 for the special case $f \equiv 1$ and $m = 1$ and change of variable give uniformly for all $t \in [h, 1-h]$,

$$
\begin{aligned}
\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i-t}{h}) &= \frac{1}{h}\int_0^1 K(\frac{s-t}{h}) + O(\frac{1}{nh^2}) \\
&= \frac{1}{h}\int_{-t/h}^{(1-t)/h} K(u) \cdot h\, du + O(\frac{1}{nh^2}) \\
&= \int_{-1}^1 K(u) du + O(\frac{1}{nh^2}) \\
&= 1 + O(\frac{1}{nh^2}),
\end{aligned}
$$

and thus the following approximation of the denominator

$$
\Big(\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i-t}{h})\Big)^l = 1 + O(\frac{1}{nh^2}) \quad \text{for each } l \in \mathbf{N}, \tag{21}
$$

uniformly in $t \in [h, 1-h]$. Since we require $nh^2 \to \infty$, we may set the denominator equal to 1. Using $E(X_i - p(t_i)) = 0$ and the independence of $X_1, \ldots, X_n$, we have for $l = 2, 3$,

$$
E(\frac{1}{(nh)^l}(\sum_{i=1}^n K(\frac{t_i-t}{h}) \cdot (X_i - p(t_i)))^l) = \frac{1}{(nh)^l}\sum_{i=1}^n K^l(\frac{t_i-t}{h}) \cdot E(X_i - p(t_i))^l.
$$

For $l = 4$ the numerator computes to

$$
\begin{aligned}
& E\Big(\frac{1}{(nh)^4}\Big(\sum_{i=1}^{n} K(\frac{t_i - t}{h}) \cdot (X_i - p(t_i))\Big)^4\Big) \\
&= \frac{1}{(nh)^4}\sum_{i=1}^{n} K^4(\frac{t_i - t}{h}) \cdot E(X_i - p(t_i))^4 \\
&\quad + \frac{3}{(nh)^4}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} K^2(\frac{t_i - t}{h}) \cdot E(X_i - p(t_i))^2 \cdot K^2(\frac{t_j - t}{h}) \cdot E(X_j - p(t_j))^2 \\
&\leq \frac{1}{(nh)^4}\sum_{i=1}^{n} K^4(\frac{t_i - t}{h}) \cdot E(X_i - p(t_i))^4 \\
&\quad + 3 \cdot \frac{1}{(nh)^2}\sum_{i=1}^{n} K^2(\frac{t_i - t}{h}) \cdot E(X_i - p(t_i))^2 \cdot \frac{1}{(nh)^2}\sum_{j=1}^{n} K^2(\frac{t_j - t}{h}) \cdot E(X_j - p(t_j))^2. \qquad (22)
\end{aligned}
$$

In order to get the desired formulas for $l = 2, 3, 4$ we consider the following expressions:

$$
\frac{1}{(nh)^l}\sum_{i=1}^{n} K^l(\frac{t_i - t}{h}) \cdot E(X_i - p(t_i))^l \quad (l = 2, 3, 4).
$$

Let $f_l(t) = E(X_t - p(t))^l$. These moments are known polynomials in $p(t)$, and by (6) they are continuous functions in $t \in [0, 1]$ with continuous derivative. Hence Lemma 6.1 can be applied. This and further arguments such as the continuity of f_l and $[-1, 1] \subset [-t/h, (1 - t)/h]$ give for $l = 2, 3, 4$,

$$
\begin{aligned}
& \frac{1}{(nh)^l}\sum_{i=1}^{n} K^l(\frac{t_i - t}{h}) f_l(t_i) \\
&= \frac{1}{n^{l-1}h^l}\int_0^1 K^l(\frac{s - t}{h}) f_l(s) ds + O(\frac{1}{n^l h^{l+1}}) \\
&= \frac{1}{(nh)^{l-1}}\int_{-t/h}^{(1-t)/h} K^l(u) f_l(t + hu) du + O(\frac{1}{n^l h^{l+1}}) \\
&= \frac{1}{(nh)^{l-1}}\int_{-1}^{1} K^l(u) f_l(t) du + \frac{1}{(nh)^{l-1}} \cdot o(1) + O(\frac{1}{(nh)^{l-1}} \cdot \frac{1}{nh^2}) \\
&= \frac{1}{(nh)^{l-1}} f_l(t)\int_{-1}^{1} K^l(u) du + o(\frac{1}{(nh)^{l-1}}) \\
&= \frac{1}{(nh)^{l-1}} E(X_t - p(t))^l \int_{-1}^{1} K^l(u) du + o(\frac{1}{(nh)^{l-1}}), \qquad (23)
\end{aligned}
$$

uniformly in $t \in [h, 1 - h]$. For $l = 2, 3$ this is the desired result. For $l = 4$, relation (23) will be used to derive the order of the terms in (22). Inserting gives

$$
\begin{aligned}
E\Big(\frac{1}{(nh)^4}\Big(\sum_{i=1}^{n} K(\frac{t_i - t}{h}) \cdot (X_i - p(t_i))\Big)^4\Big) &= O(\frac{1}{(nh)^3}) + (O(\frac{1}{nh}))^2 \\
&= O(\frac{1}{(nh)^2}),
\end{aligned}
$$

uniformly in $t \in [h, 1-h]$, which is the result for $l = 4$ $\square$

Proof of Lemma 3.2.

We consider $E(\hat{p}_h(t) - p(t))^l$ separately for $l = 1, \ldots, 4$.

In order to derive the bias approximation ($l = 1$), let $t \in [h, 1-h]$ and consider

$$E(\hat{p}_h(t)) = \frac{\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i - t}{h}) \cdot E(X_i)}{\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i - t}{h})} = \frac{\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i - t}{h}) \cdot p(t_i)}{\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i - t}{h})}.$$

By Lemma 3.1, relation (21), the denominator is $1 + O(1/(nh^2))$. Since the derivative of p is continuous, we obtain for the numerator, as in the proof of Lemma 3.1, and using Lemma 6.1,

$$\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i - t}{h}) \cdot p(t_i) \quad = \quad \int_{-t/h}^{(1-t)/h} K(u)p(t+hu)du + O(\frac{1}{nh^2}),$$

uniformly in $t \in [h, 1-h]$. By a second order Taylor expansion of $p(t+hu)$ around t, the right-hand integral is

$$p(t)\int_{-t/h}^{(1-t)/h} K(u)du + p'(t)\int_{-t/h}^{(1-t)/h} K(u)hu\,du + \frac{1}{2}\int_{-t/h}^{(1-t)/h} K(u)p''(t+\xi_u hu)h^2u^2du,$$

with $\xi_u \in [0,1]$ depending on u. Using the symmetry of K, we obtain for h sufficiently small

$$\begin{aligned}
&\frac{1}{nh}\sum_{i=1}^n K(\frac{t_i - t}{h}) \cdot p(t_i) \\
&= p(t) \cdot 1 + \frac{1}{2}\int_{-1}^{1} K(u)p''(t+\xi_u hu)h^2u^2du + O(\frac{1}{nh^2}) \\
&= p(t) + \frac{h^2}{2}\int_{-1}^{1} u^2K(u)p''(t)du + \frac{h^2}{2}\int_{-1}^{1} u^2K(u)(p''(t+\xi_u hu) - p''(t))du + O(\frac{1}{nh^2}) \\
&= p(t) + \frac{h^2}{2}p''(t)\int_{-1}^{1} u^2K(u)du + h^2o(1) + O(\frac{1}{nh^2}),
\end{aligned}$$

uniformly in $t \in [h, 1-h]$. In the last equation, the continuity of the second derivative of p was used.

The formula for the mean squared error can be derived from the preceding results by considering the following decomposition in variance and squared bias part,

$$E((\hat{p}_h(t) - p(t))^2) \quad = \quad Var(\hat{p}_h(t)) + (E(\hat{p}_h(t)) - p(t))^2.$$

Lemma 3.1 and the bias approximation (8) just proved give for these terms, uniformly in $t \in [h, 1-h]$,

$$\begin{aligned}
Var(\hat{p}_h(t)) &= \frac{1}{nh}p(t)(1-p(t))R(K) + o(\frac{1}{nh}), \\
(E(\hat{p}_h(t)) - p(t))^2 &= \Big(\frac{h^2}{2}p''(t)\mu_2(K) + o(h^2) + O(\frac{1}{nh^2})\Big)^2 \\
&= \frac{h^4}{4}p''(t)^2\mu_2(K)^2 + o(h^4) + O(\frac{1}{n}) + O(\frac{1}{n^2h^4}) \\
&= \frac{h^4}{4}p''(t)^2\mu_2(K)^2 + o(h^4) + o(\frac{1}{nh}) + O(\frac{1}{n^2h^4}),
\end{aligned}$$

which establishes (9).

The order of the error in the case $l = 3$ can be obtained using the same results, i.e. (8) and the variance formula. These and Lemma 3.1 give, uniformly in $t \in [h, 1-h]$,

$$\begin{aligned} E(\hat{p}_h(t) - p(t)) &= O(h^2) + O(\frac{1}{nh^2}), \\ E(\hat{p}_h(t) - E(\hat{p}_h(t)))^2 &= O(\frac{1}{nh}), \\ E(\hat{p}_h(t) - E(\hat{p}_h(t)))^3 &= O(\frac{1}{(nh)^2}). \end{aligned}$$

This establishes, uniformly in $t \in [h, 1-h]$,

$$\begin{aligned} &E(\hat{p}_h(t) - p(t))^3 \\ &= E(\hat{p}_h(t) - E(\hat{p}_h(t)))^3 + (E(\hat{p}_h(t) - p(t)))^3 + 3E(\hat{p}_h(t) - E(\hat{p}_h(t)))^2 \cdot (E(\hat{p}_h(t) - p(t))) \\ &= O(\frac{1}{(nh)^2}) + (O(h^2) + O(\frac{1}{nh^2}))^3 + O(\frac{1}{nh}) \cdot (O(h^2) + O(\frac{1}{nh^2})) \\ &= o(\frac{1}{nh}) + O(h^6) + O(\frac{1}{(nh^2)^3}). \end{aligned}$$

The above arguments also apply for $l = 4$. Additionally,

$$E(\hat{p}_h(t) - E(\hat{p}_h(t)))^4 = O(\frac{1}{(nh)^2})$$

(Lemma 3.1) will be used. Then,

$$\begin{aligned} & E(\hat{p}_h(t) - p(t))^4 \\ = \; & E(\hat{p}_h(t) - E(\hat{p}_h(t)))^4 + (E(\hat{p}_h(t) - p(t)))^4 \\ & +6E(\hat{p}_h(t) - E(\hat{p}_h(t)))^2 \cdot (E(\hat{p}_h(t) - p(t)))^2 + 4E(\hat{p}_h(t) - E(\hat{p}_h(t)))^3 \cdot E(\hat{p}_h(t) - p(t)) \\ = \; & O(\frac{1}{(nh)^2}) + (O(h^2) + O(\frac{1}{nh^2}))^4 \\ & +O(\frac{1}{nh}) \cdot (O(h^2) + O(\frac{1}{nh^2}))^2 + O(\frac{1}{(nh)^2}) \cdot (O(h^2) + O(\frac{1}{nh^2})) \\ = \; & o(\frac{1}{nh}) + O(h^8) + O(\frac{1}{(nh^2)^4}), \end{aligned}$$

uniformly in $t \in [h, 1-h]$, which completes the proof. □

References

Benzi R., Sutera A. and 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, 2, 1798 – 1808.

Chialvo D.R., Longtin A., Müller-Gerking, J. (1997b). Stochastic resonance in models of neuron ensembles revisited. To appear in: *Phys. Rev. E (3)*.

Collins J.J., Chow, C.C., Capela, A.C. and Imhoff, T.T. (1996). Aperiodic stochastic resonance. *Phys. Rev. E (3)*, 54, 5, 5575 – 5584.

Collins J.J., Chow, C.C. and Imhoff, T.T. (1995). Stochastic resonance without tuning. *Nature*, 376, 236 – 239.

Douglass, J.K., Wilkens L., Pantazelou E. and 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, Inc., New York and Basel.

Gammaitoni, L., Hänggi, P., Jung, P. and Marchesoni, F. (1998). Stochastic resonance. *Rev. Modern Phys.*, 70, 1, 223 – 287.

Gasser, T., Müller H.–G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Stat. Soc. Ser. B*, 47, 238 – 252.

Greenwood, P.E., Ward, L.M. and Wefelmeyer W. (1999). Statistical analysis of stochastic resonance in a simple setting. To appear in *Phys. Rev. E*.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

Härdle, W. and Marron, J.S. (1995). Fast and simple scatterplot smoothing. *Comput. Statist. Data Anal.*, 20, 1 – 17.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Henegan, C., Chow, C.C., Collins, J.J., Imhoff, T.T., Lowen, S.B. and Teich M.C. (1996). Information measures quantifying aperiodic stochastic resonance. *Phys. Rev. E (3)*, 54, 3, R2228 – R2231.

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5, 115 – 133.

Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Springer–Verlag, Berlin, Heidelberg.

Müller, U.U. and Ward, L.M. (1999). Stochastic resonance in a statistical model of a time-integrating detector. Preprint.

Ruppert, D., Sheather S.J. and Wand, M.P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Stat. Assoc.*, 90, 432, 1257 – 1270.

Stemmler, M. (1996). A single spike suffices: the simplest form of stochastic resonance in model neurons. *Network: Computations in Neural Systems*, 7, 4, 687 – 716.

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

Wiesenfeld, K., and Moss, F. (1995). Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. *Nature*, 373, 33 – 36.

August 1999