PHYSICAL REVIEW E VOLUME 60, NUMBER 4 OCTOBER 1999

Statistical analysis of stochastic resonance in a simple setting

Priscilla E. Greenwood*
Department of Mathematics,† University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z2

Lawrence M. Ward‡
Department of Psychology,§ University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

Wolfgang Wefelmeyer‖
FB 6 Mathematics,¶ University of Siegen, 57068 Siegen, Germany

(Received 11 December 1998; revised manuscript received 24 June 1999)

A subthreshold signal may be detected if noise is added to the data. We study a simple model, consisting of a constant signal to which at uniformly spaced times independent and identically distributed noise variables with known distribution are added. A detector records the times at which the noisy signal exceeds a threshold. There is an optimal noise level, called stochastic resonance. We explore the detectability of the signal in a system with one or more detectors, with different thresholds. We use a statistical detectability measure, the asymptotic variance of the best estimator of the signal from the thresholded data, or equivalently, the Fisher information in the data. In particular, we determine optimal configurations of detectors, varying the distances between the thresholds and the signal, as well as the noise level. The approach generalizes to nonconstant signals. [S1063-651X(99)17110-1]

PACS number(s): 87.10.+e, 02.50.Ph, 05.40.Ca, 07.05.Mh

I. INTRODUCTION

A detector with a threshold cannot detect a subthreshold signal. If noise is added to the signal, then information about the signal can be obtained from output of the detector. There is an optimal noise level, beyond which information about the signal deteriorates again. This phenomenon is known as stochastic resonance. For a recent review, see [1].

If the signal is *periodic* and observed over a relatively long time interval, then a common measure of detectability of the signal is the signal-to-noise ratio; see [2–5]. Instead of looking at the power spectrum, one may also look at the (empirical) residence-time probability distribution, or interspike interval histogram; see [6–8].

If an *aperiodic* signal is observed over a relatively long time interval, then detectability has been measured by a correlation measure; see [9–13]. The last reference also uses the interspike interval histogram.

If a signal is to be reconstructed without much delay, the identification must be based on observations over a relatively short time interval, in which the signal may be nearly constant. Then the signal-to-noise ratio and correlation measures break down. The model reduces to a parametric one, and information measures such as Fisher information can still be used; see [14–18], and in particular [19,20].

The inverse of the Fisher information is the minimal asymptotic variance of estimators. Here we show how, in simple specific settings, optimal estimators of a constant signal can be constructed. We explore the detectability of the signal in a system with one or more detectors. In the case of several detectors, we assume that the same noise is fed into each detector. This is always true for external noise but may also happen if the noise is internal, e.g., when neurons receive background noise from other neurons. Different detectors may well have different thresholds, or a detector may have more than one threshold; see [21–23]. We determine optimal configurations of detectors, varying the distances between the thresholds and the signal, as well as the noise level. We study the simplest possible model of signal plus noise. The signal s is constant over some time interval, say [0,1]. At uniformly spaced times $t_i = i/n$, independent and identically distributed ε_i are introduced. The noisy signal is $s+\varepsilon_i$, $i=1,\dots,n$.

If the signal is observed over a longer time interval, or if the noise has ''higher frequency'' in the sense that the times t_i are more densely spaced, or if there are several detectors each of which receives internal noise independently of the others, then the number n of observations is increased, and the variance of the estimator for the signal is reduced correspondingly. For large n, the signal can be estimated well for a large range of noise variances. This effect of the law of large numbers had first been observed in a different setting in [10] as *stochastic resonance without tuning*; see also [12,13,24,25].

Our approach differs from the literature on stochastic resonance in that we study detectability of the signal from a statistical point of view: we study optimal reconstruction of the signal from the data in terms of the variance of *rescaled* estimators for the signal, i.e., of $n^{1/2}(\hat{s}-s)$ rather than of $\hat{s}$. By the central limit theorem, the variance of $n^{1/2}(\hat{s}-s)$ is

*Electronic address: pgreenw@math.ubc.ca
†URL: http://www.math.ubc.ca/people/faculty/pgreenw/pgreenw.html
‡Electronic address: lward@cortex.psych.ubc.ca
§URL: http://www.psych.ubc.ca/~ward/personnel/ward/
‖Electronic address: wefelmeyer@mathematik.uni-siegen.de
¶URL: http://www.math.uni-siegen.de/statistik/wefelmeyer.html

about the same for all (sufficiently large) n, whereas the variance of $\hat{s}$ tends to zero as $1/n$. This is why we see stochastic resonance for arbitrarily large n, whereas the effect diminishes with increasing n in the previous treatments. Stochastic resonance without tuning is an aspect of the diminishing effect.

We compare four different types of observation of the noisy signal:

(i) The noisy signal $X_i = s + \varepsilon_i$ is fully observed. We need the information in the noisy signal to measure how much information is lost when the noisy signal is not completely observed.

(ii) Those times t_i are recorded at which the noisy signal $s+\varepsilon_i$ exceeds a single threshold, $a>0$. The observations are then the indicators $X_i^a = 1(s+\varepsilon_i > a)$. This scheme was proposed in [26] as a model of a neuron.

(iii) It is recorded when and which of a finite number of thresholds $0 < a_1 < \cdots < a_r$ are exceeded. Let $A = \{a_1, \ldots, a_r\}$ denote the set of thresholds. The observations can then be written as

$$X_i^A = \begin{cases} 0, & s+\varepsilon_i \leqslant a_1, \\ j, & a_j < s+\varepsilon_i \leqslant a_{j+1} \quad \text{for } j=1,\ldots,r-1, \\ r, & s+\varepsilon_i > a_r. \end{cases}$$

Such observations arise with r detectors with different thresholds, and common background or internal noise.

(iv) Whenever the single threshold a is exceeded, the noisy signal itself is observed. Then the observations are

$$X_i^{>a} = (s+\varepsilon_i) 1(s+\varepsilon_i > a).$$

Case (iv) is approximated by case (iii) for a large number of closely spaced thresholds above a.

Let us now explain in which sense the inverse of the Fisher information is the minimal asymptotic variance of estimators. We refer to [28], Sections 2.1 to 2.3, for the following results. We note first that in our setting, exactly unbiased estimators for s will not exist, and the concept of uniformly minimum variance unbiased estimators is not applicable. Instead, we use an *asymptotic* optimality concept. For the four types of observation described above, we have *local asymptotic normality* in the following sense. Let P_s denote the distribution of X_i, and let P_s^n denote the joint distribution of the observations $X_1, \ldots, X_n$. Then the log-likelihood ratio admits the stochastic expansion

$$\log \frac{dP^n_{s+n^{-1/2}t}}{dP^n_s} = n^{-1/2} t \sum_{i=1}^n \ell_s(X_i) - \frac{1}{2} I + o_p(1).$$

Here $\ell_s = \partial_{t=0} dP_{s+t}/dP_s$ is the *score function*, and $I = \mathrm{E}_s \ell_s^2$ is the *Fisher information*. Both ℓ_s and I will be calculated in the following sections. Call an estimator $\hat{s}$ for s *regular* with limit L if

$$n^{1/2}[\hat{s} - (s+n^{-1/2}t)] \Rightarrow L \quad \text{under } P^n_{s+n^{-1/2}t} \quad \text{for all } t.$$

Here convergence is meant in distribution. Regularity means that the distribution of $n^{1/2}(\hat{s}-s)$ converges *continuously* in s, in a rather weak sense, to some limit distribution. We do not assume $\hat{s}$ to be unbiased or asymptotically normal. The convolution theorem now says that $L = M + I^{-1/2} N$, with N a standard normal random variable, and M independent of N. If the rescaled estimator $n^{1/2}(\hat{s}-s)$ is asymptotically normal with variance I^{-1}, then $\hat{s}$ is *efficient*, i.e., asymptotically maximally concentrated in symmetric intervals,

$$P(-c \leqslant I^{-1/2} N \leqslant c) \geqslant P(-c \leqslant L \leqslant c) \quad \text{for all } c>0.$$

Equivalently, $\hat{s}$ has minimum asymptotic risk for all symmetric and bowl-shaped loss functions b,

$$\mathrm{E}b(I^{-1/2}N) \leqslant \mathrm{E}b(L).$$

In particular, the inverse I^{-1} of the Fisher information is the minimal asymptotic variance among regular estimators. Finally, if

$$n^{1/2}(\hat{s}-s) = n^{-1/2} \sum_{i=1}^n \ell_s(X_i)/I + o_p(1), \qquad (1.1)$$

then $\hat{s}$ is regular and efficient. In the following sections, we will construct estimators which have such a stochastic approximation.

The paper is organized as follows. In Sec. II we consider a single threshold. We assume that the noise distribution is known. If we observe indicators $X_1^a, \ldots, X_n^a$, an efficient estimator for the signal is obtained as a function of the empirical estimator for the probability that the noisy signal exceeds the threshold. The efficient estimator is exactly equal to the maximum likelihood estimator based on $X_1^a, \ldots, X_n^a$. Its asymptotic variance equals the inverse of the Fisher information.

We calculate the Fisher information for arbitrary (positive) noise distribution. As a function of the noise variance, the information has, in general, several local maxima, i.e., it exhibits *stochastic multiresonance*. With normal noise, the function is unimodal with a very pronounced resonance point.

We determine the proportion of information retained by thresholding, i.e., the ratio of the information in $X_1^a, \ldots, X_n^a$ and in $X_1, \ldots, X_n$. For normal noise, the proportion of information is a unimodal and symmetric function of the distance between signal and threshold. Hence the proportion of information retained by thresholding is maximal if the signal is at the threshold. The maximal value is 0.636 620, i.e., equal to the relative efficiency of the sample median in the normal location model.

In Sec. III we consider several thresholds. We assume again that the noise distribution is known. If we observe $X_1^A, \ldots, X_n^A$, an efficient estimator for the signal is, again, the maximum likelihood estimator. However, when there is more than one threshold, the maximum likelihood estimator cannot be represented as a function of the empirical estimators for the probability that the noisy signal exceeds one of the thresholds.

We calculate the Fisher information for two thresholds and arbitrary noise distribution. The information gain by a threshold $b>a$ for a constant signal $s<a$ is small.

When more and more thresholds are introduced above a fixed threshold a, the information increases to that of $X_1^{>a}, \ldots, X_n^{>a}$. The information in these observations still exhibits stochastic resonance.

We determine the proportion of information retained by $X_1^{>a}, \ldots, X_n^{>a}$ relative to $X_1, \ldots, X_n$. For normal noise and signal equal to threshold, the proportion of information retained is 0.818 310.

If the noise distribution is known only up to a scale parameter, the signal cannot be identified from the times at which a *single* threshold is exceeded, i.e., from $X_1^a, \ldots, X_n^a$. We show that with *two* thresholds, $A=\{a,b\}$, both the signal and the scale parameter of the noise distribution are estimated consistently from $X_1^{ab}, \ldots, X_n^{ab}$ by the maximum likelihood estimator.

We do not treat the case of several detectors with a different source of (internal) noise for each of them. If the sources generate noise independently of each other, the joint information is simply the sum of the informations in the separate detectors. The joint information is then considerably larger than with a single source of noise. An additional advantage of such a setting is that the noise variance may be different for different detectors.

Suppose that the signal is not constant and changes noticeably in the time interval in which the observations are made (which we have taken to be the unit interval). Then the noisy signal $X_i = s_{t_i} + \varepsilon_i$ follows a nonparametric regression model, with known error distribution, and the signal can be estimated, e.g., by a kernel estimator. Reconstruction of the signal from the corresponding thresholded data $X_i^a = 1(s_{t_i} + \varepsilon_i > a)$ is studied in [27]. The mean squared error shows stochastic resonance. The bias term of the kernel estimator affects the optimal noise variance and leads to results that are quantitatively, but not qualitatively, different from the results for *constant* signal obtained here. We will assume that the regularity conditions needed for our calculations are satisfied.

II. ONE THRESHOLD

Let a be a threshold and s a constant signal. We think of s as being non-negative and below the threshold, but the calculations will not depend on this assumption. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent with distribution function F. Write P_s for the distribution of $X_i = s + \varepsilon_i$, and E_s for expectations under this distribution. We assume that the only information we have about the signal is whether it exceeds the threshold a. Equivalently, we observe

$$X_i^a = 1(s + \varepsilon_i > a), \quad i = 1, \ldots, n.$$

The observations are independent Bernoulli random variables with probabilities

$$p_s = P(X_i^a = 1) = P_s(a, \infty) = 1 - F(a - s). \tag{2.1}$$

In this section, we consider a single threshold a and suppress a in the notation. Indeed, by choosing an appropriate scale, we may take a equal to 1.

A. Efficient recovery of the signal

We can write the signal as a function of p_s,

$$s = a - F^{-1}(1 - p_s).$$

The usual estimator for p_s is the empirical estimator

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i^a = \frac{\hat{n}}{n}, \tag{2.2}$$

with $\hat{n} = \#\{i : X_i^a = 1\}$. The estimator $\hat{p}$ is unbiased and consistent for p_s. The standardized error $n^{1/2}(\hat{p} - p_s)$ is asymptotically normal with variance $p_s(1-p_s)$. We obtain an estimator for the signal as a function of the empirical estimator,

$$\hat{s} = a - F^{-1}(1 - \hat{p}). \tag{2.3}$$

The estimator is not unbiased. Since $\hat{s}$ is a continuous function of $\hat{p}$, the estimator $\hat{s}$ is consistent for s. Since $\hat{s}$ is a continuously differentiable function of $\hat{p}$, it follows that $n^{1/2}(\hat{s} - s)$ is also asymptotically normal, with variance

$$v_s = \frac{p_s(1-p_s)}{f[F^{-1}(1-p_s)]^2} = \frac{F(a-s)[1-F(a-s)]}{f(a-s)^2}. \tag{2.4}$$

It is well known and easy to check that $\hat{p}$ is regular and efficient for p_s. Since continuously differentiable functions of regular and efficient estimators are again regular and efficient, the estimator $\hat{s}$ is regular and efficient for the signal, and v_s is the minimal asymptotic variance of regular estimators of s.

B. Variance bound and Fisher information

As pointed out in the Introduction, the minimal asymptotic variance v_s can be calculated as the inverse of the Fisher information for s, the variance of the score function for s. The score function is the logarithmic derivative, with respect to s, of the probabilities,

$$\ell_s(1) = \frac{\dot{p}_s}{p_s}, \quad \ell_s(0) = -\frac{\dot{p}_s}{1-p_s};$$

here and in the following, the dot denotes the derivative with respect to the parameter s. The Fisher information is therefore

$$I_s^a = \mathrm{E}_s \ell_s^2 = \frac{\dot{p}_s^2}{p_s} + \frac{\dot{p}_s^2}{1-p_s} = \frac{\dot{p}_s^2}{p_s(1-p_s)}. \tag{2.5}$$

Since $\dot{p}_s = -f(a-s)$, the Fisher information (2.5) is indeed equal to the inverse v_s^{-1} of the minimal asymptotic variance (2.4),

$$I_s^a = \frac{f(a-s)^2}{F(a-s)[1-F(a-s)]} = v_s^{-1}. \tag{2.6}$$

This Fisher information is also given in [19], relation (5.1). The Fisher information has been used as a measure of the transmitted information in other models; see [29,30,19]. It

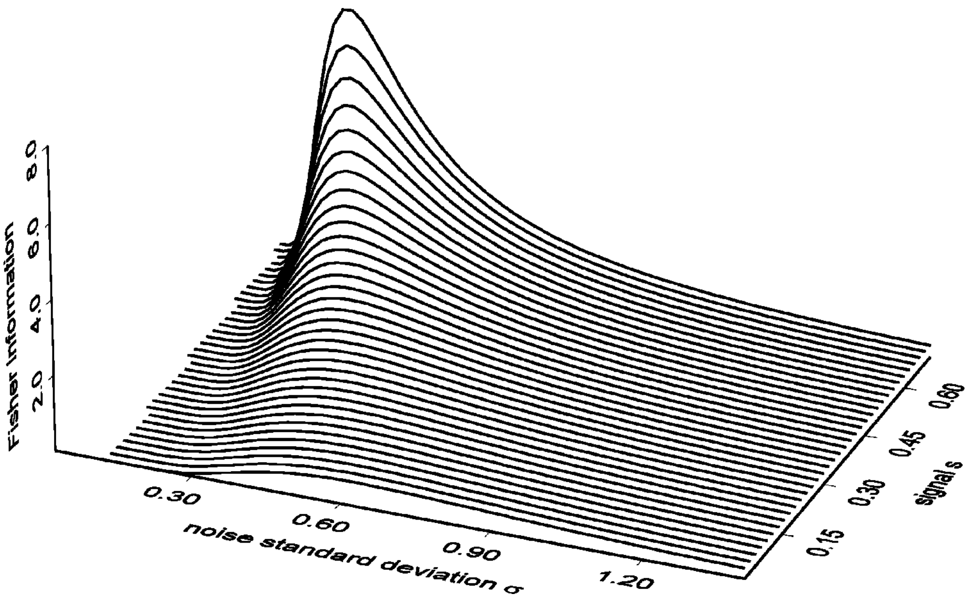

FIG. 1. Fisher information $I^1_{s\sigma}$.

will be useful to rewrite I^a_s as follows. Using integration by parts, or taking the derivative of $p_s = \int_a^\infty f(x-s)\,dx$ under the integral, we obtain

$$\dot{p}_s = -\int_a^\infty f'(x-s)\,dx = \mathrm{E}_s 1_{(a,\infty)} m_s. \tag{2.7}$$

Since $\mathrm{E}_s m_s = 0$,

$$I^a_s = \frac{(\mathrm{E}_s 1_{(a,\infty)} m_s)^2}{P_s(a,\infty)[1-P_s(a,\infty)]} = \frac{[\mathrm{cov}_s(1_{(a,\infty)}, m_s)]^2}{\mathrm{var}_s 1_{(a,\infty)}}. \tag{2.8}$$

C. Stochastic resonance

Suppose that the errors ε_i have distribution function $F_\sigma(x) = F(x/\sigma)$ with scale parameter σ, where F is standardized to variance 1. Given a threshold a and a signal s, which error variance maximizes the information in the indicators $X^a_i = 1(s+\varepsilon_i > a)$? The information (2.8) becomes

$$I^a_{s\sigma} = \frac{\left(\int_{(a-s)/\sigma}^\infty m(x) f(x)\,dx\right)^2}{\sigma^2 F\left(\frac{a-s}{\sigma}\right)\left[1-F\left(\frac{a-s}{\sigma}\right)\right]}.$$

The information $I^a_{s\sigma}$ typically tends to zero for σ tending to zero or to infinity. In general, there will not be a unique maximum. In particular, if the noise distribution has several modes, so will $I^a_{s\sigma}$ as a function of the noise variance σ. Several local maxima arise also in other threshold systems, and with other measures of signal detectability. In [31] this property is called *stochastic multiresonance*. See also [23].

If F is the standard normal distribution function Φ, we have

$$I^a_{s\sigma} = \frac{\varphi\left(\frac{a-s}{\sigma}\right)^2}{\sigma^2 \Phi\left(\frac{a-s}{\sigma}\right)\left[1-\Phi\left(\frac{a-s}{\sigma}\right)\right]}.$$

This is a unimodal function of σ with a very pronounced resonance point. The function is symmetric in $a-s$. Hence a superthreshold signal produces the same stochastic resonance property as a subthreshold signal. Figure 1 shows $I^1_{s\sigma}$ as a function of s and σ. The optimal σ decreases with the distance from the signal to the threshold; at the same time the maximal information goes to infinity. For example, if $a=1$ and the signal is low, $s=0$, then the optimal σ is 0.635 00, and the maximal value of $I_{1\sigma}$ is 0.608 42.

D. The estimator $\hat{s}$ equals the maximum likelihood estimator

The maximum likelihood estimator based on X^a_i is the solution in s of

$$0 = \sum_{i=1}^n \ell_s(X^a_i) = \hat{n}\frac{\dot{p}_s}{p_s} - (n-\hat{n})\frac{\dot{p}_s}{1-p_s}$$
$$= \frac{\dot{p}_s}{p_s(1-p_s)}[\hat{n}(1-p_s) - (n-\hat{n})p_s], \tag{2.9}$$

i.e., the solution in s of $p_s = \hat{n}/n$ or, equivalently, $1-F(a-s) = \hat{p}$. The estimator $\hat{s}$ was determined as solution of the last equation.

E. Loss of information through thresholding

How much information is lost by observing the indicators $X^a_i = 1(s+\varepsilon_i > a)$ only, rather than the noisy signal $s+\varepsilon_i$? The density of $s+\varepsilon_i$ is $f(x-s)$. Hence the score function for the noisy signal is

$$m_s(x) = -\frac{f'(x-s)}{f(x-s)} = m(x-s), \tag{2.10}$$

with $m = f'/f$, and the Fisher information is

$$I = \mathrm{E}_s m_s^2 = \mathrm{var}_s m_s = \int m(x)^2 f(x)\,dx. \tag{2.11}$$

The information I in the fully observed noisy signal X_i can be compared with the information I_s^a for X_i^a in the form (2.8). We have $I_s^a \leq I$ by the Schwarz inequality. The proportion of information retained is

$$\frac{I_s^a}{I} = [\mathrm{corr}_s(1_{(a,\infty)}, m_s)]^2.$$

We note that the information retained increases with the correlation between the indicator function $1_{(a,\infty)}$ and the score function m_s. This tells us for which noise densities f thresholding does not lose much information.

The proportion I_s^a/I is a function of the distance $a-s$ of the signal from the threshold,

$$\frac{I_s^a}{I} = R(a-s),$$

with

$$R(u) = \frac{\left(\int_u^\infty m(x) f(x) dx\right)^2}{F(u)[1-F(u)] \int m(x)^2 f(x) dx}.$$

For what error variance does X_i^a retain the most information? Suppose that ε_i has distribution function $F_\sigma(x) = F(x/\sigma)$. Then the score function for the noisy signal is $m_{s\sigma} = m((a-s)/\sigma)/\sigma$, and the proportion of information retained by X_i^a is

$$\frac{I_{s\sigma}^a}{I_\sigma} = R\left(\frac{a-s}{\sigma}\right).$$

If the signal is at the threshold, $s=a$, then $I_{s\sigma}^a/I_\sigma = R(0)$, which is independent of the noise variance σ. If the signal is below the threshold, $s<a$, we expect $I_{s\sigma}^a/I_\sigma$ to be large for large σ because the X_i^a are most informative if the noisy signal is with equal probabilities above and below the threshold. For the same reason, we expect the same behavior for $s>a$.

If F is the standard normal distribution function Φ, we have $m(x)=x$ and integration by parts gives

$$\int_a^b x\varphi(x) dx = -[\varphi(b) - \varphi(a)]. \qquad (2.12)$$

Therefore,

$$R(u) = \frac{\left(\int_u^\infty x\varphi(x) dx\right)^2}{\Phi(u)[1-\Phi(u)]} = \frac{\varphi(u)^2}{\Phi(u)[1-\Phi(u)]}.$$

The function R is unimodal and symmetric around 0. We have $R(0)=0.636\,620$. This happens to be the relative efficiency of the sample mean in the normal location model. Hence X_i^a retains about two thirds of the information if the signal is at the threshold, and considerably less if it is above or below and σ is small. Figure 2 shows $I_{s\sigma}^1/I_\sigma = R[(1-s)/\sigma]$ as a function of s and σ.

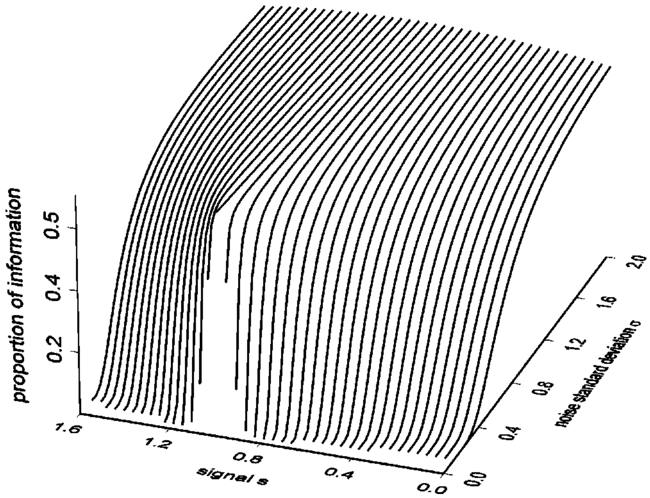

FIG. 2. Proportion of information, $I_{s\sigma}^1/I_\sigma$, retained by X_i^1.

Remark. For certain noise distributions, thresholding may not only reduce the information but even the *rate* at which the signal can be estimated. An example is a one-sided noise distribution like the exponential distribution. The corresponding location family consists of distributions that are not absolutely continuous with respect to each other. On the basis of the noisy signal $s+\varepsilon_i$, the signal s can be estimated at a rate $(n \log n)^{1/2}$. See Chapter VI in [32]. On the other hand, as long as the signal is below the threshold, $s<a$, the distributions of the noisy signal above the threshold, $(s+\varepsilon_i)1(s+\varepsilon_i>a)$, are mutually absolutely continuous for different s. In particular, the distributions of X_i^a are always mutually absolutely continuous for different s, as long as the probability of exceeding the threshold remains strictly between 0 and 1. The optimal rate for estimators of s on the basis of $(s+\varepsilon_i)1(s+\varepsilon_i>a)$ or X_i^a is therefore $n^{1/2}$.

Remark. A widely used measure for the quality of a degraded (nonconstant) signal is the signal-to-noise ratio. Unlike the Fisher information, it has the counterintuitive property that degrading the signal may improve the signal-to-noise ratio; see [33,23].

III. SEVERAL THRESHOLDS

Consider r thresholds, $0<a_1<\cdots<a_r$, a constant signal s, and a noisy signal $s+\varepsilon_i$, with $\varepsilon_1,\ldots,\varepsilon_n$ independent with distribution function F and density f. We observe which thresholds are exceeded by the noisy signal. Equivalently, we observe

$$X_i^A = \begin{cases} 0, & s+\varepsilon_i \leq a_1, \\ j, & a_j < s+\varepsilon_i \leq a_{j+1} \quad \text{for } j=1,\ldots,r-1, \\ r, & s+\varepsilon_i > a_r. \end{cases}$$

Here A stands for the set of thresholds, $\{a_1,\ldots,a_r\}$. The observations $X_1^A,\ldots,X_n^A$ are independent, with probabilities

$$p_{s0} = P(X_i^A = 0) = F(a_1 - s),$$

$$p_{sj} = P(X_i^A = j) = F(a_{j+1} - s) - F(a_j - s)$$

$$\text{for } j=1,\ldots,r-1,$$

$$p_{sr}=P(X_i^A=r)=1-F(a_r-s).$$

The observations follow a distribution on $\{0,\dots,r\}$, with a one-dimensional parameter s. For $r=1$, the family of distributions consists of *all* distributions on $\{0,1\}$, and an efficient estimator for s is obtained as a function of the empirical estimator for $p_s=P_s(a_1,\infty)$; see Sec. II. For $r>1$, we do not get such a simple efficient estimator, but the maximum likelihood estimator is, of course, still efficient.

A. The maximum likelihood estimator based on $X_1^A,\dots,X_n^A$

The score function of X_i^A is

$$\ell_s(x)=\sum_{j=0}^{r}\frac{\dot p_{sj}}{p_{sj}}1(x=j).$$

Hence the maximum likelihood estimator $\hat s$ is the solution in s of

$$0=\sum_{i=1}^{n}\ell_s(X_i^A)=\sum_{j=0}^{r}n_j\frac{\dot p_{sj}}{p_{sj}},\tag{3.1}$$

with $n_j=\#\{i:X_i^A=j\}$. The Fisher information is

$$I_s^A=\mathrm{E}_s\ell_s^2=\sum_{j=0}^{r}\frac{\dot p_{sj}^2}{p_{sj}}.\tag{3.2}$$

Here A stands for the set of thresholds $\{a_1,\dots,a_r\}$. The estimator $\hat s$ is not unbiased. A Taylor expansion of the equation around the true parameter s shows that

$$n^{1/2}(\hat s-s)=n^{-1/2}\sum_{i=1}^{n}\ell_s(X_i^A)/I_s^A+o_p(1).$$

Hence $\hat s$ is regular and efficient by the characterization (1.1).

B. Optimal choice of a second threshold

Suppose that the errors ε_i have distribution function $F_\sigma(x)=F(x/\sigma)$ with scale parameter σ, where F is standardized to variance 1. Choose two thresholds $0<a<b$ and a signal s. The Fisher information in observing which of the two thresholds is exceeded by the noisy signal is obtained from (3.2) as

$$I_{s\sigma}^{ab}=\frac{1}{\sigma^2}\left(\frac{\left(\int_{-\infty}^{(a-s)/\sigma}m(x)f(x)dx\right)^2}{F\left(\frac{a-s}{\sigma}\right)}+\frac{\left(\int_{(a-s)/\sigma}^{(b-s)/\sigma}m(x)f(x)dx\right)^2}{F\left(\frac{b-s}{\sigma}\right)-F\left(\frac{a-s}{\sigma}\right)}+\frac{\left(\int_{(b-s)/\sigma}^{\infty}m(x)f(x)dx\right)^2}{1-F\left(\frac{b-s}{\sigma}\right)}\right).\tag{3.3}$$

Of course, $I_{s\sigma}^{ab}$ reduces to $I_{s\sigma}^{a}$ for $b=a$.

Assume, for simplicity, that the distribution of ε_i is symmetric around 0. Suppose that s and a are given. By a symmetry argument, the optimal choice of threshold b is symmetrically opposite a with respect to s, namely $b=2s-a$. For $s>a$ we have $b=2s-a>a$ and

$$I_{s\sigma}^{a,2s-a}=\frac{\left(2\int_{-\infty}^{(a-s)/\sigma}m(x)f(x)dx\right)^2}{\sigma^2F\left(\frac{a-s}{\sigma}\right)}.$$

We see that $I_{s\sigma}^{a,2s-a}>I_{s\sigma}^{a}$. The information is nearly doubled by the second threshold if $F((a-s)/\sigma)$ is considerably smaller than $\frac{1}{2}$. The information gain is small if $F[(a-s)/\sigma]$ is close to $\frac{1}{2}$.

In the applications we have in mind, we will not be able to choose any of the thresholds dependent on the signal. Moreover, there will be a limit to the sensitivity of the detectors. Suppose the minimal threshold is a, so that the second threshold must be chosen above a. Suppose also that the signal is below the threshold, $s<a$. Then the information gain through the second threshold, or even through further thresholds above a, is small regardless of the configuration of signal, thresholds and noise variance. The reason is the following. For b close to a, the noisy signal $X_i=s+\varepsilon_i$ is most of the time either below both thresholds or above both thresholds, and the indicator X_i^{ab} does not say much more about the location of the signal than with a single threshold. On the other hand, for b far above a, the noisy signal rarely exceeds b, and we rarely learn more about s than with the single threshold a.

If F is the standard normal distribution function Φ, we have $m(x)=x$, and by Eq. (2.12),

$$I_{s\sigma}^{ab}=\frac{1}{\sigma^2}\left(\frac{\varphi\left(\frac{a-s}{\sigma}\right)^2}{\Phi\left(\frac{a-s}{\sigma}\right)}+\frac{\left(\varphi\left(\frac{b-s}{\sigma}\right)-\varphi\left(\frac{a-s}{\sigma}\right)\right)^2}{\Phi\left(\frac{b-s}{\sigma}\right)-\Phi\left(\frac{a-s}{\sigma}\right)}+\frac{\varphi\left(\frac{b-s}{\sigma}\right)^2}{1-\Phi\left(\frac{b-s}{\sigma}\right)}\right).$$

Suppose in particular that $a=1$. We have seen in Sec. II that X_i^1 retains the most information, as a function of s, at $s=1$: we have $I_{1\sigma}^1/I_\sigma=R(0)=0.636\,620$. The value does not depend on the noise variance σ^2, and we may take $\sigma=1$. Now we add a second threshold, $b>1$. The information retained by X_i^{1b} is

$$\frac{\varphi(0)^2}{\Phi(0)}+\frac{[\varphi(b-1)-\varphi(0)]^2}{\Phi(b-1)-\Phi(0)}+\frac{\varphi(b-1)^2}{1-\Phi(b-1)};$$

see Fig. 3. The maximum is 0.759 57, which is attained for $b=1.98$.

For thresholds in symmetric positions around the signal we obtain

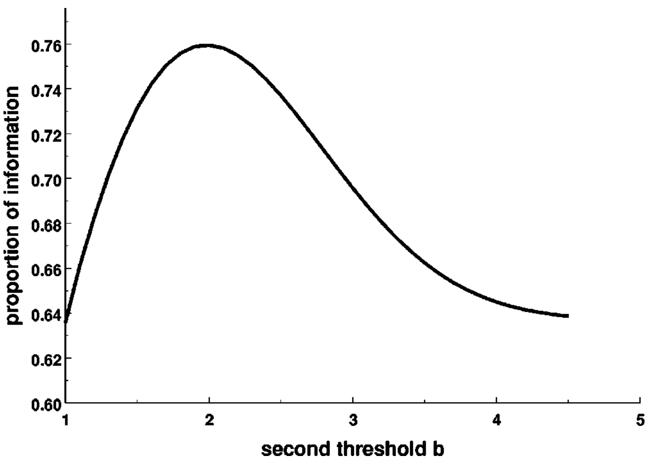

FIG. 3. Proportion of information, I_{11}^{1b}/I_1, retained by two thresholds, at 1 and b, for noise variance 1.

$$I_{s\sigma}^{a,2s-a}=\frac{2\varphi\left(\frac{a-s}{\sigma}\right)^2}{\sigma^2\Phi\left(\frac{a-s}{\sigma}\right)}.$$

C. Loss of information through thresholding

How much information is lost by observing the indicators X_i^A rather than the noisy signal $X_i=s+\varepsilon_i$? We have seen in Eq. (2.11) that the Fisher information for X_i is $I=\mathrm{var}_s m_s$. To compare with the Fisher information I_s^A for X_i^A defined in Eq. (3.2), we rewrite the latter like Eq. (2.8) in the case of one threshold. Similarly as in Eq. (2.7) the derivative of p_{sj} with respect to the parameter is

$$\dot{p}_{sj}=-\int_{a_j}^{a_{j+1}}f'(x-s)dx=\mathrm{E}_s 1_{(a_j,a_{j+1}]}m_s,$$

where m_s is the score function (2.10) for the noisy signal. The Fisher information (3.2) is then

$$I_s^A=\sum_{j=0}^{r}\frac{(\mathrm{E}_s 1_{(a_j,a_{j+1}]}m_s)^2}{P_s(a_j,a_{j+1}]}, \tag{3.4}$$

with $a_0=-\infty$ and $a_{r+1}=+\infty$. By the Schwarz inequality,

$$I_s^A\leqslant\sum_{j=0}^{r}\frac{P_s(a_j,a_{j+1}]\mathrm{E}_s 1_{(a_j,a_{j+1}]}m_s^2}{P_s(a_j,a_{j+1}]}=\mathrm{E}_s m_s^2=I.$$

D. The information contained in additional thresholds

It is clear that additional thresholds will improve the detectability of the signal. To quantify the information gain, we consider the Fisher information I_s^A in the form (3.4). Suppose that there is an additional threshold c between the thresholds a_j and a_{j+1}. The jth term in I_s^A is then replaced by

$$\frac{(\mathrm{E}_s 1_{(a_j,c]}m_s)^2}{P_s(a_j,c]}+\frac{(\mathrm{E}_s 1_{(c,a_{j+1}]}m_s)^2}{P_s(c,a_{j+1}]}.$$

This expression is, indeed, larger than the jth term in Eq. (3.4),

$$\frac{(\mathrm{E}_s 1_{(a_j,a_{j+1}]}m_s)^2}{P_s(a_j,a_{j+1}]}=\frac{(\mathrm{E}_s 1_{(a_j,c]}m_s+\mathrm{E}_s 1_{(c,a_{j+1}]}m_s)^2}{P_s(a_j,c]+P_s(c,a_{j+1}]}$$

since, in general, by the Schwarz inequality,

$$\left(\sum a_i\right)^2=\left(\sum b_i^{1/2}\frac{a_i}{b_i^{1/2}}\right)^2\leqslant\sum b_i\sum\frac{a_i^2}{b_i}$$

if $b_i\geqslant 0$ for all i.

E. Observing the noisy signal above the threshold

Consider a single threshold a. In Sec. II we have studied the situation where one observes whether the noisy signal $s+\varepsilon_i$ exceeds the threshold. Suppose now that we also observe the *size* of the noisy signal whenever it exceeds the threshold. The observations are then $X_i^{>a}=(s+\varepsilon_i)1(s+\varepsilon_i>a)$. They contain more information about the signal than the indicators $X_i^a=1(s+\varepsilon_i>a)$. The distribution of the $X_i^{>a}$ is

$$P_s(-\infty,a]\varepsilon_0(dx)+f(x-s)1(x>a)dx,$$

where ε_0 is the Dirac measure in 0. Hence, the score function of $X_i^{>a}$ is

$$m_s^{>a}(x)=\frac{\dot{q}_s}{q_s}1(x=0)+m_s(x)1(x>a),$$

with m_s the score function (2.10) of $s+\varepsilon_i$, and $q_s=P_s(-\infty,a]$ with derivative

$$\dot{q}_s=\mathrm{E}_s 1_{(-\infty,a]}m_s;$$

compare Eqs. (2.1) and (2.7). The Fisher information of $X_i^{>a}$ is therefore

$$I_s^{>a}=\frac{(\mathrm{E}_s 1_{(-\infty,a]}m_s)^2}{P_s(-\infty,a]}+\mathrm{E}_s 1_{(a,\infty)}m_s^2. \tag{3.5}$$

An efficient estimator for s is the maximum likelihood estimator. It is a solution in s of the equation

$$0=\sum_{i=1}^{n}m_s^{>a}(X_i^{>a})=\sum_{X_i^{>a}>a}m_s(X_i^{>a})+(n-\hat{n})\frac{\dot{q}_s}{q_s}, \tag{3.6}$$

with $\hat{n}=\#\{i:X_i^{>a}>a\}$ To compare $I_s^{>a}$ with the Fisher information (2.5) of the indicator X_i, we rewrite the latter as

$$I_s^a=\frac{(\mathrm{E}_s 1_{(a,\infty)}m_s)^2}{P_s(a,\infty)}+\frac{(\mathrm{E}_s 1_{(-\infty,a]}m_s)^2}{P_s(-\infty,a]}$$

and obtain $I_s^a\leqslant I_s^{>a}$ from the Schwarz inequality

$$(\mathrm{E}_s 1_{(a,\infty)}m_s)^2\leqslant P_s(a,\infty)\mathrm{E}_s 1_{(a,\infty)}m_s^2=\mathrm{E}_s m_s^2=I.$$

The proportion $I_s^{>a}/I$ of information retained by $X_i^{>a}$ is a function of $a-s$,

$$\frac{I_s^{>a}}{I}=R^{>}(a-s),$$

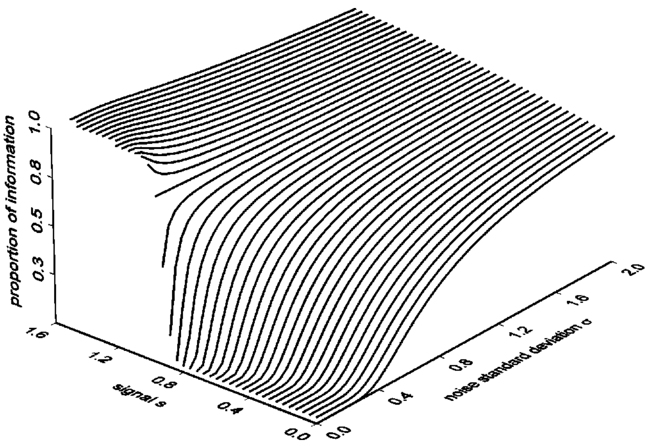

FIG. 4. Proportion of information, $I_{s\sigma}^{>1}/I_\sigma$, retained by $X_i^{>1}$.

with

$$R^>(u) = \frac{\left(\int_{-\infty}^{u} m(x) f(x)dx\right)^2}{F(u)\int m(x)^2 f(x)dx} + \frac{\int_u^\infty m(x)^2 f(x)dx}{\int m(x)^2 f(x)dx}.$$

For what error variance does $X_i^{>a}$ retain the most information? Suppose that ε_i has distribution function $F_\sigma(x) = F(x/\sigma)$. Then

$$\frac{I_{s\sigma}^{>a}}{I_\sigma} = R^>\left(\frac{a-s}{\sigma}\right).$$

If the signal is at the threshold, $s=a$, then $I_{s\sigma}^{>a}/I_\sigma = R^>(0)$, which is independent of the noise variance σ. If the signal is below the threshold, $s<a$, we expect $I_{s\sigma}^{>a}/I_\sigma$ to be large for large σ because the $X_i^{>a}$ are most informative if the noisy signal is with high probability above the threshold. For the same reason, $I_{s\sigma}^{>a}/I_\sigma$ is large for *small* σ if $s>a$.

If F is the standard normal distribution function Φ, then

$$R^>(u) = \frac{\left(\int_{-\infty}^{u} x\varphi(x)dx\right)^2}{\Phi(u)} + \int_u^\infty x^2 \varphi(x)dx$$
$$= \frac{\varphi(u)^2}{\Phi(u)} + 1 - \Phi(u) + u\varphi(u).$$

We have $R^>(0) = 0.818\,310$. Hence $X_i^{>a}$ retains about four fifths of the information if the signal is at the threshold, considerably less if it is below and σ is small, and most of the information if $s>a$ and σ is small. Figure 4 shows $I_{s\sigma}^{>1}/I_\sigma = R^>[(1-s)/\sigma]$ as a function of s and σ.

F. The limit of dense thresholds

Suppose we fix a lowest threshold a and add more and more thresholds above a such that in the limit they become dense above a. We expect that the information in observing which thresholds are exceeded by the noisy signal converges to the information in seeing the noisy signal above the threshold. To see this, choose thresholds $a_1, \ldots, a_n > a$ such that the gaps between them tend to zero and their maximum tends to infinity with n. The thresholds partition (a,∞) into $n+1$ intervals $B_1, \ldots, B_{n+1}$, generating a σ-field $\mathcal{B}_n$ which tends to the restriction of the Borel field to (a,∞). Since the score function m_s of $X_i = s+\varepsilon_i$ is in the space L_2 of P_s-square-integrable functions, the martingale convergence theorem gives

$$\mathrm{E}_s(1_{(a,\infty)} m_s | \mathcal{B}_n) \to 1_{(a,\infty)} m_s \quad \text{in } L_2. \tag{3.7}$$

We have

$$\mathrm{E}_s(1_{(a,\infty)} m_s | \mathcal{B}_n) = \sum_{j=1}^{n+1} \mathrm{E}_s 1_{B_j} \mathrm{E}_s(m_s | B_j).$$

The variance of the conditional expectation is

$$\mathrm{E}_s[\mathrm{E}_s(1_{(a,\infty)} m_s | \mathcal{B}_n)]^2 = \sum_{j=1}^{n+1} P_s B_j [\mathrm{E}_s(m_s|B_j)]^2$$
$$= \sum_{j=1}^{n+1} \frac{(\mathrm{E}_s 1_{B_j} m_s)^2}{P_s B_j}. \tag{3.8}$$

The Fisher information in observing which of the thresholds $a, a_1, \ldots, a_n$ is exceeded is obtained from Eq. (3.4) as

$$I_s^{A_n} = \frac{(\mathrm{E}_s 1_{(-\infty,a]} m_s)^2}{P_s(-\infty,a]} + \sum_{j=1}^{n+1} \frac{(\mathrm{E}_s 1_{B_j} m_s)^2}{P_s B_j}.$$

Here A_n stands for the set of thresholds $\{a, a_1, \ldots, a_n\}$. The martingale convergence theorem (3.7) and relation (3.8) then imply

$$\sum_{j=1}^{n+1} \frac{(\mathrm{E}_s 1_{B_j} m_s)^2}{P_s B_j} \to \mathrm{E}_s 1_{(a,\infty)} m_s^2.$$

Hence $I_s^{A_n}$ converges to the Fisher information (3.5) of $X_i^{>a} = (s+\varepsilon_i)1(s+\varepsilon_i>a)$.

G. Identifying the noise variance

Suppose we have one threshold a and observe whether the noisy signal exceeds it, $X_i^a = 1(s+\varepsilon_i>a)$. Suppose that the noise distribution function is $F_\sigma(x) = F(a-s/\sigma)$. Then the observations $X_1, \ldots, X_n$ are independent Bernoulli random variables with

$$p_{s\sigma} = P(X_i^a = 1) = 1 - F\left(\frac{a-s}{\sigma}\right).$$

We see that if signal s and noise variance σ^2 are unknown, they are not identifiable.

The situation is different, in general, if there is a second threshold, say $b>a$. Then the observations are

$$X_i^{ab} = \begin{cases} 0, & s+\varepsilon_i \leq a \\ 1, & a < s+\varepsilon_i \leq b \\ 2, & s+\varepsilon_i > b. \end{cases}$$

The observations $X_1^{ab}, \ldots, X_n^{ab}$ are independent, with probabilities

$$p_{s0}=P(X_i^{ab}=0)=F\left(\frac{a-s}{\sigma}\right),$$

$$p_{s1}=P(X_i^{ab}=1)=F\left(\frac{b-s}{\sigma}\right)-F\left(\frac{a-s}{\sigma}\right),$$

$$p_{s2}=P(X_i^{ab}=2)=1-F\left(\frac{b-s}{\sigma}\right).$$

The score function of X_i^{ab} with respect to s is

$$\ell_{s\sigma}^{s}(x)=\sum_{j=0}^{2}\frac{p_{s\sigma}^{s}}{p_{s\sigma}}1(x=j).$$

Similarly, the score function of X_i^{ab} with respect to σ is

$$\ell_{s\sigma}^{\sigma}(x)=\sum_{j=0}^{2}\frac{p_{s\sigma}^{\sigma}}{p_{s\sigma}}1(x=j).$$

Here the superscripts s and σ denote partial derivatives with respect to s and σ. We obtain

$$\ell_{s\sigma}^{s}(x)=-\frac{f_\sigma(a-s)}{F_\sigma(a-s)}1(x=0)-\frac{f_\sigma(b-s)-f_\sigma(a-s)}{F_\sigma(b-s)-F_\sigma(a-s)}\times 1(x=1)+\frac{f_\sigma(b-s)}{1-F_\sigma(b-s)}1(x=2),$$

$$\ell_{s\sigma}^{\sigma}(x)=-(a-s)\frac{f_\sigma(a-s)}{F_\sigma(a-s)}1(x=0)-\frac{(b-s)f_\sigma(b-s)-(a-s)f_\sigma(a-s)}{F_\sigma(b-s)-F_\sigma(a-s)}\times 1(x=1)+(b-s)\frac{f_\sigma(b-s)}{1-F_\sigma(b-s)}1(x=2).$$

We have

$$\ell_{s\sigma}^{\sigma}(0)=(a-s)\ell_{s\sigma}^{s}(0),\quad \ell_{s\sigma}^{\sigma}(2)=(b-s)\ell_{s\sigma}^{s}(2).$$

Hence, $\ell_{s\sigma}^{s}$ and $\ell_{s\sigma}^{\sigma}$ are linearly independent, and $|\mathrm{corr}_s(\ell_{s\sigma}^{s},\ell_{s\sigma}^{\sigma})|<1$, for $a<b$. This means that the Fisher information matrix for s and σ is nonsingular, and s and σ can be estimated jointly, and efficiently, by the maximum likelihood estimator.

ACKNOWLEDGMENTS

The work was supported by the NSERC, Canada, and the Crisis Points group of the Peter Wall Institute for Advanced Studies at UBC. The research was done while the third author was visiting the Department of Mathematics at UBC.

[1] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).
[2] K. Wiesenfeld and F. Moss, Nature (London) **373**, 33 (1995).
[3] Z. Gingl, L. B. Kiss, and F. Moss, Europhys. Lett. **29**, 191 (1995).
[4] P. Jung, Phys. Lett. A **207**, 93 (1995).
[5] K. Loerincz, Z. Gingl, and L. B. Kiss, Phys. Lett. A **224**, 63 (1996).
[6] T. Zhou, F. Moss, and P. Jung, Phys. Rev. A **42**, 3161 (1990).
[7] A. Longtin, A. Bulsara, and F. Moss, Phys. Rev. Lett. **67**, 656 (1991).
[8] A. R. Bulsara, S. B. Lowen, and C. D. Rees, Phys. Rev. E **49**, 4989 (1994).
[9] J. J. Collins, C. C. Chow, and T. T. Imhoff, Phys. Rev. E **52**, R3321 (1995).
[10] J. J. Collins, C. C. Chow, and T. T. Imhoff, Nature (London) **376**, 236 (1995).
[11] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, Phys. Rev. E **54**, 5575 (1996).
[12] D. R. Chialvo, A. Longtin, and J. Müller-Gerking, Phys. Rev. E **55**, 1798 (1997).
[13] D. R. Chialvo, A. Longtin, and J. Müller-Gerking (unpublished).
[14] J. Levin and J. P. Miller, Nature (London) **380**, 165 (1996).
[15] A. R. Bulsara and A. Zador, Phys. Rev. E **54**, R2185 (1996).
[16] J. J. Collins, T. T. Imhoff, and P. Grigg, J. Neurophysiol. **76**, 642 (1996).
[17] C. Heneghan, C. C. Chow, J. J. Collins, T. T. Imhoff, S. B. Lowen, and M. C. Teich, Phys. Rev. E **54**, R2228 (1996).
[18] A. Neiman, A. B. Shulgin, V. S. Anishchenko, W. Ebeling, L. Schimansky-Geier, and J. Freund, Phys. Rev. Lett. **76**, 4299 (1996).
[19] M. Stemmler, Network Comput. Neural Syst. **7**, 687 (1996).
[20] F. Chapeau-Blondeau, Phys. Lett. A **232**, 41 (1997).
[21] L. Gammaitoni, Phys. Rev. E **52**, 4691 (1995).
[22] L. Gammaitoni, Phys. Lett. A **208**, 315 (1995).
[23] F. Chapeau-Blondeau and X. Godivier, Phys. Rev. E **55**, 2016 (1997).
[24] A. Neiman, L. Schimansky-Geier, and F. Moss, Phys. Rev. E **56**, R9 (1997).
[25] P. C. Gailey, A. Neiman, J. J. Collins, and F. Moss, Phys. Rev. Lett. **79**, 4701 (1997).
[26] W. S. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115 (1943).
[27] U. U. Müller,Can. F. Stat. (to be published).
[28] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models* (Johns Hopkins University Press, Baltimore, 1993).
[29] M. A. Paradiso, Biol. Cybern. **58**, 35 (1988).
[30] H. S. Seung and H. Sompolinsky, Proc. Natl. Acad. Sci. USA **90**, 10 749 (1993).
[31] J. M. G. Vilar and J. M. Rubí, Phys. Rev. Lett. **78**, 2882 (1997).
[32] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation. Asymptotic Theory*, Applications of Mathematics Vol. 16 (Springer, New York, 1981).
[33] F. Moss, D. Pierson, and D. O'Gorman, Int. J. Bifurcation Chaos Appl. Sci. Eng. **4**, 1383 (1994).