Acta Universitatis Matthiae Belii ser. Mathematics, 16 (2009), 53–63.
Received: 15 July 2009, Accepted: 2 February 2010.

FISHER INFORMATION AS THE MEASURE OF SIGNAL OPTIMALITY IN OLFACTORY NEURONAL MODELS

ONDŘEJ POKORA

Abstract. Some new approximations of Fisher information are introduced and their properties are derived. These approximations are computed and applied to locate the optimal odorant concentration in two simple theoretical models for coding of odor intensity in olfactory sensory neurons. The results are compared with the deterministic criterion and with results based on Fisher information measure.

1. Introduction

Characterization of the input-output properties of sensory neurons and their models is commonly done by using the so called input-output response functions, $R(s)$, in which the response is plotted against the input s. The output is usually the spiking frequency, or rate of firing, but it can be also concentration of activated receptors as presented e.g. in [7, 8, 9] and also in this contribution. The response curves are usually monotonously increasing functions (most often of sigmoid shape) assigning a unique response to an input signal (see Fig. 1 for illustration).

The intuitive concept of "just noticeable difference", which has been deeply studied in psychophysics, is also implicitly involved in understanding of signal optimality in neurons. Having the transfer function $R(s)$ and minimum detectable increment ϵ of the response, we can calculate Δ_s which is the just noticeable difference in the signal. If the response curve is nonlinear (for example sigmoidal as in Fig. 1) we can see that Δ_s varies along D and the smallest values of the just noticeable difference in the signal are achieved where the response curve is steepest. The stimulus intensity for which the signal is optimal, that is the best detectable, is where the slope of the transfer function is highest.

2000 *Mathematics Subject Classification.* Primary 94A17; Secondary 62P10.
Key words and phrases. Approximations of Fisher information, olfactory neuron, optimal signal.

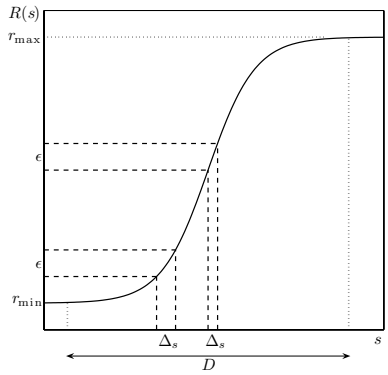

FIGURE 1. A schematic example of transfer function $R(s)$ (solid curve). The dynamic range D, threshold response $r_{\min}$, maximal discharge $r_{\max}$ and just noticeable difference Δ_s in the signal corresponding to the just noticeable difference ϵ in the response are given.

However, in practice, an identical signal does not always yield the same response. The presence of noise complicates the concept of signal optimality based on the just noticeable difference. Not only a fixed response is assigned to every level of the stimulus (as in the classical frequency coding schema), but also a probability distribution of the responses.

In [9], Fisher information was used as a general measure of signal optimality in the case of "noisy response" and applied on theoretical models. The aim of this contribution is to extend a known approximation of Fisher information to a sequence of approximations, apply the same approach on introduced approximations of Fisher information and compare these new optimality measures with known results.

2. Fisher information and its approximation

In this section, some necessary facts about Fisher information measure are recalled. Then, some approximations of Fisher information are introduced and their properties are derived.

2.1. **Fisher information and its properties.** Let us assume, it is dealt with real random variables upon the same probability space $(\Omega, \mathcal{A}, \mathsf{P})$, which have finite second moments and probability density function with respect to some countably additive measure μ. The probability density function $f(x;\theta)$ is assumed to be dependent on a scalar parameter $\theta \in \Theta$.

Regular class. Class of probability density functions $\{f(x;\theta);\theta\in\Theta\}$ is called regular if following conditions hold:

(R1) parametric space Θ is nonempty open set,
(R2) support $M=\{x\in(-\infty,\infty);f(x;\theta)>0\}$ does not depend on θ,
(R3) for almost all $x\in M$ (with respect to μ), finite derivative $\frac{\partial f(x;\theta)}{\partial\theta}$ exists,
(R4) for all $\theta\in\Theta:\int_M\frac{\partial f(x;\theta)}{\partial\theta}\,\mathrm{d}\mu(x)=0$,
(R5) $J^X(\theta)=\int_M\left(\frac{\partial\ln f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)\mathrm{d}\mu(x)$ holds $0<J^X(\theta)<\infty$.

Regular estimator. Estimator $\hat\theta=H(X)$ of parameter θ in random variable X with p.d.f. $f(x;\theta)$ is called regular if following conditions hold:

(R6) the class $\{f(x;\theta);\theta\in\Theta\}$ is regular,
(R7) $\hat\theta$ is unbiased,
(R8) for all $\theta\in\Theta:\int_M H(x)\frac{\partial f(x;\theta)}{\partial\theta}\,\mathrm{d}\mu(x)=\frac{\partial}{\partial\theta}\int_M H(x)f(x;\theta)\mathrm{d}\mu(x)$.

Fisher information. The value

$$J^X(\theta)=\mathrm{E}\left(\left(\frac{\partial\ln f(X;\theta)}{\partial\theta}\right)^2\right)=\int_M\left(\frac{\partial\ln f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)\mathrm{d}\mu(x) \tag{1}$$

is called Fisher information about parameter θ in random variable X. Fisher information is not measure of information in the sense of the theory of information (e.g. like entropy). However, it gives how much "information" is transferred into the distribution of X when the parameter θ changes. In other words, it indicates how precisely the change in parameter can be identified (estimated) from the knowledge of the changed distribution. This point of view is induced by following well-known result published in [2].

Cramér-Rao inequality. Let $\hat\theta=H(X)$ be regular estimator of parameter θ with finite second moment. Then, for all $\theta\in\Theta$ following inequality is fulfilled,

$$\frac{1}{J^X(\theta)}\le\mathrm{Var}\left(\hat\theta\right)\quad. \tag{2}$$

Hence, it gives the lower bound for variance of any regular estimator of the parameter. The proof is based on Cauchy-Schwarz inequality for variables $\hat\theta-\mathrm{E}\left(\hat\theta\right)$ and $\frac{\partial\ln f(X;\theta)}{\partial\theta}$.

Assuming we know the best estimator $\hat\theta=H(X)$ of θ in the sense of minimal variance, Cramér-Rao inequality (2) can be seen as relation which gives the quality of estimator $\hat\theta$ as a function of the true value of parameter θ. The idea of analyzing Fisher information $J^X(\theta)$ as a function of θ to find the "optimal" value of θ, i.e. the value for which the best estimator $\hat\theta$ has the lowest variance, was one of the reasons, for which the Fisher information has become a common tool in computational neuroscience (see e.g. [6, 10, 3]).

2.2. **Approximation of Fisher information.** In general, it is difficult task to compute the Fisher information analytically. Usually the integral has to be computed numerically. Moreover, having only measured data without the knowledge of their distribution (which is a typical situation), it is impossible to compute the Fisher information without estimation of the probability density function (e.g. using kernel estimators). These reasons lead to search for some approximation of the Fisher information. Following definition introduce a sequence of such approximations. It is an extension of definition of approximation $J_2^X(\theta)$, which was already used by several authors, see e.g. [6].

Approximations of Fisher information. For $k = 2, 3, \ldots$, let us define sequence of approximations

$$J_k^X(\theta) = \frac{1}{\mathrm{Var}\left(X^{k-1}\right)} \left(\frac{\partial \mathrm{E}\left(X^{k-1}\right)}{\partial \theta} \right)^2 \tag{3}$$

and sequence of conditions

(R9) $\int_M \frac{\partial}{\partial \theta} \left(x^{k-1} f(x;\theta) \right) \mathrm{d}\mu(x) = \frac{\partial}{\partial \theta} \int_M x^{k-1} f(x;\theta) \mathrm{d}\mu(x)$.

Following theorems say that, in general, approximations $J_k^X(\theta)$ are lower bounds for Fisher information $J^X(\theta)$ and that for some special distributions of X there is equality achieved.

Theorem 1. If the class $\{f(x;\theta); \theta \in \Theta\}$ satisfies regularity conditions (R1)–(R5), then, for those $k = 2, 3, \ldots$ for which condition (R9) is satisfied for all $\theta \in \Theta$, there is inequality

$$J_k^X(\theta) \leq J^X(\theta) \qquad \text{for all} \quad \theta \in \Theta \ . \tag{4}$$

The principal idea of proof of this inequality uses Cauchy-Schwarz inequality for variables $X^{k-1} - \mathrm{E}\left(X^{k-1}\right)$ and $\frac{\partial \ln f(X;\theta)}{\partial \theta}$.

Theorem 2. Under the same conditions as in Theorem 1 the equality

$$J_k^X(\theta) = J^X(\theta) \qquad \text{for all} \quad \theta \in \Theta \tag{5}$$

is fulfilled if and only if the probability density function $f(x;\theta)$ of random variable X has the form

$$f(x;\theta) = \exp\left\{ x^{k-1} c(\theta) - b(\theta) + a(x) \right\} \tag{6}$$

for some functions $a(x), b(\theta), c(\theta)$. The main way of proof follows the idea of sequential equivalent conditions published in [6] for the case $k = 2$. This result might be useful in further work for expressing the accuracy of the approximation in terms of a distance between the real distribution and form (6). The previous condition leads to introducing of following definition.

Exponential class with natural power. Random variable X has a distribution belonging to exponential class with respect to parameter θ and with power

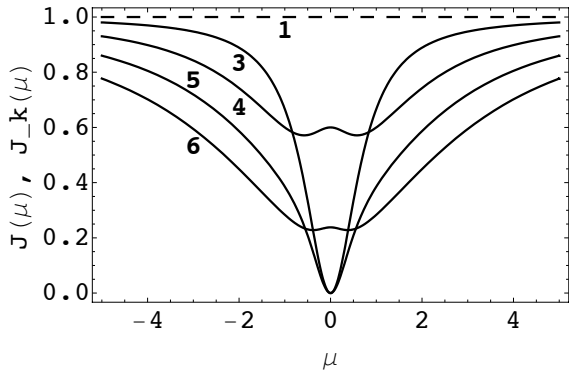

FIGURE 2. Fisher information $J^X(\mu)$ (dashed curve 1) and its approximations $J_k^X(\mu)$ for $k = 3, 4, 5, 6$ (curves 3–6) computed from random variable $X \sim \mathsf{N}(\mu, \sigma^2 = 1)$.

k, $k = 1, 2, \ldots$, if probability density function of X takes the form

$$f(x;\theta) = \exp\left\{x^k c(\theta) - b(\theta) + a(x)\right\} \tag{7}$$

for some functions $a(x), b(\theta), c(\theta)$.

Example. Let us suppose that random variable X has Gaussian distribution $X \sim \mathsf{N}(\mu, \sigma^2)$ with known variance σ^2. Fisher information about the unknown mean value μ, $J^X(\mu) = \frac{1}{\sigma^2}$ does not depend on the true value μ; it means, all values of mean are estimable with equal accuracy, which only depends on the variance. Both Fisher information $J^X(\mu)$ and its approximations $J_k^X(\mu)$ for $k = 3, 4, 5, 6$ are depicted in Fig. 2. Approximation $J_2^X(\mu) = J^X(\mu) = \frac{1}{\sigma^2}$ is accurate. This corresponds with Theorem 2, because Gaussian distribution belongs to the exponential class with respect to parameter μ with power $k = 1$, e.g. for $\sigma^2 = 1$ probability density function is $f(x;\mu) = \exp\left\{x^1\mu - \frac{\mu^2}{2} - \frac{x^2}{2} - \frac{\ln 2\pi}{2}\right\}$.

3. Theoretical models of olfactory neurons

Signal processing in olfactory systems is initialized by binding of odorant molecules to receptor molecules embedded in the membranes of sensory neurons. Binding of odorants and receptor activation trigger a sequence of biochemical events that result in the opening of ionic channels, the generation of receptor potential which triggers a train of action potentials. Studied models of the binding and activation of receptor sites are based on models proposed by [4, 7, 8].

3.1. **Methods.** Searching for "optimal odorant concentration", we aim to investigate how precisely the odorant concentration, s, can be determined from a knowledge of the response, concentration of activated receptors, $C(s)$, and which concentration levels are optimal, that means can be well determined from the knowledge of a random sample of $C(s)$. In other words, we consider an experiment in which a fixed concentration is applied and steady-state responses of the system are observed. These are independent (it is the random sample) realizations of random variable $C(s)$ from which we wish to determine s.

Deterministic approach to determine the optimal concentration is based on shape of the input-output function, $R(s)$, and it uses the optimality criterion

$$J_1(s) = \frac{\partial \mathrm{E}\left(C(s)\right)}{\partial s} \quad . \tag{8}$$

From the stochastic point of view, the determination of the concentration, s, from sampling responses of $C(s)$ corresponds to its estimation, $\hat{s}$, in chosen family of probability density functions. For reasons explained in Section 2, as measures of optimality, the Fisher information (1), $J^X(s)$, is commonly used. Here, we focus on approximations (3), $J_k^X(s)$, and on their application as another optimality criteria in search of optimal odorant concentration in investigated theoretical models.

3.2. **Models and results.** In general, the models consider interaction between odorant molecules and receptors on the surface of olfactory receptor neurons. We assume that there is only one odorant substance, that each receptor molecule possesses only one binding site and that the total number of the receptors on the surface of the membrane is fixed and equal to N. Let A denote the odorant molecules in perireceptor space, with concentration $A = \exp(s)$ which is assumed to be fixed until the olfactory system achieves the steady state. We distinguish three states in which the receptors can appear: unbound (free) state, R, bound inactive state (inactive complex of the odorant molecule and the receptor), C^*, and bound activated state (activated complex of the odorant molecule and the receptor) C. Only activated receptors trigger the response.

Optimality criteria $J_k(s)$ given by (3) are applied on two simple theoretical models of olfactory neurons. The stochastic description of both the models and results of application of $J(s)$, $J_1(s)$ and $J_2(s)$ criteria are already known. Detailed description, derivation of the steady-state (stationary) distribution of number of activated receptors $C(s)$ and results of the criteria can be found in [9]. The results of new criteria are also compared with these previous results.

Basic model.

In the simplest model each occupied receptor becomes activated instantaneously with its occupation. It is assumed that each receptor is occupied and

released independently of others in accordance with stochastic reaction schema

$$\mathrm{A} + \mathrm{R} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{C} \ , \tag{9}$$

where k_1 and k_{-1} are fixed reaction rates coefficients of association and dissociation of the odorant molecules. The ratio $K_1 = k_{-1}/k_1$ is commonly called the dissociation constant. The model can be fully described by birth and death process (see [9] for details). Using this stationary distribution to derive the mean and variance of the count of activated receptors in steady state, $C(s)$, we obtain

$$\mathrm{E}\left(C(s)\right) = \frac{N}{1 + K_1 \mathrm{e}^{-s}} \ , \tag{10}$$

$$\mathrm{Var}\left(C(s)\right) = \frac{N K_1 \mathrm{e}^{-s}}{\left(1 + K_1 \mathrm{e}^{-s}\right)^2} \ , \tag{11}$$

$$\mathrm{E}\left(C^2(s)\right) = \frac{N^2 + N K_1 \mathrm{e}^{-s}}{\left(1 + K_1 \mathrm{e}^{-s}\right)^2} \ , \tag{12}$$

$$\mathrm{E}\left(C^4(s)\right) = \frac{N\left(K_1 \mathrm{e}^{-s}\right)^{N-1}}{\left(1 + K_1 \mathrm{e}^{-s}\right)^N} \ {}_4\mathrm{F}_3\left(2, 2, 2, 1{-}N; 1, 1, 1; -\frac{\mathrm{e}^s}{K_1}\right) \ , \tag{13}$$

where ${}_p\mathrm{F}_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x)$ stands for generalized hypergeometric function (see [1]). We have ${}_4\mathrm{F}_3(2, 2, 2, 1{-}N; 1, 1, 1; x) = 1 + \sum_{k=1}^{\infty} \frac{x^k}{k!}(k+1)^3 \prod_{i=1}^{k}(i - N)$.

Assuming the normal distribution of $C(s)$, criteria of optimality $J_1(s)$, $J(s)$ and $J_2(s)$ are directly derived (see[9]),

$$J_2(s) = J_1(s) = \frac{N K_1 \mathrm{e}^{-s}}{\left(1 + K_1 \mathrm{e}^{-s}\right)^2} \ , \tag{14}$$

$$J(s) = \frac{1}{2} + \frac{(N-2) K_1 \mathrm{e}^{-s}}{\left(1 + K_1 \mathrm{e}^{-s}\right)^2} = \frac{1}{2} + \frac{N-2}{N} J_2(s) \ . \tag{15}$$

The new approximations $J_k(s)$ can be computed using relation (3) via higher moments of $C(s)$ and can be expressed in terms of hypergeometric functions.

The shapes of optimality criteria are plotted in Fig. 3. The criteria $J_1(s)$ and $J_2(s)$ are equal and have unimodal shape. For $N > 2$ (which is natural in reality), the Fisher information $J(s)$ is also unimodal and it is very close to $J_1(s)$. As stated in [9], all these criteria attain maximum value $N/4$ for the odorant log--concentration

$$s_0 = \ln K_1 \ . \tag{16}$$

The approximations $J_3(s)$ and $J_4(s)$ has also unimodal shape, but their maxima are slightly shifted from s_0 to higher odorant concentrations. This shift, however, is small and depends only on N (the shift rises with increasing N). For extremely

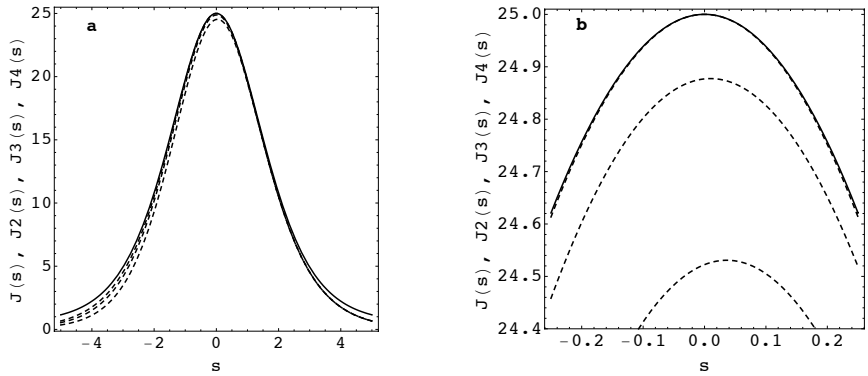

FIGURE 3. (a) Optimality criteria in the basic model: Fisher information $J(s)$, and the criteria $J_1(s) = J_2(s)$, $J_3(s)$ and $J_4(s)$ (dashed curves, up to bottom). Parameters are $K_1 = 1$ and $N = 100$. Criteria J, J_1, J_2 attain maximum value $N/4 = 25$ for the odorant log-concentration $s_0 = \ln K_1 = 0$. (b) Detail of (a); note, that the maxima of J_3 and $J_4(s)$ criteria are slightly biased.

low as well as high odorant concentrations all the criteria decrease. Both the deterministic and Fisher information criteria give the same result and locate the optimal concentration of odorant in the region around the concentration s_0 (see Fig. 3). The criteria based on approximations are slightly biased in positive direction.

Model with simple activation.

Considering the model where not every bound receptor is activated immediately, the receptors really appear in three different states: unbound, R, occupied but not activated, C*, and occupied activated, C. Model described by [5] supposes that each occupied receptor can either become activated, C, with probability $p \in (0, 1)$, or stay inactive, C*, with probability $1 - p$, independently of its past behavior and of the behavior of other receptors. Such an interaction corresponds to the following reaction schema,

$$\mathrm{C}^* \underset{k_{1\mathrm{N}}}{\overset{k_{-1}}{\rightleftarrows}} \mathrm{A} + \mathrm{R} \underset{k_{-1}}{\overset{k_{1\mathrm{A}}}{\rightleftarrows}} \mathrm{C} \ , \tag{17}$$

where $k_{1\mathrm{A}} = pk_1$ and $k_{1\mathrm{N}} = (1 - p)k_1$ are association rates for the activated and inactive state and k_1, k_{-1} have the same meaning as in basic model (9).

It can be proved (see [9]) that the steady-state number of activated receptors has binomial distribution $C(s) \sim \mathsf{Bi}(N, q(s))$ with $q(s) = p/(1 + K_1 \mathrm{e}^{-s})$ and its

moments are equal to

$$\mathrm{E}\left(C(s)\right) = \frac{Np}{1 + K_1 \mathrm{e}^{-s}} \quad , \tag{18}$$

$$\mathrm{Var}\left(C(s)\right) = \frac{NpK_1\mathrm{e}^{-s}}{\left(1 + K_1\mathrm{e}^{-s}\right)^2} + \frac{Np(1-p)}{\left(1 + K_1\mathrm{e}^{-s}\right)^2} \quad , \tag{19}$$

$$\mathrm{E}\left(C^2(s)\right) = \frac{Np\left(1 + p(N-1) + K_1\mathrm{e}^{-s}\right)}{\left(1 + K_1\mathrm{e}^{-s}\right)^2} \quad , \tag{20}$$

$$\mathrm{E}\left(C^4(s)\right) = \frac{Np\left(1 - \frac{p\,\mathrm{e}^s}{K_1 + \mathrm{e}^s}\right)^N {}_4\mathrm{F}_3\left(2, 2, 2, 1-N; 1, 1, 1; \frac{p\,\mathrm{e}^s}{\mathrm{e}^s(p-1) - K_1}\right)}{K_1\mathrm{e}^{-s} - (p-1)} \quad . \tag{21}$$

Criteria J_1 and J_2 are derived analytically,

$$J_1(s) = \frac{pNK_1\mathrm{e}^{-s}}{\left(1 + K_1\mathrm{e}^{s}\right)^2} \quad , \tag{22}$$

$$J_2(s) = \frac{pNK_1^2\mathrm{e}^{-s}}{\left(1 + K_1\mathrm{e}^{-s}\right)^2\left(K_1 + (1-p)\mathrm{e}^{s}\right)} \quad , \tag{23}$$

Fisher information $J(s)$ and its approximations $J_k(s)$ for Gaussian distributed $C(s)$ are evaluated numerically.

As well as in basic model (9), maximum value of the criterion $J_1(s)$ is located at odorant log-concentration $s_1 = \ln K_1$, independently on the value of activation probability p. According to [9], criterion $J_2(s)$ achieves its maximum for the odorant log-concentration

$$s_2 = \ln K_1 - \ln \frac{4(1-p)}{\sqrt{9 - 8p} - 1} \quad . \tag{24}$$

For lower activation probabilities p the location of maximum of $J_2(s)$ is shifted to lower concentrations of odorant.

As shown in Fig. 4, the shape and location of maxima of Fisher information criterion $J(s)$ and approximations $J_2(s)$, $J_3(s)$ and $J_3(s)$ are similar, but different from the maximum of deterministic criterion $J_1(s)$. The deterministic and statistical approaches can give different results, the optimum from statistical point of view is located at lower concentrations of odorant than that obtained with the approach based on the slope of the input-output function. In comparison with Fisher information, the maxima of approximations $J_k(s)$ are slightly biased in positive sense, i.e. locate the optimal signal in higher odorant concentrations than criterion $J(s)$ does. Nevertheless, these maxima are less than the deterministic optimum (as already known for maximum of $J_2(s)$).

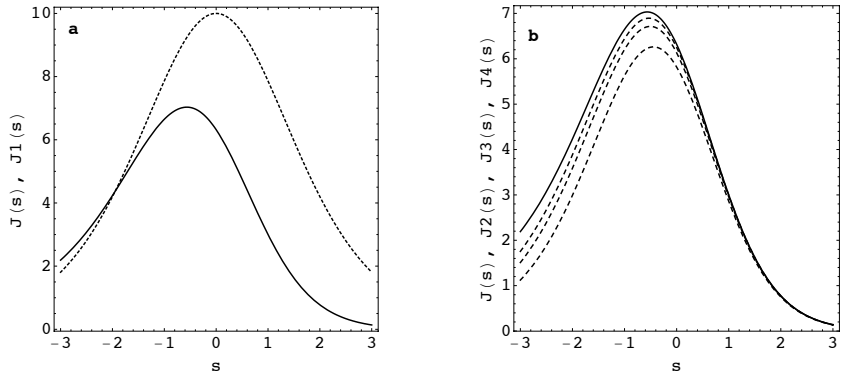

FIGURE 4. Optimality criteria in the model with simple activation: (a) first derivative of the input-output function $J_1(s)$ (dotted curve) and Fisher information $J(s)$ (solid), (b) Fisher information $J(s)$ (solid curve) and its approximations $J_k(s)$ (dashed curves, up to bottom for $k = 2, 3, 4$) as functions of the odorant log-concentration, s, in the perireceptor space. Maximum of $J_1(s)$ is located at $s_1 = 0$. Maximum of $J(s)$ is located at $s \approx -0.565$. Maxima of approximations $J_k(s)$ are shifted to higher concentrations. Parameters are $K_1 = 1$, $N = 100$ and $p = 0.4$.

4. Conclusions

Two theoretical models of theoretical models of olfactory sensory neurons were searched for the optimal signal, s, as defined by the application of approximations $J_k(s)$ of Fisher information $J(s)$. In both models, the approximations $J_k(s)$ have similar shape as Fisher information $J(s)$. In comparison with optimal concentration defined by Fisher information, the maxima of the approximations are biased in positive sense, it means the corresponding odorant concentration determined as optimal are located in higher values. In the model with simple activation, the optimal odorant concentration defined in the sense of approximations $J_k(s)$ is different (less) than the deterministically determined value. Interesting is, that in investigated models the approximations seem to be ordered, even thought there is no clear ordering of functions $J_k(s)$ in general.

Acknowledgement

This work was supported by Grant 201/05/H007 MŠMT ČR and by Center for Theoretical and Applied Statistics LC06024.

References

[1] Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions*, Dover Publications, New York, 1972.

[2] Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[3] Johnson, D. H. and Ray, W., Optimal stimulus coding by neural populations using rate codes, *Journal of Computational Neuroscience* **16** (2004), 129–138.

[4] Kaissling, K. E., Flux detectors vs. concentration detectors: two types of chemoreceptors, *Chemical Senses* **23** (1998), 99–111.

[5] Lánský, P. and Rospars J.-P., Coding of odor intensity, *BioSystems* **31** (1993), 15–38.

[6] Lánský, P. and Greenwood, P. E., Optimal signal in sensory neurons under extended rate coding concept, *BioSystems* **89** (2007), 10–15.

[7] Mankin, R. W. and Mayer, M. S., A phenomenological model of the perceived intensity of single odorants, *Journal of Theoretical Biology* **100** (1983), 123–138.

[8] Maurin, F., A mathematical model of chemoreception for odours and taste, *Journal of Theoretical Biology* **215** (2002), 297–303.

[9] Pokora, O. and Lánský, P., Statistical approach in search for optimal signal in simple olfactory neuronal models, *Mathematical Biosciences* **214** (2008), 100–108.

[10] Sanchez-Montanes, M. A. and Pearce, T. C., Fisher information and optimal odor sensors, *Neurocomputing* **38** (2001), 335–341.

Department of Mathematics and Statistics, Faculty of Science, Masaryk University; Kotlářská 2; 611 37 Brno; Czech Republic

E-mail address: pokora@math.muni.cz