Humboldt-Universität zu Berlin
Institute for Theoretical Biology

Exercise Set 4
Summer 2006, May 15

Computational Neuroscience IV: Analysis of Neural Systems

Dr. R. Kempter, Prof. Dr. A.V.M. Herz

Maximum Likelihood Estimation

The Maximum Likelihood (ML) estimate $\hat{\boldsymbol{\theta}}_{ML}$ of the parameter vector $\boldsymbol{\theta}$ is chosen to be the value $\hat{\boldsymbol{\theta}}_{ML}$ that maximizes the conditional probability or 'likelihood function'

$$p(\mathbf{x}_T|\boldsymbol{\theta}) = p(x(1), x(2), \ldots, x(T)|\boldsymbol{\theta})$$

of the T measurements $\mathbf{x}_T = (x(1), x(2), \ldots, x(T))^T$. The ML estimator corresponds to the value $\hat{\boldsymbol{\theta}}_{ML}$ that makes the obtained measurements most likely. Often it is convenient to deal with the log-likelihood function $\log p(\mathbf{x}_T|\boldsymbol{\theta})$. The ML estimator is usually found from the solutions of the likelihood equation

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathbf{x}_T|\boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML}} = 0 \ .$$

1. Let the probability density function (pdf) p of a scalar-valued random variable z be

$$p(z|\theta) = \theta^2 \, z \, \exp(-\theta \, z), \quad z \geq 0, \quad \theta > 0 \ .$$

Determine the ML estimate of the parameter θ. There are available T independent measurements $z(1), \ldots, z(T)$ on z.

In the ML framework of ICA, we denote by $\tilde{p}_i$ the assumed densities of the independent components, and by $G_i'(s_i) := \partial/\partial s_i \, [\log \tilde{p}_i(s_i)]$ the derivative of non-polynomial moments G_i used for the estimation. Constrain the estimates of the independent components to be uncorrelated, zero-mean, and to have unit variance. Then the ML estimator is locally consistent[1], if the assumed densities $\tilde{p}_i$ fulfill

$$E\{s_i \, G_i'(s_i) - G_i''(s_i)\} > 0 \ . \tag{1}$$

2. Why is it not correct to use $G(y) = y^4$ within the ML framework? What would be a possible choice for a subgaussian distribution?

3. Take $G'(y) = -y$. What is the interpretation of this choice in the ML framework? Conclude from condition (1) that G' must be nonlinear.

4. Show that for a gaussian random variable s_i of zero mean and unit variance, above condition (1) on the non-polynomial moments G cannot be fulfilled for any G. Hint: show that the expectation in (1) is zero.

[1] ML estimators are called consistent if $\hat{\boldsymbol{\theta}}_{ML}$ converges to the true value of $\boldsymbol{\theta}$ when the number of measurements grows infinitely large.

Kullback-Leibler divergence and Jensen's inequality

The Kullback-Leibler divergence between two n-dimensional pdf's $p^{(1)}$ and $p^{(2)}$ is defined as

$$\delta(p^{(1)}, p^{(2)}) = \int \mathrm{d}\boldsymbol{\xi}\, p^{(1)}(\boldsymbol{\xi}) \, \log \frac{p^{(1)}(\boldsymbol{\xi})}{p^{(2)}(\boldsymbol{\xi})} \; .$$

The Kullback-Leibler distance is always nonnegative, and zero if and only if the two distributions are equal. It can thus be considered as a kind of (non-symmetric) distance between the two pdf's. This feature is a consequence of the (strict) convexity of the negative logarithm and Jensen's inequality

$$E\{f(y)\} \geq f(E\{y\}) \; ,$$

which holds for any strictly convex function f and any random variable y.

5. Using Jensen's inequality, show that the Kullback-Leibler divergence is nonnegative. Hint: take $f(y) = -\log y$ and $y = p^{(2)}(\boldsymbol{\xi})/p^{(1)}(\boldsymbol{\xi})$.

Infomax principle

6. The input x and the output y of a so-called 'noisy graded-response neuron' is given by

$$y = \phi(x) + n$$

where n is gaussian noise and ϕ is some sigmoidal scalar function, i.e., ϕ is monotonically increasing and fulfills $\lim_{\xi \to -\infty} \phi(\xi) = 0$ and $\lim_{\xi \to +\infty} \phi(\xi) = 1$. In order to have an efficient information flow from input to output the transfer function ϕ can be adapted. Efficient information transmission requires that we maximize the mutual information of input and output. To derive the corresponding optimal ϕ proceed as indicated below.

a) Argue that the joint pdf $p_{x,y}(\xi, \eta)$ of the input and the output can be written as $p_{x,y}(\xi, \eta) = p_n(\eta - \phi(\xi))\, p_x(\xi)$, where p_n is the pdf of the noise.

b) Show that the joint entropy $H(x, y)$ of input and output is constant, i.e., independent of ϕ.

c) Show that maximizing the mutual information $I(x, y)$ of input and output is equivalent to maximizing the entropy $H(y)$ of the output.

d) Calculate the entropy $H(y)$ of the output. Hint: use the entropy of an invertible transformation $u = \psi(v)$: $H(u) = H(v) + E\{\log |\psi'(v)|\}$ where ψ' is the derivative of ψ.

e) Prove that the maximum of the entropy of the output is reached if and only if $\phi'(x) = p_x(x)$.

DR. R. KEMPTER, phone 2093-8925, room 2315, `r.kempter(AT)biologie.hu-berlin.de`
PROF. DR. A.V.M. HERZ, phone 2093-9112, room 2325, `a.herz(AT)biologie.hu-berlin.de`

Problems handed out on Monday, 15.05.2006.
Solutions to be handed in by Monday, 22.05.2006, 12^{15} pm.
Discussion and presentation of the problems on Friday, 26.05.2006, $8^{30} - 10^{00}$ am in front of room 2317 I-W (Zwischengeschoß).