Andrej Pázman

Distribution of estimators and optimal experimental design in nonlinear regression

Kybernetika, Vol. 32 (1996), No. 5, 491--499

Persistent URL: http://dml.cz/dmlcz/124813

KYBERNETIKA — VOLUME 32 (1996), NUMBER 5, PAGES 491 – 499

DISTRIBUTION OF ESTIMATORS AND OPTIMAL EXPERIMENTAL DESIGN IN NONLINEAR REGRESSION[1]

ANDREJ PÁZMAN

In this survey we present in a condensed exposition good approximations of probability densities of LS estimators in nonlinear regression models. This includes also marginal densities or densities of scalar parametric functions. Further, the main ideas of two approaches how to use these densities for optimal experimental design are presented: the approach using the second order approximation technique, and the approach using optimality criteria in an integral form.

1. INTRODUCTION

We consider a nonlinear regression experiment

$$y(x_i) = \eta(x_i, \theta) + \epsilon_i; \qquad i = 1, \dots N, \tag{1}$$

$$\epsilon_1, \dots, \epsilon_N \quad \text{i.i.d.} \sim N(0, \sigma^2).$$

which is performed according to some (exact) design

$$\vec{X} = (x_1, \dots, x_N).$$

The vector $\theta \in \Theta$ is the vector of unknown parameters of dimension p. The (Fisher) information matrix corresponding to the design $\vec{X}$ and to the parameter θ has the form

$$M_{ij}(\vec{X}, \theta) = \sum_{k=1}^{N} \frac{\partial \eta(x_k, \theta)}{\partial \theta_i} \frac{\partial \eta(x_k, \theta)}{\partial \theta_j}. \tag{2}$$

A standard optimality criterion in model (1) is given by a function (cf. [2], [4], [6])

$$\vec{X} \to \Phi[M^{-1}(\vec{X}, \theta)]$$

[1] Paper presented at the International Workshop: Advances in Model-Oriented Data Analysis, Spetses, Greece, June 5 – 9, 1995. Research supported by the Slovak Grant Agency.

where Φ is equal to $\Phi[M^{-1}] = \ln\det[M^{-1}]$ (D-optimality) or has any other form known from the classical theory of design in linear models. This approach can be criticized for two reasons:

i) Such criteria depend on θ. But this is a principal property of any optimality criterion in nonlinear models, since the amount of information obtained from the observations may depend very much on θ. Consequently, for designing nonlinear experiments we need some prior knowledge about the parameter value, e.g. a prior density or a guess about the value of θ. The last approach is used also in this paper.

ii) In contrast to linear models, the expression $\sigma^2 M^{-1}(\vec{X}, \theta)$ is just a first order approximation of the variance of the LS estimator

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \|y - \eta(\theta)\|^2 \tag{3}$$

and it corresponds to the asymptotic normal distribution of $\hat{\theta}$. (We note that by $\|\ \|$ we denote the Euclidean norm, i.e. the square root of the sum of squares.) This approximation may be misleading in small sample nonlinear regression, unless we restrict the set of designs to those giving models with small curvatures (cf. [5]). Another approach is the use of second order approximations of the model or of a confidence region (cf. [6] for a survey).

In this paper we present alternative approaches which are based exclusively on the probability density of $\hat{\theta}$. *This is justified by the evident fact that the whole information about the statistical properties of $\hat{\theta}$ is contained in its probability distribution.*

2. APPROXIMATE JOINT PROBABILITY DENSITY OF $\hat{\theta}$

Before discussing experimental design, we present approximate probability densities of estimators. In this section for the vector $\hat{\theta}$, in the next section for the estimator of any scalar function of θ.

From this place up to the last paragraph of Section 4 the design $\vec{X}$ is fixed, and we omit to write it in the notation.

Unless some particular cases, we do not know the exact probability density of $\hat{\theta}$, but there are very good approximations. For surveys the reader is referred to [9], and recently to [7]. Some results discussed in [7] require the knowledge of moments (Edgeworth expansions), or their precision is studied only asymptotically.

The approximate density of $\hat{\theta}$ considered in the present paper is given by the formula (4) without using moments, just the first and second order derivatives of $\eta(x_i, \theta)$ are required. The precision of this approximation can be evaluated for small samples by a geometric insight into the model (cf. [8, 11, 12]).

Here is the formula for the approximate density (cf. [12]):

$$q(\hat{\theta}|\theta) = \frac{\mathcal{F}(\hat{\theta}, \theta)}{(2\pi)^{m/2}\sigma^m \det^{1/2}[M(\hat{\theta})]} \exp\left\{-\frac{1}{2\sigma^2}\|P(\hat{\theta})[\eta(\hat{\theta}) - \eta(\theta)]\|^2\right\} \tag{4}$$

where

$$P(\theta) = \frac{\partial\eta(\theta)}{\partial\theta^{\mathrm{T}}}\, M^{-1}(\theta)\, \frac{\partial\eta^{\mathrm{T}}(\theta)}{\partial\theta}$$

is an orthogonal projector. The term $\mathcal{F}(\hat{\theta}, \theta)$ is rather complicated in general nonlinear models, however, it can be expressed in simple terms if the model is "linear in a certain sense":

i) We have

$$\mathcal{F}(\hat{\theta}, \theta) = \det[M(\hat{\theta})],$$

if model (1) is linear in θ, or if it is intrinsically linear. The last assumption means that the intrinsic curvature of Bates–Watts (cf. [3]) is equal to zero identically. In this case the density (4) is exact, as is well known.

ii) If on the other hand, only the parameter effect curvature is zero identically, or equivalently, if $M(\theta) = \text{const}$ on Θ, one takes (cf. [8, 9])

$$\mathcal{F}(\hat{\theta}, \theta) = \det\left[M_{ij}(\hat{\theta}) + [\eta(\hat{\theta}) - \eta(\theta)]^{\mathrm{T}} H_{ij}(\hat{\theta})\right]$$

where

$$H_{ij}(\theta) = \frac{\partial^2 \eta(\theta)}{\partial \theta_i \partial \theta_j}.$$

iii) In more general nonlinear models we need the expressions

$$R_{ikjl}(\theta) = H_{ij}^{\mathrm{T}}(\theta)\,[I - P(\theta)]\,H_{kl}(\theta) - H_{il}^{\mathrm{T}}(\theta)\,[I - P(\theta)]\,H_{kj}(\theta); \quad i, j, k, l = 1, \ldots, p,$$

which are know in differentiable geometry as components of the Riemannian curvature tensor. (Cf. [1] for the definition of this tensor in general statistical models). If a regression model has either a zero intrinsic curvature or a zero parameter curvature, then the Riemannian curvature tensor is zero identically, i.e. for every θ. The reverse is not true, for instance, if $p = 1$, then evidently $R(\theta) \equiv 0$, although the intrinsic and the parameter curvatures may be non-zero. Another example is the classical Michaelis–Menten model which also has $R(\theta) \equiv 0$. Important is that regression models having $R(\theta) = 0$ identically, still maintain some properties of linear models. (Cf. [10] for a corresponding classification of nonlinear regression models.)

In models with $R(\theta) \equiv 0$ we have (cf. [8, 9])

$$\mathcal{F}(\hat{\theta}, \theta) = \det[Q(\hat{\theta}, \theta)], \tag{5}$$

where

$$Q_{ij}(\hat{\theta}, \theta) = M_{ij}(\hat{\theta}) + [\eta(\hat{\theta}) - \eta(\theta)]^{\mathrm{T}}\,[I - P(\hat{\theta})]\,H_{ij}(\hat{\theta})$$

is a modified information matrix.

iv) In general models the term $\mathcal{F}(\hat{\theta}, \theta)$ can be expressed as a polynomial in the components of $Q(\hat{\theta}, \theta)$ and of $R(\hat{\theta})$. (Cf. [12] for the explicit form of this polynomial). The formula (4) is an approximate density as soon as the model is not intrinsically linear, but its precision is "almost exact" in the following sense: one can take a subset S of the sample space R^N so that the probability of S is nearly one, and that (4) is the exact density obtained from samples y belonging to the set S. For more details concerning the meaning of "almost exact" cf. [11].

However, it seems that the choice of $\mathcal{F}(\hat{\theta}, \theta)$ given by (5) is sufficient in many cases. This choice is equivalent to neglecting the value of the Riemannian curvature tensor.

Moreover, the density (4) with $\mathcal{F}(\hat{\theta},\theta) = \det[Q(\hat{\theta},\theta)]$ is equal to the saddlepoint approximation of the density of $\hat{\theta}$ (cf. [7]). So besides a geometrical insight of the approximation (4) we have also an asymptotical justification based on the properties of the saddlepoint approximation.

3. DENSITIES OF ESTIMATORS OF SCALAR FUNCTION OF θ

Sometimes an experiment giving the maximum precision of the estimator of the response function is required. Or, the aim of the experiment is the estimation of a given function of θ. Consequently we need also densities of estimators of scalar, linear or nonlinear, parametric functions $\theta \in \Theta \to \Gamma(\theta)$. In particular, $\Gamma(\theta)$ may be equal to the ith parameter θ_i.

The aim of this section is to present a recent method for computing the approximate probability densities of such estimators. Unfortunately, the approximation, although still very good in many cases, is not "almost exact" in the sense mentioned in Section 2. (Cf. [14] for computing the densities of each parameter separately in the two-dimensional Michaelis–Menten model.)

Denote

$$G(y) \equiv \Gamma(\hat{\theta}(y)).$$

Here $\hat{\theta}(y)$ is the estimator (3) as a function of $y = (y(x_1), \ldots, y(x_N))^{\mathrm{T}}$. We remind the reader that the design $\vec{X} = (x_1, \ldots, x_N)$ is fixed. Denote further by $f_G(\gamma)$ the density of $G(y)$.

The main idea here is to use the Dirac function $\delta(\cdot)$. We write for any fixed $\gamma \in \mathcal{R}^1$

$$\begin{aligned} f_G(\gamma) &= \int_{\mathcal{R}^1} \delta(\gamma - u) f_G(u)\,\mathrm{d}u \\ &= \lim_{\lambda \downarrow 0} \int_{\mathcal{R}^N} \frac{1}{\sqrt{2\pi}\lambda} e^{-\frac{1}{2\lambda^2}[G(y)-\gamma]^2} f(y|\theta)\,\mathrm{d}y. \end{aligned} \tag{6}$$

If the model and the function Γ are linear, then $[G(y) - \gamma]^2$ is a quadratic form in the variable y. In the general case we approximate it by a quadratic form $Q_\gamma(y)$, to obtain approximately

$$f_G(\gamma) \doteq \lim_{\lambda \downarrow 0} \int_{\mathcal{R}^N} \frac{1}{\sqrt{2\pi}\lambda} e^{-\frac{1}{2\lambda^2} Q_\gamma(y)} f(y|\theta)\,\mathrm{d}y$$

where $f(y|\theta)$ is the normal density of y. Important is a good choice of $Q_\gamma(y)$ to obtain a simple limit and a good accuracy of the approximation. Two different choices are considered in [14]. The first consists in computing

$$\theta_\gamma := \operatorname*{argmin}_{\theta \in \Theta, \Gamma(\theta) = \gamma} \|\bar{\eta} - \eta(\theta)\|^2$$

for each γ. Here $\bar{\eta}$ is the true value of $\eta(\theta)$. The form $Q_\gamma(y)$ is then obtained as the quadratic Taylor formula of the function $y \to [G(y) - y]^2$ at the point $y = \eta(\theta_\gamma)$.

The resulting approximate density is (cf. [14]):

$$f_G(\gamma) \doteq \frac{1}{\sqrt{2\pi}\sigma\|b_\gamma\|} \exp\left\{-\frac{1}{2\sigma^2}\|P_\gamma[\eta(\theta_\gamma) - \bar{\eta}]\|^2\right\}, \tag{7}$$

where

$$b_\gamma = \left.\frac{\partial \Gamma(\hat{\theta}(y))}{\partial y}\right|_{y=\eta(\theta_\gamma)},$$

and

$$P_\gamma = \frac{b_\gamma b_\gamma^{\mathrm{T}}}{\|b_\gamma\|^2}.$$

The required derivative of $\hat{\theta}(y)$ with respect to y is obtained by the use of the implicit function theorem, as described in Section 4.

In the second approach, which gives a small improvement of (7), we take for $Q_\gamma(y)$ the quadratic form which appears when using the Laplace approximation of the integral (6) (before taking the limit in (6); for details cf. [14]). The resulting formulae are more complicated, and are not presented here.

4. DESIGN BASED ON THE SECOND ORDER APPROXIMATION OF THE ENTROPY

In fact, we do not need to know the density of $\hat{\theta}$ when computing the second order approximation of the moments of this distribution, because the integrals which define such moments can be taken with respect to the density of the observed vector y. Such approximations are well known (cf. e. g. [16]). However, to approximate the entropy of the distribution of $\hat{\theta}$ one can not avoid the use of $q(\hat{\theta}|\theta)$.

The entropy of $q(\hat{\theta}|\theta)$ is

$$Ent = -\int_\Theta [\ln q(\hat{\theta}|\theta)] q(\hat{\theta}|\theta)\, \mathrm{d}\hat{\theta},$$

and it can be written, by the use of the Taylor formula, in the form

$$\begin{aligned} Ent &= E_\theta[-\ln q(\hat{\theta}|\theta)] = -\int_{\mathcal{R}^N} \ln q(\hat{\theta}(y)|\theta) f(y|\theta)\, \mathrm{d}y \\ &\doteq -\int_{\mathcal{R}^N} \left\{ \ln q(\hat{\theta}(y)|\theta)_{y=\eta(\theta)} + \left.\frac{\partial \ln q(\hat{\theta}(y)|\theta)}{\partial y^{\mathrm{T}}}\right|_{y=\eta(\theta)} (y - \eta(\theta)) \right. \\ &\quad + \left. \frac{1}{2}(y-\eta(\theta))^{\mathrm{T}} \left.\frac{\partial^2 \ln q(\hat{\theta}(y)|\theta)}{\partial y \partial y^{\mathrm{T}}}\right|_{y=\eta(\theta)} (y - \eta(\theta)) \right\} f(y|\theta)\, \mathrm{d}y. \end{aligned}$$

To express the last integral we need the moments of y and the derivatives of $\hat{\theta}(y)$. Since $f(y|\theta)$ is normal, the moments are trivial. To obtain the derivatives, we use the normal equation

$$0 = \frac{\partial}{\partial \theta}\|y - \eta(\theta)\|^2 = 2[\eta(\theta) - y]^{\mathrm{T}} \frac{\partial \eta(\theta)}{\partial \theta^{\mathrm{T}}}$$

and the evident equality

$$\hat{\theta}(y)\Big|_{y=\eta(\theta)} = \theta.$$

Since the equality in the normal equation holds after setting $\theta = \hat{\theta}(y)$, the mapping $y \to \hat{\theta}(y)$ is defined implicitly by this equation, at least in a neighbourhood of the point $y = \eta(\theta)$. This allows to compute the derivatives of $\hat{\theta}(y)$. For that, let us denote by $F(y, \theta)$ the right hand side of the normal equation. According to the implicit function theorem

$$\frac{\partial \hat{\theta}^{\mathrm{T}}(y)}{\partial y} = -\frac{\partial F(y,\theta)}{\partial y}\Big|_{\theta=\hat{\theta}(y)} \left[\frac{\partial F(y,\theta)}{\partial \theta^{\mathrm{T}}}\right]^{-1}\Big|_{\theta=\hat{\theta}(y)}.$$

By a direct differentiation we obtain the higher order derivatives.

By the described procedure we obtain the second order approximation of the entropy, which, expressed as a function of $\vec{X}$, has then the form (cf. [15]):

$$Ent_{\vec{X}} \doteq ent_{\vec{X}} - \frac{\sigma^2}{2} M^{-1}_{ij}(\vec{X},\theta)\Big\{M^{-1}_{ab}(\vec{X},\theta)\Big[R_{ajbi}(\vec{X},\theta) + U^b_{aij}(\vec{X},\theta)\Big] - \Gamma^d_{ai}(\vec{X},\theta)\,\Gamma^a_{dj}(\vec{X},\theta) - \Gamma^a_{ac}(\vec{X},\theta)\,\Gamma^c_{ij}(\vec{X},\theta)\Big\} \qquad (8)$$

where

$$ent_{\vec{X}} = -\frac{1}{2}\ln\det M(\vec{X},\theta) + \text{const.}$$

is the first-order approximation of the entropy, and where

$$\Gamma^d_{ai}(\vec{X},\theta) = \sum_{k=1}^{N} \frac{\partial\eta(x_k,\theta)}{\partial\theta_d}\,\frac{\partial^2\eta(x_k,\theta)}{\partial\theta_a\partial\theta_i},$$

$$U^b_{aij}(\vec{X},\theta) = \sum_{k=1}^{N} \frac{\partial\eta(x_k,\theta)}{\partial\theta_b}\,\frac{\partial^3\eta(x_k,\theta)}{\partial\theta_a\partial\theta_i\partial\theta_j}.$$

Similarly, the expression $R_{ajbi}(\vec{X},\theta)$ is a component of the Riemannian curvature tensor for the given design $\vec{X}$.

The value of $Ent_{\vec{X}}$ has been used in [15] to compute the optimum design for the Michaelis–Menten model, and has been compared with the classical D-optimum design.

5. DESIGNS WITH OPTIMALITY CRITERIA IN AN INTEGRAL FORM

The second order approximations of moments or of the entropy, etc., can be wrong if for example the probability density of $\hat{\theta}$ is bimodal, or if a large part of the distribution of $\hat{\theta}$ is on the boundary of the parameter space Θ. In such cases

i) we need criteria of an integral form, as considered in this section, and

ii) we have to modify the density (4) to include also the probability on the boundary of Θ.

Let us consider the second problem. To take into account the probability distribution of $\hat{\theta}$ on the boundary of the parameter space Θ, we modify the estimator itself. Instead of (3) we consider

$$\tilde{\theta}(y) = \operatorname*{argmin}_{\theta} \{\|y - \eta(\theta)\|^2 + \omega(\theta)\}$$

where $\omega(\theta)$ is some smooth penalty function, which is infinite on the boundary of Θ, and zero in an "interior part" of $int(\Theta)$, say $\Theta^\star \subset int(\Theta)$. The particular choice of this penalty function is not decisive, and some possibilities are presented in [13]. One can easy see that we have $\hat{\theta} = \tilde{\theta}$ on $\Theta^\star$, but instead of the estimates $\hat{\theta}$ on the boundary of Θ, we have estimates $\tilde{\theta}$, which are in the region $\Theta - \Theta^\star$ close to the boundary of Θ. So there are shifts of estimates from the boundary to a neighbourhood region. What is important is that such shifts of estimates do not influence very much the values of the optimality criteria given below.

The approximate probability density of $\tilde{\theta}$ (which is on the same level of precision as (4) for $\mathcal{F}(\hat{\theta}, \theta) = \det[Q(\hat{\theta}, \theta)]$) has the form

$$q(\tilde{\theta}|\theta) = \frac{\mathcal{T}(\tilde{\theta}, \theta)}{(2\pi)^{m/2}\sigma^m \det^{1/2}[M(\tilde{\theta})]} \exp\left\{-\frac{1}{2\sigma^2}\left\|P(\tilde{\theta})[\eta(\tilde{\theta}) + u(\tilde{\theta}) - \eta(\theta)]\right\|^2\right\} \quad (9)$$

where

$$\mathcal{T}(\tilde{\theta}, \theta) = \det\left[Q_{ij}(\tilde{\theta}, \theta) - \frac{1}{\sigma^2} u^{\mathrm{T}}(\tilde{\theta})\, H_{ij}(\tilde{\theta}) + \frac{\partial^2 \omega(\tilde{\theta})}{\partial\theta_i \partial\theta_j}\right]; \qquad i, j = 1, \ldots, p,$$

and where $u(\theta)$ is a vector

$$u(\theta) = \frac{\partial \eta(\theta)}{\partial \theta^{\mathrm{T}}} M^{-1}(\theta) \frac{\partial \omega(\theta)}{\partial \theta}.$$

In this section we consider criteria functions of the form

$$\vec{X} \to \Phi[Q_{\vec{X}}]$$

where Φ is a classical criterion function known from linear models (e. g. $\ln\det(\cdot)$ for D-optimality, $\mathrm{tr}(\cdot)$, for A-optimality, etc.), and where $Q_{\vec{X}}$ is the $p \times p$ mean square error matrix of $\hat{\theta}$ with entries

$$\{Q_{\vec{X}}\}_{ij} = \int_{\Theta} (\tilde{\theta} - \theta)_i\, (\tilde{\theta} - \theta)_j\, q_{\vec{X}}(\tilde{\theta}|\theta)\, \mathrm{d}\tilde{\theta},$$

Since the moments in these entries are given by integrals, we speak about integral criteria functions. Sometimes the criterion function is given by a unique integral, as is in the case of a generalized A-optimality criterion where

$$\Phi[Q_{\vec{X}}] = \sum_{i=1}^{p} \{Q_{\vec{X}}\}_{ii} = \int_{\Theta} \|\tilde{\theta} - \theta)\|^2\, q_{\vec{X}}(\tilde{\theta}|\theta)\, \mathrm{d}\tilde{\theta} \quad (10)$$

considered in details in [13]. Similarly, the criterion function is given by just one integral if it is a generalized L-criterion function, i.e. if it has the form $\Phi(\cdot) = \mathrm{tr}[W(\cdot)]$ where W is a given positive semidefinite matrix. On the other hand, in the case of the generalized D-optimality criterion with $\Phi(\cdot) = \ln\det(\cdot)$, the criterion function can not be reduced to a unique integral.

To minimize the integral (10) (or any other criterion function given by one integral) with respect to the design $\vec{X}$, we proceed by using directly the method of stochastic approximation.

In general, the stochastic approximation means to minimize:

$$\min_x \varphi(x)$$

where $\varphi(x)$ is an unknown function. The information about the values of $\varphi(x)$ is obtained by observing some random variables $z(x_i)$; $i = 1, \ldots$ such that

$$E[z(x_i)] = \varphi(x_i)$$

where the points $x_1, x_2, \ldots$ are chosen according to a certain rule.

In our case instead of $x_1, x_2, \ldots$ the sequence of designs $\vec{X}_1, \vec{X}_2, \ldots$ is taken according to a rule, which in fact is a modification of the gradient method. The limit of this sequence is the optimal design. We take $\varphi(\cdot)$ of the form

$$\varphi(\vec{X}) = \int_\Theta ||\tilde{\theta} - \theta||^2 \, q_{\vec{X}}(\tilde{\theta}|\theta) \, \mathrm{d}\tilde{\theta},$$

and we take

$$z(\vec{X}_i) = ||\theta^{(i)} - \theta||^2 \, q_{\vec{X}_i}(\theta^{(i)}|\theta).$$

Here the design $\vec{X}_i$ is the design at the ith step of the iteration, and $\theta^{(i)}$ is taken at random from a uniform distribution on Θ. So instead of observing, we simulate vectores $\theta^{(i)}$, and compute $z(\vec{X}_i)$. Hence, in our problem we combine a Monte–Carlo simulation and a gradient minimization. For details and an example cf. [13].

For criteria functions which can not be reduced to one integral (say for D-optimality), such a direct use of stochastic approximation is not possible. An indirect use may be possible, however has not been checked.

Conclusion. Having good approximations of the probability distribution of the least squares estimator on the whole parameter space, including its boundary, one can build meaningful optimality criteria, and compute optimum designs. In contrast to linear model, the criteria functions are not convex, but one can use standard numerical procedures to minimize the criteria functions, at least when the number of parameters and the number of considered design points is not very large.

(Received December 12, 1995.)

REFERENCES

[1] S. Amari: Differential-Geometrical Methods in Statistics. (Lecture Notes in Statist. *28.*) Springer-Verlag, Berlin 1985.

[2] A. C. Atkinson and A. N. Donev: Optimum Experimental Designs. Clarendon Press, Oxford 1992.

[3] D. M. Bates and D. G. Watts: Relative curvature measures of nonlinearity. J. Roy. Statist. Soc. *B 42* (1980), 1–25.

[4] P. Chaudhuri and P. A. Mykland: Nonlinear experiments: Optimal design and inference based on likelihood. J. Amer. Statist. Assoc. *88* (1993), 538–546.

[5] M. A. Clyde: Bayesian designs for approximate normality. In: Advances in Model-Oriented Data Analysis (Ch. P. Kitsos and W. Müller, eds.), Physica-Verlag, Heildelberg 1995.

[6] I. Ford, C. P. Kitsos and D. M. Titterington: Recent advances in nonlinear experimental design. Technometrics *31* (1989), 49–60.

[7] P. Hougaard: Nonlinear regression and curved exponential families. Improvement of the approximation to the asymptotic distribution. Metrika *42* (1995), 191–202.

[8] A. Pázman: Probability distribution of the multivariate nonlinear least squares estimator. Kybernetika *20* (1984), 209–230.

[9] A. Pázman: Small sample distributional properties of nonlinear regression estimators (a geometric approach). Statistics *21* (1990), 323–367.

[10] A. Pázman: A classification of nonlinear regression models and parameter confidence regions. Kybernetika *28* (1992), 444–453.

[11] A. Pázman: Nonlinear Statistical Models. Kluwer, Dordrecht 1993.

[12] A. Pázman: Higher dimensional nonlinear regression – a statistical use of the Riemannian curvature tensor. Statistics *25* (1994), 17–25.

[13] A. Pázman and L. Pronzato: Nonlinear experimental design based on the distribution of estimators. J. Stat. Plann. Inference *33* (1992), 385–402.

[14] A. Pázman and L. Pronzato: A Dirac-function method for densities of nonlinear statistics and for marginal densities in nonlinear regression. Statist. Probab. Lett. *26* (1996), 159–167.

[15] L. Pronzato and A. Pázman: Second-order approximation of the entropy in nonlinear least-squares estimation. Kybernetika *30* (1994), 187–198. Erratum. Kybernetika *32* (1996), 104.

[16] B. C. Wei: Some second order asymptotics in nonlinear regression. Austral. J. Statist. *33* (1991), 75–84.

Prof. RNDr. Andrej Pázman, DrSc., Faculty of Mathematics and Physics, Comenius University, Mlynská dolina, 842 15 Bratislava. Slovak Republic.