ELSEVIER

Chemometrics and Intelligent Laboratory Systems 49 (1999) 41–48

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Confidence intervals for nonlinear regression extrapolation

Alexey L. Pomerantsev *

Institute of Chemical Physics of RAS, Polycert, 4 Kosygin Street, 117977 Moscow, Russian Federation

Received 20 June 1998; accepted 4 April 1999

Abstract

The various methods of confidence intervals construction for nonlinear regression are considered. The new method named by a method of associated simulation (the AS-method) is proposed. Using computerized simulation, it is shown on the example that only two methods, the bootstrap and the AS-method, give a satisfactory accuracy. The advantage of the AS-method is the speed. In comparison with the bootstrap, the prize is at least 10 000 times. This method may be applied when regression parameters estimates are obtained by the maximum likelihood method. It was proposed to use the AS-method when extrapolation of complicated physico-chemical model is performed to predict the behavior of the model in the area far from observation.

Keywords: Nonlinear regression model; Confidence interval; Forecast; Extrapolation

1. Introduction

The nonlinear regression is a popular tool for describing the behavior of complex physico-chemical systems [1]. It enables to create hard models, which are based on the objective reasons about the investigated processes. Only such models allow to predict the value of response far from the area of observation [2]. The error of forecast is determined first of all by extrapolation, and lesser by measurement error. Therefore, the estimation of forecasted value solely is not enough, but the confidence interval forecast is required. In this paper, the various ways of constructing such confidence intervals are investigated.

The methods under consideration differ both in complexity of execution and on accuracy that is a closeness to the 'true' confidence interval. To check the accuracy, we used a model example that will be presented later. All the methods set forth here are known, except one offered by the author. The new method of confidence estimation named 'associated simulation' is a variety of the Monte-Carlo method.

In the course of investigation of the confidence intervals, the same variable frequently appears as a random and as a determined one, being a realization of this random variable. For convenience, all random variables will be designated by Greek letters and determined ones by Latin letters. The elimination was made only for the variance of the normal law, which is traditionally designated as σ^2, and its estimate as s^2.

* Tel.: +7-95-939-7483 (office), +7-95-283-4757 (home); fax: +7-95-283-4757 (home); E-mail: polycert@chph.ras.ru (office), al.pom@g23.relcom.ru (home)

0169-7439/99/$ - see front matter
PII: S0169-7439(99)00026-X

Let:

$$\boldsymbol{\eta} = f(\boldsymbol{a},\boldsymbol{x}) + \boldsymbol{\varepsilon}, \tag{1}$$

be a nonlinear regression, where $\boldsymbol{a} = (a_1,\ldots,a_p)^t$ is the p-dimensional vector of unknown parameters, $\boldsymbol{x}$ is the factor vector (the dimension of $\boldsymbol{x}$ is not essential to us), $\boldsymbol{\varepsilon}$ is the error vector, which we suppose to be distributed normally and independently with zero mean and variance σ^2. Let $\boldsymbol{\eta}_0$ be an experimental n-dimensional data set, which is obtained by measurements in the points $\boldsymbol{x}_i \in O$ (for $i = 1\ldots, n$), where O is an area of observations.

For searching the estimates $\boldsymbol{\alpha}$ of parameters $\boldsymbol{a}$, we shall use least squares method, so:

$$\boldsymbol{\alpha}(\boldsymbol{\eta}) = \arg \min_{a} Q(\boldsymbol{\eta},\boldsymbol{a}), \tag{2}$$

where $Q(\boldsymbol{\eta},\boldsymbol{a}) = [\boldsymbol{\eta} - f(\boldsymbol{a},\boldsymbol{x})]^t[\boldsymbol{\eta} - f(\boldsymbol{a},\boldsymbol{x})]$ is the sum of squares. The estimation of parameter σ^2 will be carried out by a traditional unbiased mode:

$$s^2 = Q(\boldsymbol{\eta},\boldsymbol{\alpha})/(n-p) \tag{3}$$

Let $g(\boldsymbol{a})$ be a differentiable scalar function of parameter vector $\boldsymbol{a}$. Then, statistics:

$$\gamma = g(\boldsymbol{\alpha}) = g(\boldsymbol{\alpha}(\boldsymbol{\eta})) \tag{4}$$

is an estimator of $g(\boldsymbol{a})$. The random variable $\gamma_+(P)$ is an upper bound of one-sided confidence interval for $g = g(\boldsymbol{a})$ at reliability P, if:

$$\mathrm{Prob}\{\gamma_+(P) > g\} \geq P. \tag{5}$$

For simplicity, we shall consider only upper bound, meaning that a low bound $\gamma-$ may be easily evaluated as $\gamma-(P) = \gamma_+(1-P)$. First of all, we shall consider the case when $g(\boldsymbol{a})$ is the regression function itself (Eq. (1)) outside the area of observation $g(\boldsymbol{a}) = f(\boldsymbol{a},\boldsymbol{x}_0)$, where $\boldsymbol{x}_0 \notin O$.

Let $F(z,\boldsymbol{a},\sigma^2)$ be a cumulative distribution function of statistics (2), i.e.:

$$F(z,\boldsymbol{a},\sigma^2) = \mathrm{Prob}\{\boldsymbol{\alpha} < z\}, \tag{6}$$

then the distribution $G(z,\boldsymbol{a},\sigma^2)$ of statistics (4) is:

$$G(z,\boldsymbol{a},\sigma^2) = \mathrm{Prob}\{\gamma < z\} = \int\!\!\ldots\!\!\int_{g(\boldsymbol{a})<z} \frac{\partial^p F(\boldsymbol{x},\boldsymbol{a},\sigma^2)}{\partial x_1 \ldots \partial x_p} \mathrm{d}x_1 \ldots \mathrm{d}x_p. \tag{7}$$

The confidence interval (5) can be constructed with the help of statistics (4) as:

$$\gamma_+(P) = G^{-1}(P,\boldsymbol{a}_0,s_0^2). \tag{8}$$

Here G^{-1} is the function inverse to function (7), i.e., $G[G^{-1}(P,\ldots),\ldots] = P$, $\boldsymbol{a}_0$ and s_0^2 are realizations of estimates (2) and (3), obtained in experiment, i.e., for $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. Certainly, in arbitrary case, the distribution (6) is unknown, therefore, for confidence interval construction, we use the distributions $\hat{F}$ and $\hat{G}$, which are approximations of distributions (6) and (7). Now we consider various modes of confidence intervals construction.

2. Methods for confidence intervals construction

2.1. Stochastic approximation (SA)

It is known [3] that for a linear normal regression, the distribution (6) is Gaussian. In a nonlinear case, this statement is correct only asymptotically when number of observations is large. Nevertheless, it is possible to construct a confidence interval basing on this fact. Let us assume that:

$$\hat{F} = \Phi(\boldsymbol{a},\sigma^2\boldsymbol{A}^{-1}) \tag{9}$$

where Φ is p-dimensional cumulative normal distribution with the mean vector $\boldsymbol{a}$ and matrix of covariance $\sigma^2\boldsymbol{A}^{-1}$. Expanding functions $f(\boldsymbol{a},\boldsymbol{x})$ and $g(\boldsymbol{a})$ in a series at a point $\boldsymbol{a} = \boldsymbol{a}_0$ and using Gauss' approximation [3] for a Hesse matrix $\boldsymbol{A}$ we receive:

$$\boldsymbol{A}_{ij} = \frac{1}{2}\frac{\partial^2 Q(\boldsymbol{a}_0)}{\partial a_i \partial a_j} \approx \sum_1^n \frac{\partial f(\boldsymbol{a}_0)}{\partial a_j}\frac{\partial f(\boldsymbol{a}_0)}{\partial a_i};$$

$$i,j = 1,\ldots,p. \tag{10}$$

For such approximation, the distribution $\hat{G}$ is also Gaussian with parameters $E(\gamma) = g_0$ and $D(\gamma) = s_0^2\boldsymbol{v}^t\boldsymbol{A}^{-1}\boldsymbol{v}$, where $g_0 = g(\boldsymbol{a}_0)$ and $\boldsymbol{v} = \nabla g(\boldsymbol{a})$. From Eq. (8) we get:

$$\gamma_+(P) = g_0 + x_P s_\gamma. \tag{11}$$

Here x_P is P-quantile of normal distribution and $s_\gamma = \sqrt{D(\gamma)}$. In some cases (for example, when $g(\boldsymbol{a}) \geq 0$), the confidence intervals may be con-

structed much more precisely, if lognormal distribution $\hat{G}$ is used. In this case, the expression for a confidence interval will be:

$$\gamma_+(P) = g_0 e^{x_P u_\gamma} \tag{12}$$

where:

$$u_\gamma^2 = \ln\left[0.5\left(1 + \sqrt{1 + 4\left(s_\gamma/g_0\right)^2}\right)\right].$$

It is easy to construct confidence interval in the form (11) or (12), as the matrix (Eq. (10)) is always calculated during the search of regression parameters. However, the accuracy of this method in a nonlinear case is not enough.

2.2. Linearization (L)

In some cases, the regression model (Eq. (1)) can be transformed to a linear one. Assume that there are such maps: $\tilde{y} = h(y)$, $\tilde{\boldsymbol{a}} = \boldsymbol{b}(\boldsymbol{a})$ and $\tilde{\boldsymbol{x}} = \boldsymbol{w}(\boldsymbol{x})$, and simultaneously $h(g(\boldsymbol{b}(\boldsymbol{a}))) = \tilde{\boldsymbol{g}}^t\tilde{\boldsymbol{a}}$ and $h(f(\boldsymbol{b}(\boldsymbol{a}), \boldsymbol{w}(\boldsymbol{x})) = \tilde{\boldsymbol{f}}^t(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{a}}$. Estimation of parameters $\tilde{\boldsymbol{a}}$ is then converted to the search of a minimum of the squared functional $Q(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{a}}) = (\tilde{\boldsymbol{\eta}} - \boldsymbol{H}\tilde{\boldsymbol{a}})^t(\tilde{\boldsymbol{\eta}} - \boldsymbol{H}\tilde{\boldsymbol{a}})$, where $\tilde{\boldsymbol{\eta}} = h(\boldsymbol{\eta})$ and $\boldsymbol{H} = (\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}_1), \ldots, \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}_n))^t$. In this case, the confidence interval (5) has the form:

$$\gamma_+(P) = h^{-1}\left(\tilde{\boldsymbol{g}}^t\tilde{\boldsymbol{a}}_0 + x_P\tilde{s}_\gamma\right) \tag{13}$$

where $\tilde{\boldsymbol{a}}_0 = (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t h(\boldsymbol{\eta}_0)$, $\tilde{s}_\gamma^2 = [(\tilde{Q}(h(\boldsymbol{\eta}_0), \tilde{\boldsymbol{a}}_0))/(n-p)]\tilde{\boldsymbol{g}}^t(\boldsymbol{H}^t\boldsymbol{H})^{-1}\tilde{\boldsymbol{g}}$, x_P is P-quantile of normal distribution.

Such mode of confidence interval construction is very widespread [4]. However, it is well known that coordinate transformation of initial regression can lead to serious bias in the estimates. Below we shall see it on an example.

2.3. Bootstrap (BS)

The methods explicated above are based on the vague approximation of the distributions of estimates $\boldsymbol{\alpha}$ and forecasted function $g(\boldsymbol{a})$. A very effective method called 'bootstrap' does not have such shortages. Apparently, for the first time it was explained and applied in out-of-the-way work of Chuev et al. [5]. Later it was independently offered and justified by Efron [6,7] and Efron and Tibishirani [8]. The basic idea of this method is to simulate the distribution of estimates of parameters (6) with the help of Monte-Carlo method and then take a percentile of distribution (7) as a confidence interval.

Let $\boldsymbol{a}_0$ and s_0^2 be realizations of estimates (2) and (3), obtained in experiment, i.e., for $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. Consider random variables $\boldsymbol{\alpha}^*$ and γ^*:

$$\boldsymbol{\alpha}^* = \arg\min_a Q(\boldsymbol{\eta}^*, \boldsymbol{a}), \tag{14}$$

$$\gamma^* = g(\boldsymbol{\alpha}^*). \tag{15}$$

Here data set vector $\boldsymbol{\eta}^* = f(\boldsymbol{a}_0, \boldsymbol{x}) + \boldsymbol{\varepsilon}^*$ and error vector $\boldsymbol{\varepsilon}^*$ are normally distributed with variance s_0^2. The main assumption of this method is that the variates (14) and (15) are 'similar' to (2) and (4) correspondingly. More precisely, it is supposed that there is a monotone function $r(\cdot)$ (its form is not important), that $r(\gamma) - r(g)$ and $r(\gamma^*) - r(g_0)$ have identical zero symmetric distribution [6]. If it is right, the confidence interval for g can be obtained as a sample percentile of distribution of the variate γ^*. Practically, it means the following algorithm:

Step 0. Obtain estimates $\boldsymbol{a}_0$ and s_0^2 by initial data sample $\boldsymbol{\eta}_0$;

Step 1. Construct a normal independent error vector $\boldsymbol{\varepsilon}^* \sim N(0, s_0^2\boldsymbol{I})$ with the help of pseudorandom number generator;

Step 2. Build the BS-data sample $\boldsymbol{\eta}^* = f(\boldsymbol{a}_0, \boldsymbol{x}) + \boldsymbol{\varepsilon}^*$;

Step 3. Estimate values $\boldsymbol{\alpha}^*$ and γ^* by formulae (14) and (15);

Last step. Repeat independently steps 1–3 by N times and gain the sample of values $\gamma_1^*, \ldots, \gamma_N^*$.

Then:

$$\gamma_+(P) = P\text{-percentile}\{\gamma_1^*, \ldots, \gamma_N^*\}. \tag{16}$$

The experience shows that the BS-method gives very exact values for confidence intervals. The only shortage is a large expenditure of time for its realization. It is obvious that the principal troubles are connected with step 3. The time expended on one minimization of the objective function (14) reaches 10 s when a complicated model with large number of nonlinear parameters is fitted. For the reliable forecast, it is necessary to execute not less than $N = 1000$ recurrings that gives $t > 3$ h of work.

2.4. Free simulation (FS)

The inexactness in confidence intervals calculated by formulae (11), (12) and (13) can arise because of an abnormality of distribution (6), as well as non-linearity of function (4). Let us specify the confidence interval (11) by taking into consideration a nonlinear dependence of function $g(\boldsymbol{\alpha})$. Consider a normal vectorial variable $\boldsymbol{\alpha}^*$ and random variable $\gamma^* = g(\boldsymbol{\alpha}^*)$, which is generated by $\boldsymbol{\alpha}^*$:

$$\text{Prob}\{\boldsymbol{\alpha}^* < z\} = \Phi\left(z, \boldsymbol{a}_0, s_0^2 \boldsymbol{A}^{-1}\right), \qquad (17)$$

where Φ is defined by Eq. (9) and matrix $\boldsymbol{A}$ is defined by Eq. (10). If we accept the same assumption about the equivalence γ^* and γ, as in the BS-method, it is possible to construct a confidence interval as a percentile of the simulated sample $\gamma_1^*, \ldots, \gamma_N^*$. To obtain this sample, it is necessary to produce the realizations of normal variate $\boldsymbol{\alpha}^*$ (Eq. (17)) multiple times. The easiest way to make it is to use Choletsky's expansion [9] for the matrix $\boldsymbol{A}^{-1} = \boldsymbol{B}^2$, where $\boldsymbol{B}$ is triangular positive definite matrix. Then, $\boldsymbol{\alpha}^* = \boldsymbol{a}_0 + s_0 \boldsymbol{B}\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I})$.

Since the stage of searching of estimations is absent, the time necessary for realization of this algorithm is considerably less than at the BS-method. However, the FS-method does not give quite exact values that is connected, on our opinion, with the use of normal distribution (17) for estimates of parameters.

2.5. Associated simulation (AS)

We offer this method as the compromise between the accuracy of the BS- and the speed of the FS-methods. The main idea is that using parameters simulation instead of data simulation, we shall try to specify the form of distribution (17) by approximating it to the real one.

Carrying out our researches, we were convinced repeatedly that in spite of nonlinearity of a model, distribution of random variable:

$$c(\boldsymbol{\alpha}) = s_0^{-2}\left[Q(\boldsymbol{\eta}_0, \boldsymbol{\alpha}) - Q(\boldsymbol{\eta}_0, \boldsymbol{a}_0)\right] \qquad (18)$$

is very close to the χ^2 law with p *df*. For a linear normal regression, the statement is true exactly because $c(\boldsymbol{\alpha}) = (\boldsymbol{\alpha} - \boldsymbol{a})^t \boldsymbol{A}^{-1}(\boldsymbol{\alpha} - \boldsymbol{a})$. In nonlinear case, the validity of it may be proved by central limit theorem [10]. The following algorithm that improves the method of FS is based on this empirical fact.

Step 0. Divide the segment [0,1] on the m not overlapped intervals by points P_k, which are $0 = P_0 < \ldots < P_m = 1$. Then, with the help of function inverse to the distribution χ^2 with p *df* (P-quantile of χ^2), map this partition to semiaxis $[0,\infty] = \cup_1^m B_k$ so, that $\text{Prob}\{\chi^2 \in B_k\} = P_k - P_{k-1}$.

Step 1. Simulate the random variable $\boldsymbol{\alpha}^*$ by formula (17).

Step 2. Calculate the value $c(\boldsymbol{\alpha}^*)$ by Eq. (18) and define what 'basket' B_k this value is put in.

Step 3. If this 'basket' is already filled up to the maximum level $N(P_k - P_{k-1})$, go back to step 1, else calculate the value γ^* by Eq. (15).

Last step. Independently repeating steps 1–3 until all 'baskets' are filled up, and we obtain the simulated sample of values $\gamma_1^*, \ldots, \gamma_N^*$. Then the confidence interval is determined in accordance with relation (16).

The experiment presented below shows that the AS-method gives a rather satisfactory interval and the time expenditure for its realization is increased not more in a comparison with the FS-method.

The sense of an offered technique becomes transparent if one takes into account that the expression (18) for $c(\boldsymbol{\alpha}^*)$ determines 'the area of the indifference' of parameter estimates [3]:

$$I(C) = \{\boldsymbol{\alpha} : Q(\boldsymbol{\eta}_0, \boldsymbol{\alpha}) - Q(\boldsymbol{\eta}_0, \boldsymbol{a}_0) < C\}, \qquad (19)$$

for the likelihood function $L(\boldsymbol{\eta}, \boldsymbol{\alpha}) = \sigma^{-n}\exp[-Q(\boldsymbol{\eta}, \boldsymbol{\alpha})/2\sigma^2]$. We suppose, that for $C = s_0^2 \chi_p^2(P)$, where $\chi_p^2(P)$ is the P-quantile of the χ^2 distribution with p *df*, statement:

$$\text{Prob}\{\boldsymbol{\alpha} \in I(P)\} \geq P, \qquad (20)$$

is fulfilled with a sufficient accuracy for nonlinear models even for a small number of degrees of freedom.

3. Experimental

To compare the above-mentioned methods, we used a model example. It was rather difficult to find the system where distinction between different confi-

dence intervals were essential. We have constructed such an example using the model investigated for the prediction of service life of rubber articles [11]. The following model was used as the equation of regression (Eq. (1)):

$$f(\boldsymbol{a},\boldsymbol{x}) = 1 - \exp\left[-\left(x_1 \exp(a_2 - 1000\, a_3/x_2)\right)^{a_1}\right]. \quad (21)$$

Here factor x_1 is time (h), and factor x_2 is temperature (K). The parameter vector $\boldsymbol{a}$ has the following sense: a_1 is a scale, a_2 is a preexponential, a_3 is activation energy. This model is a result of simple scaling transformation of the common kinetics model $y = 1 - \exp[-(kt)^{a_1}]$ with well known Arrhenius dependence of constant rate $k = k_0 \exp(-E/RT)$, where $a_2 = \ln(k_0)$ and $a_3 = 0.001\ E/R$.

The data (see Fig. 1) include three series of measurements conducted for values of the factor x_2 equal accordingly to 383 K, 368 K and 353 K. The values of the factor x_1 have been varied with a step equals to 24 h. For simulation of the 'real' data, the following values of parameters of function (21) were set ('true' values):

$$a_1 = 1.5,\ a_2 = 17,\ a_3 = 8,\ \sigma^2 = 0.005. \quad (22)$$

Their estimates obtained in accordance with formulae (2) and (3), are:

$$\boldsymbol{\alpha}_1 = 1.3865,\ \boldsymbol{\alpha}_2 = 18.8578,\ \boldsymbol{\alpha}_3 = 8.6898,$$

$$s^2 = 0.0067. \quad (23)$$

All necessary calculations were carried out using integrated computer system 'Kinetic Trunk' [12] on the computer Pentium-100.

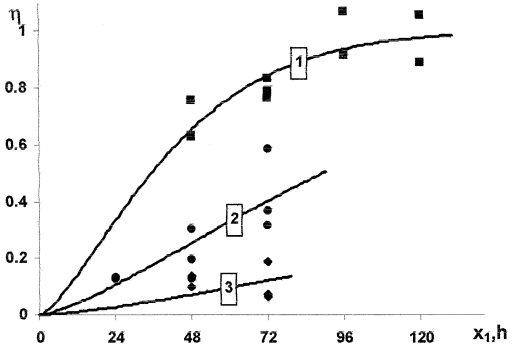

Fig. 1. Example: data and fitting. 1(■) $x_2 = 383$ K, 2(●) $x_2 = 368$ K, 3(♦) $x_2 = 353$ K.

We took the result of extrapolation of function (21) to conditions $x_1 = 8760$ h and $x_2 = 293$ K as the function g. Thereby, we evaluated the confidence interval for forecast of the response for 1 year at room temperature $g(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}, 8760, 293)$. The 'true' forecast, calculated for values (22), is:

$$g(1.5, 17, 8) = 0.1472, \quad (24)$$

and its estimate calculated for values (23) is $g_0 = 0.0877$.

How far is the extrapolation done? It is a very difficult question, since we have two different factors to be extrapolated — x_1 factor is changed from 120 to 8760 and x_2 factor is changed from 353 to 293. Moreover, all these values depend on the scale of dimensions. Our experience shows that the alteration of response value y is more important. In the example data values $\boldsymbol{\eta}_0$ lay within the interval [0.06,1.07] and the 'true' value of forecast is equal to 0.1472. That is why we can state that there is no extrapolation to the response y. We have observed all the points y along the curve $y(x_1)$ at fixed x_2 factor values and there is no extrapolation by factor x_1. In such case, we can consider the prediction as an extrapolation of the whole curve $y(x_1)$ by the temperature factor x_2 from interval [383,353] to the point $x_2 = 293$. Such issue arises when we need to predict the aging properties of polymer materials. The results of so-called 'accelerated aging tests' performed at higher temperatures (stress conditions) should be processed and extrapolated to the lower temperature (normal conditions). Of course we should trust in the model we used. We believe in it because we have observed its form up to the deep degrees of transformation at fixed temperatures, and we believe in the Arrhenius law to be used for the temperature extrapolation.

Applying all the methods determined above, the one-sided confidence intervals $\gamma_+(P)$ for probability P varying from 0.001 up to 0.999 with step 0.001 have been constructed. The greatest distinctions were observed in the area $P > 0.8$, which is plotted in Fig. 2. Besides it, the 'exact' values (points 6 in Fig. 2) of confidence interval for $P = 0.8678$, 0.9143, 0.9402, 0.9599, 0.9728, 0.9887, were determined by the method described below. Hereafter, talking about a confidence interval, we shall intend the whole curve of $\gamma_+(P)$, for all admissible values of probability P.

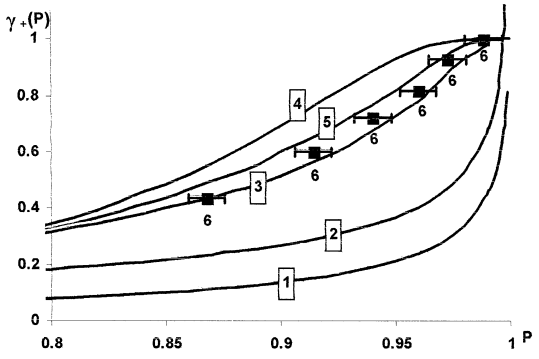

Fig. 2. Upper bounds of confidence intervals γ_+ vs. probability P for various methods: (1) the SA-method, (2) the L-method, (3) the BS-method, (4) the FS-method, (5) the AS-method, (6) (■) 'exact' values.

The function (21) is nonnegative, therefore, confidence interval of the SA method was calculated using formula (12). These values are presented in Fig. 2 (curve 1). One can see that this interval lays much below the 'exact' values.

By simple transformations $\tilde{y} = \ln[-\ln(1-y)]$, $\tilde{x}_1 = \ln x_1$, $\tilde{x}_2 = 1/x_2$, $\tilde{a}_1 = a_1$, $\tilde{a}_2 = -a_1 a_3$, $\tilde{a}_3 = a_1 a_2$, it is possible to linearize function (21) and reduce it to the form $\tilde{f}(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{x}}) = \tilde{a}_1 \tilde{x}_1 + \tilde{a}_2 \tilde{x}_2 + \tilde{a}_3$. However, it is necessary to sacrifice two points $\boldsymbol{\eta}(96,383) = 1.0698$ and $\boldsymbol{\eta}(120,383) = 1.0584$, which are over $\boldsymbol{\eta} = 1$ (see Fig. 1). The confidence interval constructed by formula (13) is already above (curve 2 in Fig. 2), but still far from points 6.

For realization of the BS procedure, we accepted estimates $\boldsymbol{a}_0$ and s_0^2 (Eq. (23)) as initial values and have conducted $N = 5000$ recurrings. The obtained confidence interval (curve 3) is wonderfully agreed with the 'exact' values. However, the expenditures of time for realization of this algorithm were about 2 h that seems to us unacceptable for such a simple regression problem.

The method of FS conducted with $N = 5000$ recurrings has brought the results (curve 4 in Fig. 2) that are overstated in correspondence with 'exact' values. It is connected, apparently, with the distribution (17) that is too wide in a comparison with true distribution. Expenditure of time on a realization of the algorithm was about 20 s.

For realization of the method of AS, the following values were specified. Number of recurrings was $N = 5000$; number of 'baskets' was $m = 28$. The first nine intervals for construction the 'baskets' were taken by size 0.1 (from 0 up to 0.9 — volume 500); the next nine intervals had size 0.01 (from 0.9 up to 0.99 — volume 50); the last 10 intervals had size 0.001 (from 0.99 up to 1.0 — volume 5). The result is represented by the curve 5 in Fig. 2. It differs from the 'exact' values, but it does not seem essential to us, since the distinction is not great and has the right direction (interval is overestimated). Expenditure of time was about 25 s.

The quality of the χ^2-distribution simulation was checked together with construction of confidence interval by the AS-algorithm. Besides it using an independent series of simulation (simultaneous with the BS procedure), the supposition that variate (18) is subject to the χ^2 law with $p = 3$ df was verified. In Fig. 3, the following values are represented:

$$\delta_1 = \hat{c}(P) - \chi_3^2(P) \text{ (curve 1) and}$$

$$\delta_2 = \tilde{c}(P) - \chi_3^2(P) \text{ (curve 2)}, \qquad (25)$$

where χ_3^2 is P-quantile of χ^2 distribution with 3 df, $\hat{c}(P)$ is sample percentile of variate (18) in the BS-algorithm and $\tilde{c}(P)$ sample percentile of variate (18) in the AS-algorithm. The smallness of deviation δ_1 characterizes a conformity of our hypothesis (20) and the smallness of deviation δ_2 characterizes the quality of reproduction of χ^2 distribution in the AS. From Fig. 3, one can see that for $P > 0.75$, ranges of deviations δ_1 begin to increase. On the other hand, the

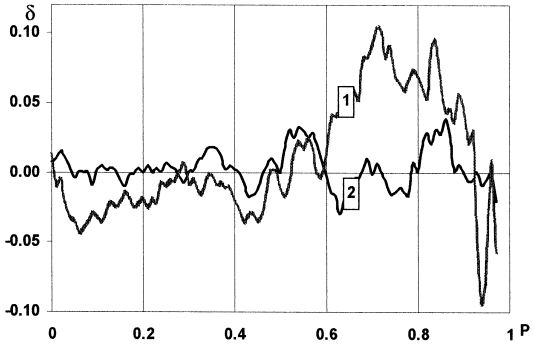

Fig. 3. Deviations of sample percentiles (Eq. (25)) of values (Eq. (20)) from quantiles $\chi_3^2(P)$ for: 1 — the BS-method, 2 — the AS-method.

Table 1
Number of failures in checking the confidence intervals constructed by methods: SA — stochastic approximation, AS — associated simulation

P	Method		Should be			P_1	$\gamma_+^0(P_1)$
	SA	AS	Min	Mean	Max		
0.99	182	35	21	31	40	0.989	0.995
0.97	279	84	77	93	108	0.973	0.926
0.95	344	124	134	155	174	0.960	0.817
0.93	410	185	193	217	239	0.940	0.720
0.9	470	265	281	309	336	0.914	0.598
0.85	569	409	431	464	496	0.868	0.434

Total number of attempts is 3093.

deviations δ_1 are small, therefore, the shape of distribution of variate (18) is close to χ_3^2. However, the sample estimate of the variance of $\tilde{\chi}_3^2$ distribution, obtained by the AS-method is equal to 6.3705, whereas theoretically, it should be 6. Such difference ($N = 5000$) is essential for the significance equals to 0.001 and testifies that performing the idea of 'baskets' in the AS-algorithm, we construct the wider distribution that is required. It agrees with data in Fig. 2.

In Fig. 2 (the points 6), 'exact' boundaries of the confidence interval are represented and their confidence regions (horizontal lines) are indicated. We have received these data using the following algorithm.

Step 0. Create data set $\boldsymbol{\eta} = f(\boldsymbol{a}, \boldsymbol{x}) + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$ using parameters $\boldsymbol{a}$ and σ^2 equal to 'true' values (22);

Step 1. Find estimates $\boldsymbol{\alpha}$ and s^2 by formulae (2) and (3);

Step 2. Construct confidence intervals $\gamma_+(P)$ by any method;

Step 3. Compare obtained values with a 'true' value of g calculated by Eq. (24);

Last Step. Independently repeating steps 1–3 (in an amount of N recurrings), obtain number of failures (cases, when $\gamma_+(P) < g$) for each value of P.

We have applied this algorithm for methods of SA and AS at $P = 0.99, 0.97, 0.95, 0.93, 0.9, 0.85$ and $N = 3093$. The outcomes are represented in Table 1. In the first column, there are the values of confidence probability P, in the 2nd and the 3rd are numbers of failures for the SA- and the AS-methods. In the 4th column, there is a minimum, in the 5th is a mean, and in the 6th is a maximum of theoretical values of failures in correspondence with binomial law for significance level 0.05. In the 7th column, the evaluation of real confidence probability P_1 is presented. It was calculated as:

$$P_1(P) = 1 - N_1(P)/N$$

where $N_1(P)$ is a number of failures for the AS-method (the 3rd column). Then the 'exact' values of confidence intervals for these probabilities (the last column of Table 1) were determined by equation:

$$\gamma_+^0(P_1) = \gamma_+(P(P_1)), \tag{26}$$

where $\gamma_+(\cdot)$ is the appropriate boundary of the confidence interval obtained by the AS-method. These values (26) are presented in Fig. 2 in points 6.

4. Conclusions

In this paper, the various methods of confidence intervals construction for nonlinear regression were considered. The new method named by a method of AS is proposed. It is shown on the example that only the BS and the AS method give a satisfactory accuracy. Advantage of the AS-method is the speed. In a comparison with the BS, the prize is at least 10 000 times. The substantiation of the AS-method (as well as other nonlinear regression methods) is asymptotic. This method may be applied when parameter estimators are obtained by the MLM. The rather high accuracy (in comparison with a method of SA) may be explained, in our opinion, by successful choice of ranging statistics (18), which distribution quickly converges to χ^2 even in a hardly nonlinear case. We consider that the further refinement of the AS-method is possible. So, in particular, we assume, that if in Eq. (18) we use MLM-estimator [4] of variance $s^2 = Q(\boldsymbol{\eta}, \boldsymbol{\alpha})/n$ instead of estimator (3), the exactitude of the interval will be improved.

References

[1] S.D. Brown, S.T. Sum, F. Despagne, B.K. Lavine, Anal. Chem. 68 (12) (1996) 21R–61R.

[2] A.L. Pomerantsev, O.N. Karpukhin, E.F. Brin, Kinet. Katal. 24 (5) (1983) 1233–1238.
[3] Y. Bard, Nonlinear Parameter Estimation, Academic Press, New York, 1974.
[4] N.R. Draper, H. Smith, Applied Regression Analysis, Wiley, New York, 1981.
[5] Yu.V. Chuev, Yu.B. Mikhailov, V.I. Kuzmin, Prediction of Quantitative Process Datum, Sov. Radio, Moscow, 1975, in Russian.
[6] B. Efron, Ann. Stat. 7 (1) (1979) 1–26.
[7] B. Efron, J. Am. Stat. Assoc. 81 (1987) .
[8] B. Efron, R. Tibishirani, Statistical Science 1 (1) (1986) 54–75.
[9] L. Fox, An Introduction to Numerical Linear Algebra, Oxford Univ. Press, Clarendon, 1964.
[10] J.R. Barra, Notions Fondamentales de Statistique Mathematique, Dunod, Paris, 1971.
[11] Ye.V. Bysrtitskaya, O.N. Karpukhin, G.A. Maksimova, in: Proc. of Int. Rubber Conf., Vol. 3, Moscow, 1994, pp. 605–612, in Russian.
[12] A.L. Pomerantsev, in: Proc. of Int. Rubber Conf., Vol. 4, Moscow, 1994, pp. 291–297, in Russian.