Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2023) 537-544

ECCOMAS
Proceedia

UNCECOMP 2023
5th ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Science and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Athens, Greece, 12-14 June 2023

EFFECT OF RANDOM PARAMETERS IN NONLINEAR REGRESSION ON THE OPTIMAL EXPERIMENTAL DESIGN (UNCECOMP 2023)

D. Jarušková

Czech Technical University, Faculty of Civil Engineering
Thákurova 7, CZ 166 29 Praha 6, Czech Republic
daniela.jaruskova@cvut.cz

Abstract. *We consider a nonlinear regression model that describes a relationship between input parameters and an output of an experiment. The output is measured repeatedly in several time points. The regression function is supposed to contain additional random parameters that remain the same in a single experiment but differ from one experiment to the other. Due to the additional random parameters the variability of least squares estimates of the parameters of interest may be large and their distribution may be not normal. For choosing a good design of experiment we might be interested in the complete joint distribution of the LS estimates. We compare three methods for approximating this distribution and illustrate the methods by some examples.*

Keywords: Nonlinear regression, LSE of parameters of interest, additional random parameter, approximate probability distribution of parameters of interest.

ISSN:2623-3339
Peer-review under responsibility of the organizing committee of UNCECOMP 2023.
doi: 10.7712/120223.10356.19596

1 INTRODUCTION

For many experiments a model can be found that sufficiently well describes a relationship between input variables and an output of an experiment. We assume that measurements of the output are performed repeatedly in times $\{t_i, i = 1, \ldots, n\}$ and the relationship between the inputs and the output at time t_i is modeled by a known function f that contains a vector of unknown parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, i.e. we denote $f_i(\boldsymbol{\beta}) = f(t_i, \boldsymbol{\beta})$. If the functions $\{f_i(\boldsymbol{\beta})\}$ are nonlinear with respect to $\boldsymbol{\beta}$ and if the outputs $\{Y_i, i = 1, \ldots, n\}$ are measured with random errors $\{e_i, i = 1, \ldots, n\}$ that are supposed to be additive, independent and identically distributed with zero mean and a variance σ_m^2 we deal with a nonlinear regression model.

In our contribution we assume that the outputs may be affected by an additional vector random parameter $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)$, independent of $\{e_i\}$, that might express for instance some material parameters. The distribution of $\boldsymbol{\gamma}$ is supposed to be known, here it is normal $\mathcal{N}(\boldsymbol{\gamma}^0, \Sigma_\gamma)$.

The goal of statistical inference is to estimate a true value $\boldsymbol{\beta}^*$ of the vector parameter of interest $\boldsymbol{\beta}$ in the model

$$Y_i = f_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) + e_i, \quad i = 1, \ldots, n. \tag{1}$$

We consider here the least squares estimate:

$$\widehat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta}} \sum \big(Y_i - f_i(\boldsymbol{\beta}, \boldsymbol{\gamma}^0)\big)^2. \tag{2}$$

Indeed, the statistical inference should not stop at obtaining a point estimate but go on by presenting $(1-\alpha)100\%$ confidence regions that express an accuracy of the estimate. Alternatively, the probability that $|\widehat{\beta}_j - \beta_j^*| \le \Delta_j, j = 1, \ldots, p$ for some given $\{\Delta_j\}$ may also be of interest. For constructing confidence regions we need to know the probability distribution of $\widehat{\boldsymbol{\beta}}$, or at least to know its good approximation. The distribution of $\{Y_i\}$ is affected not only by random behavior of error measurements $\{e_i\}$ but also by the additional random parameter that causes a correlation between them. Indeed, the same is true for a distribution of $\widehat{\boldsymbol{\beta}}$. Especially, it is important to know at least an approximate distribution of $\widehat{\boldsymbol{\beta}}$ when several models indexed by a design parameter d are compared with the aim to find a model that enables to estimate the parameter $\boldsymbol{\beta}^*$ with the best possible accuracy.

In the design experiment problems a decision for selecting an optimal model is based on variances of coordinates of $\widehat{\boldsymbol{\beta}}$, or on a function of its variance-covariance matrix (as D-optimal design, A-optimal design etc.), see [1]. The idea for using these criteria comes from an assumption that the distribution of the estimate $\widehat{\boldsymbol{\beta}}$ is approximately normal. This is indeed true, when n is large and the variances $Var\gamma_j, j = 1, \ldots, k$ as well as the variance σ_m^2 of the measurement errors $\{e_i\}$ are all small and the regression functions are only slightly nonlinear with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Unfortunately, this is not necessarily true as it is illustrated by the following example when a random parameter is present.

Example 1.

The following nonlinear regression model is considered

$$Y_i = (1-d)\log\beta + d\beta + (1 + a \cdot d)\gamma + e_i, i = 1, \ldots, n,$$

where $\{e_i\}$ are i.i.d. with $\mathcal{N}(0, \sigma_m^2)$ and γ is distributed according to $\mathcal{N}(0, \sigma_\gamma^2)$, $a > 0$. We are to decide whether to estimate the one-dimensional parameter β from a nonlinear model $(d = 0)$ with a smaller random noise:

$$Y_i = \log\beta + \gamma + e_i, i = 1, \ldots, n,$$

or from a linear model $(d = 1)$ with a larger random noise:

$$Y_i = \beta + (1 + a)\gamma + e_i, i = 1, \ldots, n.$$

In this simple case we can find an exact distribution of least squares estimates in both models. For $d = 0$ it is a two-parameters log-normal distribution:

$$\widehat{\beta}_0 = \exp\left(\sum Y_i/n\right) \sim \mathcal{LN}(\log \beta^*, \sigma^2 + \sigma_m^2/n),$$

while for $d = 1$ it is normal

$$\widehat{\beta}_1 = \sum Y_i/n \sim \mathcal{N}(\beta^*, \sigma^2(1 + a)^2 + \sigma_m^2/n).$$

Suppose that $\beta^* = 1.1$, $\sigma_\gamma^2 = (1/3.3)^2$, $\sigma_m^2 = 0.01^2$, $a = 0.178$ and $n = 10$, then the variances of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the same being 0.127. The probability density functions of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are plotted in Figure 1.

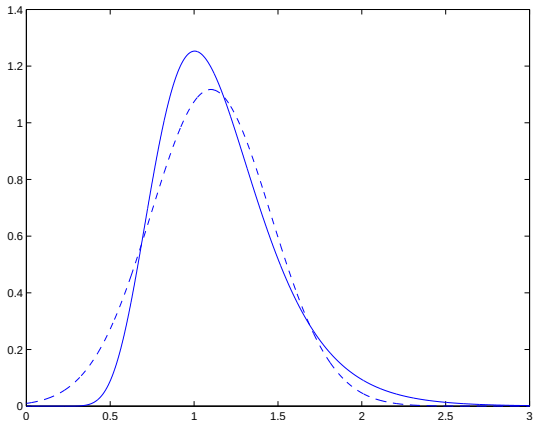

Figure 1: Probability density function for LSE under two considered models.

When we calculate probabilities that the estimate belongs to a given interval, we get some confusing results

$$P(1.1 - 0.4 \leq \widehat{\beta}_0 \leq 1.1 + 0.4) = 0.78,$$
$$P(1.1 - 0.4 \leq \widehat{\beta}_1 \leq 1.1 + 0.4) = 0.74,$$

while

$$P(1.1 - 0.9 \leq \widehat{\beta}_0 \leq 1.1 + 0.9) = 0.97,$$
$$P(1.1 - 0.9 \leq \widehat{\beta}_1 \leq 1.1 + 0.9) = 0.99.$$

2 Three methods for approximating a distribution of $\widehat{\boldsymbol{\beta}}$

Our paper presents three methods for obtaining an approximate distribution of $\widehat{\boldsymbol{\beta}}$. The more detailed description may be found in [2].

It is well known that the distribution of an estimate $\widehat{\boldsymbol{\beta}}$ depends on a true value of $\boldsymbol{\beta}^*$. This means that the optimal design depends on a true value of parameter as well. In practice, the

design is chosen in two steps. First, $\boldsymbol{\beta}^*$ is estimated only roughly and then, the design is selected with respect to this estimate. The procedure may be repeated.

In all presented examples the regression functions are supposed to be linear with respect to the additional random parameter:

$$f_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = g_{i1}(\boldsymbol{\beta})\gamma_1 + \cdots + g_{ik}(\boldsymbol{\beta})\gamma_k.$$

This is indeed a strong assumption. In case this assumption is not valid we may replace the original regression function by its linear approximation at $\boldsymbol{\gamma}^0$. It works quite well when all variances $Var\gamma_j$ are not too big and the nonlinearity of regression functions with respect to the coordinates of $\boldsymbol{\gamma}$ is small.

Method based on linearization with respect to parameter of interest

The method is based on the linearization of the regression functions at the true value $\boldsymbol{\beta}^*$.The approximate distribution is normal with a mean $\boldsymbol{\beta}^*$ and a variance-covariance matrix

$$(\boldsymbol{F}^{*T}\boldsymbol{F}^*)^{-1}\boldsymbol{F}^{*T}\boldsymbol{\Sigma}\boldsymbol{F}^*(\boldsymbol{F}^{*T}\boldsymbol{F}^*)^{-1},$$

with $\boldsymbol{F}^* = ||\frac{\partial f_i}{\partial \beta_l}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^0)||_{i=1,l=1}^{n\ p}$, $\boldsymbol{\Sigma} = \boldsymbol{G}\Sigma_\gamma\boldsymbol{G}^T + \sigma_m^2\boldsymbol{I}$, where $\boldsymbol{G} = ||g_{ij}(\boldsymbol{\beta}^*)||_{i=1,j=1}^{n\ k}$, provided the parameter space of $\boldsymbol{\beta}$ is $\mathcal{R}^k$. If the parameter space is bounded, we obtain a trimmed normal distribution. The method is very fast. It works very well in many situations but it may fail sometimes as we will illustrate later.

Monte Carlo simulations method

We repeatedly generate realizations of $\boldsymbol{\gamma}$ and $\{e_i\}$ from their distributions and numerically find arguments minimizing least squares function. The quality of this method depends highly on our ability to find a true minimum. For iterative numerical method we have to choose a starting value. The choice of the starting value may affect the obtained distribution. The method is very time-consuming.

Finite sample approximation

The method has been derived in [3] using a theory of projection in differential geometry. Motivated by problems coming from nonlinear regression with additional random parameter, the results were generalized in [4].

Under the assumption that measurements errors have a normal distribution the method provides us with an approximate probability density function in the form:

$$h_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{\beta}) = \frac{det\,\boldsymbol{Q}(\boldsymbol{\beta}, \boldsymbol{\beta}^*)}{(2\pi)^{p/2}\big(det\big(\boldsymbol{M}(\boldsymbol{\beta})\big)\big)^{1/2}} \exp\Big\{ -\frac{1}{2}\big(\boldsymbol{f}(\boldsymbol{\beta}) - \boldsymbol{f}(\boldsymbol{\beta}^*)\big)^T\boldsymbol{A}(\boldsymbol{\beta})\big(\boldsymbol{f}(\boldsymbol{\beta}) - \boldsymbol{f}(\boldsymbol{\beta}^*)\big)\Big\}, \quad (3)$$

where

$$\begin{aligned}
\boldsymbol{M}(\boldsymbol{\beta}) &= \boldsymbol{F}(\boldsymbol{\beta})^T\boldsymbol{\Sigma}\boldsymbol{F}(\boldsymbol{\beta}),\\
\boldsymbol{A}(\boldsymbol{\beta}) &= \boldsymbol{F}(\boldsymbol{\beta})\big(\boldsymbol{M}(\boldsymbol{\beta})\big)^{-1}\boldsymbol{F}(\boldsymbol{\beta})^T,\\
\boldsymbol{P}(\boldsymbol{\beta}) &= \boldsymbol{\Sigma}\boldsymbol{F}(\boldsymbol{\beta})\big(\boldsymbol{M}(\boldsymbol{\beta})\big)^{-1}\boldsymbol{F}(\boldsymbol{\beta})^T,\\
\boldsymbol{Q}(\boldsymbol{\beta}, \boldsymbol{\beta}^*) &= \boldsymbol{F}(\boldsymbol{\beta})^T\boldsymbol{F}(\boldsymbol{\beta}) + \big(\boldsymbol{f}(\boldsymbol{\beta}) - \boldsymbol{f}(\boldsymbol{\beta}^*)\big)^T\big(\boldsymbol{I} - \boldsymbol{P}(\boldsymbol{\beta})\big)^T\boldsymbol{H}(\boldsymbol{\beta})
\end{aligned}$$

and $\boldsymbol{H}(\boldsymbol{\beta}) = ||\frac{\partial^2 f_i}{\partial\beta_j\partial\beta_\kappa}||_{i=1,j=1,\kappa=1}^{n\ p\ p}$ is a three-dimensional array.

The approximate density is calculated on a grid in the parameter space. It is relatively fast when $p \leq 2$ but it is time demanding for $p > 2$.

In many situations the results of all three methods coincide but there are examples when the first method fails as in the following example.

Example 2.

We suppose that an experiment consists in obtaining two observations that fulfill:

$$Y_1 = \cos(\beta) + \gamma + e_1,$$
$$Y_2 = \sin(\beta) + e_2.$$

Note that since $\sin\beta = \cos(\beta - \pi/2)$ the model is in the form (1) for $t_1 = 0$ and $t_2 = \pi/2$. The parameter space for β is $[0, \pi]$. The additional random parameter has a normal distribution $\mathcal{N}(0, \sigma_\gamma^2 = 0.9^2)$ and the measurements errors e_1 and e_2 are i.i.d. with a normal distribution $\mathcal{N}(0, \sigma_m^2 = 0.1^2)$. Suppose that the true value of β is $\beta^* = \pi/2$ so that the observations Y_1 and Y_2 satisfy the model $Y_1 = \gamma + e_1$, $Y_2 = 1 + e_2$. Figure 2 presents generated data from this model together with the corresponding regression functions path. The LS estimate of β is the point on the path that is closest to (Y_1, Y_2).

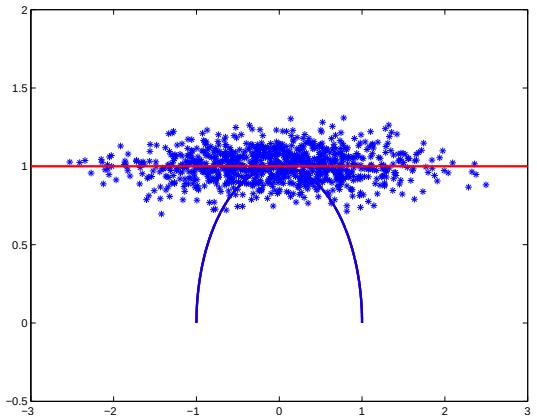

Figure 2: Generated data from the true model together with corresponding regression functions path.

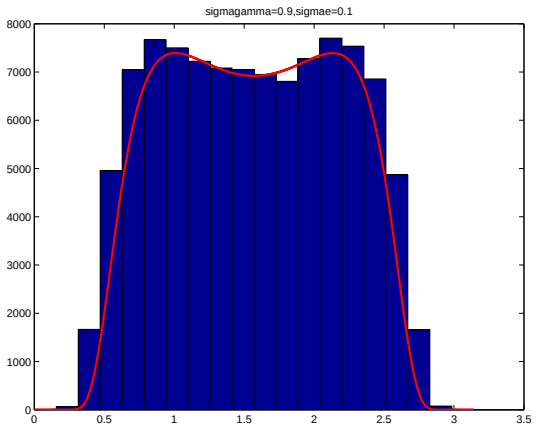

Figure 3: Histogram of estimates of parameter of interest calculated from generated data (4) together with density function (3).

For this example we may find the exact solution being

$$\widehat{\beta} = \text{atan}\Big(\frac{Y_2}{Y_1}\Big), \quad \text{for} \quad Y_1 > 0, \qquad \widehat{\beta} = \pi - \text{atan}\Big(-\frac{Y_2}{Y_1}\Big), \quad \text{for} \quad Y_1 < 0 \tag{4}$$

and generate values $\widehat{\beta}$. Figure 3 shows a histogram of estimates calculated from generated data using (4) together with a density (3). The first method fails here completely as after the linearization with respect to β at its true value $\pi/2$ the variables (Y_1, Y_2) satisfy:

$$Y_1 = -\big(\beta - \tfrac{\pi}{2}\big) + \gamma + e_1,$$
$$Y_2 = 1 + e_2.$$

Denote the LS estimate in this model by $\tilde{\beta}$. As $\tilde{\beta} = \frac{\pi}{2} - Y_1$ for $-\frac{\pi}{2} \le Y_1 \le \frac{\pi}{2}$, and $\tilde{\beta} = 0$ for $Y_1 > \frac{\pi}{2}$, $\tilde{\beta} = \pi$ for $Y_1 < -\frac{\pi}{2}$. The distribution of $\tilde{\beta}$ is a trimmed normal distribution with a probability $P(\tilde{\beta} = 0) = P(\tilde{\beta} = \pi) = 0.04$. Figure 4 presents generated values from this distribution.

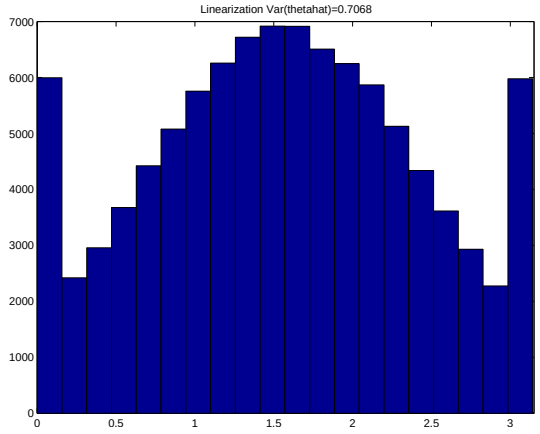

Figure 4: Histogram of LS estimates in the model obtained by linearization.

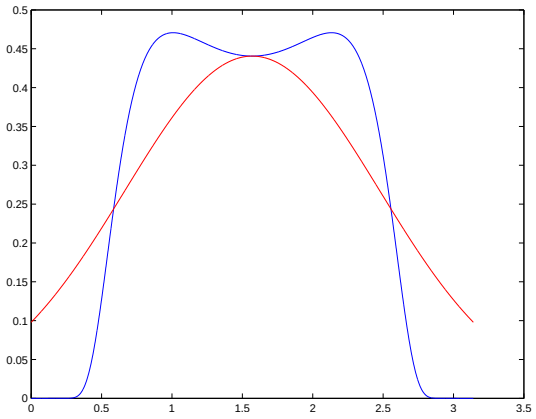

Figure 5: Density of estimates obtained by linearization and density calculated by (3).

The density function of the trimmed normal distribution and the density function obtained by the third method are plotted in Figure 5. The discrepancy between these two is obvious. The density function obtained by a third method approximate the true density function very well.

The following example illustrates that our decision how to select an optimal design might be affected by our knowledge of the probability distribution of parameter estimate.

Example 3.

In an example introduced by [5] and studied by [6] we have encounter an event that demonstrates how important is to approximate a probability distribution of a parameter estimate for selecting a reasonably good design of an experiment.

The aim of the model described in [5] is to estimate thermal properties of a material, i.e. thermal conductivities λ_x, λ_y and a volumetric capacity C, from an experiment when a square sample $[0, 0.05] \times [0, 0.05]$ (m^2) is exposed to a constant and uniform heat flux φ on the left and bottom boundaries while the right and top edges are insulated. It is supposed that measurements are taken at equidistant times by one sensor whose position is to determined.

In [5] it was suggested that the regression function relating temperature to time and a sensor position may be given as follows:

$$T(t, x, y; \lambda_x, \lambda_y, C; \varphi) = \theta_x(t; \lambda_x; C; \varphi; x) + \theta_y(t; \lambda_y; C; \varphi; y),$$

$$\theta_x(t; \lambda_x; C; \varphi; x) = \frac{2\varphi}{\sqrt{C\lambda_x}}\sqrt{t}F\left(\frac{\tilde{x}}{\sqrt{t}}\right),$$
$$\theta_y(t; \lambda_y; C; \varphi; y) = \frac{2\varphi}{\sqrt{C\lambda_y}}\sqrt{t}F\left(\frac{\tilde{y}}{\sqrt{t}}\right)$$

with $\widetilde{x} = (x/2)\sqrt{C/\lambda_x}$, $\widetilde{y} = (y/2)\sqrt{C/\lambda_y}$ and

$$F(z) = \frac{\exp(-z^2)}{\sqrt{\pi}} - z\big(1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-v^2}\, dv\big), \quad z \geq 0.$$

The heat flux φ was supposed to be an additional random parameter distributed according to a normal distribution $\mathcal{N}(25\,000, 100^2)$ (Wm^{-2}), while λ_x, λ_y and C were to be estimated. In our simplified version $\lambda_y = 4.7$ $(Wm^{-1}K^{-1})$ and $C = 1700000$ $(Jm^{-3}K^{-1})$ are also known and the only estimated parameter is λ_x. Moreover, we were looking for an optimal position of a sensor on a grid in the bottom boundary with a distance 0.0001 (m) between two neighboring points. The bottom boundary was chosen because the first and the third method applied to a sparse grid in the entire square sample suggested that the bottom boundary is a region when an optimal position should be looking for.

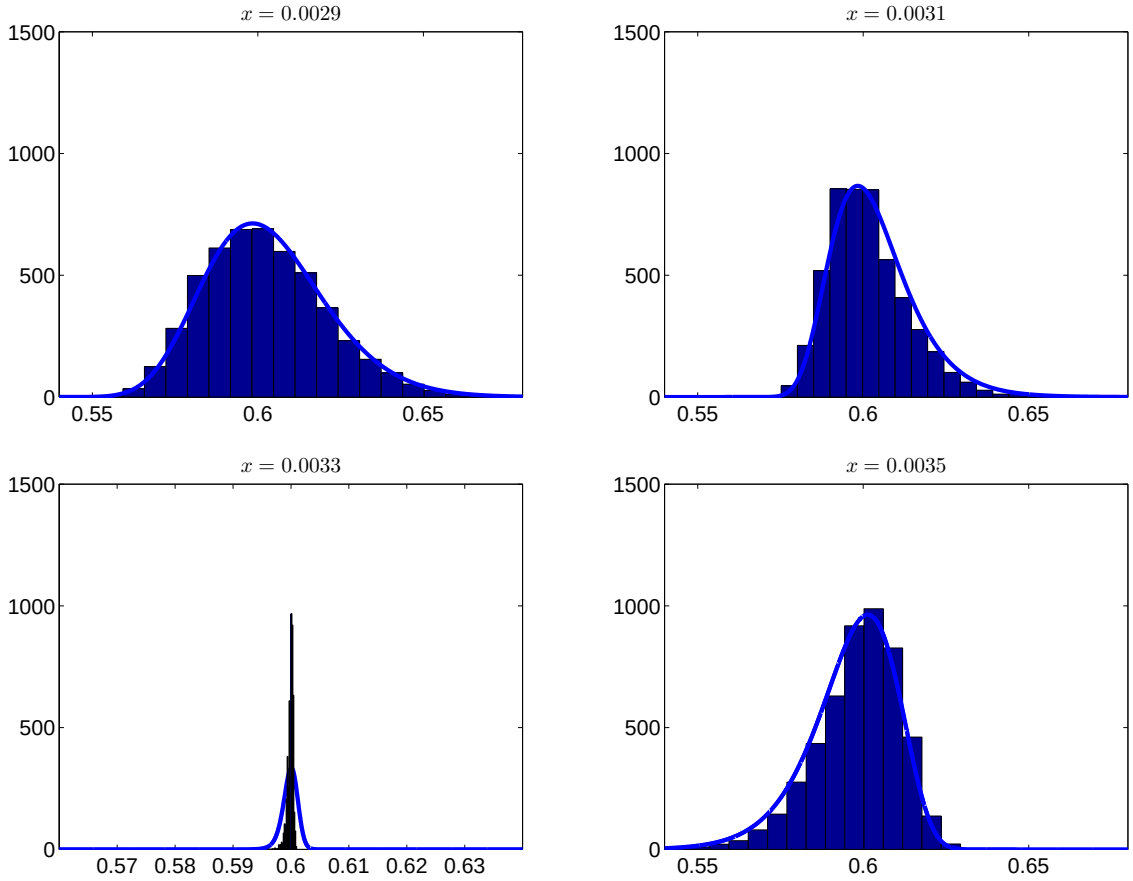

Figure 6: Histogram of LS estimates of $\widehat{\lambda_x}$ obtained by the second method and corresponding approximate densities obtained by the third method for several positions of a sensor.

We started with computing variances for all grid points of the bottom boundary using all three suggested methods. Surprisingly, all methods provided us with variances that were close to each other. Applying a criterion of smallest variance all methods decided that the point $x = 0.0033$ (m) was the best position. Then, we studied the probability distribution of the estimate $\widehat{\lambda}_x$ not only for the optimal position but also in its neighborhood. The density function corresponding to an optimal position was almost symmetric but when the sensor was shifted only negligibly,

i.e. 2 mm to the right or to the left, the density of the estimate became extremely skewed as it could be seen in Figure 6. This effect was clearly detected by the second as well as by the third method. The position $x = 0.0033$ (m) seems to be unstable and to put a sensor there is risky. Indeed, this information cannot be obtained when the entire probability distributions are not studied and only the variance is considered.

3 Conclusions

An additional random parameter may cause that a least squares estimate of the parameter of interest may be not normal. The reason is that due to its presence there exists a correlation between the subsequent observations as was pointed out in [6]. We get much better information on the required estimate behavior if we are able to approximate its distribution. When parameter is one or two-dimensional we may use two suggested method. Unfortunately, when the parameter has more than two-dimensional the suggested methods are time-consuming.

REFERENCES

[1] L. Pronzato L., A. Pázman, *Design of experiments in nonlinear models*, New York: Springer, 2013.

[2] D. Jarušková, A. Pázman, Methods for approximating distribution of unknown parameter estimates with application in material thermophysics. *International Journal for Uncertainty Quantification*, **11**, 31–47, 2021.

[3] A. Pázman, *Nonlinear statistical models*. Kluwer Academic Publishers, 1993.

[4] A. Pázman, Distribution of multivariate nonlinear LS estimator under an uncertain input. *Statistical Papers*, **60**, 179–194, 2019.

[5] E. Ruffio, D. Saury, D. Petit, Robust experiment design for the estimation of thermophysical parameters using stochastic algorithms. *Int. Heat Mass Transfer*, **55**, 2901-2915, 2012.

[6] D. Jarušková, A. Kučerová, Estimation of thermophysical parameters revisited from the point of view of nonlinear regression with random parameters. *Int. Heat Mass Transfer*, **106**, 135–141, 2017.