Effect of random parameters in nonlinear regression on the optimal experimental design

Daniela JARUŠKOVÁ

Czech Technical University in Prague

UNCECOMP 2023

Nonlinear regression with an additional random parameter.

$$Y_i = f(i/n; \boldsymbol{\beta}, \boldsymbol{\gamma}) + e_i, \; i = 1, \ldots, n.$$

$Y_1, \ldots, Y_n$ data
$\boldsymbol{\beta}$ is a parameter of interest ($\boldsymbol{\beta}^*$ its true value)
$\boldsymbol{\gamma} \sim N(\boldsymbol{\gamma}^0, \boldsymbol{\Sigma}_\gamma)$ is an additional random parameter
$\{e_i\}$ are i.i.d. $N(0, \sigma_m^2)$ measurements errors.

Problem with an additional random parameter

Linear models with respect to γ

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

$$\begin{aligned} Y_i = {} & g_{i1}(\boldsymbol{\beta}^*)\gamma_1^0 + \cdots + g_{ik}(\boldsymbol{\beta}^*)\gamma_k^0 + \\ & + g_{i1}(\boldsymbol{\beta}^*)(\gamma_1 - \gamma_1^0) + \cdots + g_{ik}(\boldsymbol{\beta}^*)(\gamma_k - \gamma_k^0) + e_i, \end{aligned}$$

$$\operatorname{Var}\boldsymbol{Y} = \boldsymbol{\Sigma} = \boldsymbol{G}\boldsymbol{\Sigma}_\gamma \boldsymbol{G}^T + \sigma_m^2 \boldsymbol{I}, \boldsymbol{G} = ||g_{ij}(\boldsymbol{\beta}^*)||_{i=1,j=1}^{n,k}.$$

Ordinary least squares method - frequentist approach

Estimate $\boldsymbol{\beta}$ by the least squares metod:

$$\widehat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(Y_i - f(i/n;\, \boldsymbol{\beta}, \boldsymbol{\gamma}^0) \right)^2.$$

Further, I denote $f_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = f(i/n;\, \boldsymbol{\beta}, \boldsymbol{\gamma})$.

The estimate $\widehat{\boldsymbol{\beta}}$ is a random variable (vector) due to measurement errors and additional random parameter(s). Fluctuation around the true value $\boldsymbol{\beta}^*$ is given by its distribution. We may use its variance (or its multidimensional analogue in case of multidimensional $\boldsymbol{\beta}$) to express an accuracy of the estimate $\widehat{\boldsymbol{\beta}}$. However, if the distribution is skewed or multimodal, we may ask if the variance is the best characteristic. This very important when we compare several models (Design of experiment).

Example 1

$$Y_i = (1-d)\,\log(\theta) + d\,\theta + (1+ad)\gamma + e_i,\ i = 1, \ldots, n.$$

$$\begin{aligned} d = 0, &\quad Y_i = \log\theta + \gamma + e_i,\ i = 1, \ldots, n, \\ d = 1, &\quad Y_i = \theta + (1+a)\gamma + e_i,\ i = 1, \ldots, n. \end{aligned}$$

$$Var\,\gamma = \sigma_\gamma = (1/3.3)^2,\ Var\,e_i = \sigma_e^2 = 0.01^2,\ \theta^* = 1.1$$

$d = 0$

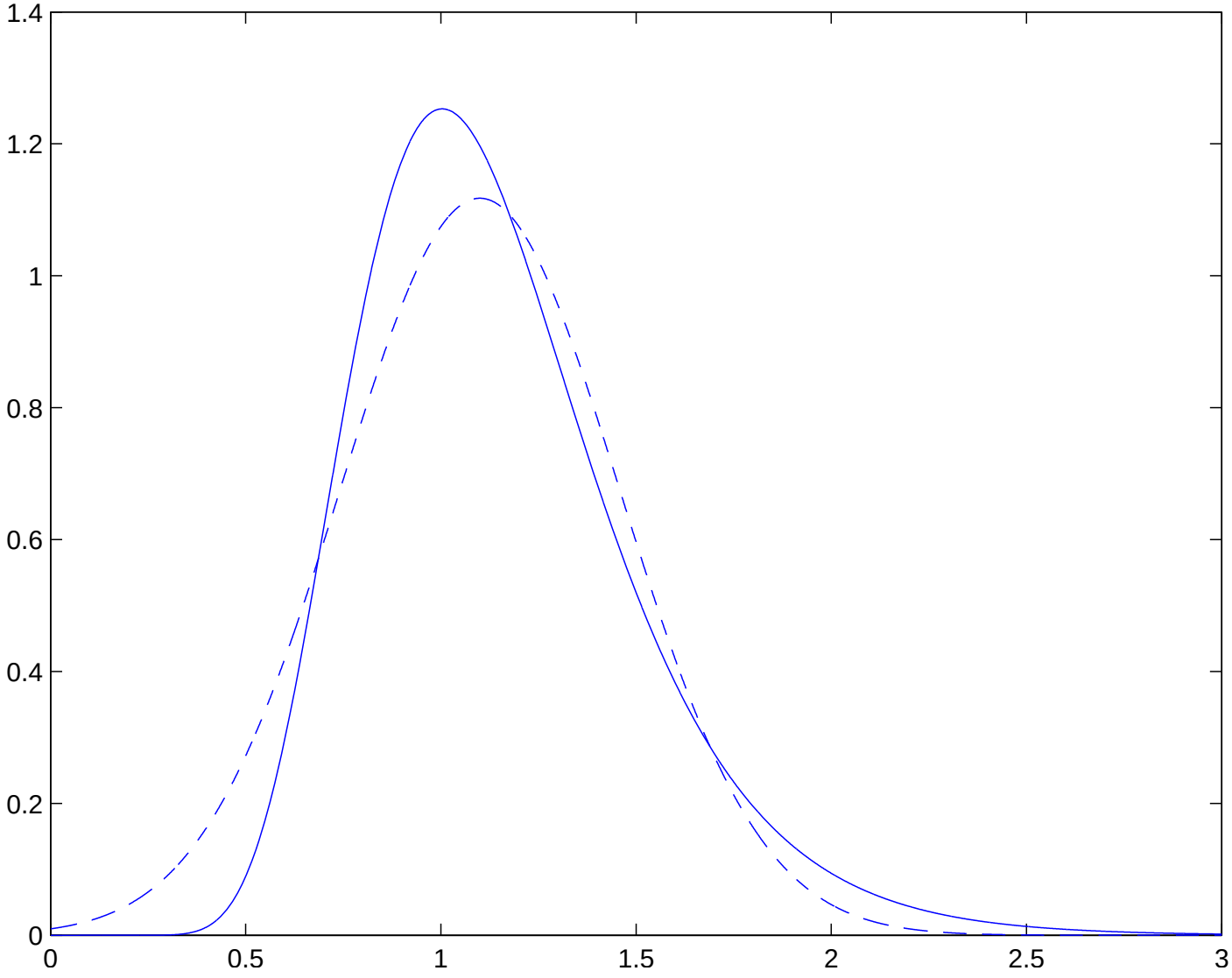
$$\widehat{\theta} = \exp\left(\frac{1}{n}\sum Y_i\right) \sim LN(\log 1.1, A = \sigma_\gamma^2 + \sigma_e^2/n)$$

$$n = 10, \quad E\,\widehat{\theta} = 1.15,\ Var\,\widehat{\theta} = 0.127$$

$d = 1,\ a = 0.178$

$$\widehat{\theta} = \frac{1}{n}\sum Y_i \sim N(1.1, A_a = \sigma_\gamma^2\,(1+a)^2 + \sigma_e^2/n)$$

$$n = 10, \quad E\,\widehat{\theta} = 1.1,\ Var\,\widehat{\theta} = 0.127$$


1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
2.5
3

$$P(1.1 - 0.4 \leq \widehat{\theta}_{LN} \leq 1.1 + 0.4) = 0.78,$$
$$P(1.1 - 0.4 \leq \widehat{\theta}_{N} \leq 1.1 + 0.4) = 0.74.$$

$$P(1.1 - 0.9 \leq \widehat{\theta}_{LN} \leq 1.1 + 0.9) = 0.97,$$
$$P(1.1 - 0.9 \leq \widehat{\theta}_{N} \leq 1.1 + 0.9) = 0.99.$$

Linearization method

$$f(i/n, \boldsymbol{\beta}) = f(i/n, \boldsymbol{\beta}^*) + \sum_{j=1}^{p} \frac{\partial f}{\partial \beta_j}(i/n, \boldsymbol{\beta}^*)(\beta_j - \beta_j^*)$$

$$Var\widehat{\boldsymbol{\beta}} \approx (\boldsymbol{F}^{*T}\boldsymbol{F}^*)^{-1}\boldsymbol{F}^{*T}\boldsymbol{\Sigma}\boldsymbol{F}^*(\boldsymbol{F}^{*T}\boldsymbol{F}^*)^{-1},$$

$\boldsymbol{F}^* = ||\frac{\partial f_i}{\partial \beta_l}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^0)||_{i=1,l=1}^{n\ \ p}$,

σ_m^2 and max $Var\gamma_j$ are small. The approximation is very easy to compute. For designing an experiment the method we start with.

Monte-Carlo simulations method

We repeatedly generate realizations of γ and $\{e_i\}$ from their distributions to numerically find the argument minimizing the least squares function. The distribution of $\widehat{\boldsymbol{\beta}}$ and its characteristics as the mean, variance, skewness, etc., may then be estimated by their sample counterparts computed from the set of the obtained values of $\{\widehat{\boldsymbol{\beta}}\}$.

The minimization of the least squares function may be performed over a dense grid in $\boldsymbol{B}$ or by a more powerful method such as the Levenberg - Marquardt algorithm.

Finite sample approximation method using differential geometry

The approximate density $h_{\widehat{\beta}}(\beta)$ of the ordinary least squares estimate $\widehat{\beta}$ has the form:

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

where

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

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

Example 2

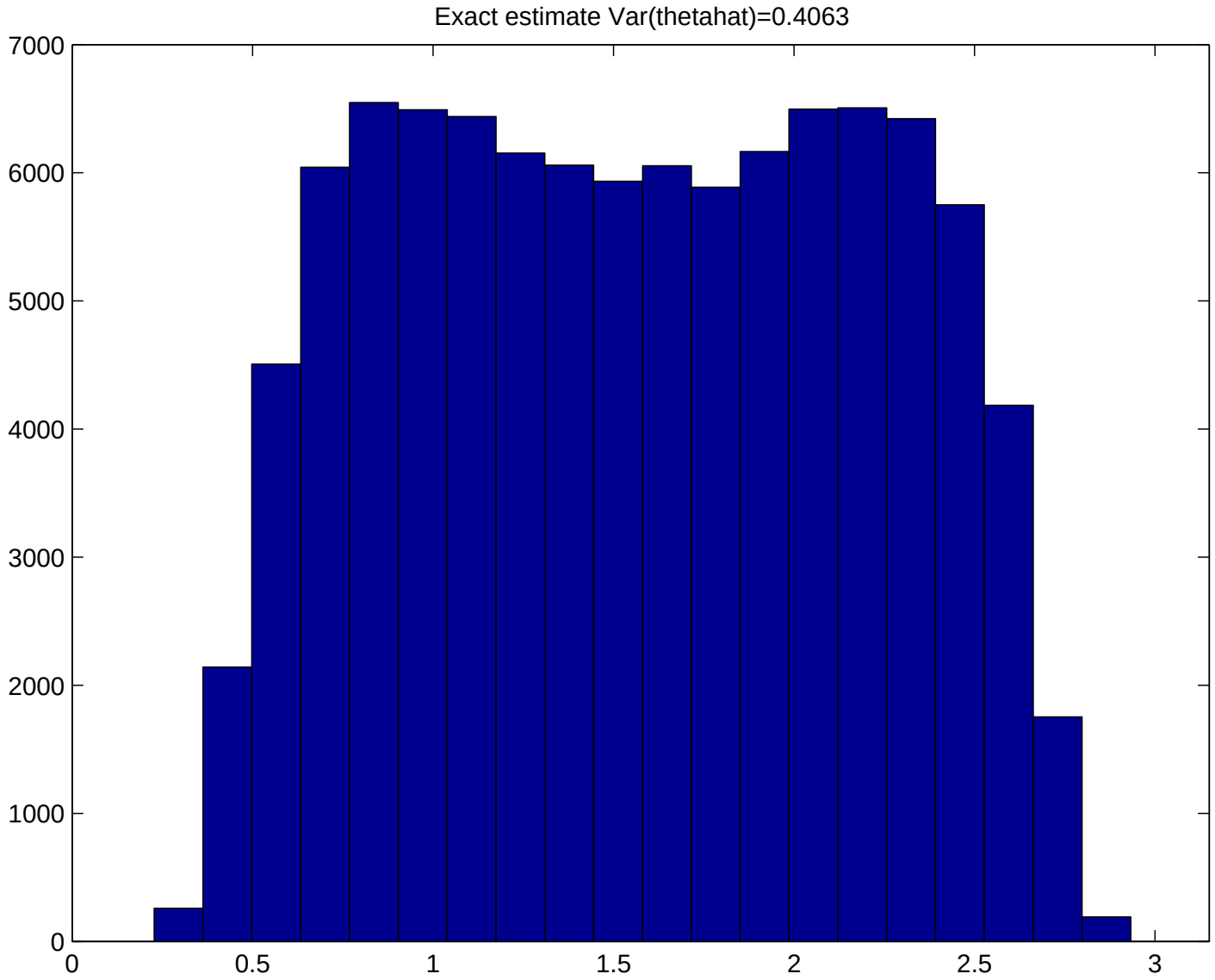
$$Y_1 = \cos(\theta) + \gamma + e_1,$$
$$Y_2 = \sin(\theta) + e_2$$

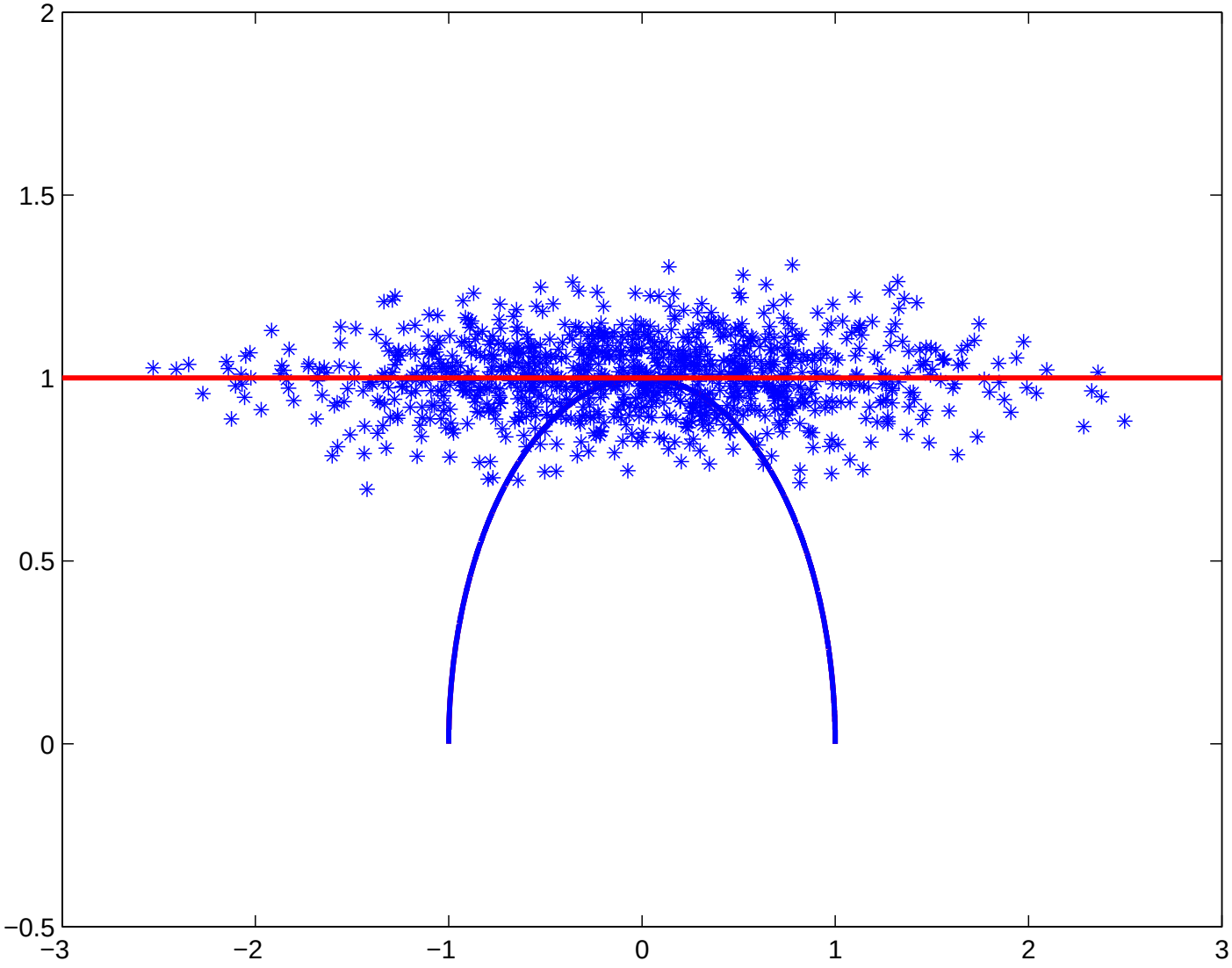
$\sin\theta = \cos(\theta - \pi/2), \quad \theta \in [0, \pi].$

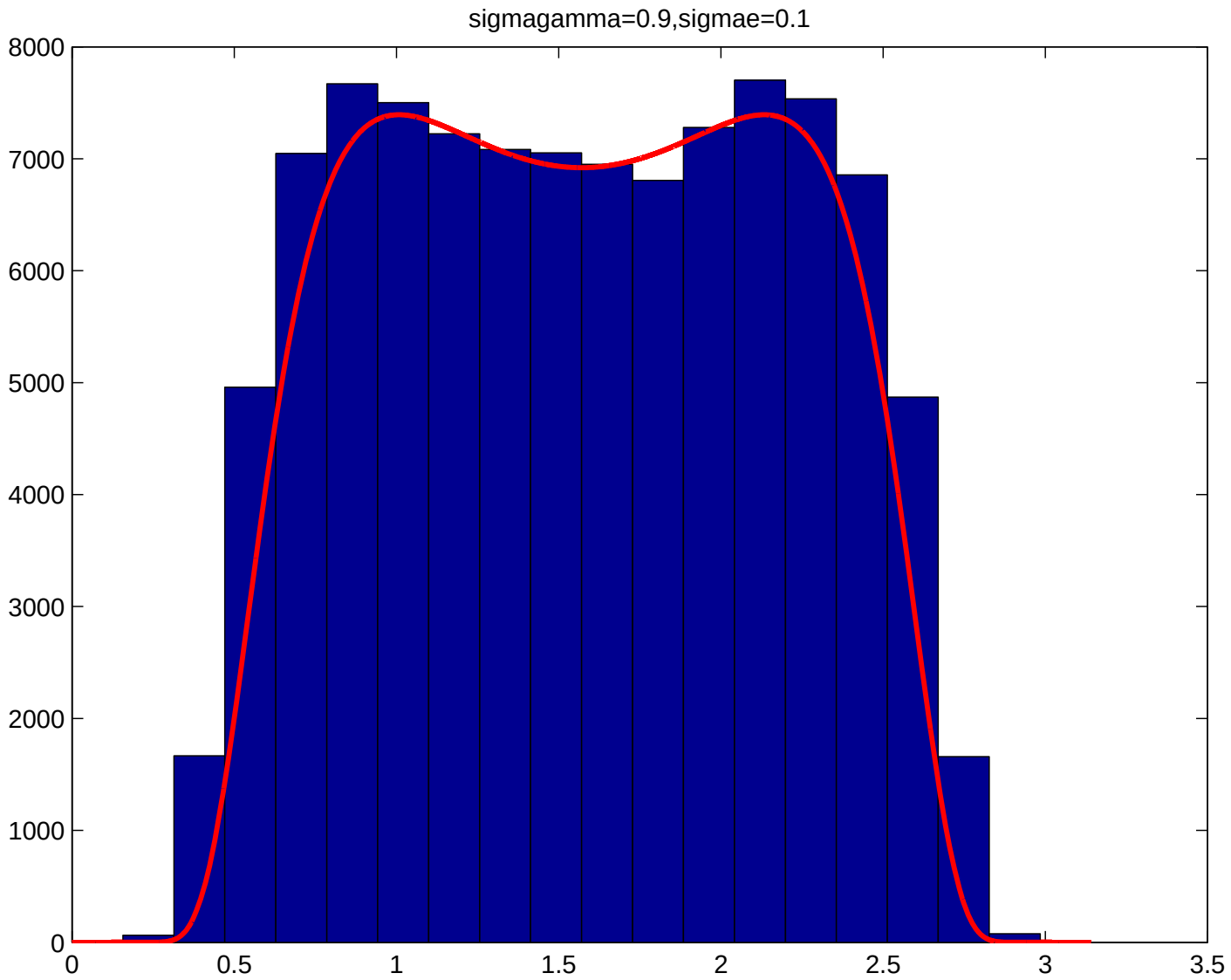
$\theta^* = \pi/2, \quad Y_1 = \gamma + e_1,\ Y_2 = 1 + e_2.$

$$\widehat{\theta} = \text{atan}\Big(\frac{Y_2}{Y_1}\Big), \quad Y_1 > 0, \qquad \widehat{\theta} = \pi - \text{atan}\Big(-\frac{Y_2}{Y_1}\Big), \quad Y_1 < 0.$$

$\gamma \sim N(0, 0.9^2), \qquad e_i \sim N(0, 0.1^2), i = 1, 2.$


Exact estimate Var(thetahat)=0.4063
7000
6000
5000
4000
3000
2000
1000
0
0
0.5
1
1.5
2
2.5
3


2
1.5
1
0.5
0
−0.5
−3
−2
−1
0
1
2
3

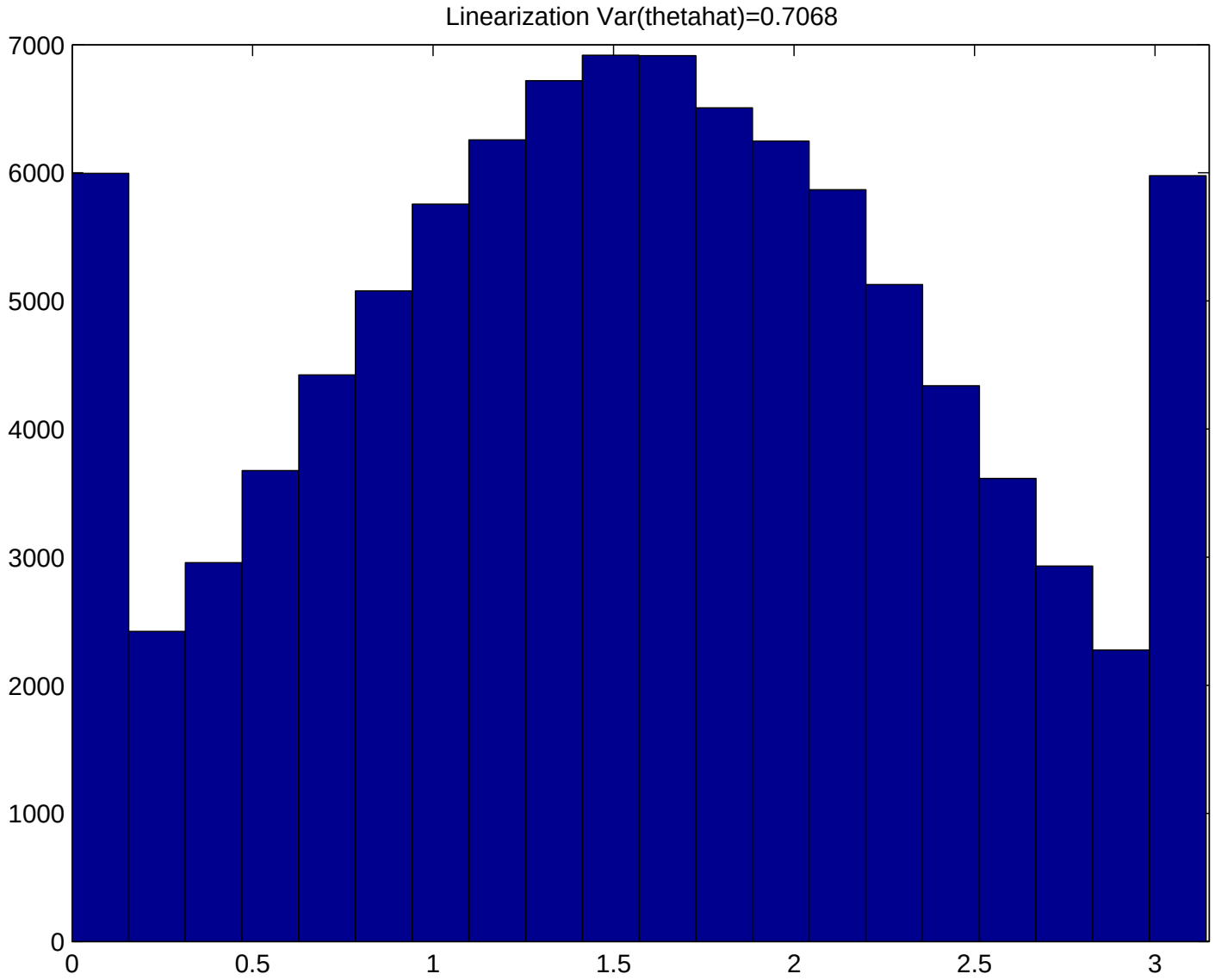

sigmagamma=0.9,sigmae=0.1
8000
7000
6000
5000
4000
3000
2000
1000
0
0
0.5
1
1.5
2
2.5
3
3.5

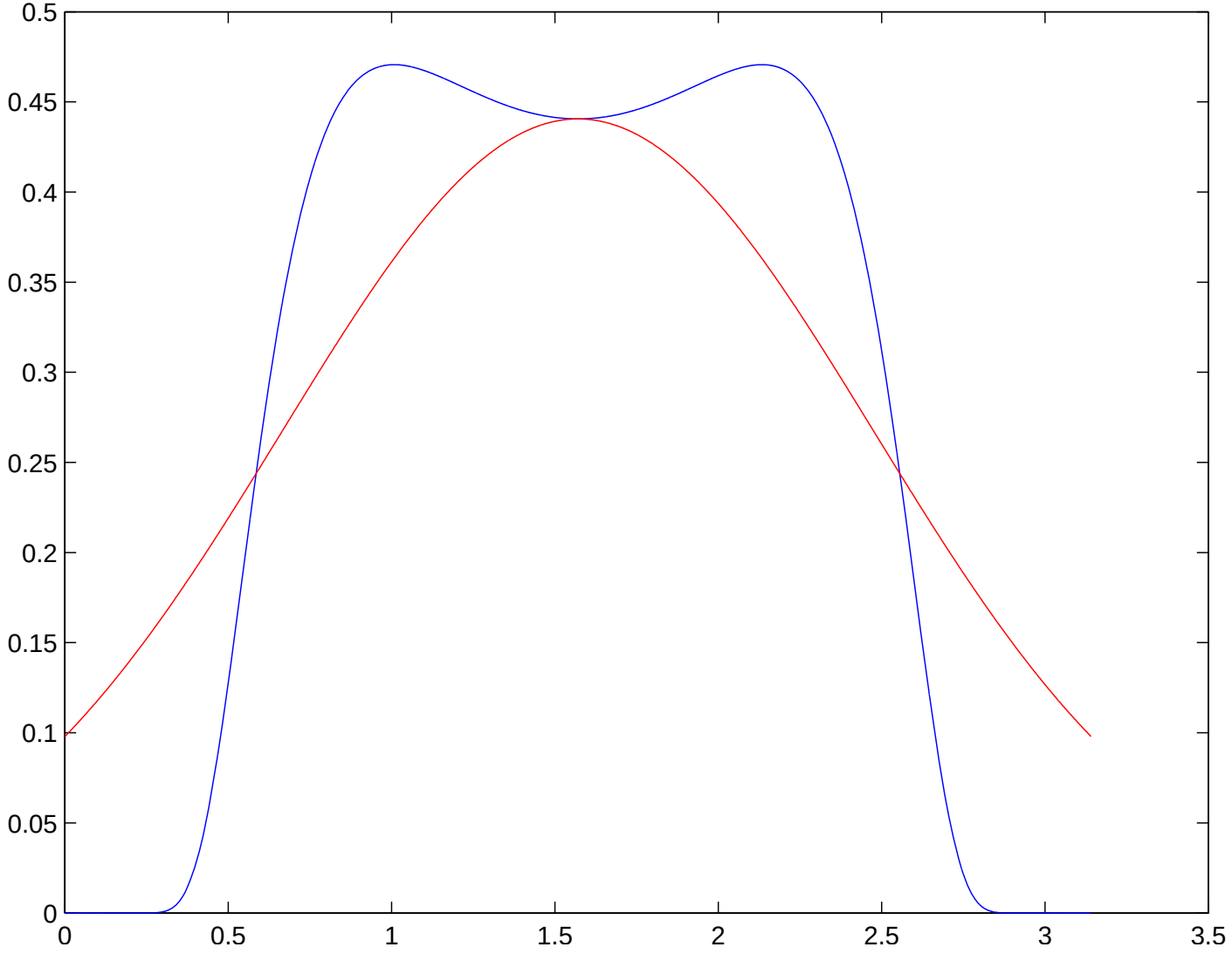
Linearization in $\pi/2$

$$Y_1 = -(\theta - \frac{\pi}{2}) + \gamma + e_1,$$
$$Y_2 = 1 + e_2.$$

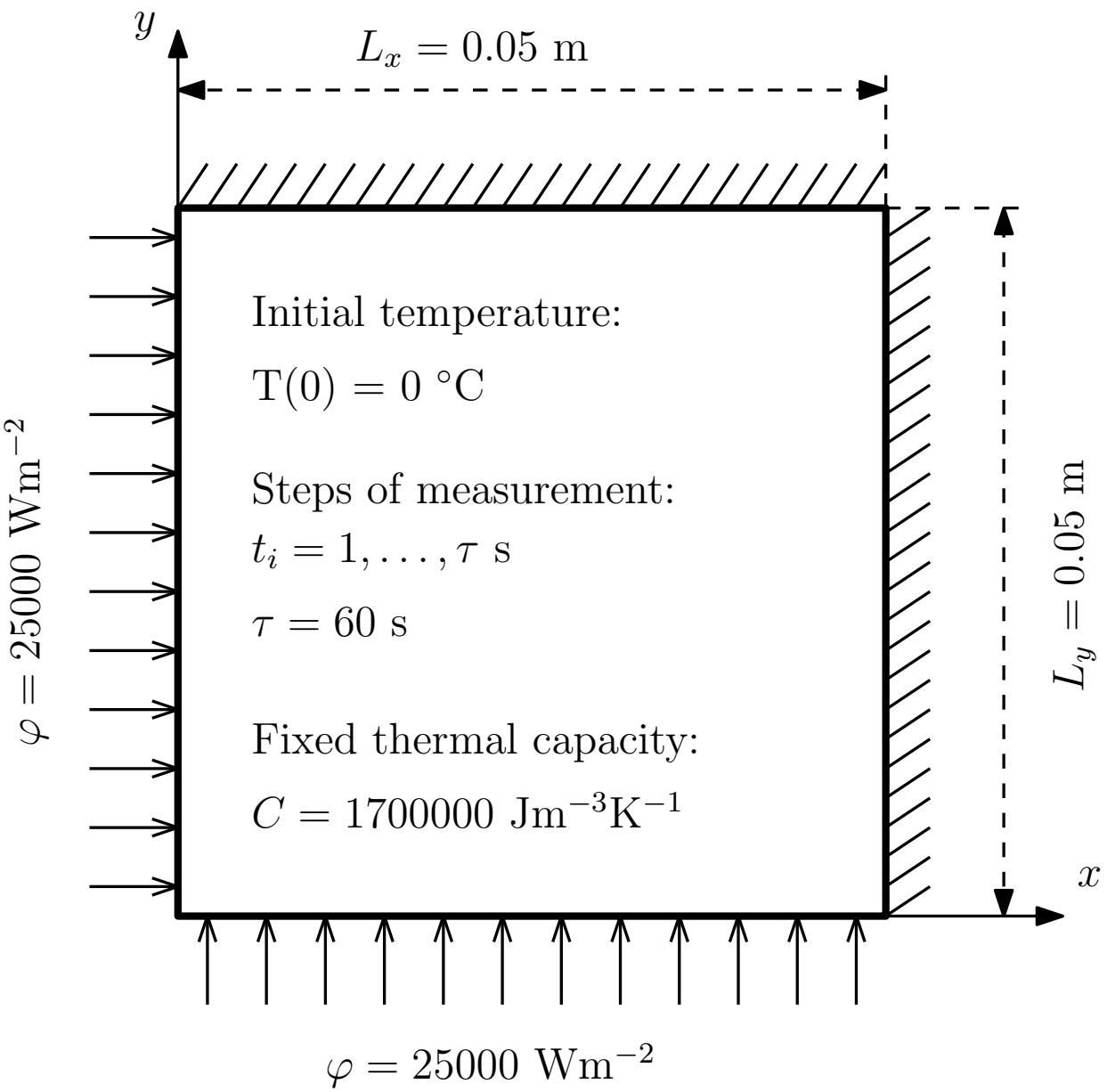
$\tilde{\theta} = \frac{\pi}{2} - Y_1, \quad -\frac{\pi}{2} \leq Y_1 \leq \frac{\pi}{2},$
$\tilde{\theta} = 0,\ Y_1 > \frac{\pi}{2}, \quad \tilde{\theta} = \pi,\ Y_1 < -\frac{\pi}{2}.$

$\tilde{\theta}$ has a trimmed normal distribution with
$P(\tilde{\theta} = 0) = P(\tilde{\theta} = \pi) = 0.04.$


Linearization Var(thetahat)=0.7068
7000
6000
5000
4000
3000
2000
1000
0
0
0.5
1
1.5
2
2.5
3


0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.5
1
1.5
2
2.5
3
3.5

Example 3


y
$L_x = 0.05$ m
Initial temperature:
$\mathrm{T}(0) = 0$ °C
Steps of measurement:
$t_i = 1, \ldots, \tau$ s
$\tau = 60$ s
Fixed thermal capacity:
$C = 1700000$ Jm^{-3}K^{-1}
$\varphi = 25000$ Wm^{-2}
$L_y = 0.05$ m
x
$\varphi = 25000$ Wm^{-2}

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

$$\theta_x(t;\lambda_x;C;\varphi;x) = \frac{2\varphi}{\sqrt{C\lambda_x}}\sqrt{t}F\Big(\frac{\widetilde{x}}{\sqrt{t}}\Big),$$

$$\theta_y(t;\lambda_y;C;\varphi;y) = \frac{2\varphi}{\sqrt{C\lambda_y}}\sqrt{t}F\Big(\frac{\widetilde{y}}{\sqrt{t}}\Big)$$

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

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

Simple application

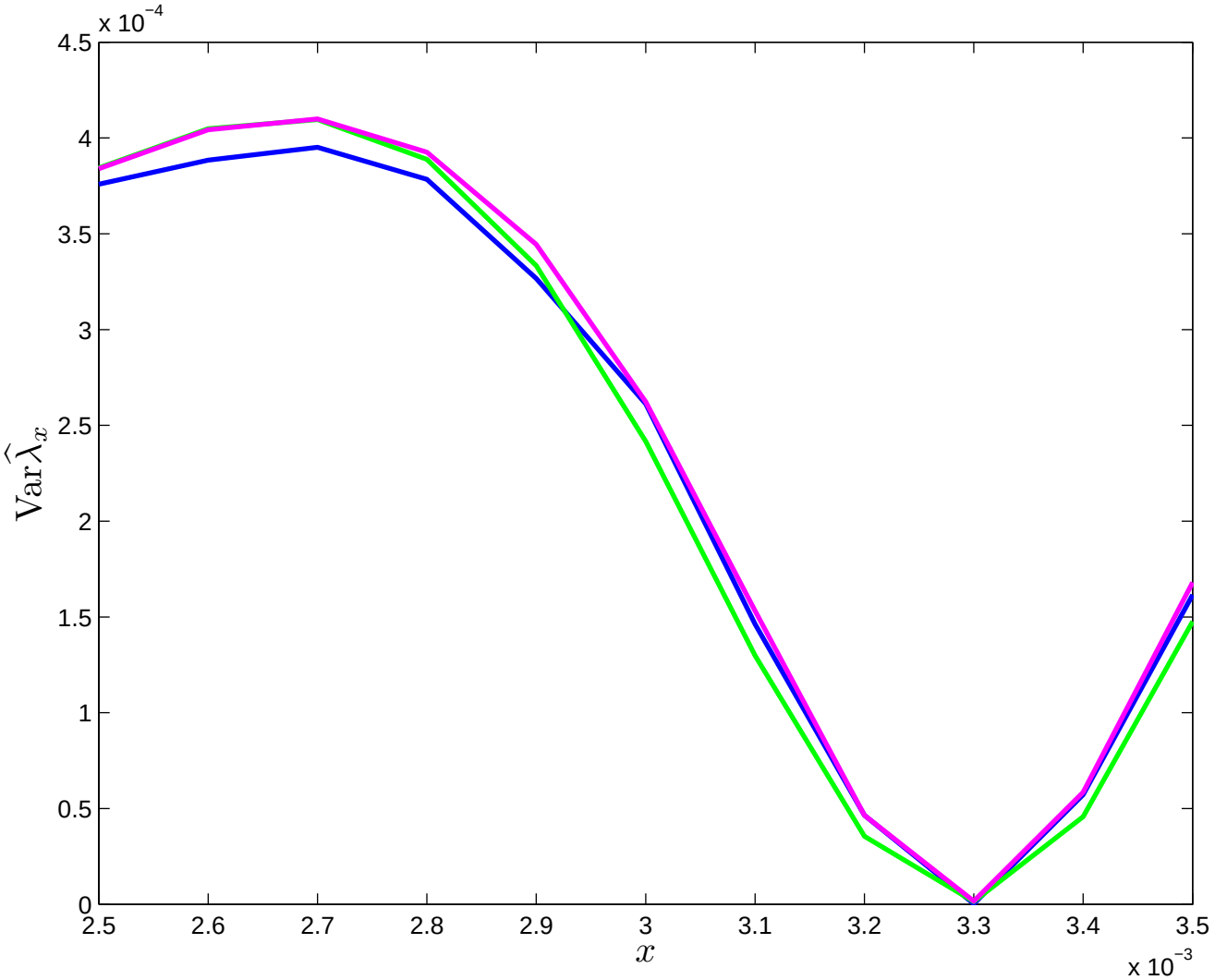
The thermal conductivity $\lambda_y = 4.7\ (\mathrm{Wm^{-1}K^{-1}})$ and the heat volumetric capacity $C = 1\,700\,000\ (\mathrm{Jm^{-3}K^{-1}})$ are known.

The conductivity $\lambda_x\ (\mathrm{Wm^{-1}K^{-1}})$ is our parameter of interest and its true value is $\lambda_x^* = 0.6\ (\mathrm{Wm^{-1}K^{-1}})$.

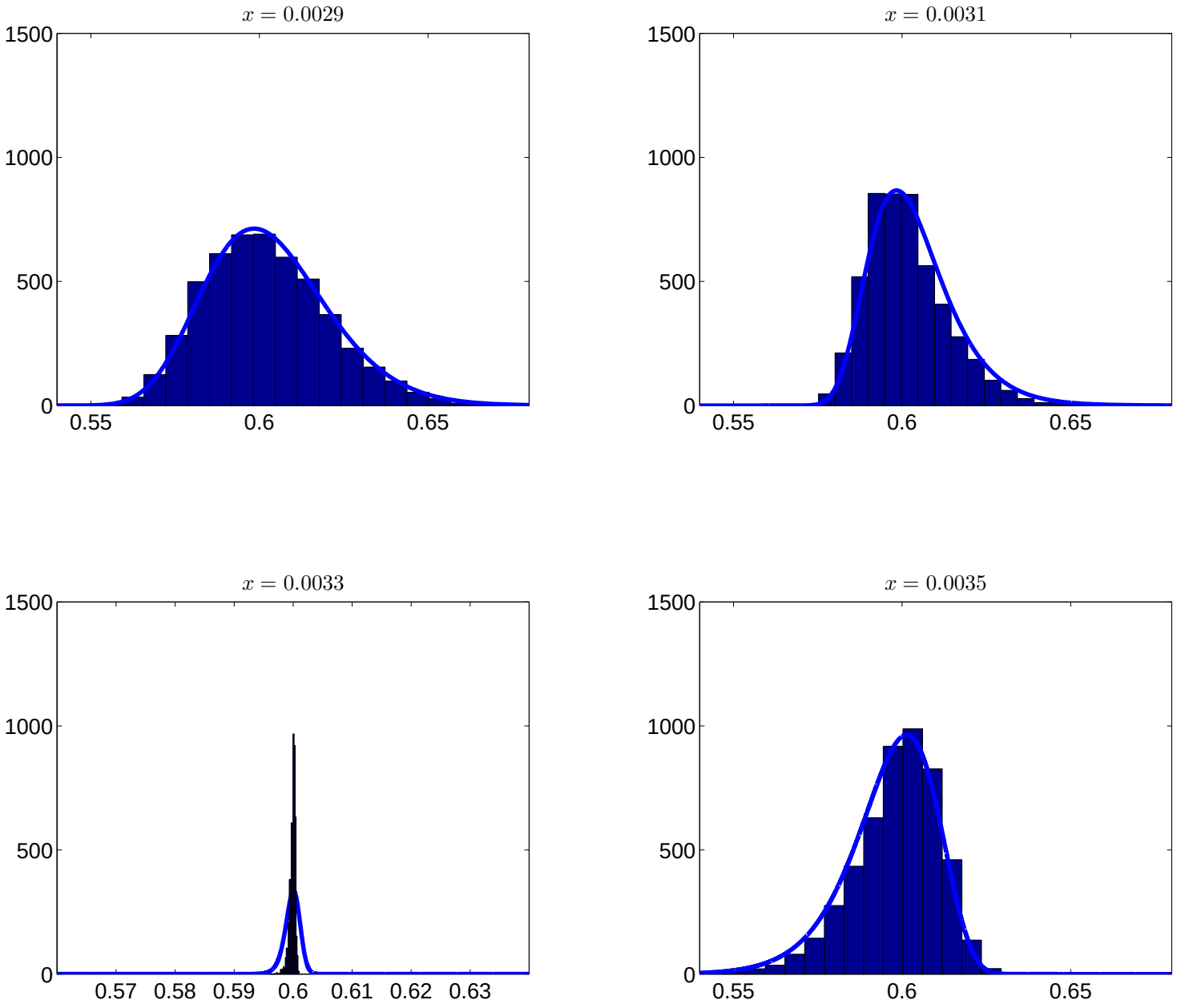
The heat flux $\varphi\ (\mathrm{Wm^{-2}})$ is an additional random parameter $\sim N(25\,000, 100^2)$.

Themperature $(^\circ\mathrm{C})$ is measured at equidistant time points $\{1, 2, \ldots, 60\}$ (s) with i.i.d. measurement errors $N(0, 0.1^2)$.

We are looking for the optimal position of the sensor on the bottom boundary, more specifically, in a grid $\{(x, y) \in [0.002 : 0.0001 : 0.05] \times \{0\}\}$.



Comparison of variances of $\widehat{\lambda}_x$ computed by the proposed methods - green line -linearization method, blue line - method of simulations and numerical minimization, magenta line - method based on differential geometry.


$x = 0.0029$
$x = 0.0031$
$x = 0.0033$
$x = 0.0035$
1500
1000
500
0
0.55
0.6
0.65
0.57
0.58
0.59
0.61
0.62
0.63

Conclusions - practical suggestions

Try all three methods if possible.

Look for a set of "almost optimal" designs if possible.

Compute a distribution of the estimate near the "optimal" designs if possible.

References:

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.

D. Jarušková and 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.

D. Jarušková and A. Kučerová, Different views on additional random parameters in experiment design for thermophysical parameters estimation, Proceedings of UNCECOMP 2017

D. Jarušková and A. Pázman, Methods for approximating of unknown parameters estimates with application in material thermophysics, *IJUQ*, **11**, 31 – 47, 2021.