Physikalisch-Technische Bundesanstalt
Braunschweig and Berlin
National Metrology Institute

Uncertainty evaluation in curve fitting and regression problems

Clemens Elster

Physikalisch-Technische Bundesanstalt (PTB)

Braunschweig and Berlin, Germany

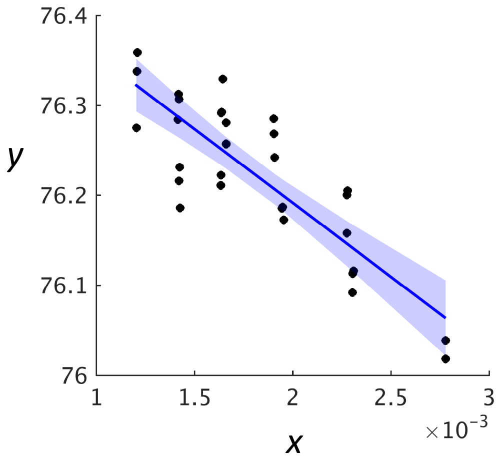
BIPM Workshop on Measurement Uncertainty

BIPM 15-16 June 2015

Motivation

- Curve fitting and regression problems relevant in metrology
- GUM and its supplements hardly give guidance
- Least-squares methods often applied
- Bayesian inference provides an alternative

Regression problem



$$y_i = f_{\boldsymbol{\theta}}(x_i) + \varepsilon_i \,,\; i = 1,\ldots,n$$

Data   Regression function   Errors

e.g. $\varepsilon_i \sim N(0,\sigma^2)$

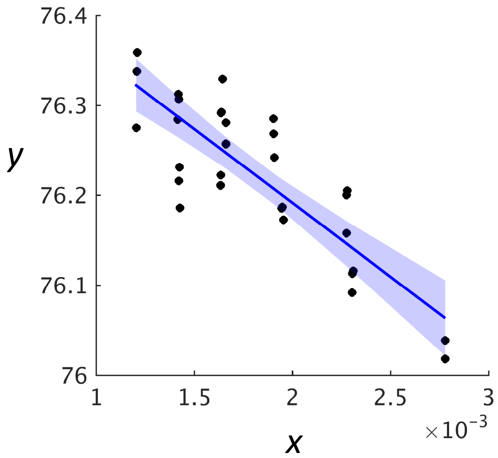
Measurand

– Regression parameters $\boldsymbol{\theta}$

– Regression function $f_{\boldsymbol{\theta}}(x)$

– Prediction of x from subsequent measurement of y

Bayesian inference

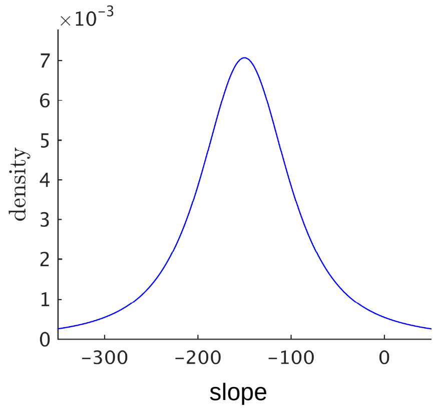


$$y_i = f_{\boldsymbol{\theta}}(x_i) + \varepsilon_i\,,\; i = 1, \ldots, n$$

Data | Regression function | Errors

e.g. $\varepsilon_i \sim N(0, \sigma^2)$

Prior knowledge

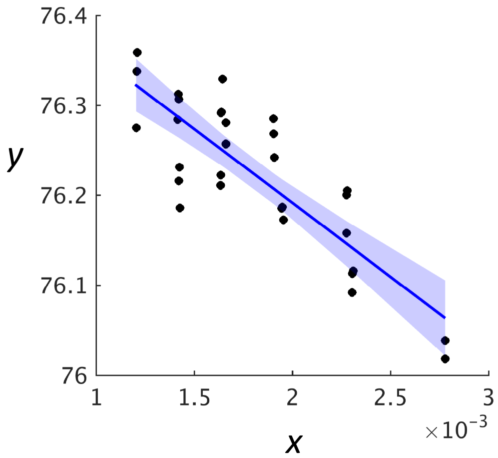


– Previous experiments

– Expert knowledge

– Known relations / constraints

Bayesian inference



$$y_i = f_{\boldsymbol{\theta}}(x_i) + \varepsilon_i\,,\ i = 1,\ldots,n$$

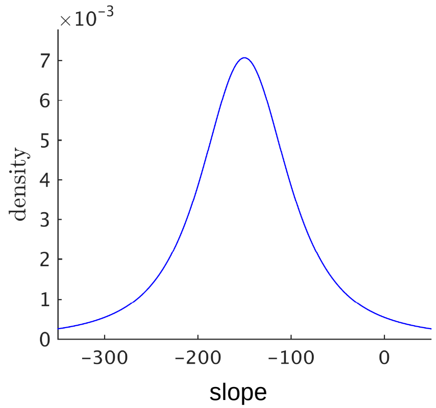
Data

Regression function

Errors

e.g. $\varepsilon_i \sim N(0,\sigma^2)$

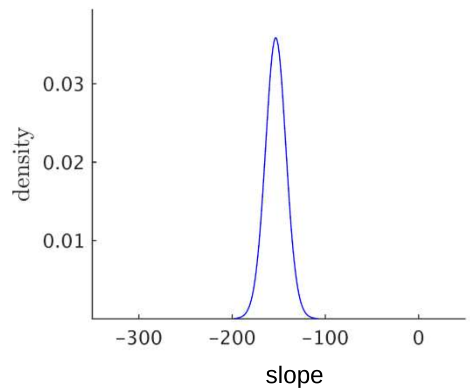
Prior knowledge



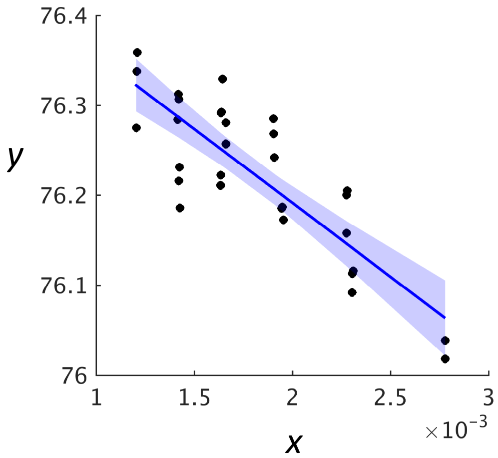
$\times$ Likelihood

Bayes theorem

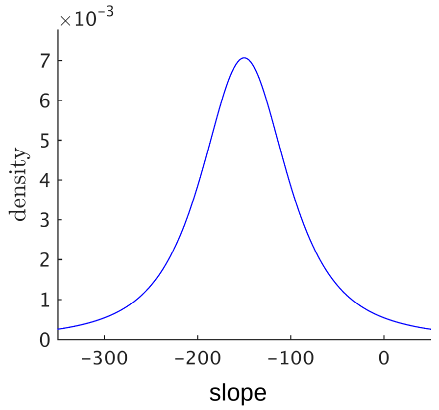
Posterior distribution



Bayesian inference



Prior knowledge

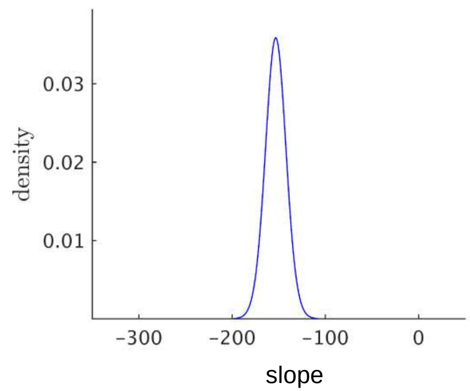


×

Likelihood

Bayes theorem

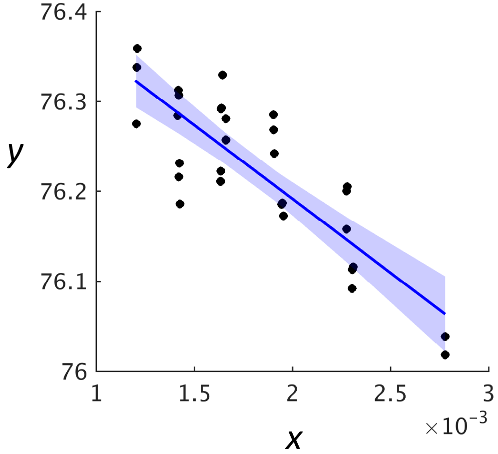
Posterior distribution



<u>Posterior distribution</u>

- Comprehensive characterization of uncertainty
- Basis for calculation of estimate and uncertainty
- Transferable (as prior in subsequent analyses)

Bayesian inference



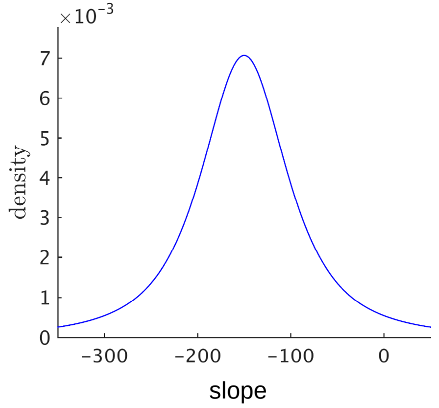
Prior knowledge

<u>Posterior distribution</u>

- Comprehensive characterization of uncertainty
- Basis for calculation of estimate and uncertainty
- Transferable (as prior in subsequent analyses)

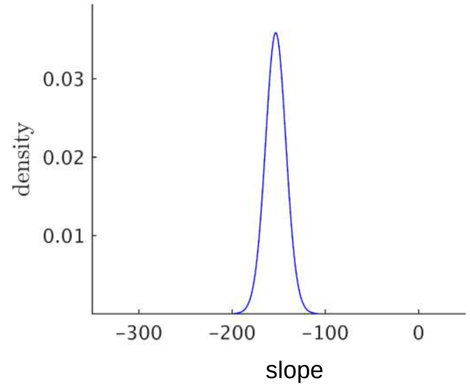
A O'Hagan and J J Forster (2004) Kendall's advanced theory of statistics, volume 2B: Bayesian inference, Oxford, Oxford University Press

C P Robert (2007) The Bayesian choice: From Decision-Theoretic Foundations to Computational Implementations, Springer, New York

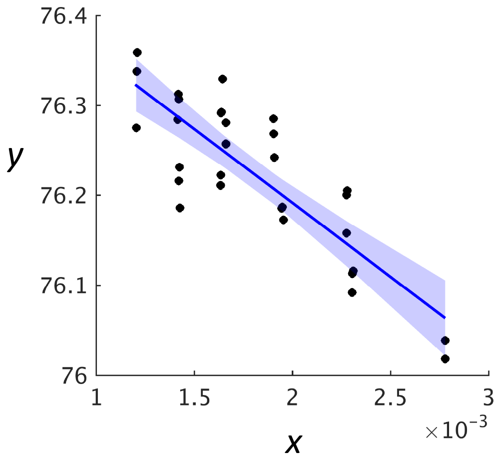


$\times$ Likelihood → Bayes theorem

Posterior distribution



Least-squares



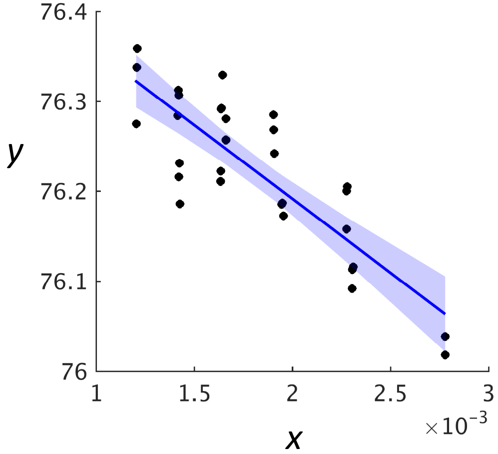
$$y_i = f_{\boldsymbol{\theta}}(x_i) + \varepsilon_i \,,\; i = 1,\ldots,n$$

Data   Regression function   Errors

e.g. $\varepsilon_i \sim N(0,\sigma^2)$

$$\boxed{\hat{\boldsymbol{\theta}}: \quad \min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(y_i - f_{\boldsymbol{\theta}}(x_i)\right)^2}$$

Least-squares



$$y_i = f_{\boldsymbol{\theta}}(x_i) + \varepsilon_i\,,\; i = 1,\ldots,n$$

Data — Regression function — Errors

e.g. $\varepsilon_i \sim N(0,\sigma^2)$

$$\boxed{\hat{\boldsymbol{\theta}}: \quad \min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(y_i - f_{\boldsymbol{\theta}}(x_i)\right)^2}$$

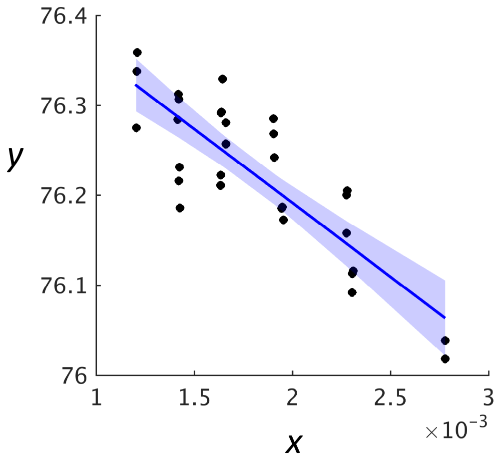
$$\mathbf{U}(\hat{\boldsymbol{\theta}}) = \hat{\sigma}^2 \left(\mathbf{J}^T\mathbf{J}\right)^{-1}, \quad u(\hat{\theta}_\alpha) = \sqrt{\hat{\sigma}^2 \left(\mathbf{J}^T\mathbf{J}\right)^{-1}_{\alpha\alpha}}\;, \quad \mathbf{J} = \partial f_{\boldsymbol{\theta}}(\mathbf{x}) / \partial \boldsymbol{\theta}$$

$$\frac{\boldsymbol{\psi}^T\hat{\boldsymbol{\theta}}}{\boldsymbol{\psi}^T\mathbf{U}(\hat{\boldsymbol{\theta}})\boldsymbol{\psi}} \sim t_{n-p}\,, \quad I_{\theta_\alpha} = \hat{\theta}_\alpha \pm t_{n-p}^{0.975} u(\hat{\theta}_\alpha)$$

G A F Seber and C J Wild (1989) **Nonlinear regression**, Wiley, New York

N R Draper and H Smith (1998) **Applied regression analysis**, Wiley, New York

Least-squares



$$y_i = f_{\boldsymbol{\theta}}(x_i) + \varepsilon_i\,,\ i = 1,\ldots,n$$

Data Regression function Errors

e.g. $\varepsilon_i \sim N(0,\sigma^2)$

$$\boxed{\hat{\boldsymbol{\theta}}: \quad \min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(y_i - f_{\boldsymbol{\theta}}(x_i)\right)^2}$$

- Cannot include general prior knowledge (but possibly constraints)
- Different meaning of uncertainties than for Bayesian approach
- Typically (much) easier to compute than Bayesian approach
- Provides point estimates and uncertainties – no distribution

LS combined with GUM-S1/S2 versus Bayesian inference

- Combining LS methods with Supplement 1 (or 2) to the GUM also yields a distribution for the regression parameters

- Is such proceeding equivalent to a Bayesian inference?

LS combined with GUM-S1/S2 versus Bayesian inference

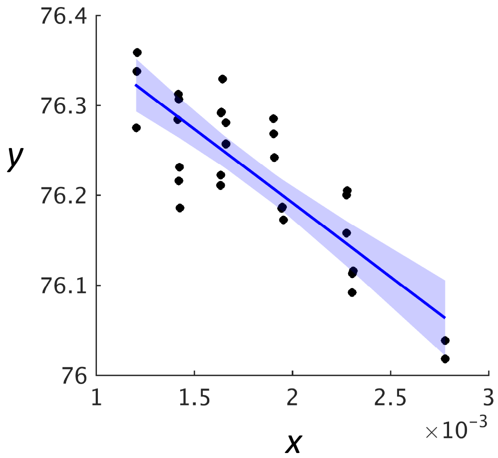
- Combining LS methods with Supplement 1 (or 2) to the GUM also yields a distribution for the regression parameters
- Is such proceeding equivalent to a Bayesian inference?
- In general it is not:
  - No account of prior knowledge (in terms of a distribution)
  - Generally even different to Bayesian inference with noninformative prior

C Elster and B Toman 2011 Bayesian uncertainty analysis for a regression model versus application of GUM Supplement 1 to the least-squares estimate *Metrologia* **48** 233-240

Normal linear straight line regression

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0,\sigma^2)$$

[1)] Noninformative prior

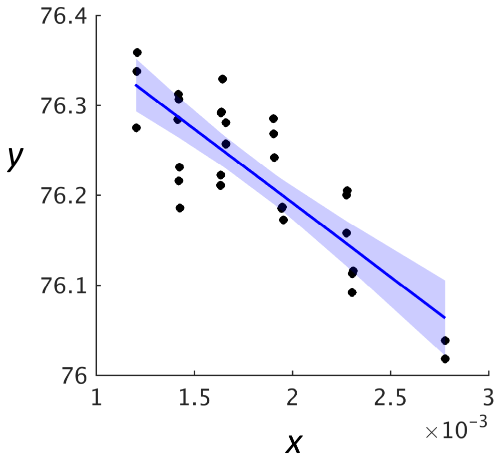


	$\hat{\theta}_2$	$u(\theta_2)$	$I_{0.95}$
LS	-164.6	19.4	$[-204.3,-124.8]$
Bayes[1)]	-164.6	20.1	$[-204.3,-124.8]$

Normal linear straight line regression

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0,\sigma^2)$$

[1)]Noninformative prior



	$\hat{\theta}_2$	$u(\theta_2)$	$I_{0.95}$
LS	-164.6	19.4	$[-204.3, -124.8]$
Bayes[1)]	-164.6	20.1	$[-204.3, -124.8]$

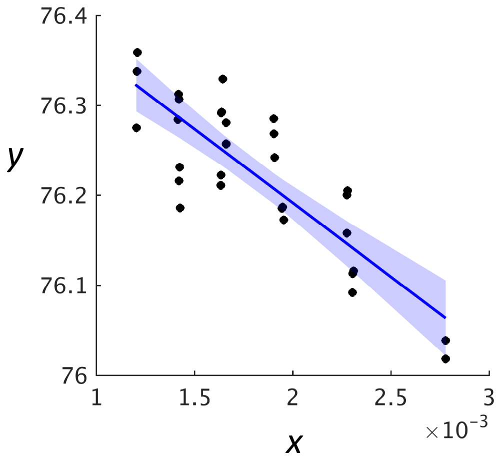
– Same estimates & intervals

– But different meaning of uncertainties

Normal linear straight line regression

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

1) Noninformative prior



	$\hat{\theta}_2$	$u(\theta_2)$	$I_{0.95}$
LS	-164.6	19.4	$[-204.3, -124.8]$
Bayes[1)]	-164.6	20.1	$[-204.3, -124.8]$

– Same estimates & intervals

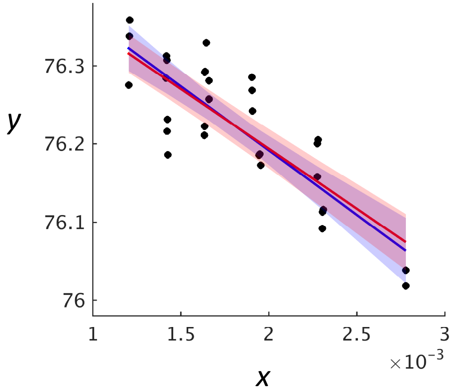
– But different meaning of uncertainties

– Different standard uncertainties

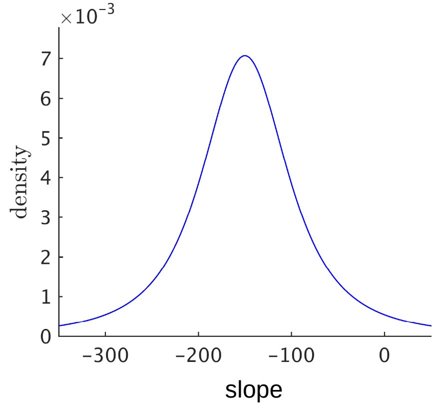
Normal linear straight line regression

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

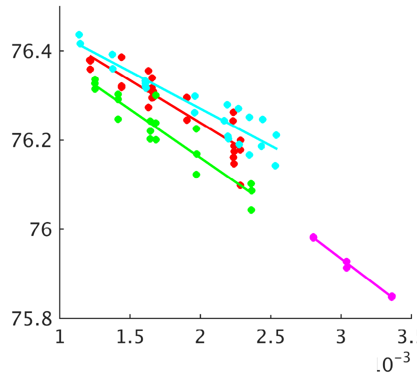
[1)]Informative prior



	$\hat{\theta}_2$	$u(\theta_2)$	$I_{0.95}$
LS	-164.6	19.4	$[-204.3, -124.8]$
Bayes[1)]	-153.2	11.4	$[-175.6, -130.8]$



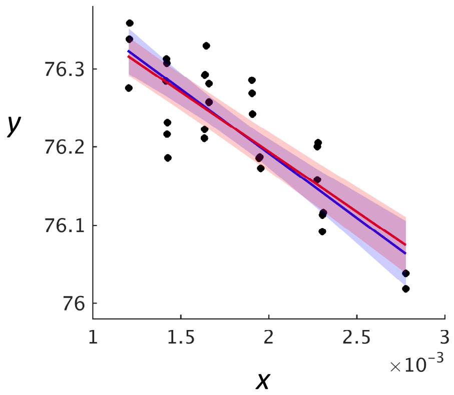
Prior knowledge from previous measurements[1]



[1]K Klauenberg, G Wübbeler, B Mickan, P M Harris and C Elster
A tutorial on Bayesian Normal linear regression, submitted to **Metrologia**

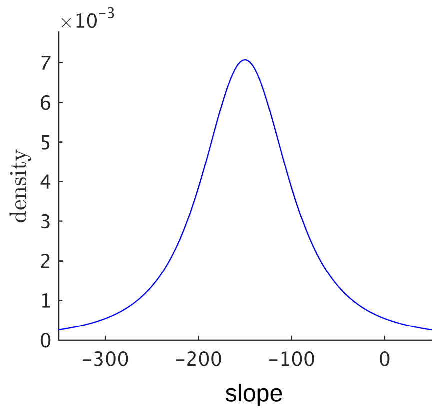
Normal linear straight line regression

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0,\sigma^2)$$



	$\hat{\theta}_2$	$u(\theta_2)$	$I_{0.95}$
LS	-164.6	19.4	$[-204.3, -124.8]$
Bayes[1)]	-153.2	11.4	$[-175.6, -130.8]$

[1)]Informative prior

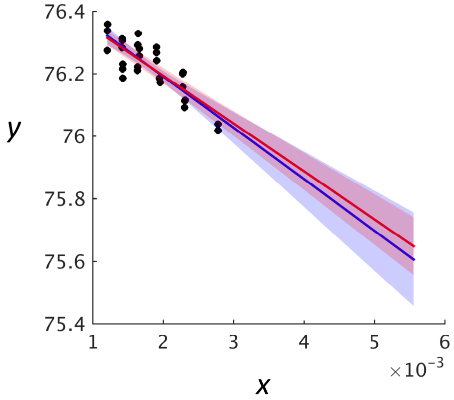


– Different estimates & intervals

– Bayesian inference can result in smaller uncertainties than LS

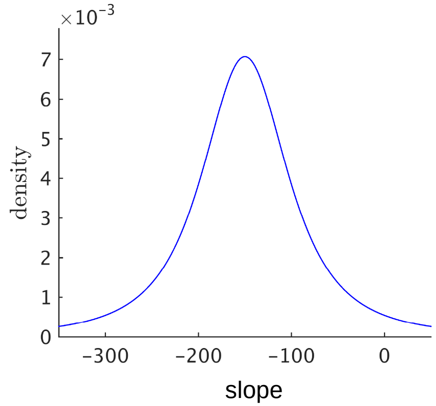
Normal linear straight line regression

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$



	$\hat{\theta}_2$	$u(\theta_2)$	$I_{0.95}$
LS	-164.6	19.4	$[-204.3, -124.8]$
Bayes[1)]	-153.2	11.4	$[-175.6, -130.8]$

[1)]Informative prior



– Different estimates & intervals

– Bayesian inference can result in smaller uncertainties than LS

– Relevant in particular for prediction beyond the range of data[1,2]

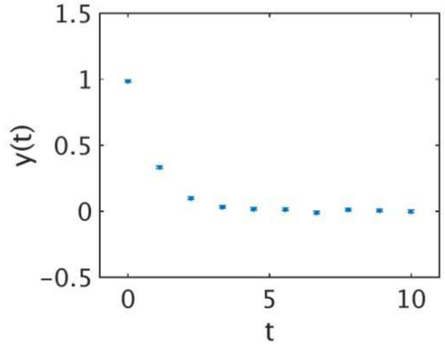
[1]G J P Kok, A M H van der Veen, P M Harris, I M Smith and C Elster (2015) *Bayesian analysis of a flow meter calibration problem*, **Metrologia**, 52, 400-405

[2]K Klauenberg, G Wübbeler, B Mickan, P M Harris and C Elster *A tutorial on Bayesian Normal linear regression*, submitted to **Metrologia**

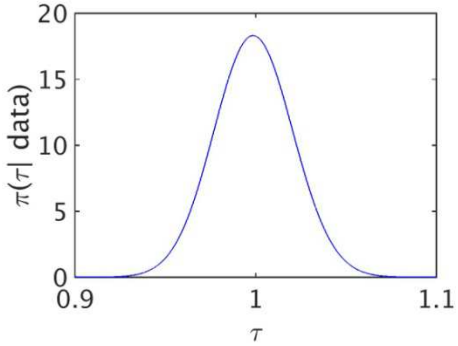
Nonlinear regression

Model $y(t_i) = ae^{-t_i/\tau} + \varepsilon_i,\ \ \varepsilon_i \sim N(0,\sigma^2),\ \ a,\sigma$ known, $\pi(\tau) \propto 1\ \ (0 < \tau < 10)$

Data



Posterior



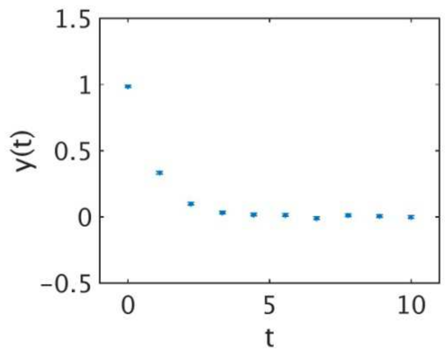
	$\hat{\tau}$	$u(\tau)$	$I_{0.95}$
Bayes	1.0	0.02	[0.96,1.04]
LS	1.0	0.02	[0.96,1.04]

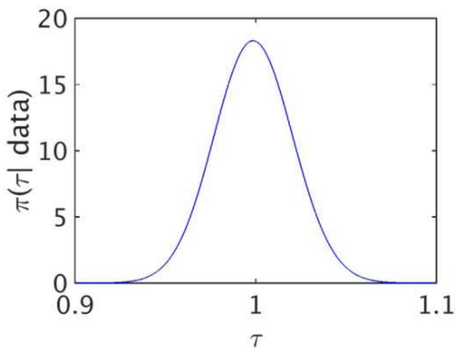
Nonlinear regression

Model $y(t_i) = a e^{-t_i/\tau} + \varepsilon_i,\ \ \varepsilon_i \sim N(0,\sigma^2),\ \ a,\sigma$ known, $\pi(\tau) \propto 1\ \ (0 < \tau < 10)$

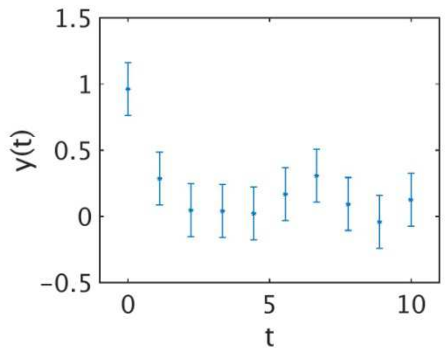
Data

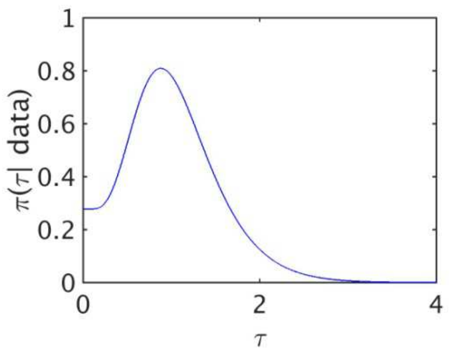
Posterior





	$\hat{\tau}$	$u(\tau)$	$I_{0.95}$
Bayes	1.0	0.02	[0.96,1.04]
LS	1.0	0.02	[0.96,1.04]





Bayes	1.00	0.54	[0.09,2.21]
LS	0.88	0.41	[0.06,1.70]

For large σ LS may underrate uncertainties due to nonlinearities

Conclusions

- LS and Bayesian inference different conceptually and typically also numerically

- Calculation of results usually easier for LS than for Bayesian inference

Conclusions

- LS and Bayesian inference different conceptually and typically also numerically
- Calculation of results usually easier for LS than for Bayesian inference
- LS yields reasonable results when
  - Prior knowledge not available or not really informative, and
  - Simple statistical model appropriate, and
  - No severe nonlinearity in problem
- Bayesian inference advantageous when
  - Prior knowledge available and informative
  - Severe nonlinearity in problem
  - Non-trivial statistical model (e.g. heteroscedastic problem)

Acknowledgment & reference

- Gerd Wübbeler and Bodo Mickan for preparing figures and providing data

- Part of this work has been funded by the European Metrology Research Program (EMRP) project NEW 04 Novel mathematical and statistical approaches to uncertainty evaluation. The EMRP is jointly funded by the EMRP participating countries within EURAMET (European Association of National Metrology Institutes) and the European Union.

- See also talks by Katy Klauenberg, Gertjan Kok, Nicolas Fischer

- C Elster, K Klauenberg, M Walzel, G Wübbeler, P Harris, M Cox, C Matthews, I Smith, L Wright, A Allard, N Fischer, S Cowen, S Ellison, P Wilson, F Pennecchi, G Kok, A van der Veen, L Pendrill (2015) *A Guide to Bayesian Inference for Regression Problems*, Deliverable of EMRP project NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation" (available at http://www.ptb.de/emrp/new04-home.html)

References

- A O‘Hagan and J J Forster (2004) **Kendall‘s advanced theory of statistics, volume 2B: Bayesian inference**, Oxford, Oxford University Press
- C P Robert (2007) **The Bayesian choice: From Decision-Theoretic Foundations to Computational Implementations**, Springer, New York
- G A F Seber and C J Wild (1989) **Nonlinear regression**, Wiley, New York
- N R Draper and H Smith (1998) **Applied regression analysis**, Wiley, New York
- G J P Kok, A M H van der Veen, P M Harris, I M Smith and C Elster (2015) Bayesian analysis of a flow meter calibration problem, **Metrologia**, 52, 400-405
- K Klauenberg, G Wübbeler, B Mickan, P M Harris and C Elster A tutorial on Bayesian Normal linear regression, submitted to **Metrologia**
- C Elster and B Toman 2011 *Bayesian uncertainty analysis for a regression model versus application of GUM Supplement 1 to the least-squares estimate* **Metrologia 48** 233-240
- C Elster, K Klauenberg, M Walzel, G Wübbeler, P Harris, M Cox, C Matthews, I Smith, L Wright, A Allard, N Fischer, S Cowen, S Ellison, P Wilson, F Pennecchi, G Kok, A van der Veen, L Pendrill (2015) ***A Guide to Bayesian Inference for Regression Problems***, Deliverable of EMRP project NEW04 “Novel mathematical and statistical approaches to uncertainty evaluation", 2015 (available at http://www.ptb.de/emrp/new04-home.html)