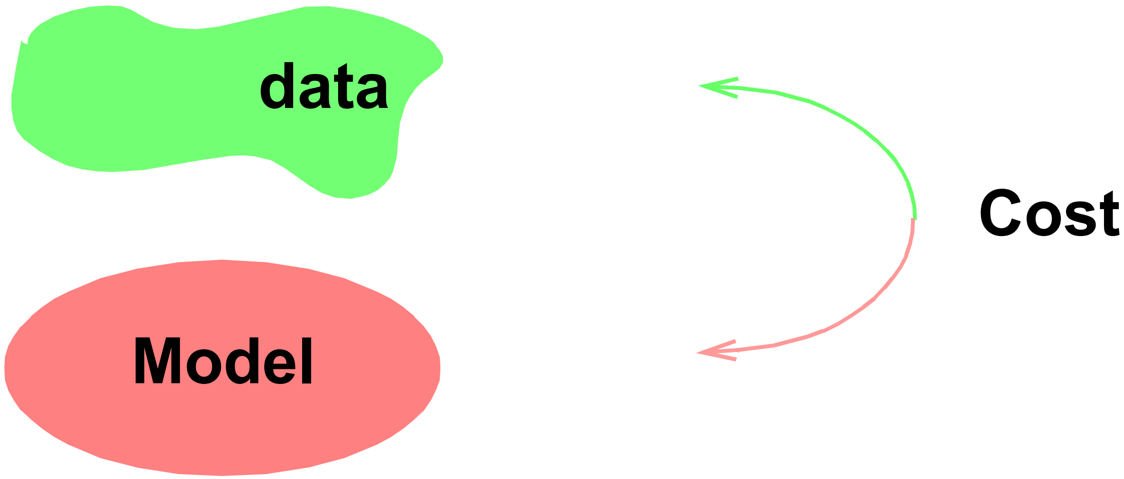
# System identification




**Goal**

A statistical framework

# Outline

A motivating example: why do *you* need system identification?!

Characterizing estimators

Basic steps of the identification process

A statistic approach to system identification

Identification in the presence of ‘input’ and ‘output’ noise

# Why do you need identification methods?

## A simple experiment

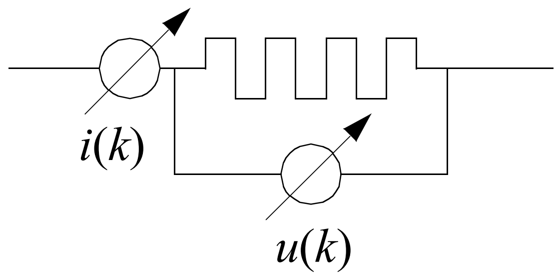
# Why do you need identification methods?

## A simple experiment

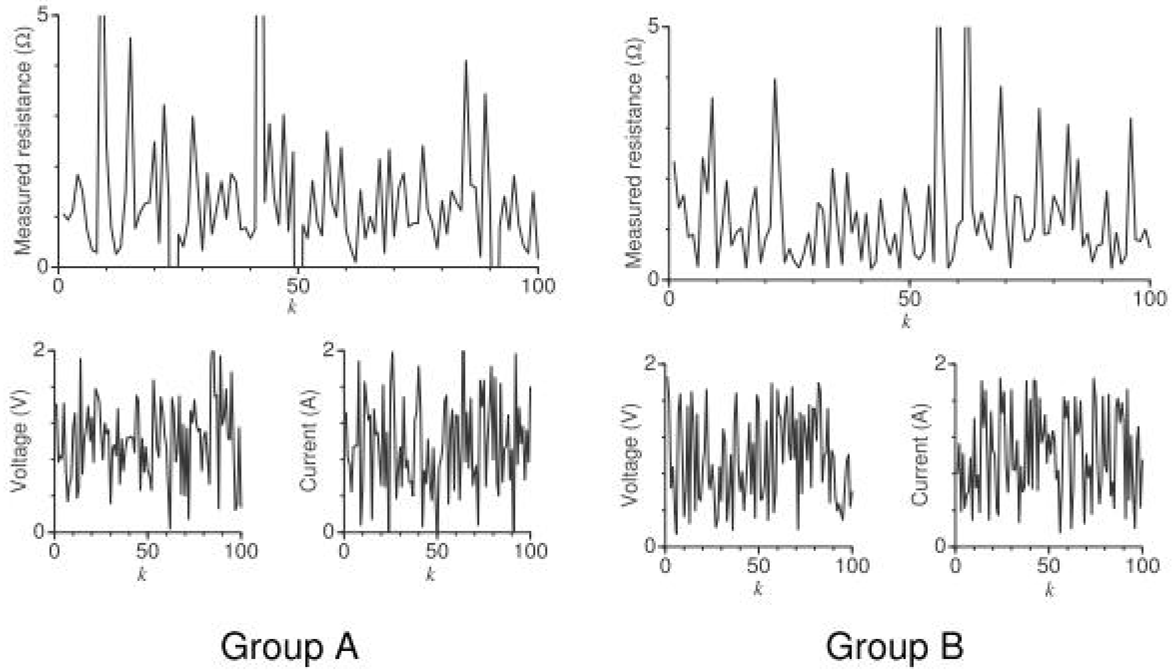
Multiple measurements lead to conflicting results.

How to combine all this information?

# Why do you need identification methods
# Measurement of a resistance




2 sets of measurements



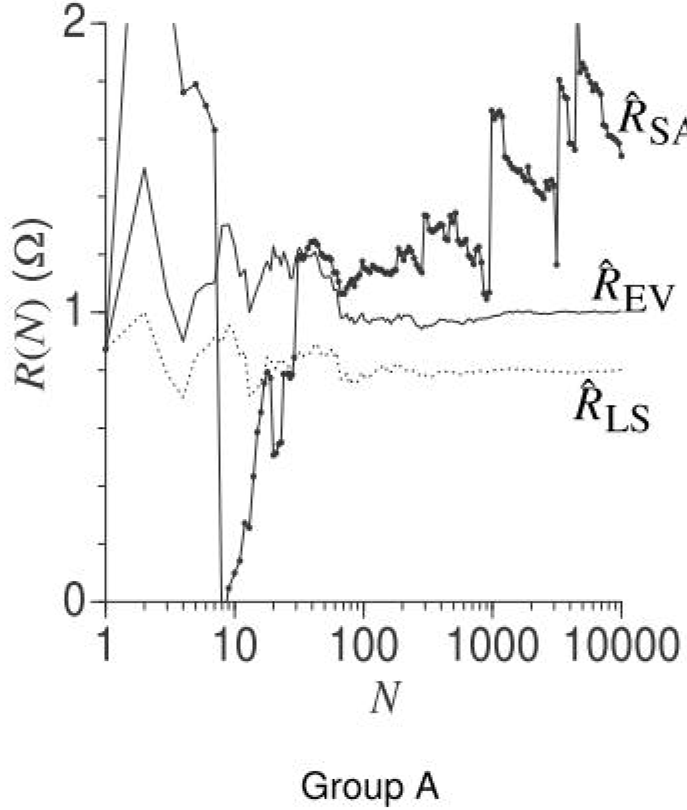
# 3 different estimators

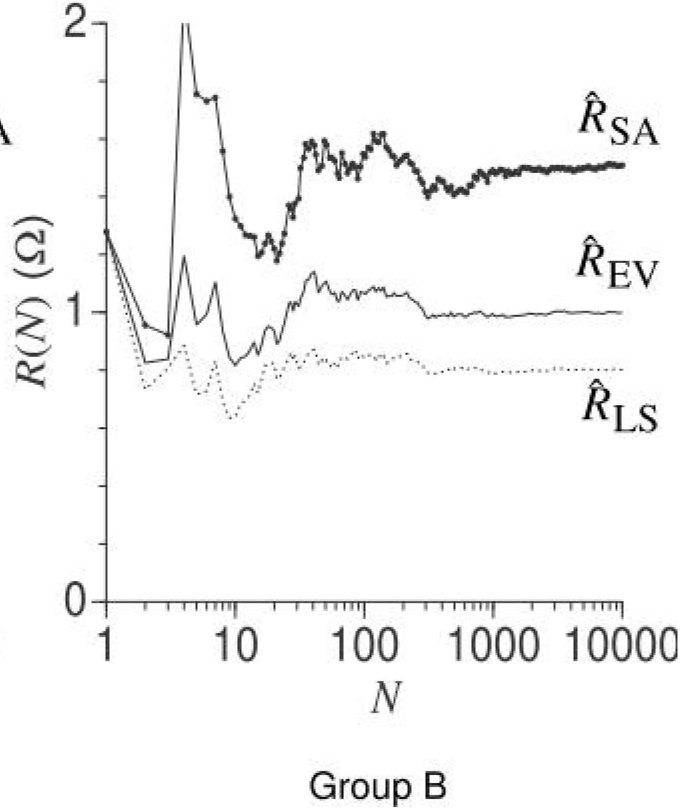
$$\hat{R}_{\mathrm{SA}}(N) = \frac{1}{N}\sum_{k=1}^{N}\frac{u(k)}{i(k)}$$

$$\hat{R}_{\mathrm{LS}}(N) = \frac{\frac{1}{N}\sum_{k=1}^{N}u(k)i(k)}{\frac{1}{N}\sum_{k=1}^{N}i(k)^2}$$

$$\hat{R}_{\mathrm{EV}}(N) = \frac{\frac{1}{N}\sum_{k=1}^{N}u(k)}{\frac{1}{N}\sum_{k=1}^{N}i(k)}$$

# and their results

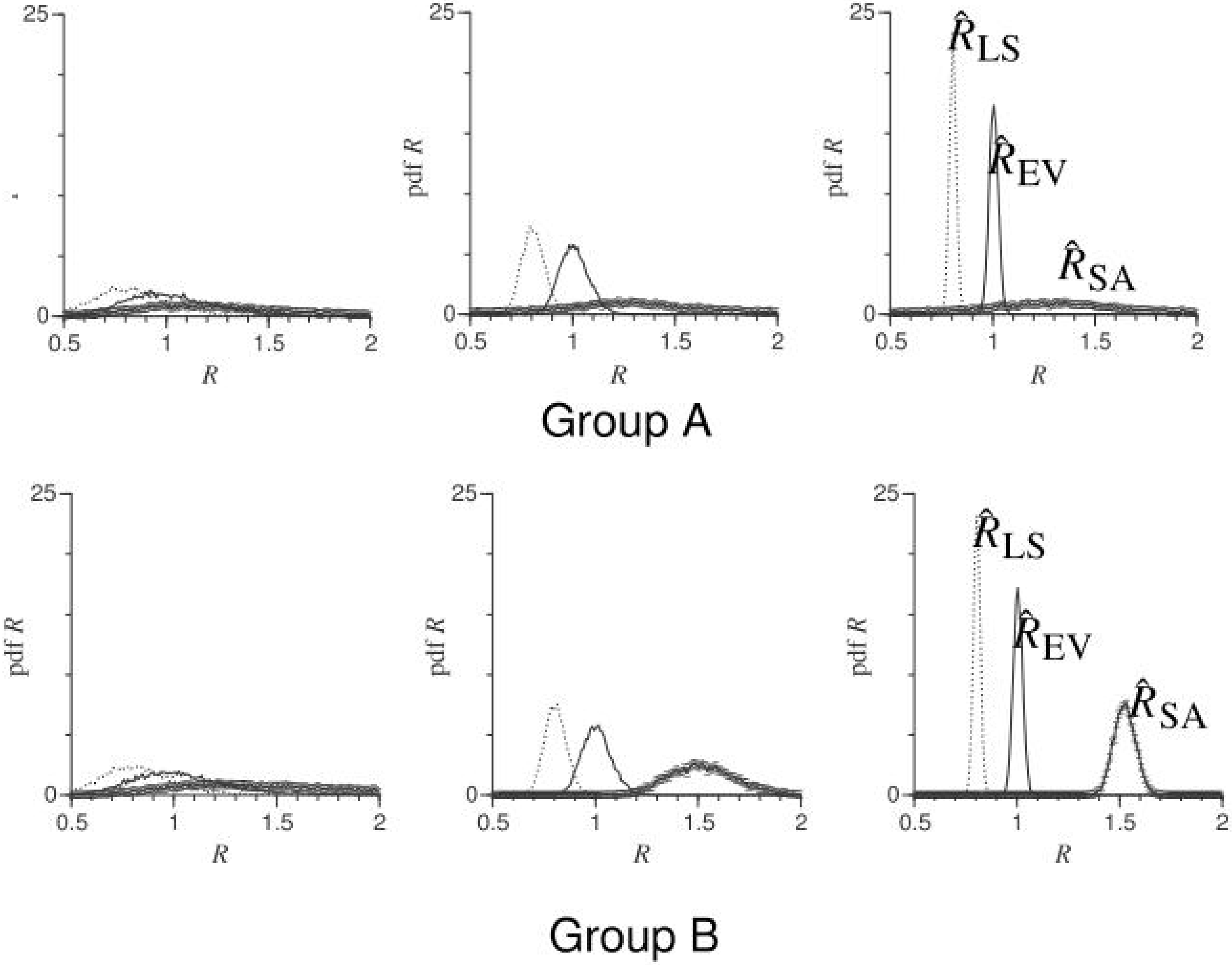



Remarks

- variations decrease as function of $N$, except for $\hat{R}_{SA}$
- the asymptotic values are different
- $\hat{R}_{SA}$ behaves 'strange'

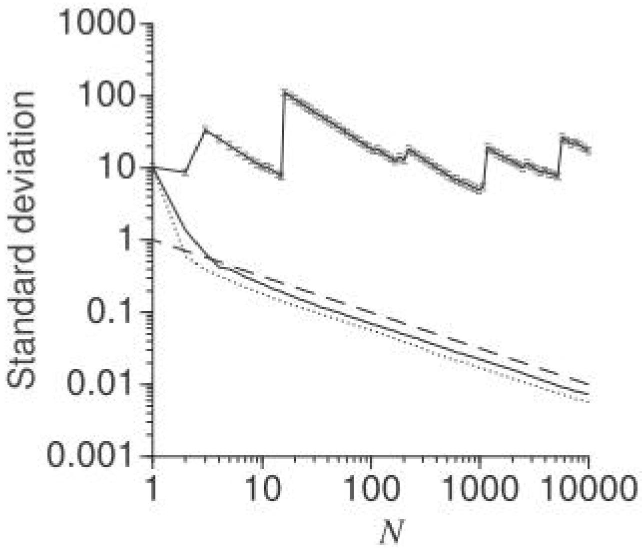
# Repeating the experiments.



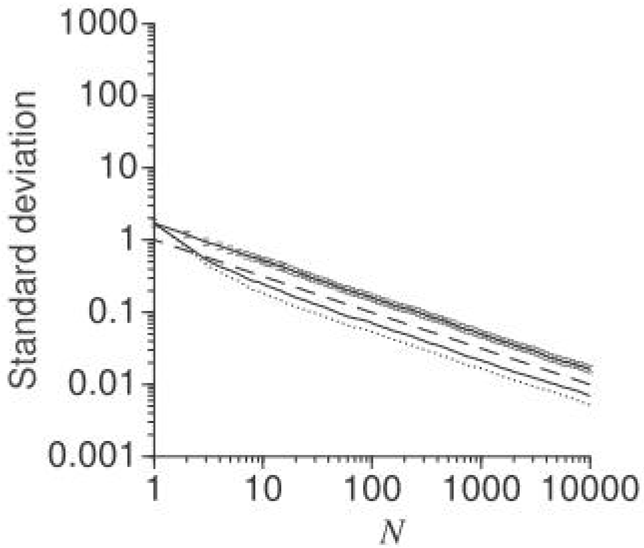

Observed pdf of $\hat{R}(N)$ for both groups, from the left tot the right $N = 10, 100$, and $1000$

- the distributions become more concentrated around their limit value
- $\hat{R}_{SA}$ behaves ‘strange’ for group A

# Repeating the experiments




GroupA




Group B

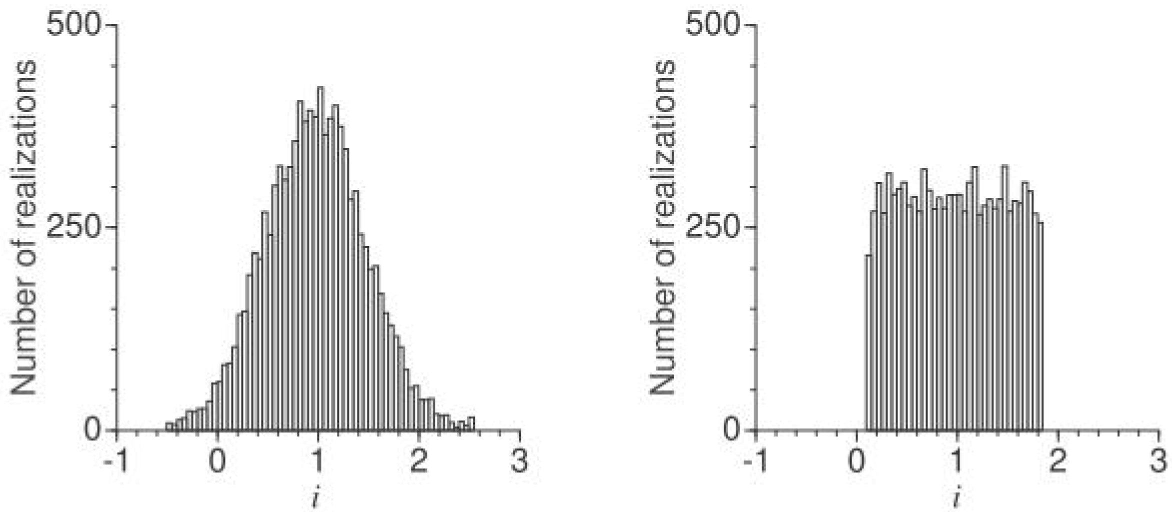
Standard deviation of $\hat{R}(N)$

full dotted line: $\hat{R}_{SA}$, dotted line: $\hat{R}_{LS}$, full line: $\hat{R}_{EV}$, dashed line $1/\sqrt{N}$.

- the standard deviation decrease in $\sqrt{N}$

- the uncertainty also depends on the estimator

# Strange behaviour of $R_{SA}$ for group A




Histogram of the current measurements.

- The current takes negative values for group A

- the estimators tend to a normal distribution
  even when the noise behaviour is completely different

# Simplified analysis

Why do the asymptotic values depend on the estimator?

Can we explain the behaviour of the variance?

Why does the $R_{\mathrm{SA}}$ estimator behave strange for group A?

**More information is needed to answer these questions**

# Noise model of the measurements

$$i(k) = i_0 + n_i(k) \qquad u(k) = u_0 + n_u(k)$$

**Assumptions**:

$n_i(k)$ and $n_u(k)$ are:

- mutually independent
- zero mean
- independent and identically distributed
- have a symmetric distribution
- variance $\sigma_u^2$ and $\sigma_i^2$.

# Statistical tools

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} x(k) = 0$$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} x(k)^2 = \sigma_x^2$$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} x(k) y(k) = 0$$

# Asymptotic value of $\hat{R}_{\mathrm{LS}}$

$$\lim_{N\to\infty}\hat{R}_{\mathrm{LS}}(N) = \lim_{N\to\infty}\left(\sum_{k=1}^{N}u(k)i(k)\right)\Big/\left(\sum_{k=1}^{N}i^2(k)\right)$$

$$= \lim_{N\to\infty}\frac{\frac{1}{N}\sum_{k=1}^{N}(u_0+n_u(k))(i_0+n_i(k))}{\frac{1}{N}\sum_{k=1}^{N}(i_0+n_i(k))^2}$$

Or

$$\lim_{N\to\infty}\hat{R}_{\mathrm{LS}}(N) =$$

$$\lim_{N\to\infty}\frac{u_0 i_0 + \frac{u_0}{N}\sum_{k=1}^{N}n_i(k) + \frac{i_0}{N}\sum_{k=1}^{N}n_u(k) + \frac{1}{N}\sum_{k=1}^{N}n_u(k)n_i(k)}{i_0^2 + \frac{1}{N}\sum_{k=1}^{N}n_i^2(k) + \frac{2i_0}{N}\sum_{k=1}^{N}n_i(k)}$$

And finally

$$\lim_{N\to\infty}\hat{R}_{\mathrm{LS}}(N) = \frac{u_0 i_0}{i_0^2+\sigma_i^2} = R_0\frac{1}{1+\sigma_i^2/i_0^2}$$

**It converges to the wrong value!!!**

# Asymptotic value of $\hat{R}_{\mathrm{EV}}$

$$\lim_{N \to \infty} \hat{R}_{\mathrm{EV}}(N) = \lim_{N \to \infty} \left(\sum_{k=1}^{N} u(k)\right) \Big/ \left(\sum_{k=1}^{N} i(k)\right)$$

$$= \lim_{N \to \infty} \frac{\frac{1}{N}\sum_{k=1}^{N} (u_0 + n_u(k))}{\frac{1}{N}\sum_{k=1}^{N} (i_0 + n_i(k))}$$

$$= \left( \lim_{N \to \infty} \frac{u_0 + \frac{1}{N}\sum_{k=1}^{N} n_u(k)}{i_0 + \frac{1}{N}\sum_{k=1}^{N} n_i(k)} \right)$$

$$= R_0$$

**It converges to the exact value!!!**

# Asymptotic value of $\hat{R}_{\mathrm{LS}}$

$$\hat{R}_{\mathrm{SA}}(N) = \frac{1}{N}\sum_{k=0}^{N}\frac{u(k)}{i(k)} = \frac{1}{N}\sum_{k=0}^{N}\frac{u_0 + n_u(k)}{i_0 + n_i(k)} = \frac{1}{N}\frac{u_0}{i_0}\sum_{k=0}^{N}\frac{1 + n_u(k)/u_0}{1 + n_i(k)/i_0}$$

The series expansion exist only for small noise distortions

$$\frac{1}{1+x} = \sum_{l=0}^{\infty}(-1)^l x^l \text{ for } |x| < 1$$

*Group A*: The expected value does not exist for the data of group A.
**The estimator does not converge**.

*Group B*: For group B the series converges and

$$\lim_{N\to\infty}\hat{R}_{\mathrm{SA}}(N) \approx R_0\left(1 + \frac{\sigma_i^2}{i_0^2}\right)$$

**The estimator converges to the wrong value!!**

# Variance expressions

First order approximation

$$\sigma^2_{R_{\mathrm{LS}}}(N) \approx \sigma^2_{R_{\mathrm{EV}}}(N) \approx \sigma^2_{R_{\mathrm{SA}}}(N) \approx \frac{R_0^2}{N}\left(\frac{\sigma_u^2}{u_0^2} + \frac{\sigma_i^2}{i_0^2}\right)$$

- variance decreases in $1/N$

- variance increases with the noise

- for low noise levels, all estimators have the same uncertainty

---> **Experiment design**

# Cost function interpretation

The previous estimates match the model $u = Ri$ as good as possible on the data.

A criterion to express the goodness of the fit is needed ----> Cost function interpretation.

$\hat{R}_{\mathrm{SA}}(N)$

$$V_{\mathrm{SA}}(R) = \frac{1}{N}\sum_{k=1}^{N}(R(k) - R)^2 .$$

$\hat{R}_{\mathrm{LS}}(N)$

$$V_{\mathrm{LS}}(R) = \frac{1}{N}\sum_{k=1}^{N}(u(k) - Ri(k))^2$$

$\hat{R}_{\mathrm{EV}}(N)$

$$V_{\mathrm{EV}}(R, i_0, u_0) = \frac{1}{N}\left(\sum_{k=1}^{N}\frac{(u(k) - u_0)^2}{\sigma_u^2} + \sum_{k=1}^{N}\frac{(i(k) - i_0)^2}{\sigma_i^2}\right) \text{ subject to } u_0 = Ri_0$$

# Conclusion

- A simple problem

- Many solutions

- How to select a good estimator?

- Can we know the properties in advance?

**Need for a general framework !!**

# Outline

A motivating example: why do *you* need system identification?!

Characterizing estimators

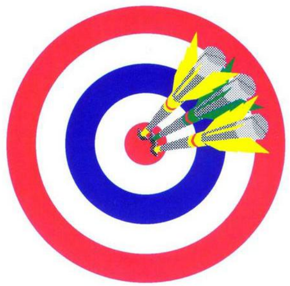
Basic steps of the identification process

A statistic approach to system identification

Identification in the presence of ‘input’ and ‘output’ noise

# Characterizing estimators

Location properties: are the parameters concentrated around the ‘exact value’ ?



Dispersion properties: is the uncertainty small or large?

# Location properties

# unbiased and consistent estimators

Unbiased estimates

the mean value equals the exact value

**Definition**

An estimator $\theta$ of the parameters $\theta_0$ is unbiased if $E\{\theta\} = \theta_0$, for all true parameters $\theta_0$. Otherwise it is a biased estimator.

Asymptotic unbiased estimates: unbiased for $N \to \infty$

# Example

## i) The sample mean

$$\bar{u}(N) = \frac{1}{N}\sum_{k=1}^{N} u(k)$$

Unbiased?

$$E\{\bar{u}(N)\} = \frac{1}{N}\sum_{k=1}^{N} E\{u(k)\} = \frac{1}{N}\sum_{k=1}^{N} u_0 = u_0$$

## ii) The sample variance

$$\sigma_u^2(N) = \frac{1}{N}\sum_{k=1}^{N} (u(k) - \bar{u}(N))^2$$

Unbiased?

$$E\{\sigma_u^2(N)\} = \frac{N-1}{N}\sigma_u^2$$

Alternative expression

$$\frac{1}{N-1}\sum_{k=1}^{N} (u(k) - \bar{u}(N))^2$$

# Example cont'd

$$\sigma_1^2 = \frac{1}{N-1}\sum_{k=1}^{N}(u(k) - \bar{u}(N))^2 \text{ and } \sigma_2^2 = \frac{1}{N}\sum_{k=1}^{N}(u(k) - \bar{u}(N))^2$$

bias $\sigma_1^2 = 0$, and bias $\sigma_2^2 = \frac{\sigma^2}{N}$

variance $\sigma_1^2$ > variance $\sigma_2^2$

RMS error $\sigma_1^2$ > RMS error $\sigma_2^2$

## Best choice?

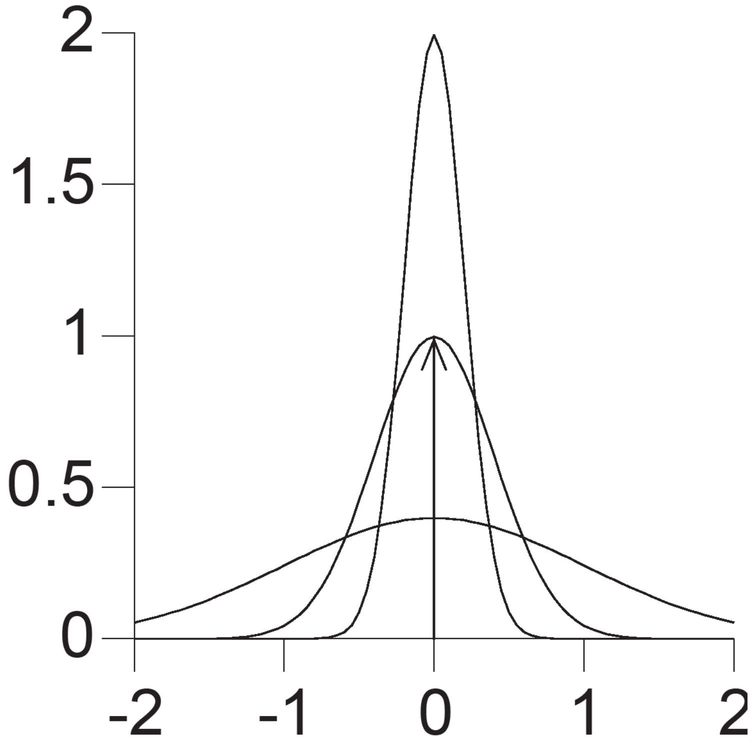
# Consistent estimates

Consistent estimates: the probability mass gets concentrated around the exact value

$$\lim_{N \to \infty} \text{Prob}\left( \left| \hat{\theta}(N) - \theta_0 \right| > \delta > 0 \right) = 0$$

or

$$\operatorname*{plim}_{N \to \infty} \hat{\theta}(N) = \theta_0$$



# Consistency

# properties

$$\operatorname{plim} f(a) = f(\operatorname{plim} a)$$

$\operatorname{plim} ab = \operatorname{plim} a \operatorname{plim} b$ if all plim exist

# Example

$$\operatorname*{plim}_{N \to \infty} \hat{R}_{\mathrm{EV}}(N) = \operatorname*{plim}_{N \to \infty} \frac{\frac{1}{N}\sum_{k=1}^{N} u(k)}{\frac{1}{N}\sum_{k=1}^{N} i(k)}$$

$$= \frac{\operatorname*{plim}_{N \to \infty}\left(\frac{1}{N}\sum_{k=1}^{N} u(k)\right)}{\operatorname*{plim}_{N \to \infty}\left(\frac{1}{N}\sum_{k=1}^{N} i(k)\right)}$$

$$= \frac{u_0}{i_0}$$

$$= R_0$$

# Dispersion properties

# efficient estimators

- Mostly the covariance matrix is used, however alternatives like percentiles exist.

- For a given data set, there exists a minimum bound on the covariance matrix:

## the Cramér-Rao lower bound.

$$CR(\theta) = Fi^{-1}(\theta_0)$$

with

$$Fi(\theta_0) = E\left\{\left(\frac{\partial}{\partial\theta}l(Z|\theta)\right)^T\left(\frac{\partial}{\partial\theta}l(Z|\theta)\right)\right\} = -E\left\{\frac{\partial^2}{\partial\theta^2}l(Z|\theta)\right\}.$$

The derivatives are calculated in $\theta = \theta_0$

# The likelihood function

1) Consider the measurements $Z \in R^N$

2)$Z$ is generated by a hypothetical, exact model with parameters $\theta_0$

3) $Z$ is disturbed by noise --> stochastic variables

4) Consider the probability density function $f(Z|\theta_0)$ with

$$\int_{Z \in R^N} f(Z|\theta_0) dZ = 1 .$$

5)Interpret this relation conversely, viz:

how likely is it that a specific set of measurements $Z = Z_m$ are generated by a system with parameters $\theta$?

--> Measurements given
Model parameters as the free variables:

$L(Z_m|\theta) = f(Z = Z_m|\theta)$, with $\theta$ the free variables

$L(Z_m|\theta)$**: likelihood function.**

# Interpretation of the Cramér-Rao lower bound

Model $y_0 = f(u_0, \theta)$

Measurement $y = y_0 + n_y$ and $n_y \sim N(0, \sigma^2)$

Likelihood function $L(y|\theta) = f_n(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - f(u_0, \theta))^2}{2\sigma^2}}$

Loglikelihood function $l(y|\theta) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y - f(u_0, \theta))^2}{2\sigma^2}$

$$\frac{\partial l}{\partial \theta} = -\frac{(y - f(u_0, \theta))}{\sigma^2} \frac{\partial}{\partial \theta} f(u_0, \theta)$$

Information matrix

$$Fi(\theta_0) = E\left\{\left(\frac{\partial l}{\partial \theta}\right)^T \left(\frac{\partial l}{\partial \theta}\right)\right\}$$

$$= E\left\{\frac{(y - f(u_0, \theta))^2}{\sigma^4}\left(\frac{\partial}{\partial \theta} f(u_0, \theta)\right)^2\right\}$$

$$= \frac{1}{\sigma^2}\left(\frac{\partial}{\partial \theta} f(u_0, \theta)\right)^2$$

# Example

Determine the flow of tap water by measuring the height $h_0(t)$ of the water in a measuring jug as a function of time $t$

**Model**

$$h_0(t) = a(t - t_{\text{start}}) = at + b \text{ with } \theta = [a, b]$$

**Measurements**

$$h(k) = at_k + b + n_h(k)$$

**Noise model**

$$n_h(k)\text{: iid zero mean normally distributed } N(0, \sigma^2)$$

**Likelihood function**

for the set of measurements $h = \{(h(1), t_1), ..., (h(N), t_N)\}$ :

$$L(h|a, b) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(h(k) - at_k - b)^2}$$

# Example Continued

**Log likelihood function**

$$l(h|a, b) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(h(k) - at_k - b)^2$$

**Information matrix**

$$Fi(\theta_0) = E\left\{\left(\frac{\partial}{\partial\theta}l(Z|\theta)\right)^T\left(\frac{\partial}{\partial\theta}l(Z|\theta)\right)\right\} = -E\left\{-\frac{\partial^2}{\partial\theta^2}l(Z|\theta)\right\}$$

$$Fi(a, b) = \frac{1}{\sigma^2}\begin{bmatrix} Ns^2 & N\mu \\ N\mu & N \end{bmatrix},$$

**Cramér-Rao lower bound**

$$CR(a, b) = \frac{\sigma^2}{N(s^2 - \mu^2)}\begin{bmatrix} 1 & -\mu \\ -\mu & s^2 \end{bmatrix}$$

with $\mu = \frac{1}{N}\sum_{k=1}^{N} t_k$ and $s^2 = \frac{1}{N}\sum_{k=1}^{N} t_k^2$.

# Example continued

Case 1: $a$ and $b$ unknown: consider $Fi^{-1}(a, b)$

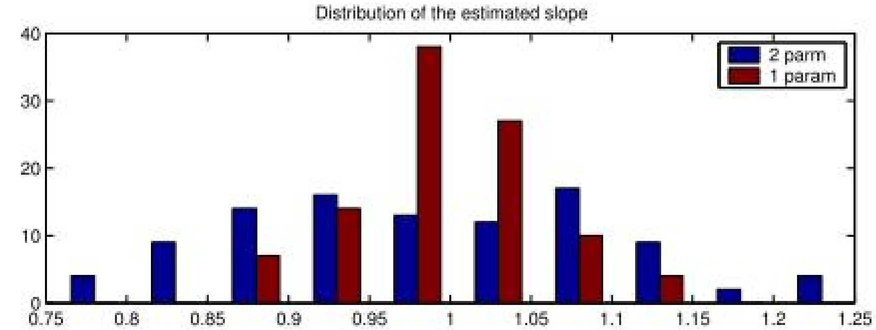
$$\sigma_a^2(a, b) = \frac{\sigma^2}{N(s^2 - \mu^2)}$$

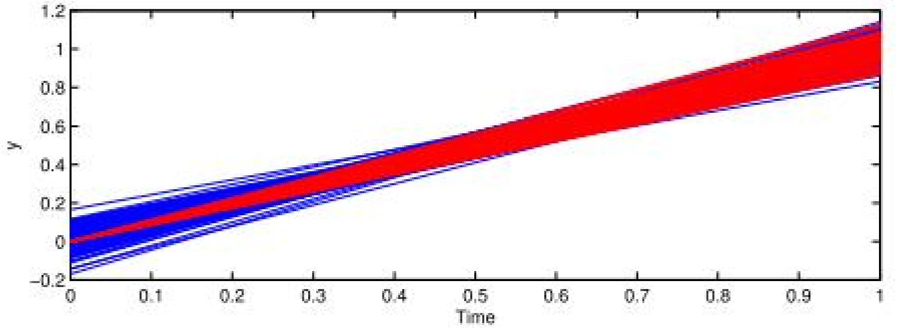
Case 2: $a$ unknown: consider $Fi^{-1}(a)$

$$\sigma_a^2(a) = \frac{\sigma^2}{Ns^2}$$

Discussion points

- impact of the number of measurements
- impact of the number of parameters
- the analysis is done without selecting an estimator

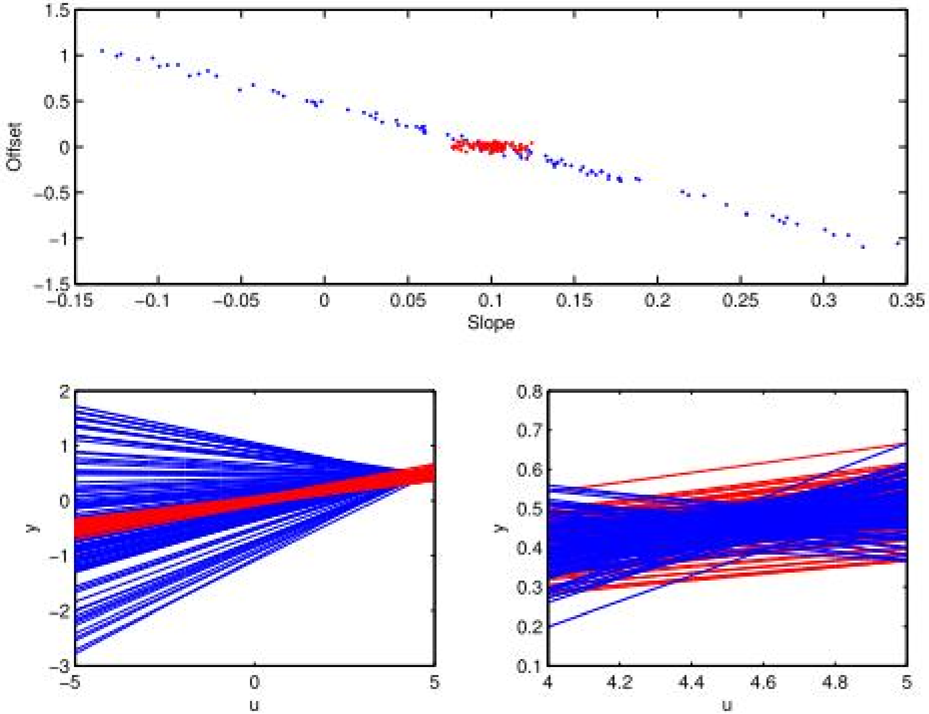



$y(t_k) = a_0 t_k + n_k$, with $t_k = [0:N]/N$ and $n_k \sim N(0, \sigma^2) = 1$, $a_0 = 1$

model 1: $y = at + b$ (two parameters)

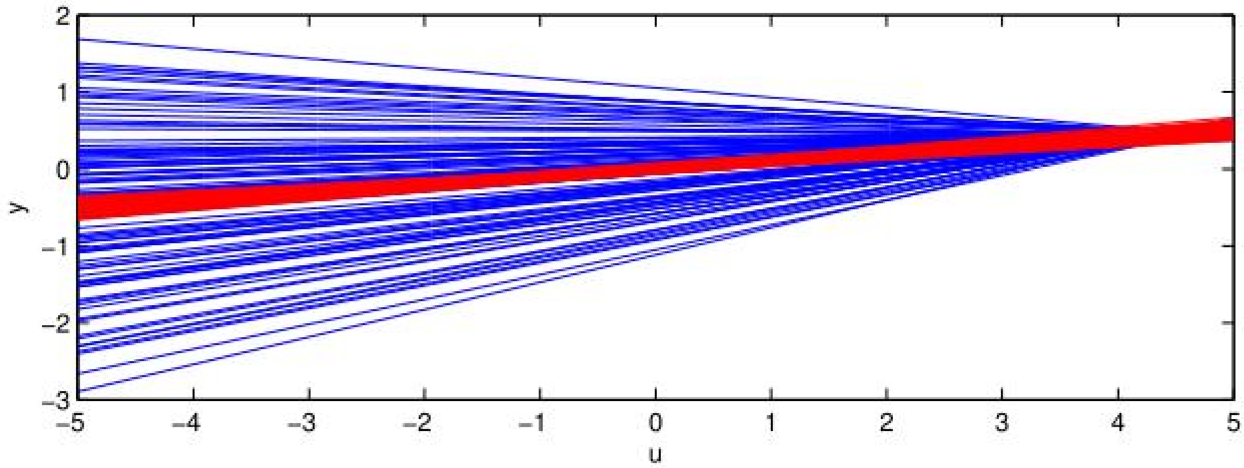
model 2: $y = at$ (one parameter)

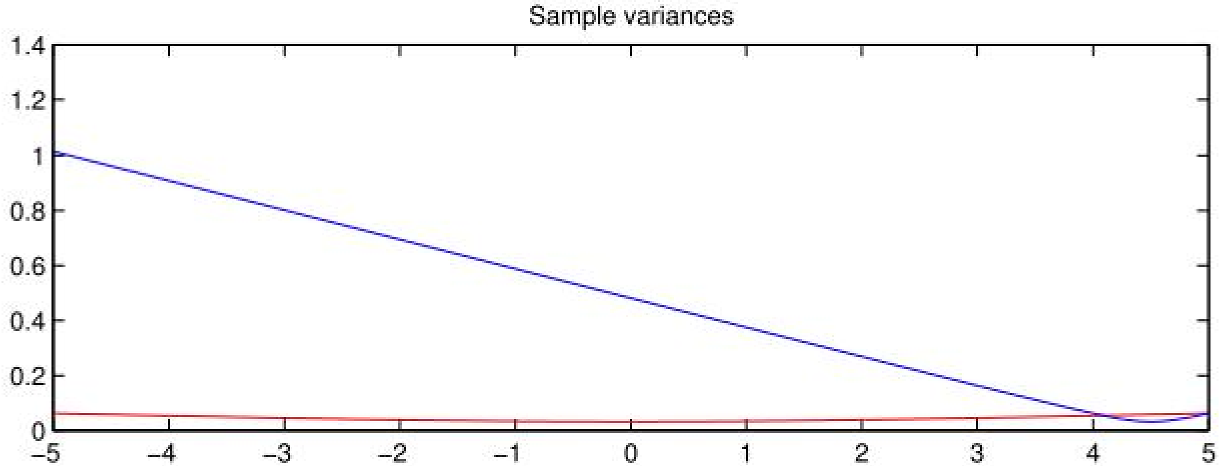



IImpact experimtent design: red [-5:5], blue [4 5]

$$C_{\mathrm{exp1}} = \begin{bmatrix} 1.5\times10^{-4} & 0.48\times10^{-4} \\ 0.48\times10^{-4} & 9.1\times10^{-4} \end{bmatrix} \quad \text{, and} \quad C_{\mathrm{exp2}} = \begin{bmatrix} 1.5\times10^{-2} & -6.7\times10^{-2} \\ -6.7\times10^{-2} & 30.1\times10^{-2} \end{bmatrix}$$

$$R_{\mathrm{exp1}} = \begin{bmatrix} 1 & 0.13 \\ 0.13 & 1 \end{bmatrix} \quad \text{, and} \quad R_{\mathrm{exp2}} = \begin{bmatrix} 1 & -0.9985 \\ -0.9985 & 1 \end{bmatrix}$$




Interpretation of the covariance matrix, and the impact of the experiment design on the model uncertainty.

# Characterizing estimators

- Goal asymptotic analysis
  - what happens with the estimate if more data are gathered?
  - hypothesis: asymptotic behaviour reflects finite sample behaviour
  - true finite sample behaviour is in general very difficult to establish (exception: linear least squares)

- Consistency
  - convergence in *stochastic* sense to the true value
  - does not exclude divergence for some realisations
  - guarantees with high probability that estimate is close to true value
  - consistency does not imply asymptotic unbiasedness
  - proof: law of large numbers

- Asymptotic unbiasedness
  - asymptotically the expected value equals the true value
  - does not guarantee that the estimate is close to the true value
  - in general very difficult to verify (exception: linear least squares)

- Asymptotic normality
  - allows to construct uncertainty bounds with a given confidence level
  - proof: linearisation around limit value + central limit theorem

# Characterizing estimators (Cont'd)

- Asymptotic variance
  - measure of the convergence rate
  - construction of uncertainty bounds
  - proof: linearisation around limit value

- Asymptotic efficiency
  - minimal uncertainty within class of asymptotically unbiased estimators
  - in practice applied to class of consistent estimators
  - Cramér-Rao lower bound does not always exist
    or may be too conservative
  - proof: comparison asymptotic variance with Cramér-Rao lower bound

- Robustness
  - sensitivity (asymptotic) properties to the noise assumptions
  - proof: validity law of large numbers and central limit theorem
    under relaxed noise assumptions

# Outline

A motivating example: why do *you* need system identification?!

Characterizing estimators

Basic steps of the identification process

A statistic approach to system identification

Identification in the presence of ‘input’ and ‘output’ noise

# Basic steps in identification

1) collect the information: experiment setup

2) select a model

parametric >< nonparametric models
white>< black box models
linear><nonlinear models
linear -in-the-parameters><nonlinear-in-the-parameters

$$\varepsilon = y - (a_1 u + a_2 u^2), \ \varepsilon(\omega) = Y(\omega) - \frac{a_0 + a_1 j\omega}{b_0 + b_1 j\omega} U(\omega)$$

3) match the model to the data

select a cost function
LS, WLS, MLE, Bayes estimation

4) validation

does the model explain the data?
can it deal with new data?

Remark

this scheme is not only valid for the classical identification theory.
It also applies to neural nets, fuzzy logic, ...

# Outline

A motivating example: why do *you* need system identification?!

Characterizing estimators

Basic steps of the identification process

A statistic approach to system identification

Identification in the presence of ‘input’ and ‘output’ noise

# A statistical framework: choice of the cost functions

$$y_0 = G(u, \theta_0),\ y = y_0 + n_y,\ e = y - G(u, \theta_0)$$

Least squares estimation

$$V_{\text{LS}}(\theta) = \frac{1}{N}\sum_{k=1}^{N} e^2(k, \theta)$$

Weighted least squares estimation

$$V_{\text{WLS}}(\theta) = \frac{1}{N} e(\theta)^T W e(\theta)$$

Maximum likelihood estimation

$$f(y|\theta_0) = f_{n_y}(y - G(u, \theta_0))$$

$$\theta_{\text{ML}} = \underset{\theta}{\text{argmax}} f(y_m|\theta)$$

# Least squares: principle

Model

$$y_0(k) = g(u_0(k), \theta)$$

with $k$ the measurement index, and

$$y(k) \in \mathrm{R},\; u(k) \in \mathrm{R}^{1xM},\; \theta \in \mathrm{R}^{n_\theta x1}$$

Measurements

$$y(k) = y_0(k) + n_y(k)$$

Match model and measurements
Choose:

$$e(k, \theta) = y(k) - y(k, \theta),$$

with $y(k, \theta)$ the modelled output.

Then

$$\theta_{\mathrm{LS}} = \arg\min_{\theta} V_{\mathrm{LS}}(\theta),$$

with

$$V_{\mathrm{LS}}(\theta) = \frac{1}{N}\sum_{k=1}^{N} e^2(k, \theta)$$

# Least squares: special case
# model that is linear-in-the-parameters

$$y_0 = K(u_0)\theta_0$$

$$e(\theta) = y - K(u)\theta, \; K = -\frac{\partial e}{\partial \theta}$$

$$\theta_{LS} = (K^T K)^{-1} K^T y$$

# Properties

$$\theta_{LS} = (K^T K)^{-1} K^T y$$

Noise assumptions

$$y = y_0 + n_y \text{ with } E[n_y] = 0$$

# Bias?

$$E[\theta_{LS}] = E\left[(K^T K)^{-1} K^T y\right]$$

$$= (K^T K)^{-1} K^T E[y]$$

$$= (K^T K)^{-1} K^T y_0$$

Note that $y_0 = K\theta_0$

$$E[\theta_{LS}] = (K^T K)^{-1} K^T K \theta_0 = \theta_0$$

# Properties

$$\theta_{LS} = (K^T K)^{-1} K^T y$$

Noise assumptions

$$y = y_0 + n_y \text{ with } E[n_y] = 0$$

# Noise sensitivity?

$$\theta_{LS} - \theta_0 = (K^T K)^{-1} K^T n_y$$

$$= \left(\frac{1}{N} K^T K\right)^{-1} \frac{1}{N} K^T n_y$$

Note

$$\lim_{N \to \infty} \frac{1}{N} K^T n_y \sim N\left(0, K^T \frac{C_y}{N^2} K\right)$$

Covariance matrix

$$C_\theta = \left(\frac{1}{N} K^T K\right)^{-1} K^T \frac{C_y}{N^2} K \left(\frac{1}{N} K^T K\right)^{-1}$$

# Asymptotical normal distributed

# Example: weight of a loaf of bread

model

$$y_0 = \theta_0$$

measurements

$$y(k) = y_0 + n_y(k)$$

estimator

$$e(k) = y(k) - \theta$$

Standard formulation

$$y = K\theta + n_y \text{ with } K = (1, 1, \dots, 1)^T$$

Solution

$$\theta_{LS} = (K^T K)^{-1} K^T y = \frac{1}{N}\sum_{k=1}^{N} y(k)$$

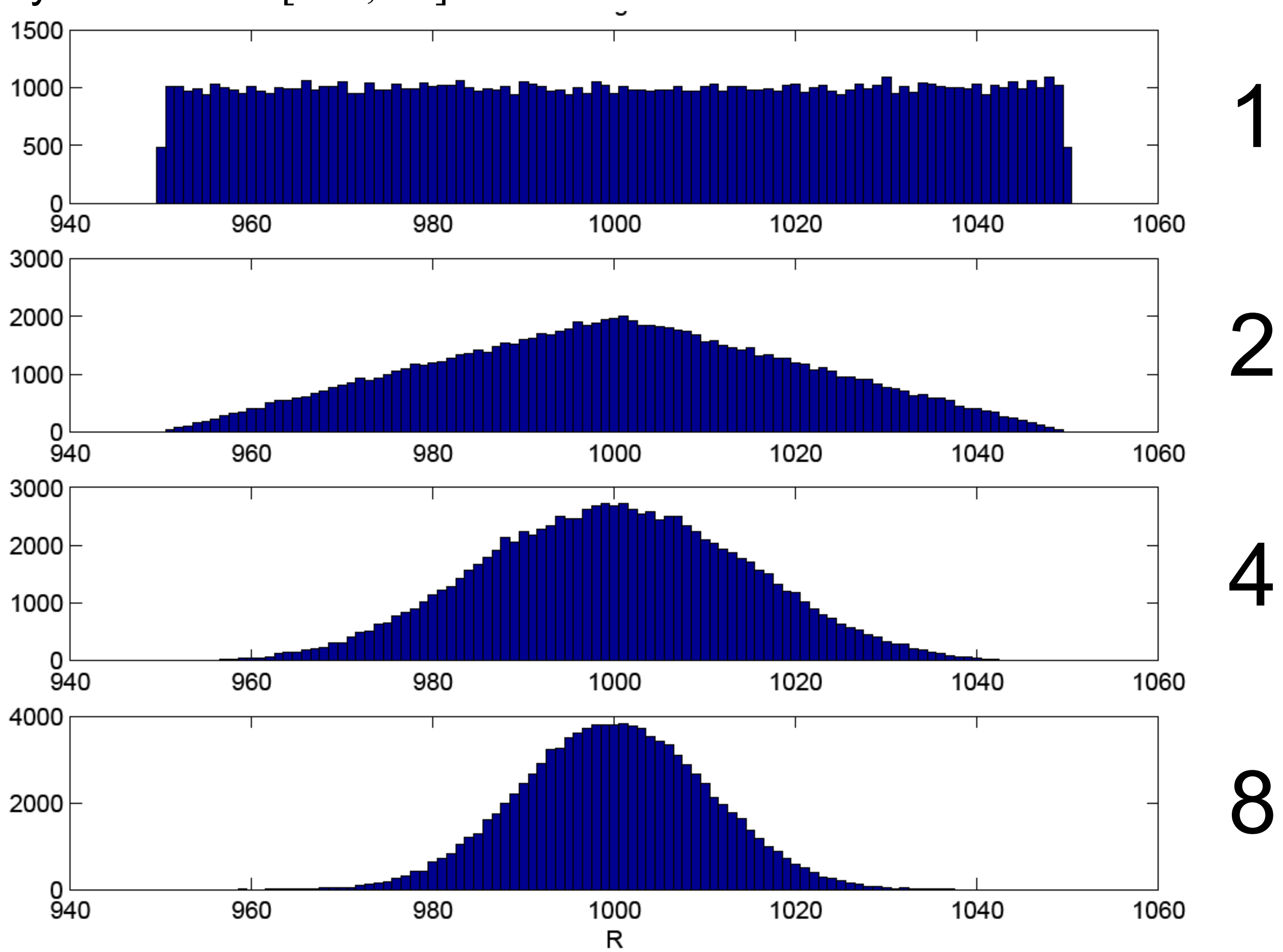
$$\sigma_\theta^2 = \left(\frac{1}{N}K^T K\right)^{-1} K^T \frac{C_y}{N^2} K \left(\frac{1}{N}K^T K\right)^{-1} = \frac{\sigma_y^2}{N}, \text{ for white noise}$$

# Example: weight of a loaf of bread (Cont'd)

measurements

$$y(k) = y_0 + n_y(k), \; k = 1, \ldots, N$$

noise uniformly distributed $[-50, 50]$



# Weighted least squares

Goal: bring your confidence in the measurements into the problem

Model

$$y_0(k) = g(u_0(k), \theta)$$

$$y(k) \in \mathrm{R},\ u(k) \in \mathrm{R}^{1xM},\ \theta \in \mathrm{R}^{n_\theta x1}$$

Measurements

$$y(k) = y_0(k) + n_y(k)$$

confidence in measurement $k$: $w(k)$

Match model and measurements

$$e(k, \theta) = y(k) - y(k, \theta),$$

Then

$$\theta_{\mathrm{LS}} = \arg\min_{\theta} V_{\mathrm{LS}}(\theta),$$

with

$$V_{\mathrm{WLS}}(\theta) = \frac{1}{N}\sum_{k=1}^{N} \frac{e^2(k, \theta)}{w(k)}$$

# Weighted least squares (continued)

Generalization: use a full matrix to weight the measurements

define

$$e(\theta) = (e(1, \theta) \ldots e(N, \theta))$$

consider a positive definite matrix $W$

Then

$$V_{\mathrm{WLS}}(\theta) = \frac{1}{N} e(\theta)^T W^{-1} e(\theta)$$

Special choice:

$$W = E\{n_y n_y^T\} = C_{n_y n_y} \in R^{NxN}$$

This choice minimizes $C_\theta$

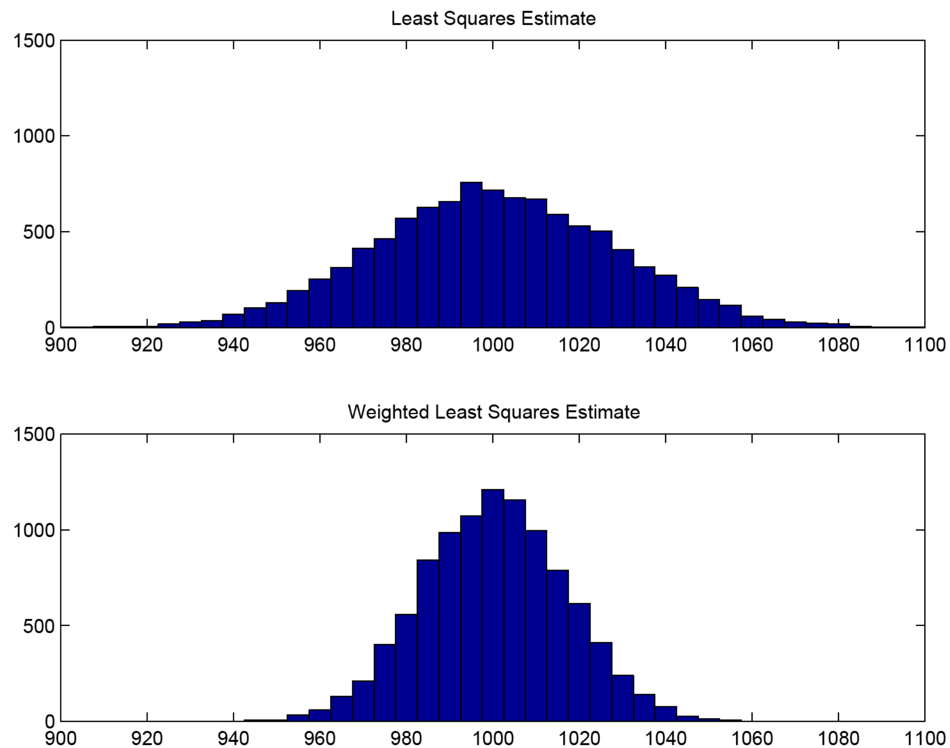
# Example: resistor measurement, 2 voltmeters

$$\begin{aligned} i(k) &= i_0(k) \\ u(k) &= u_0(k) + n_u(k) \end{aligned}, \quad k = 1, 2, \ldots, 100$$

$i_0$ uniformly distributed in $[0, 0.01]$

$R_0 = 1000$

Voltmeter 1: $N(0, \sigma_u^2 = 1)$, Voltmeter 2: $N(0, \sigma_u^2 = 3^2)$



# Maximum likelihood estimation

Model

$$y_0(k) = g(u_0(k), \theta)$$

$$y(k) \in \mathrm{R},\ u(k) \in \mathrm{R}^{1xM},\ \theta \in \mathrm{R}^{n_\theta x1}$$

Measurements

$$y(k) = y_0(k) + n_y(k)$$

with $f_{n_y}$ the pdf of the noise $n_y$

Match model and measurements
Choose the experiments such that the model becomes most likely:

Then

$$\theta_{\mathrm{ML}} = \arg\max_{\theta} f(y_m|\theta, u)$$

with

$$f(y|\theta, u) = f_{n_y}(y - G(u, \theta))$$

# Maximum likelihood: example
# weight of a loaf of bread

Model:

$$y_0 = \theta_0$$

Measurements:

$$y(k) = y_0 + n_y(k)$$

Additional information

The distribution $f_y$ of $n_y$ is normal with zero mean and standard deviation $\sigma_y$

Likelihood function:

$$f(y|\theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(y-\theta)^2}{2\sigma_y^2}} = \frac{1}{\sqrt{2\pi\sigma_y^2}^N} e^{-\frac{1}{2\sigma_y^2}\sum_{k=1}^{N}(y(k)-\theta)^2}$$

Maximum likelihood estimator:

$$\theta_{ML} = \frac{1}{N}\sum_{k=1}^{N} y(k)$$

# Resistance example with Gaussian and Laplace noise

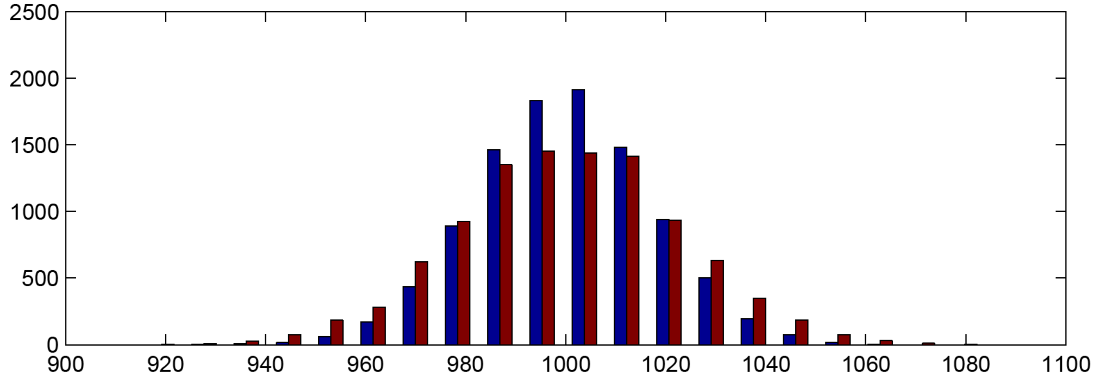
white Gaussian noise

$$\arg\min_{R} \sum_{k=1}^{N} |u(k) - Ri(k)|^2$$ --> least squares

white Laplace noise

$$\arg\min_{R} \sum_{k=1}^{N} |u(k) - Ri(k)|$$ --> least absolute values

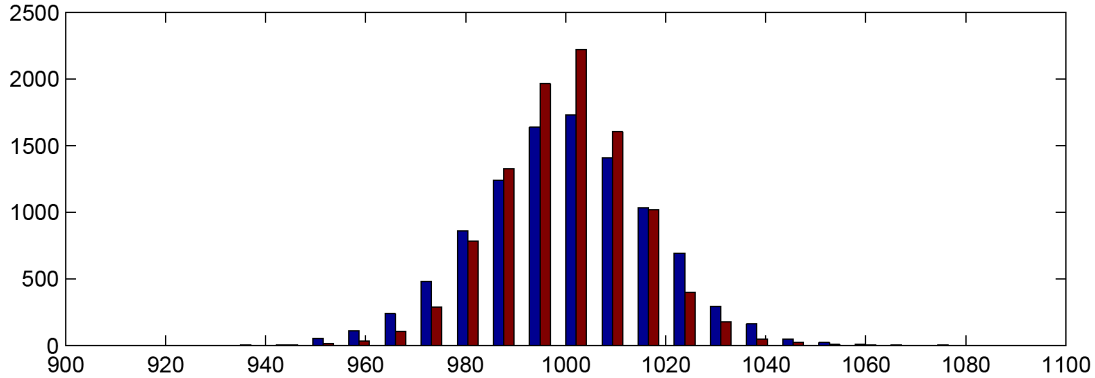
Gaussian




Least Squares

Least Abs Values

Laplace



# Properties of the Maximum likelihood estimator

***principle of invariance***: if $\theta_{\mathrm{ML}}$ is a MLE of $\theta \in \mathrm{R}^{n_\theta}$, then $\theta_g = g(\theta_{\mathrm{ML}})$ is a MLE of $g(\theta)$ where $g$ is a function, $\theta_g \in \mathrm{R}^{n_g}$ and $n_g \leq n_\theta$, with $n_\theta$ a finite number.

***consistency***: if $\theta_{\mathrm{ML}}(N)$ is an MLE based on $N$ iid random variables, with $n_\theta$ independent of $N$, then $\theta_{\mathrm{ML}}$ converges to $\theta_0$ almost surely: $\mathrm{a.s.}\lim_{N \to \infty} \theta_{\mathrm{ML}}(N) = \theta_0$.

***asymptotic normality***: if $\theta_{\mathrm{ML}}(N)$ is a MLE based on $N$ iid random variables, with $n_\theta$ independent of $N$, then $\theta_{\mathrm{ML}}(N)$ converges in law to a normal random variable with the Cramér-Rao lower bound as covariance matrix.

# Bayes estimator: principle

Choose the parameters that have the highest probability:

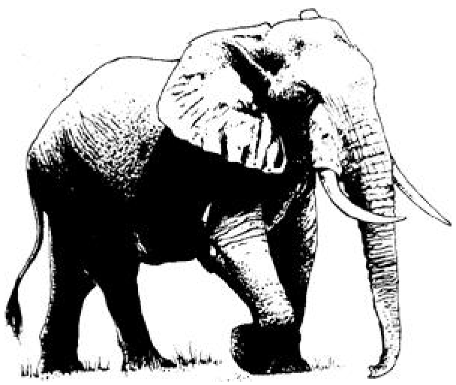
$$\theta = \arg\max_{\theta} f(\theta | u, y)$$

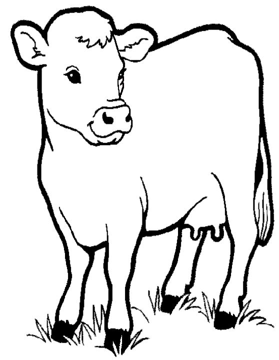
Problem: prior distribution of the parameters is required

$$f(\theta | u, y) = \frac{f(y | \theta, u) f(\theta)}{f(y)}$$

# Bayes estimator: example 1

Use of Bayes estimators in our daily life





# Bayes estimator: example 2

weight of a loaf of bread

Model:

$$y_0 = \theta_0$$

Measurements:

$$y(k) = y_0 + n_y(k)$$

Additional information 1: disturbing noise

The distribution $f_y$ of $n_y$ is $N(0, \sigma_y^2)$

Additional information 2: prior distribution of the parameters

The bread is normally distributed: $N(800\text{gr}, \sigma_w)$

Bayes estimator:

$$f(y|\theta)f(\theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(y-\theta)^2}{2\sigma_y^2}} \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{(\theta-w)^2}{2\sigma_w^2}}$$

$$\theta = \frac{z/\sigma_y^2 + w/\sigma_w^2}{1/\sigma_y^2 + 1/\sigma_w^2}$$

# Example continued

After making several independent measurements $y(1), \ldots, y(N)$

$$f(y|\theta)f(\theta) = \frac{1}{\left(\sqrt{2\pi\sigma_y^2}\right)^N} e^{-\sum_{k=1}^{N} \frac{(y(k)-\theta)^2}{2\sigma_y^2}} \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{(\theta-w)^2}{2\sigma_w^2}}$$

the Bayes estimator becomes

$$\theta = \frac{\sum_{k=1}^{N} y(k)/\sigma_y^2 + w/\sigma_w^2}{N/\sigma_y^2 + 1/\sigma_w^2}$$

For a large number of measurements:

$$\theta = \frac{\sum_{k=1}^{N} y(k)/\sigma_y^2}{N/\sigma_y^2} = \frac{1}{N}\sum_{k=1}^{N} y(k)$$

# Outline

A motivating example: why do *you* need system identification?!

Characterizing estimators

Basic steps of the identification process

A statistic approach to system identification

Identification in the presence of ‘input’ and ‘output’ noise

# Identification in the presence of input and output noise

Model

$$y_0(k) = g(u_0(k), \theta_0)$$

Measurements

$$u(k) = u_0(k) + n_u(k)$$
$$y(k) = y_0(k) + n_y(k)$$

Multiple solutions

- MLE formulation --> errors-in-variables EIV
- instrumental variables
- total least squares

# Problem

Noise on the regressor --> systematic errors

$$\hat{R}_{LS}(N) = \frac{\frac{1}{N}\sum_{k=1}^{N} u(k)i(k)}{\frac{1}{N}\sum_{k=1}^{N} i^2(k)}$$

$$\lim_{N \to \infty} \hat{R}_{LS}(N) = R_0 \frac{1}{1 + \sigma_i^2 / i_0^2} \qquad (0\text{-}1)$$

or in general

$$\theta_{LS}(N) = (K^T K)^{-1} K^T y$$

quadratic terms --> bias

# Errors-in-variables (MLE)

**Model**

$$y_0(k) = g(u_0(k), \theta_0)$$

**Measurements**

$$u(k) = u_0(k) + n_u(k)$$
$$y(k) = y_0(k) + n_y(k)$$

, pdf of the noise $n_u$ --> $f_{n_u}$, $n_y$ --> $f_{n_y}$

for simplicity: $f(n_u, n_y) = f_{n_u} f_{n_u}$

**Cost:** likelihood function

$$f((y, u)|(y_0, u_0, \theta_0)) = f_{n_y}(y - y_0|y_0, \theta_0) f_{n_u}(u - u_0|u_0, \theta_0)$$

with

$$y_0(k) = g(u_0(k), \theta_0)$$

**Parameters**

the model parameters $\theta_0$
the unknown, true input and output: $u_0(k), y_0(k)$
Note that the number of parameters depends on $N$ !!!!!

# EIV

## Example: Resistance

model

$$u_0(k) = Ri_0(k)$$

measurements

$$u(k) = u_0(k) + n_u(k)$$

$$i(k) = i_0(k) + n_i(k)$$

noise model

$n_u(k), n_i(k)$ i.i.d. zero mean normally distributed

$$n_u(k) \dashrightarrow N(0, \sigma_u^2)$$

$$n_i(k) \dashrightarrow N(0, \sigma_i^2)$$

# Example EIV (cont'd)

likelihood function

$$f((i, u)|(i_0, u_0, \theta_0)) = f_{n_i}(i - y_0|y_0, \theta_0) f_{n_u}(u - u_0|u_0, \theta_0)$$

with

$$u_0(k) = R i_0(k)$$

or

$$\frac{1}{(\sqrt{2\pi\sigma_i^2})^N} e^{-\sum_{k=1}^{N} \frac{(i(k) - i_0(k))^2}{\sigma_i^2}} \frac{1}{(\sqrt{2\pi\sigma_u^2})^N} e^{-\sum_{k=1}^{N} \frac{(u(k) - u_0(k))^2}{\sigma_u^2}}$$

The cost function becomes

$$V_{\mathrm{MLE}}(i, u, \theta) = \sum_{k=1}^{N} \frac{(i(k) - i_0(k))^2}{\sigma_i^2} + \frac{(u(k) - u_0(k))^2}{\sigma_u^2}$$

with

$$u_0(k) = R i_0(k)$$

# Example EIV (cont'd)

Elimination of $i_0, u_0$

$$V_{\text{MLE}}(i, u, \theta) = \sum_{k=1}^{N} \frac{(Ri(k) - u(k))^2}{\sigma_u^2 + R\sigma_i^2}$$

Solution

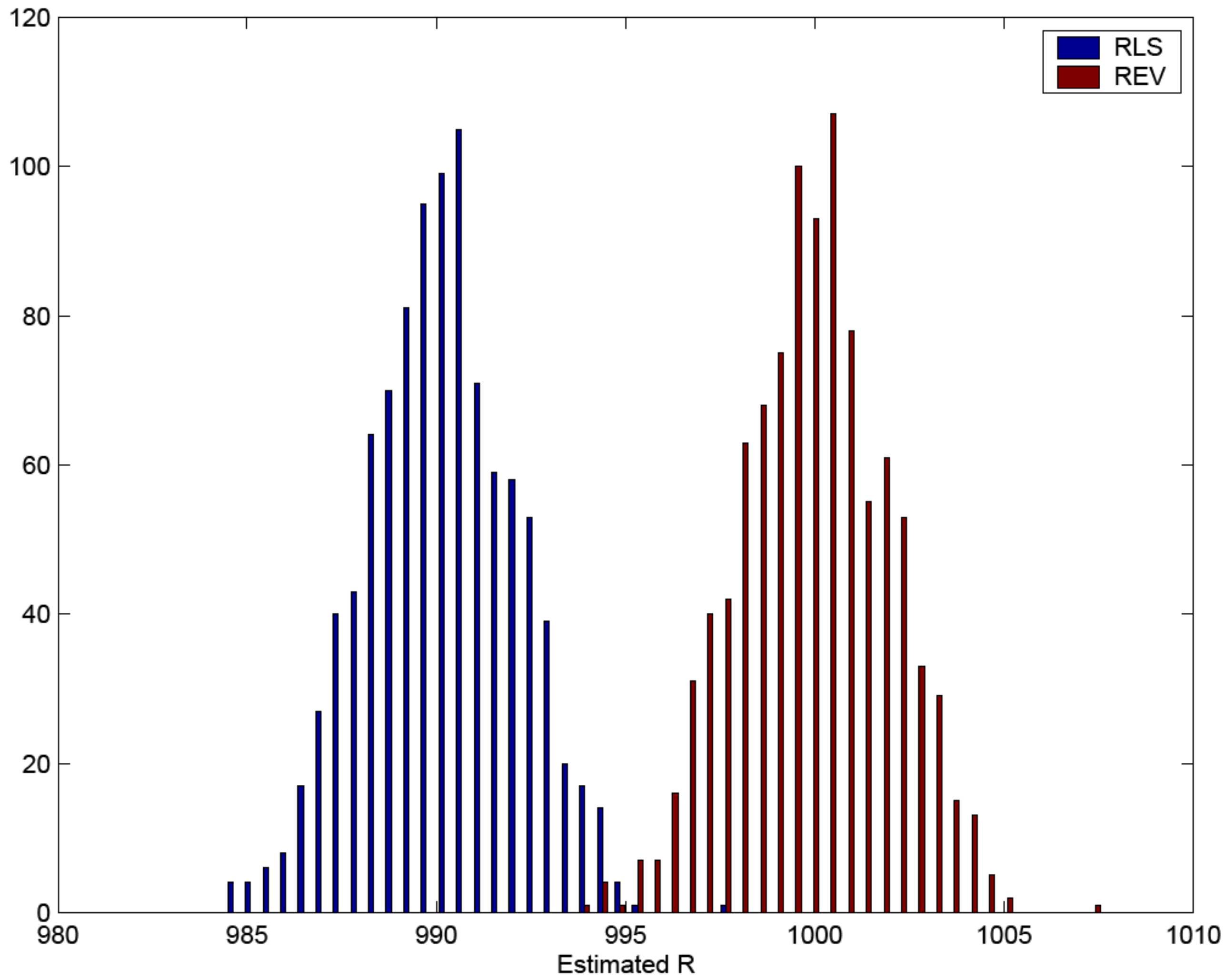
$$\hat{R}_{\text{EIV}} = \frac{\frac{\sum u(k)^2}{\sigma_u^2} - \frac{\sum i(k)^2}{\sigma_i^2} + \sqrt{\left(\frac{\sum u(k)^2}{\sigma_u^2} - \frac{\sum i(k)^2}{\sigma_i^2}\right)^2 + 4\frac{(\sum u(k)i(k))^2}{\sigma_u^2\sigma_i^2}}}{\frac{\sum u(k)i(k)}{\sigma_u^2}},$$

compare to the Least Squares estimator

$$\hat{R}_{\text{LS}} = \frac{\sum_{k=1}^{N} u(k)i(k)}{\sum_{k=1}^{N} i(k)^2}$$

# Example: Resistance

$$R_0 = 1000,\ i_0 : N(0, 0.01^2),\ \sigma_i^2 = 0.001^2,\ \sigma_u^2 = 1$$



# Outline

A motivating example: why do *you* need system identification?!

Characterizing estimators

Basic steps of the identification process

A statistic approach to system identification

Identification in the presence of ‘input’ and ‘output’ noise