INSTITUTE OF ECONOMIC STUDIES, FACULTY OF SOCIAL SCIENCES
CHARLES UNIVERSITY IN PRAGUE (*established 1348*)

ROBUST STATISTICS AND ECONOMETRICS

INSTITUTE OF ECONOMIC STUDIES
FACULTY OF SOCIAL SCIENCES
CHARLES UNIVERSITY IN PRAGUE

JAN ÁMOS VÍŠEK

Week 5

Content of lecture

1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here

Content of lecture

1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here
2. Estimators alternative to the classical ones
   - Locatin parameter
   - Scale parameter
   - General parameter

Content

1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here
2. Estimators alternative to the classical ones
   - Locatin parameter
   - Scale parameter
   - General parameter

Content

1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here
2. Estimators alternative to the classical ones
   - Locatin parameter
   - Scale parameter
   - General parameter

Analysis of the export from the Czech republic to EU in 1994

Number of industries 91

X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest producers,*
$TFPW_\ell$	-	*by wages normed productvity in the i-th industry,*
Bal_ℓ	-	*Balasa index in the i-th industry,,*
DP_ℓ	-	*cost discontinuity in 1993 in the i-th industry*
	-	*etc., about 20 explanatory variables*

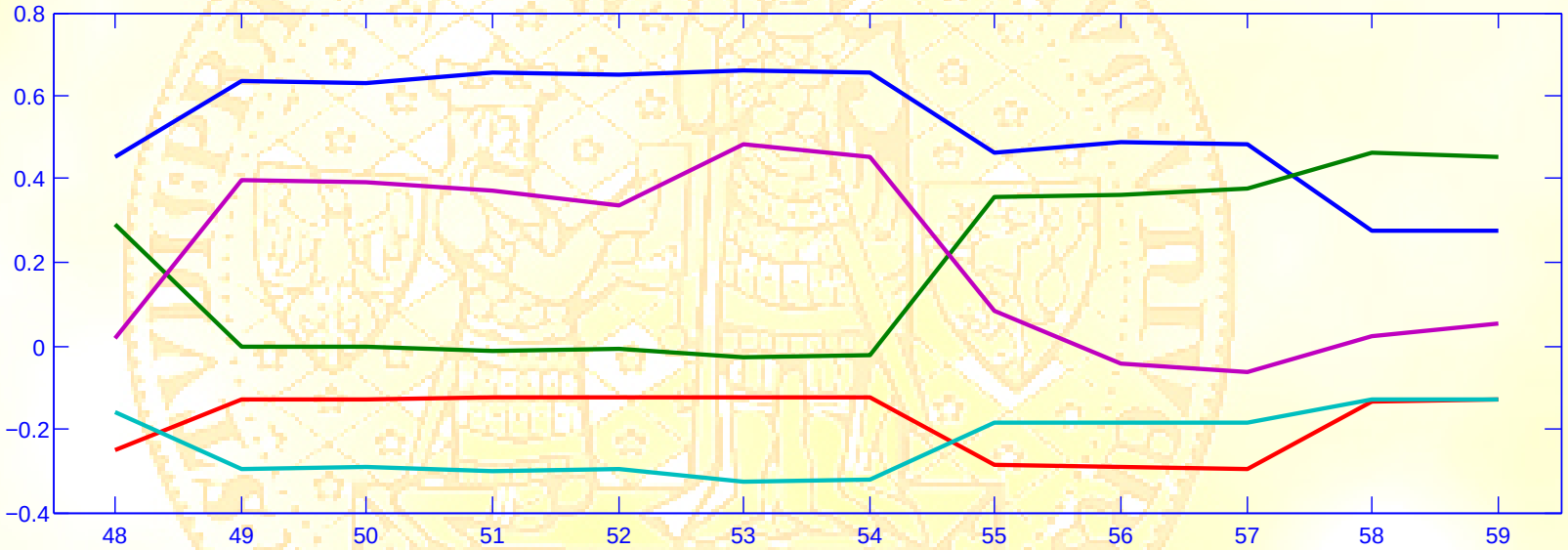
Analysis of the export from the Czech republic to EU in 1994

Number of industries 91

X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest producers,*
$TFPW_\ell$	-	*by wages normed productvity in the i-th industry,*
Bal_ℓ	-	*Balasa index in the i-th industry,,*
DP_ℓ	-	*cost discontinuity in 1993 in the i-th industry*
	-	*etc., about 20 explanatory variables*

No reasonable model by OLS - coefficient of determination 0.28

ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *least trimmed squares*.



The development of the estimates of regression coefficients. The blue line represents $\hat{\beta}_1^{(LTS,n,h)}$ (down-scaledby $\frac{1}{10}$), the purple is $\hat{\beta}_8^{(LTS,n,h)}$, the green is $\hat{\beta}_3^{(LTS,n,h)}$, the red is $\hat{\beta}_4^{(LTS,n,h)}$ and light blue (the lowest curve) is $\hat{\beta}_6^{(LTS,n,h)}$ (down-scaled again by $\frac{1}{10}$). There is an evidnet break at 54.

Analysis of the export from the Czech republic to EU in 1994 by means of the *least trimmed squares*

has found:

Main subgroup

with number of industries 54 and model

$$\frac{X_\ell}{S_\ell} = 4.64 - 0.032 \cdot \frac{US_\ell}{VA_\ell} - 0.022 \cdot \frac{HS_\ell}{VA_\ell} - 0.124 \cdot \frac{K_\ell}{VA_\ell} + 1.035 \cdot CR_\ell$$
$$-3.199 \cdot TFPW_\ell + 1.048 \cdot BAL_\ell + 0.452 \cdot DP_\ell + \varepsilon_\ell$$

X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest producers,*
$TFPW_\ell$	-	*by wages normed productvity in the i-th industry,*
Bal_ℓ	-	*Balasa index in the i-th industry,,*
DP_ℓ	-	*cost discontinuity in 1993 in the i-th industry*

with coefficient of determination 0.97 and stable submodels

Analysis of the export from the Czech republic to EU in 1994 by means of the *least trimmed squares*

has found:

Complementary subgroup

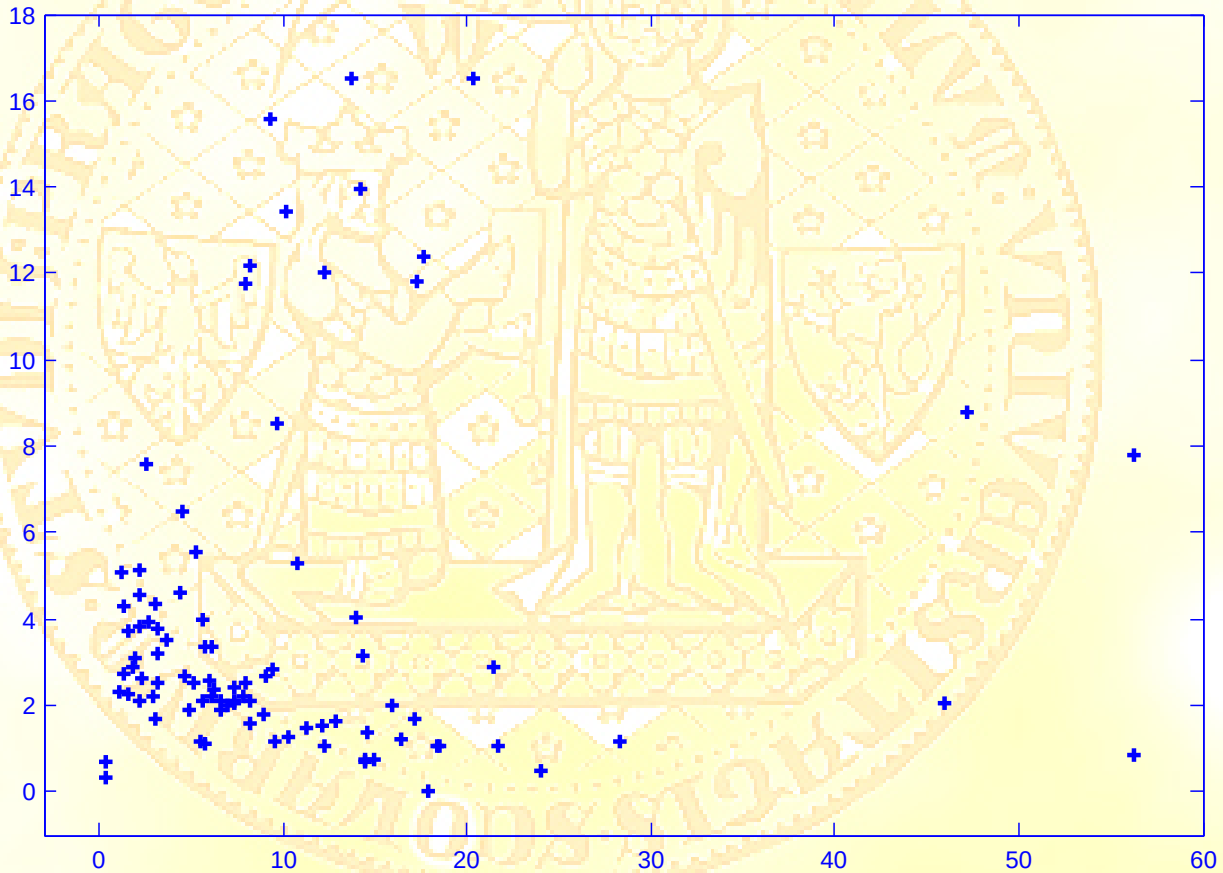
with number of industries 33 and model

$$\frac{X_\ell}{S_\ell} = -0.634 + 0.089 \cdot \frac{US_\ell}{VA_\ell} + 0.235 \cdot \frac{HS_\ell}{VA_\ell} + 0.249 \cdot \frac{K_\ell}{VA_\ell} + 1.174 \cdot CR_\ell$$
$$+0.690 \cdot TFPW_\ell + 2.691 \cdot BAL_\ell - 0.051 \cdot DP_\ell + \varepsilon_\ell$$

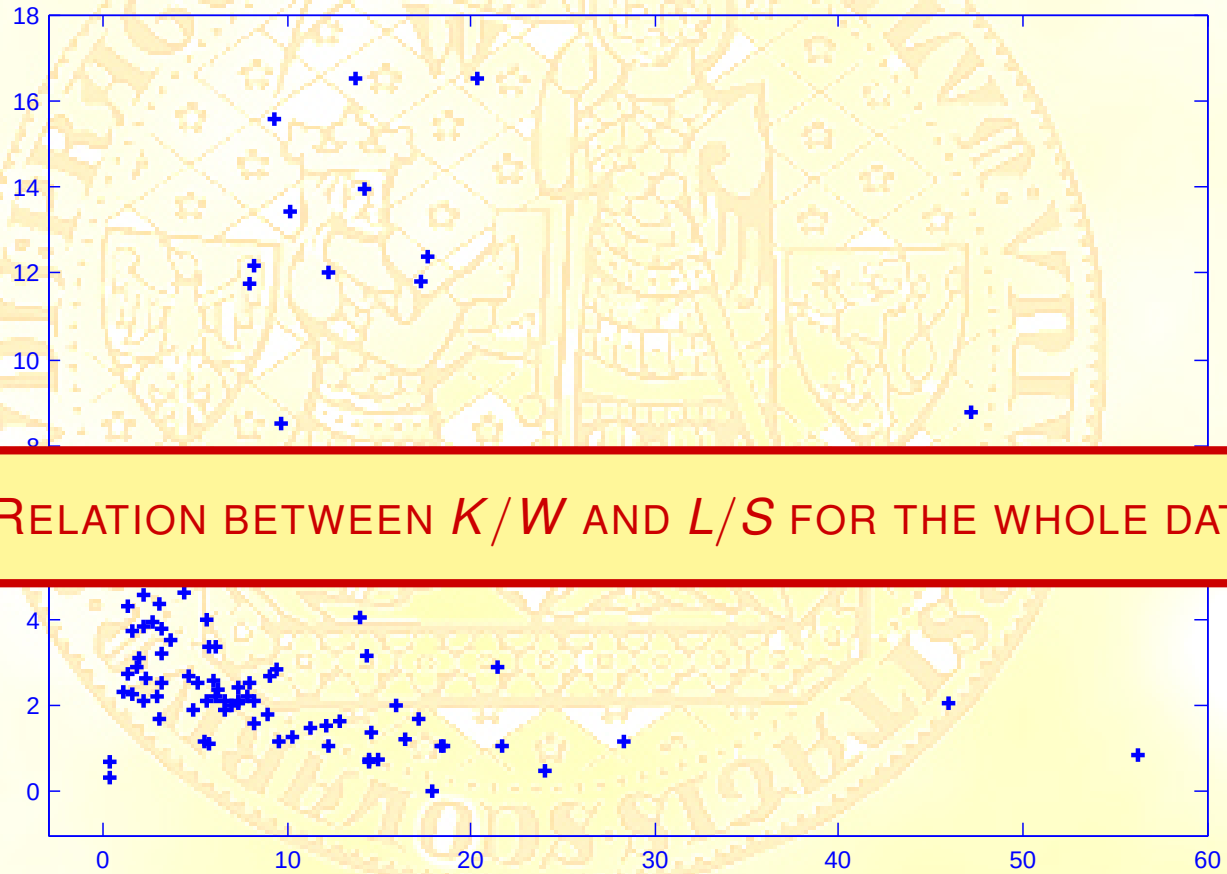
X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest producers,*
$TFPW_\ell$	-	*by wages normed productvity in the i-th industry,*
Bal_ℓ	-	*Balasa index in the i-th industry,,*
DP_ℓ	-	*cost discontinuity in 1993 in the i-th industry*

with coefficient of determination 0.93 and stable submodels

ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *least trimmed squares*.

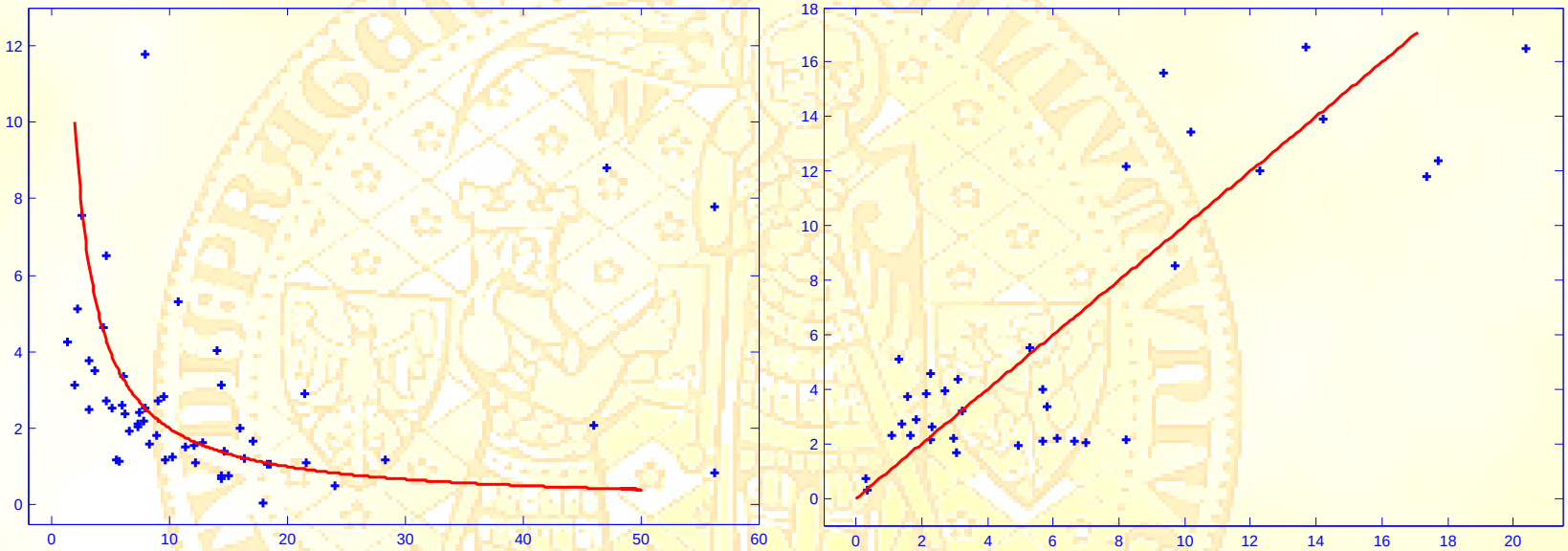


ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *least trimmed squares*.



RELATION BETWEEN K/W AND L/S FOR THE WHOLE DATA.

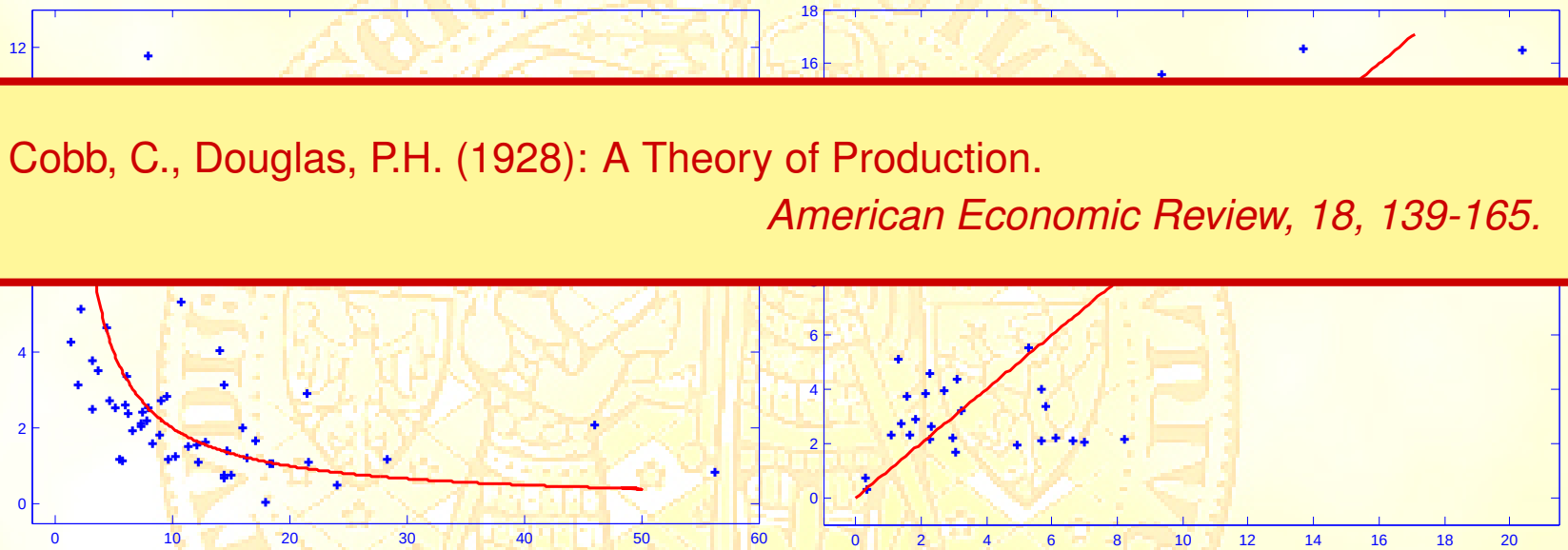
ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *least trimmed squares*.



RELATION BETWEEN K/W AND L/S FOR THE *Main subpopulation* (LEFT PICTURE)

AND FOR THE *Complementary subpopulation* (RIGHT PICTURE).

ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *least trimmed squares*.



Cobb, C., Douglas, P.H. (1928): A Theory of Production.
American Economic Review, 18, 139-165.

RELATION BETWEEN K/W AND L/S FOR THE *Main subpopulation* (LEFT PICTURE)
AND FOR THE *Complementary subpopulation* (RIGHT PICTURE).

Let us put:

The least trimmed squares

Let us put:

The least trimmed squares

Residuals $\forall \beta \in R \quad \rightarrow \quad r_i(\beta) = Y_i - X_i'\beta$

Order statistics of squared residuals, i. e.

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq \ldots \leq r^2_{(n)}(\beta)$$

Let us put:

The least trimmed squares

Residuals $\forall \beta \in R \quad \rightarrow \quad r_i(\beta) = Y_i - X_i'\beta$

Order statistics of squared residuals, i. e.

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq \ldots \leq r^2_{(n)}(\beta)$$

Definition

Let $\frac{n}{2} < h \leq n$. Then

$$\hat{\beta}^{(LTS,n,h)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{h} r^2_{(i)}(\beta)$$

will be called *the Least Trimmed Squares (LTS)*.

Properties of LTS

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley *&* Son.

the Least Trimmed Squares

$$\hat{\beta}^{(LTS,n,h)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{h} r^2_{(i)}(\beta) \qquad \frac{n}{2} < h \leq n,$$

(Notice the order of words, remember there is also the Trimmed Least Squares.)

Properties of LTS

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley *&* Son.

the Least Trimmed Squares

$$\hat{\beta}^{(LTS,n,h)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{h} r^2_{(i)}(\beta) \qquad \frac{n}{2} < h \leq n,$$

(Notice the order of words, remember there is also the Trimmed Least Squares.)

Many advantages - e. g.

1. the breakdown point equal to $([\frac{n-p}{2}]+1)n^{-1}$ if $h = [\frac{n}{2}] + [\frac{p+1}{2}]$

Properties of LTS

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley *&* Son.

the Least Trimmed Squares

$$\hat{\beta}^{(LTS,n,h)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{h} r^2_{(i)}(\beta) \qquad \frac{n}{2} < h \leq n,$$

(Notice the order of words, remember there is also the Trimmed Least Squares.)

Many advantages - e. g.

1. the breakdown point equal to $([\frac{n-p}{2}]+1)n^{-1}$ if $h = [\frac{n}{2}] + [\frac{p+1}{2}]$
2. *scale- and regression equivariant*

Properties of LTS

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley *&* Son.

the Least Trimmed Squares

$$\hat{\beta}^{(LTS,n,h)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{h} r^2_{(i)}(\beta) \qquad \frac{n}{2} < h \leq n,$$

(Notice the order of words, remember there is also the Trimmed Least Squares.)

Many advantages - e. g.

1. the breakdown point equal to $([\frac{n-p}{2}]+1)n^{-1}$ if $h = [\frac{n}{2}] + [\frac{p+1}{2}]$
2. *scale- and regression equivariant*
3. $\sqrt{n}\left(\hat{\beta}^{(LTS,n,h)} - \beta^0\right) = \mathcal{O}_p(1)$

Content

1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here
2. Estimators alternative to the classical ones
   - Locatin parameter
   - Scale parameter
   - General parameter

Hint for test

Od zadavatele jsme dostali data $(Y, X) = \{(y_1, x_1), (y_2, x_2), ..., (y_n, x_n)\}$ a spočetli odhad regresních koeficientů pro jednoduchý regresní model (simple regression), řekněme

$$\hat{\beta}_0^{(OLS,n)}(Y, X) = 3, \quad \hat{\beta}_1^{(OLS,n)}(Y, X) = \frac{1}{2},$$

kde "(Y, X)" naznačuje, že odhady jsou vyčísleny pro data (Y, X). Jinými slovy, odhadli jsme model

$$\hat{y} = 3 + \frac{1}{2} \cdot x. \tag{1}$$

Při předávání výsledků zadavateli se zjistí, že došlo k chybě, že byla prohozena vysvětlující a vysvětlovaná proměnná. Stačí prostě převést ve vztahu (1) x nalevo a y napravo, tj. říci, že správný odhad je

$$\hat{x} = -6 + 2 \cdot y \tag{2}$$

nebo ne?

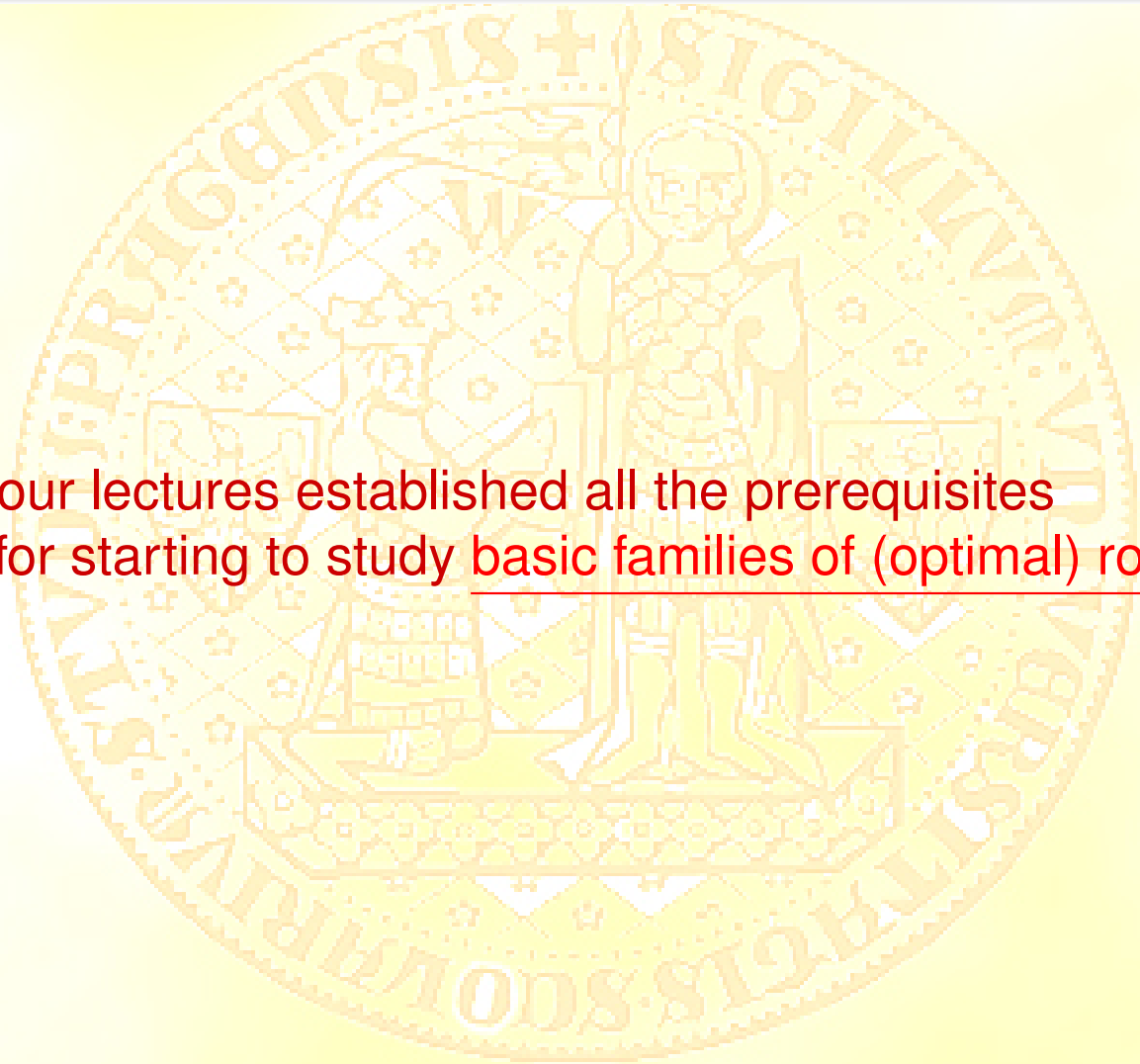
Hint for test

Nápověda. *Lidé často odpovídají, že to nestačí, že se v prvé a druhé situaci minimalizují jiné součty čtverců, "vertikální", respektive "horizontální" součty čtverců. NenÃ ale jasně, zda je to pravda, neboť zmíněné součty čtverců jsou převoditelné pomocí násobení* $\hat{\beta}_1 = \tan(\alpha)$*, kde* α *je sklon regresní přímky. Jinými slovy vertikální (vr) a horizontální (hr) rezidua jsou pro bod* (y, x) *a libovolné* $\beta = (\beta_0, \beta_1) \in R^2$ *ve vztahu*

$$vr = y - \beta_0 - \beta_1 \cdot x \quad \text{a} \quad hr = x + \frac{\beta_0}{\beta_1} - \frac{y}{\beta_1}, \quad \Rightarrow \quad vr = -hr \cdot \beta_1,$$

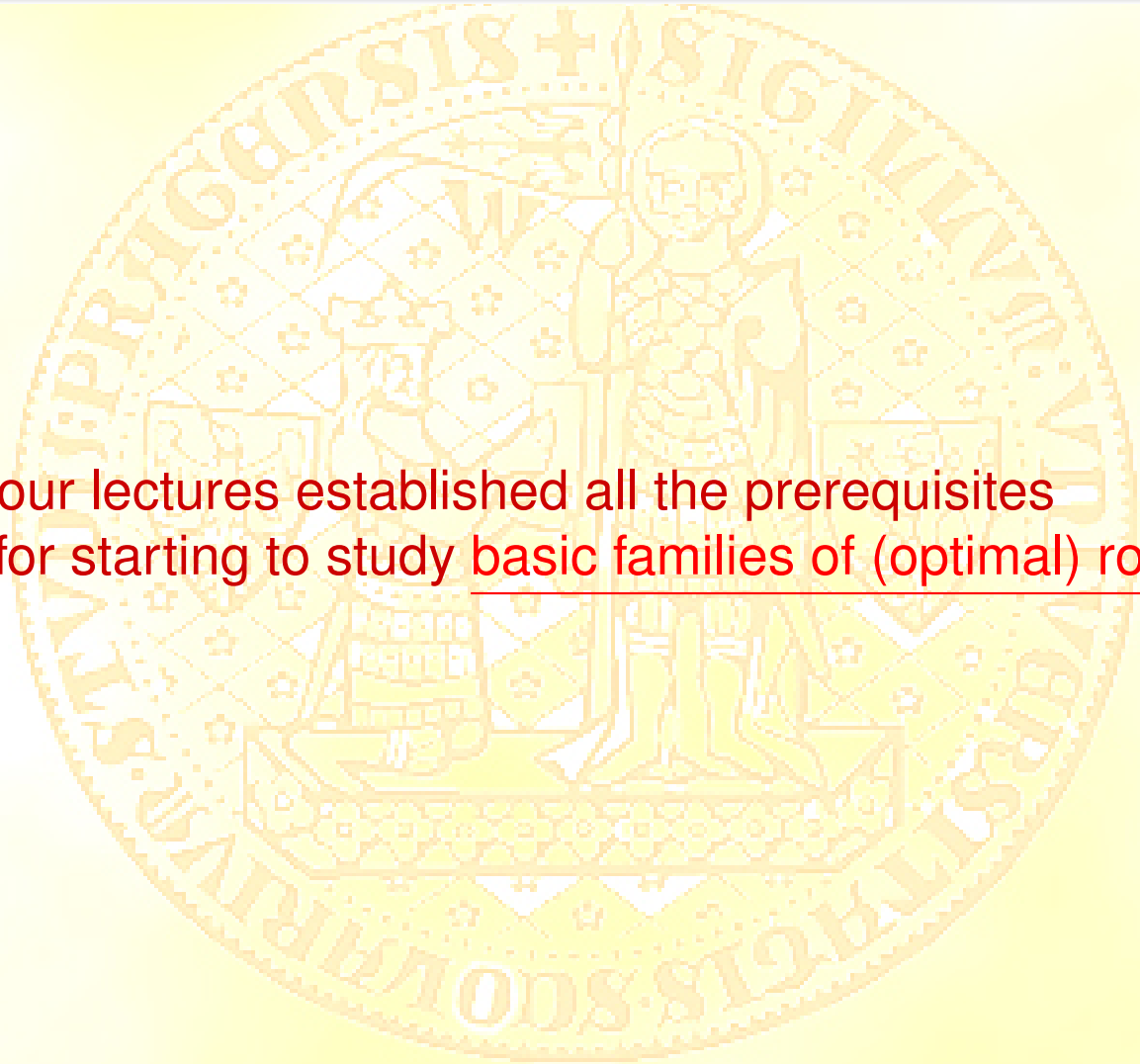
tj. $vr^2 = (hr \cdot \beta_1))^2$*. Dosáhneme-li tedy minima u sumy čtverců vertikálních reziduí, dosáhneme i minima u sumy čtverců horizontalních reziduí. Tak jak to vlastně je? Lze použít (2) či nikoliv?*

The most popular families of robust estimators



The first four lectures established all the prerequisites
for starting to study basic families of (optimal) robust estimators.

The most popular families of robust estimators



The first four lectures established all the prerequisites
for starting to study basic families of (optimal) robust estimators.

Content

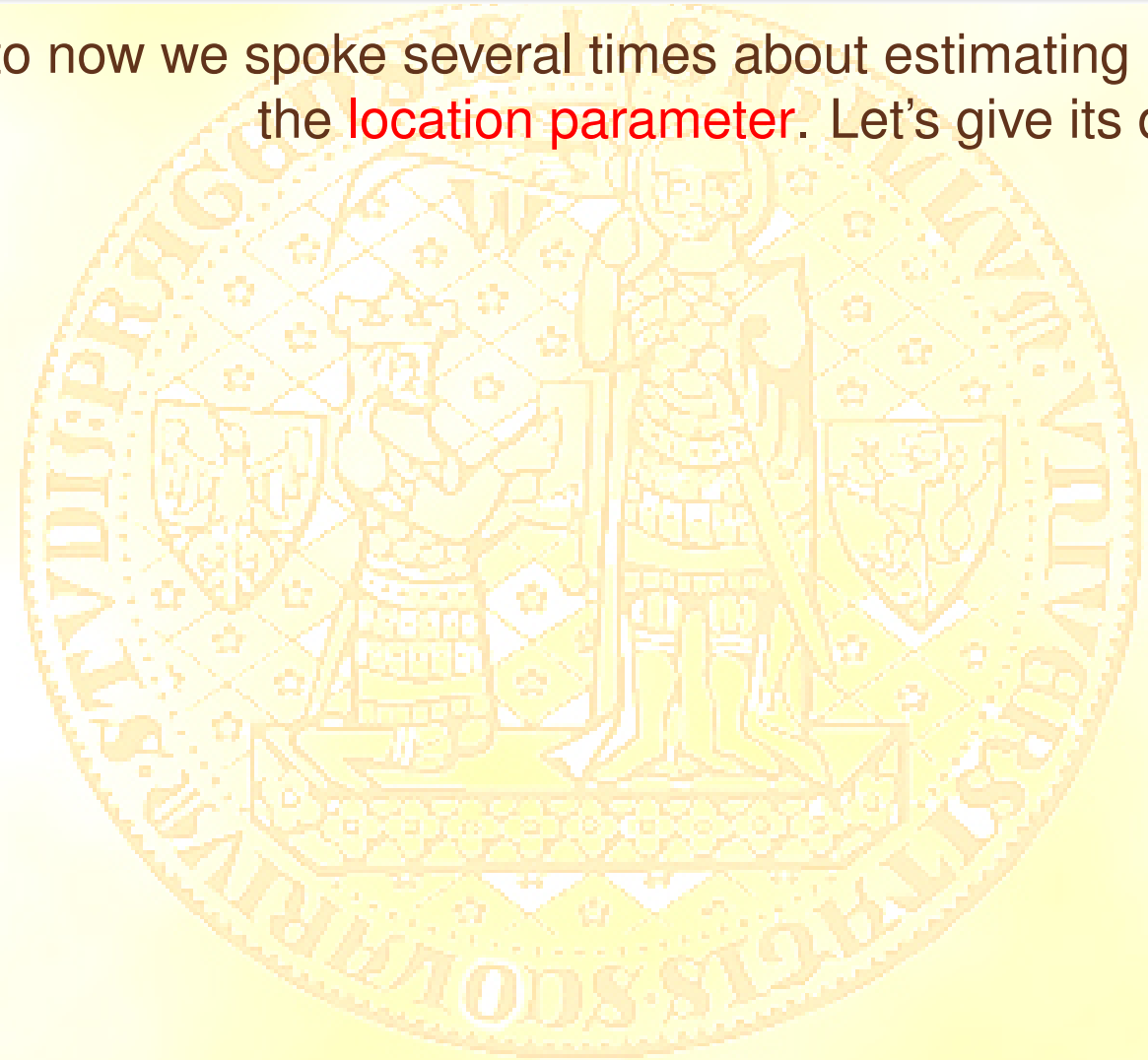
1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here
2. Estimators alternative to the classical ones
   - Locatin parameter
   - Scale parameter
   - General parameter

Content

1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here
2. Estimators alternative to the classical ones
   - Locatin parameter
   - Scale parameter
   - General parameter

Reviewing the basic families of robust estimators - location.

Up to now we spoke several times about estimating the location parameter. Let's give its definition:



Reviewing the basic families of robust estimators - location.

Up to now we spoke several times about estimating the location parameter. Let's give its definition:

Let $F(x)$ be a (parent) d. f. . Then $\{F(x-\mu)\}_{\mu \in R}$ is called the family with location parameter.

Reviewing the basic families of robust estimators - location.

Up to now we spoke several times about estimating the location parameter. Let's give its definition:

Let $F(x)$ be a (parent) d. f. . Then $\{F(x-\mu)\}_{\mu \in R}$ is called the family with location parameter.

Let's start with estimating the location parameter:

The solution of the extremal problem

$$\hat{\mu}^{(M,n)} = \arg\min_{\mu \in R} \sum_{i=1}^{n} \rho\left(x_i - \mu\right)$$

is called *Maximum likelihood-like estimators of location* or *M-estimators of location*, for short.

Reviewing the basic families of robust estimators - location.

Up to now we spoke several times about estimating the location parameter. Let's give its definition:

Let $F(x)$ be a (parent) d. f. . Then $\{F(x-\mu)\}_{\mu \in R}$ is called the family with location parameter.

Let's start with estimating the location parameter:

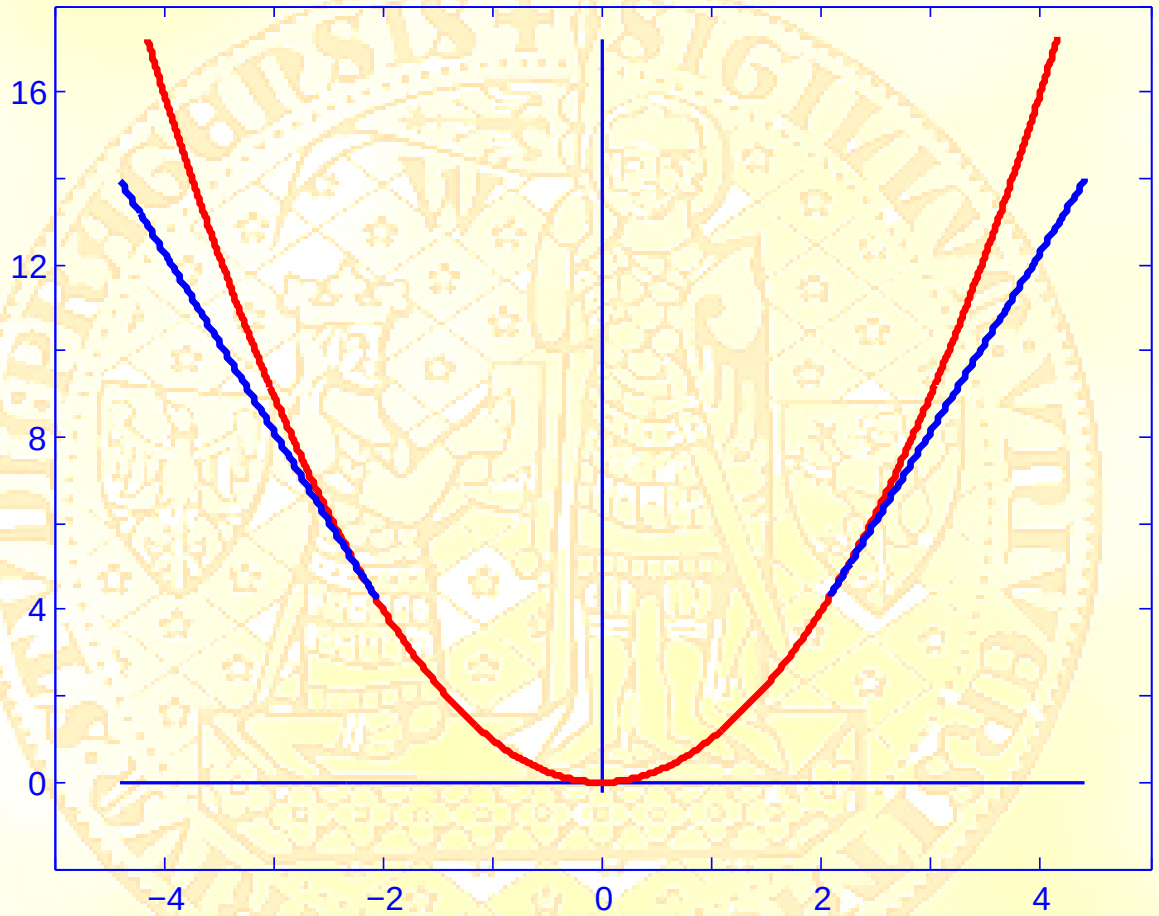
The solution of the extremal problem

$$\hat{\mu}^{(M,n)} = \underset{\mu \in R}{\arg\min} \sum_{i=1}^{n} \rho\left(x_i - \mu\right)$$

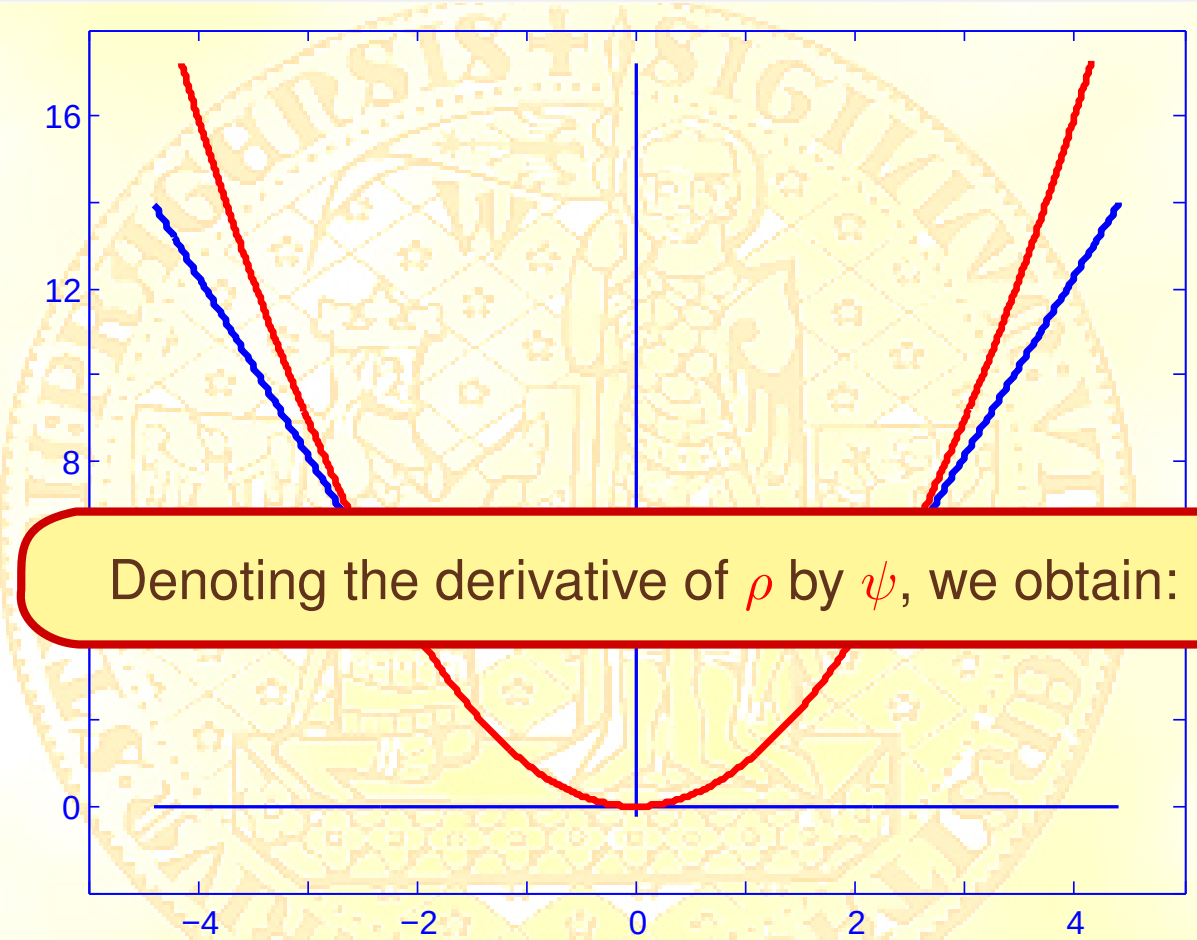
is called *Maximum likelihood-like estimators of location* or *M-estimators of location*, for short.

(Example of ρ is on the next slide.)

Reviewing the basic families of robust estimators - location.

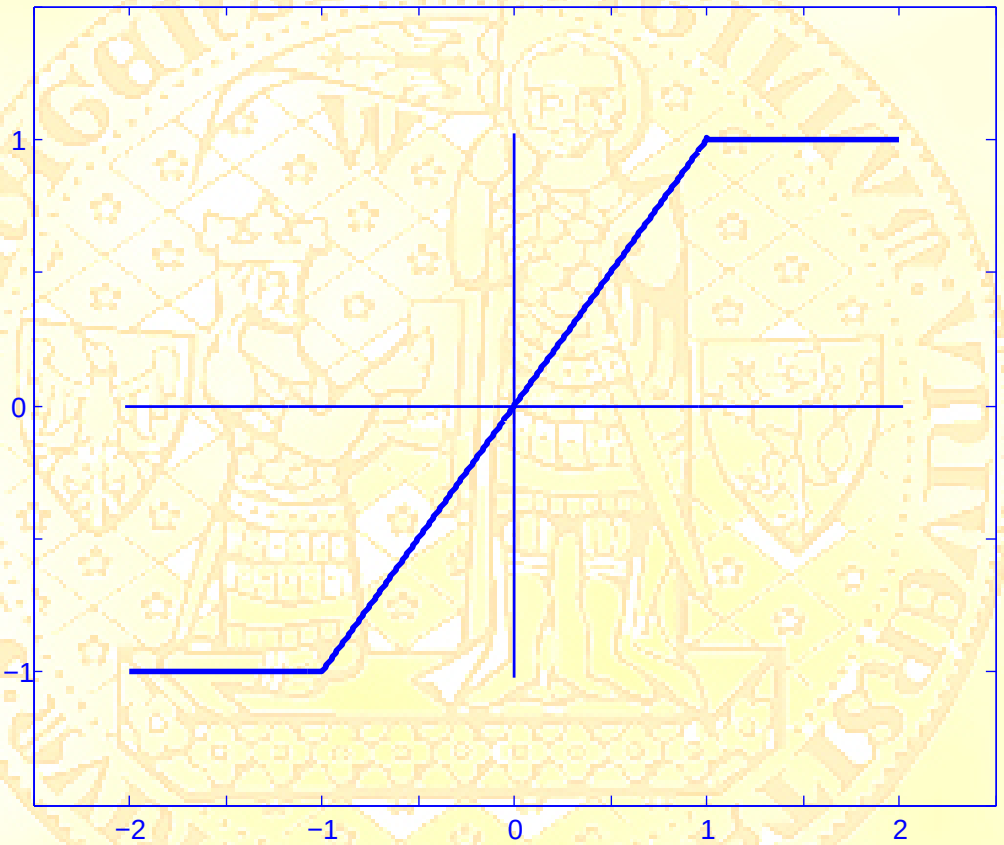


Reviewing the basic families of robust estimators - location.

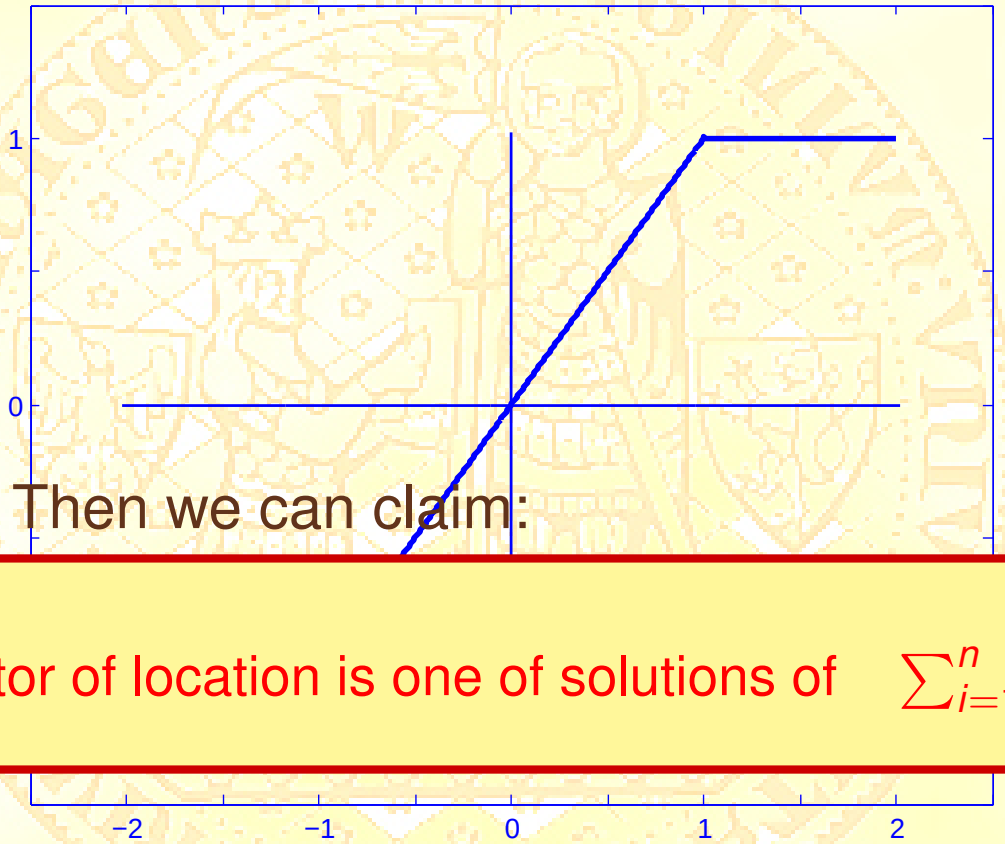


Denoting the derivative of ρ by ψ, we obtain:

Reviewing the basic families of robust estimators - location.



Reviewing the basic families of robust estimators - location.

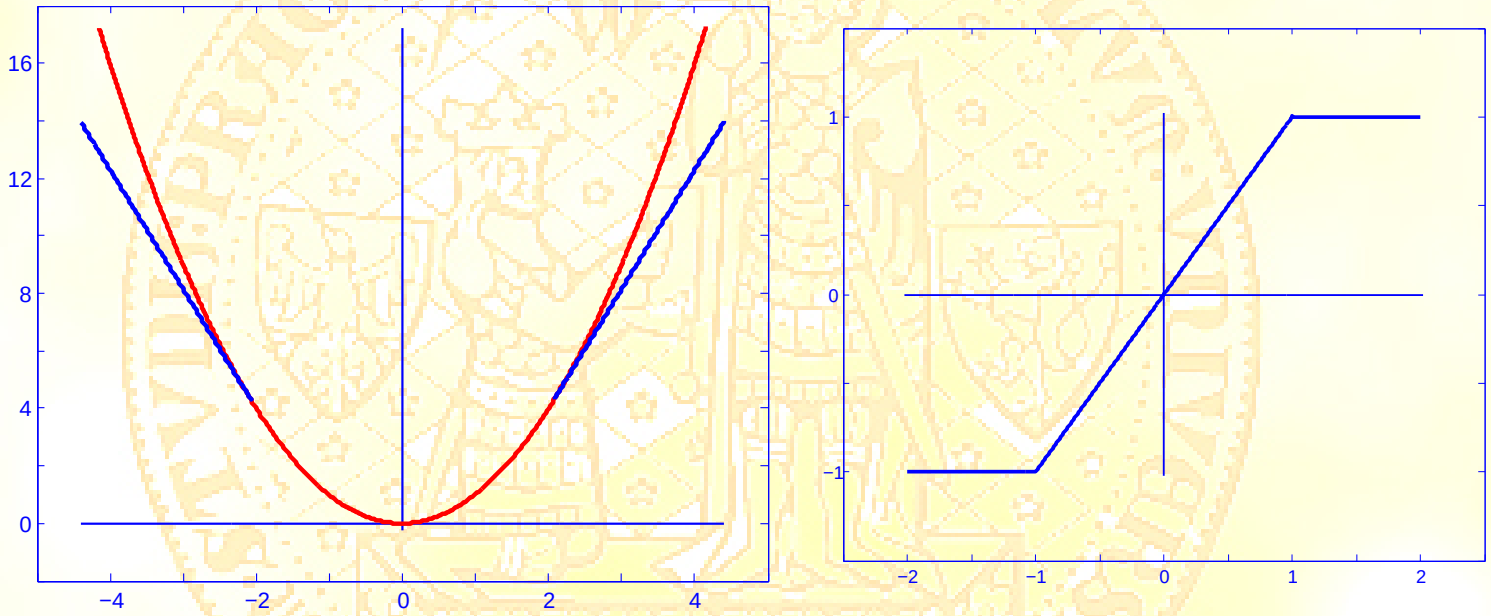


Then we can claim:

M-estimator of location is one of solutions of $\sum_{i=1}^{n} \psi(x_i - \mu) = 0$.

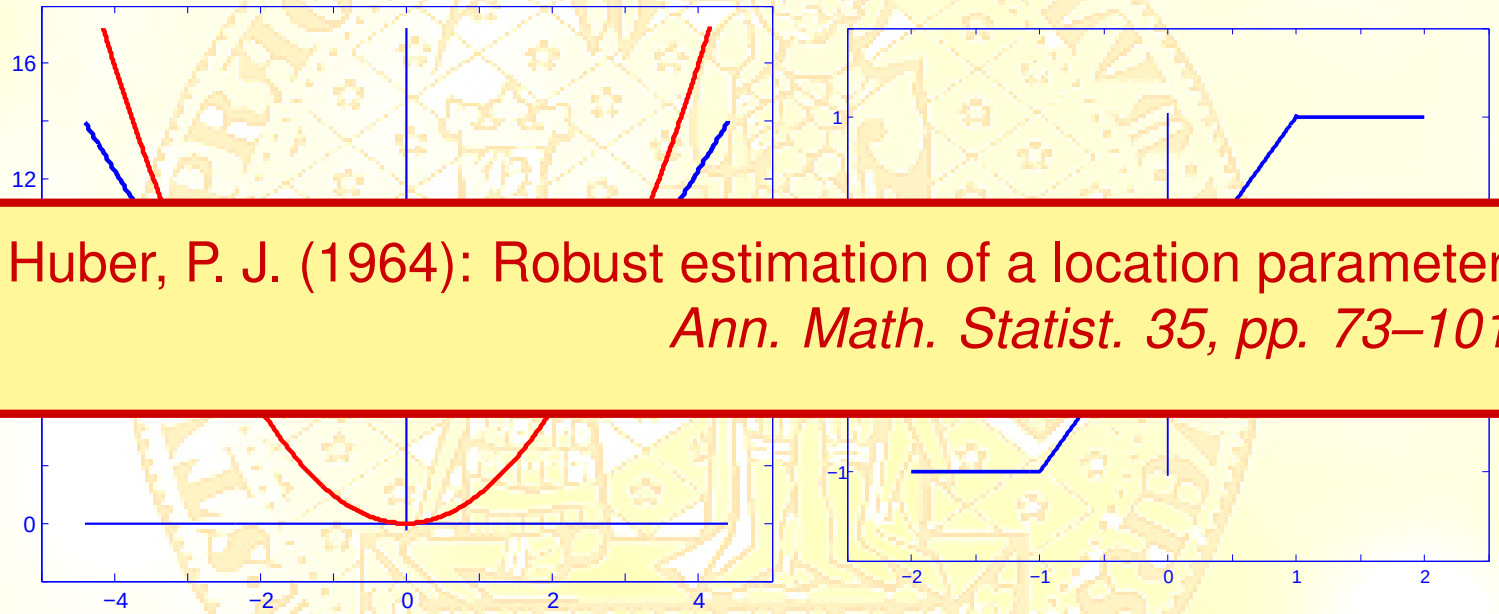
Reviewing the basic families of robust estimators - location and scale.

The functions ρ and ψ were firstly proposed in:



Reviewing the basic families of robust estimators - location and scale.

The functions ρ and ψ were firstly proposed in:



Huber, P. J. (1964): Robust estimation of a location parameter.
Ann. Math. Statist. 35, pp. 73–101.

Hence they are usually referred to as Huber's ρ and Huber's ψ.

Content

1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here
2. Estimators alternative to the classical ones
   - Locatin parameter
   - Scale parameter
   - General parameter

Reviewing the basic families of robust estimators - scale.

Up to now we spoke also several times about estimating the scale parameter. Let's give its definition: .... .

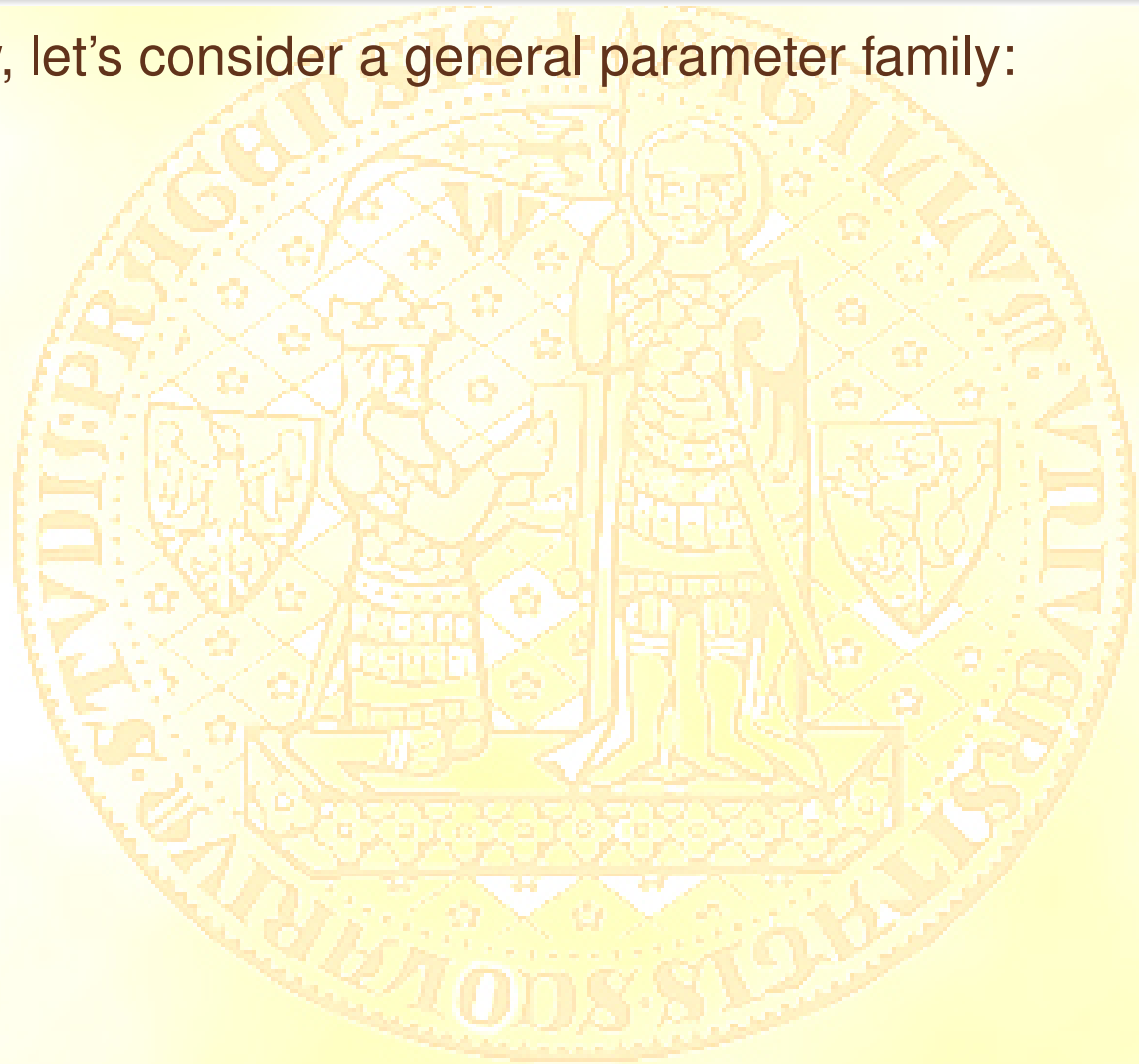
We'll skip it, you are syrely able to go it yourself.

Content

1. Let's fulfil the promise
   - Example of estimator with 50% of breakdown point
   - The first hint for test, i. e. an advantage for whose who are here
2. Estimators alternative to the classical ones
   - Locatin parameter
   - Scale parameter
   - General parameter

Reviewing the basic families of robust estimators - general parameter.

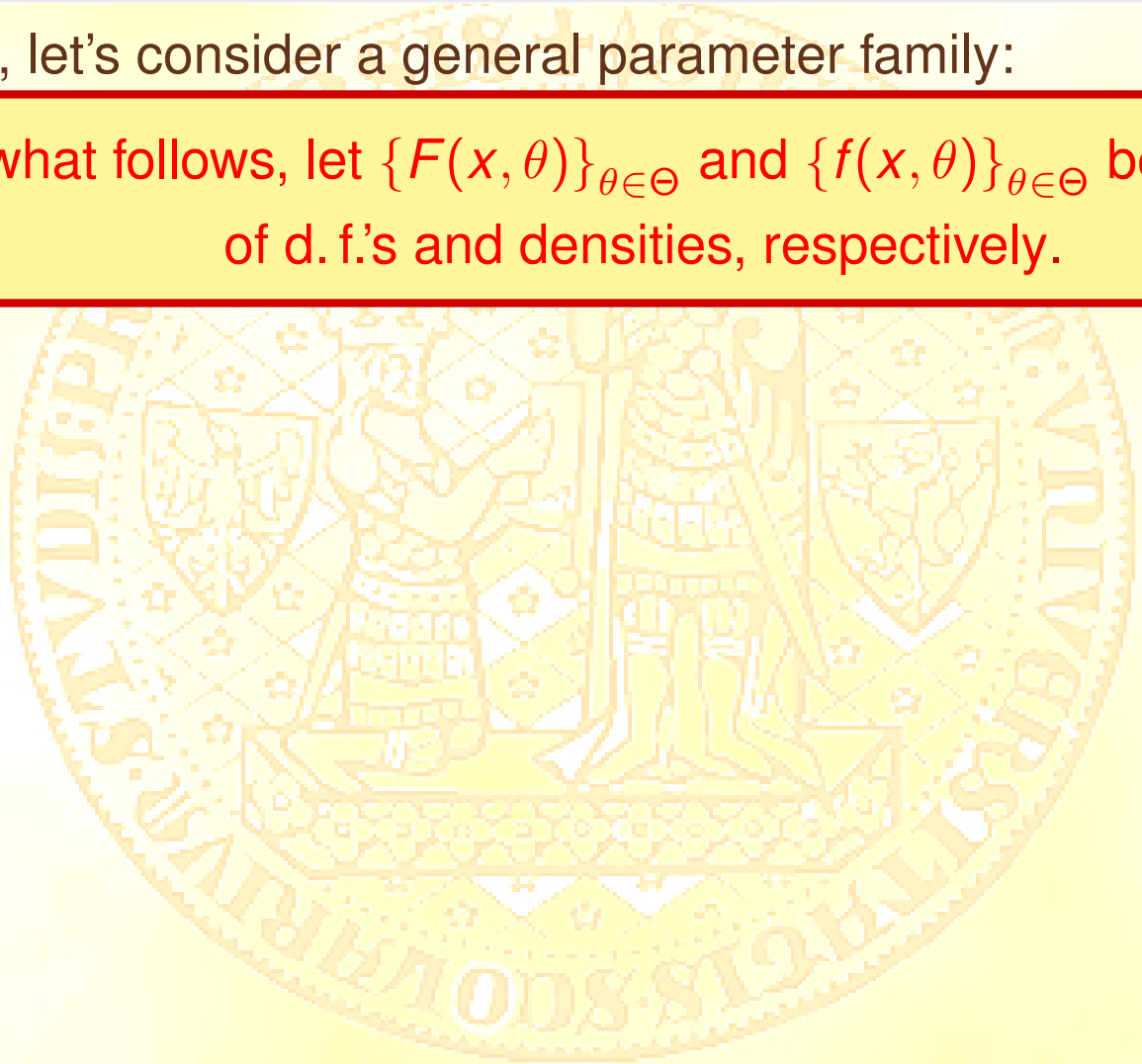
Now, let's consider a general parameter family:



Reviewing the basic families of robust estimators - general parameter.

Now, let's consider a general parameter family:

In what follows, let $\{F(x,\theta)\}_{\theta\in\Theta}$ and $\{f(x,\theta)\}_{\theta\in\Theta}$ be families of d. f.'s and densities, respectively.



Reviewing the basic families of robust estimators - general parameter.

Now, let's consider a general parameter family:

In what follows, let $\{F(x,\theta)\}_{\theta\in\Theta}$ and $\{f(x,\theta)\}_{\theta\in\Theta}$ be families of d. f.'s and densities, respectively.

Then:

The solution of the extremal problem

$$\hat{\theta}^{(M,n)} = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \rho\left(x_i,\theta\right)$$

is called *Maximum likelihood-like estimators of the parameter* θ or *M-estimators of* θ, for short.

Reviewing the basic families of robust estimators - general parameter.

Now, let's consider a general parameter family:

In what follows, let $\{F(x,\theta)\}_{\theta\in\Theta}$ and $\{f(x,\theta)\}_{\theta\in\Theta}$ be families of d. f.'s and densities, respectively.

Then:

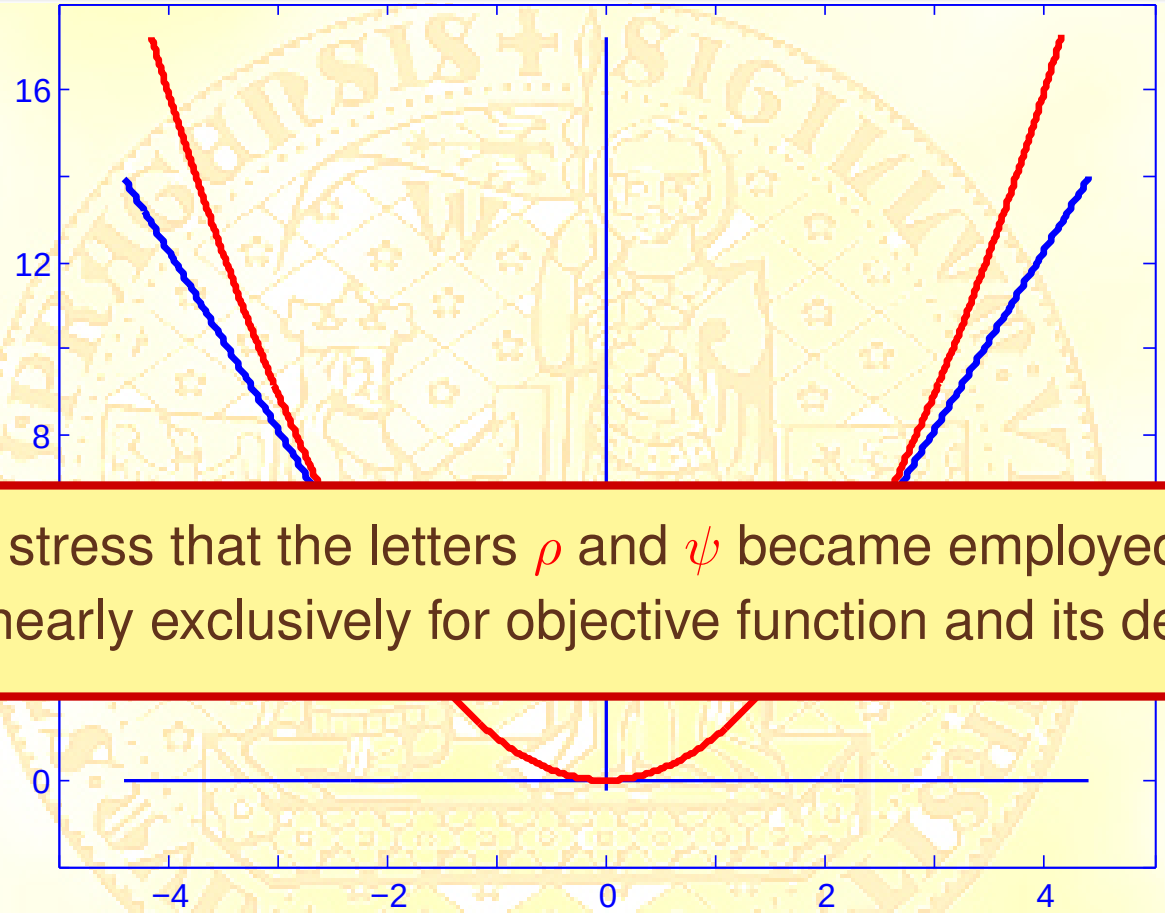
The solution of the extremal problem

$$\hat{\theta}^{(M,n)} = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \rho\left(x_i,\theta\right)$$

is called *Maximum likelihood-like estimators of the parameter* θ or *M-estimators of* θ, for short.

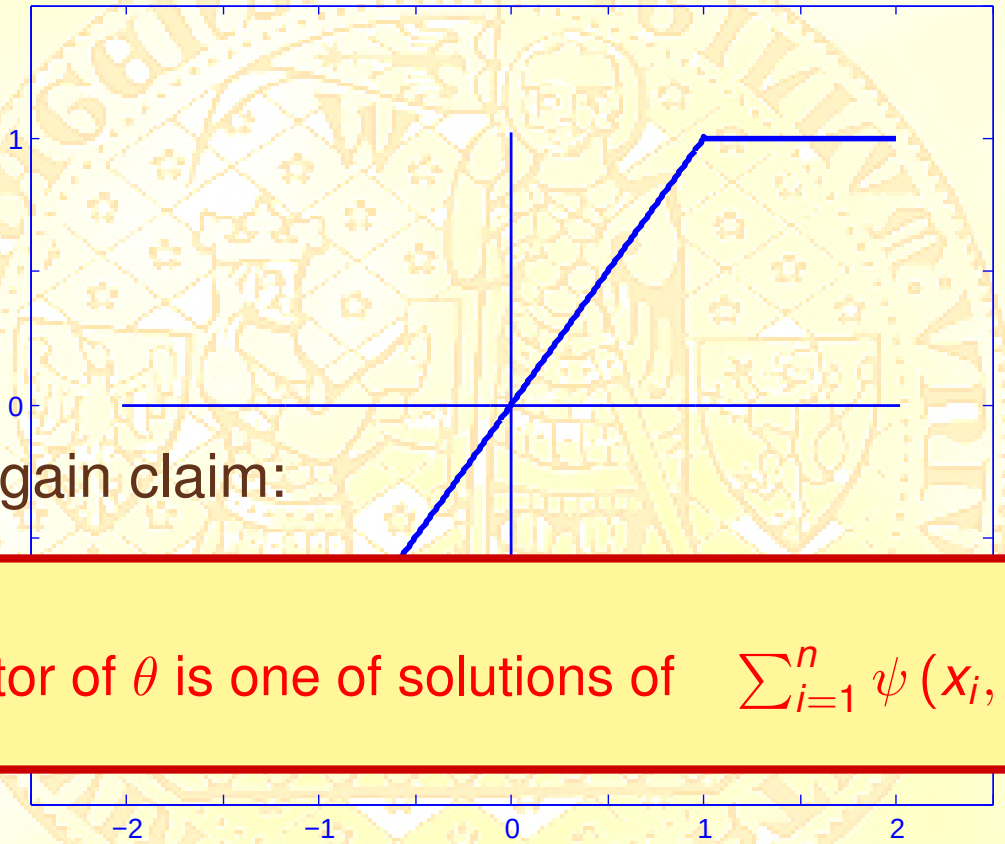
(We can use the same ρ as for location and scale.)

Reviewing the basic families of robust estimators - general parameter.



Let's stress that the letters ρ and ψ became employed nearly exclusively for objective function and its derivative.

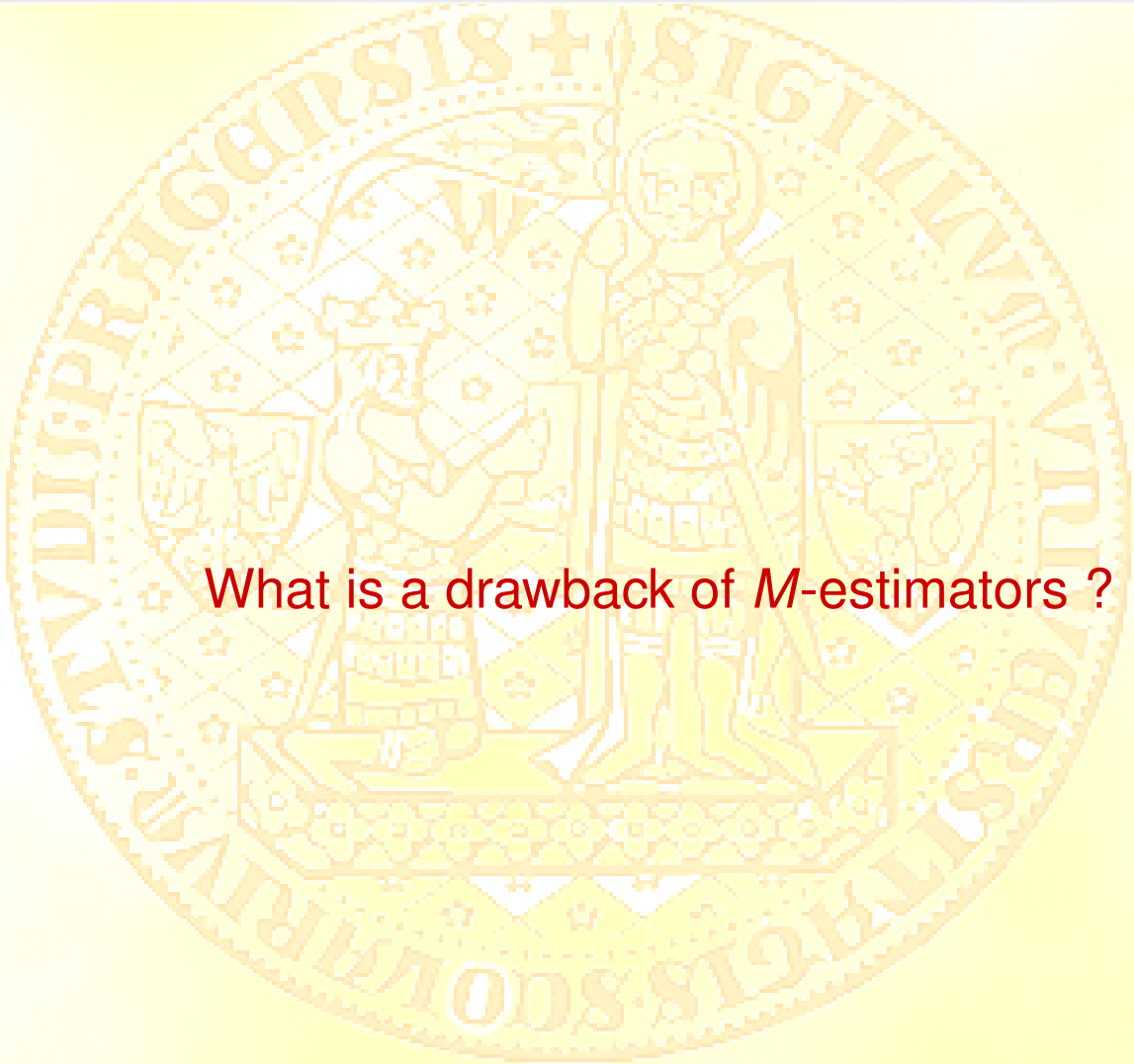
Reviewing the basic families of robust estimators - general parameter.



Then we can again claim:

M-estimator of θ is one of solutions of $\sum_{i=1}^{n} \psi(x_i, \theta) = 0$.

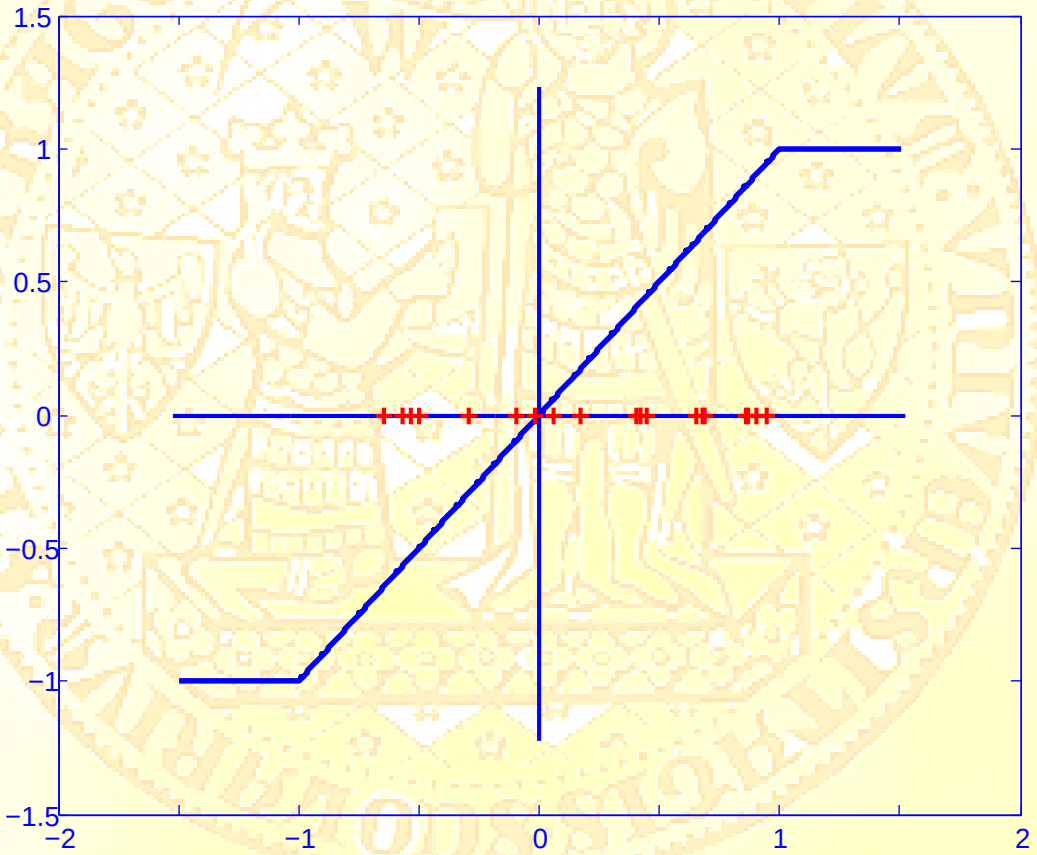
M-estimators - general parameter.



What is a drawback of *M*-estimators ?

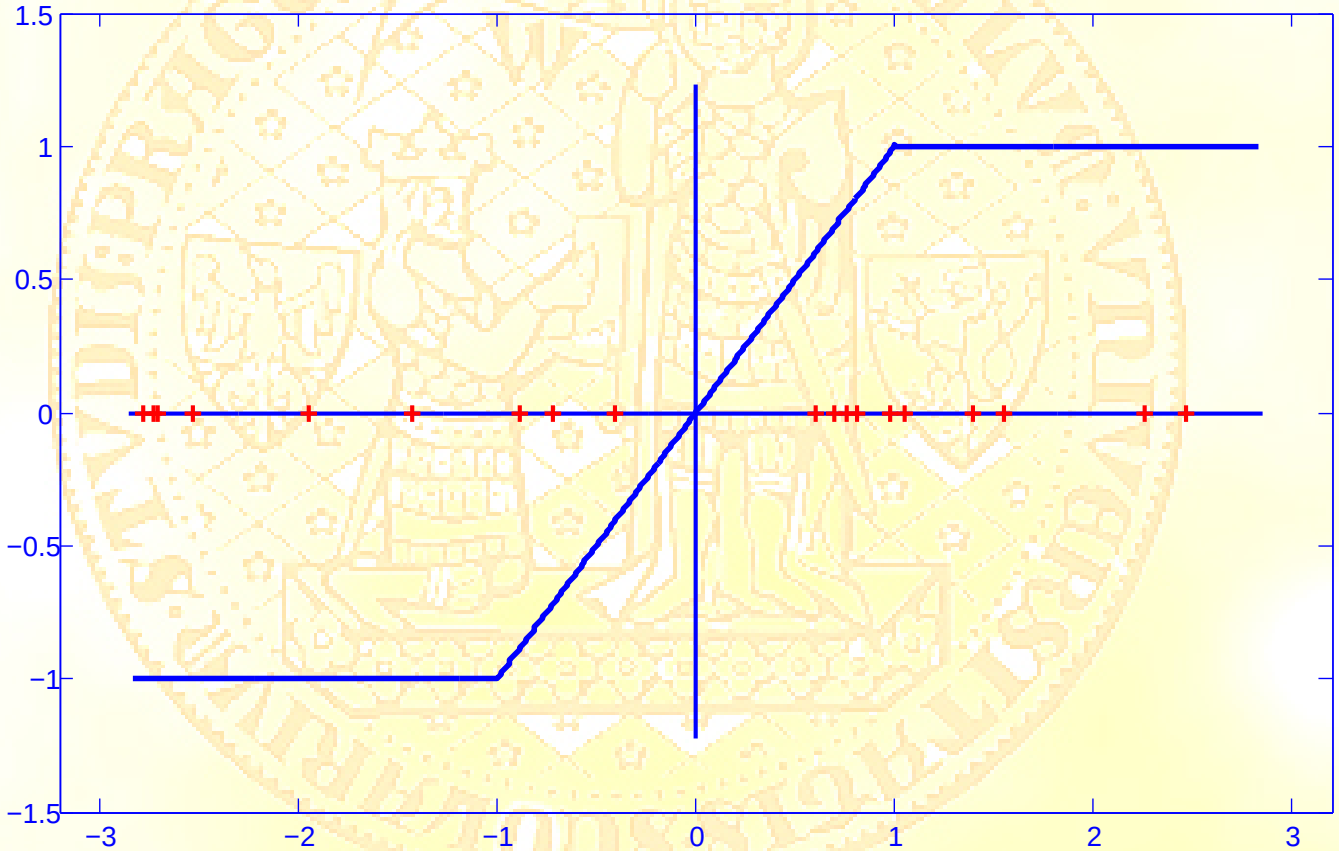
M-estimators - general parameter.

To learn it, let's consider the following data:

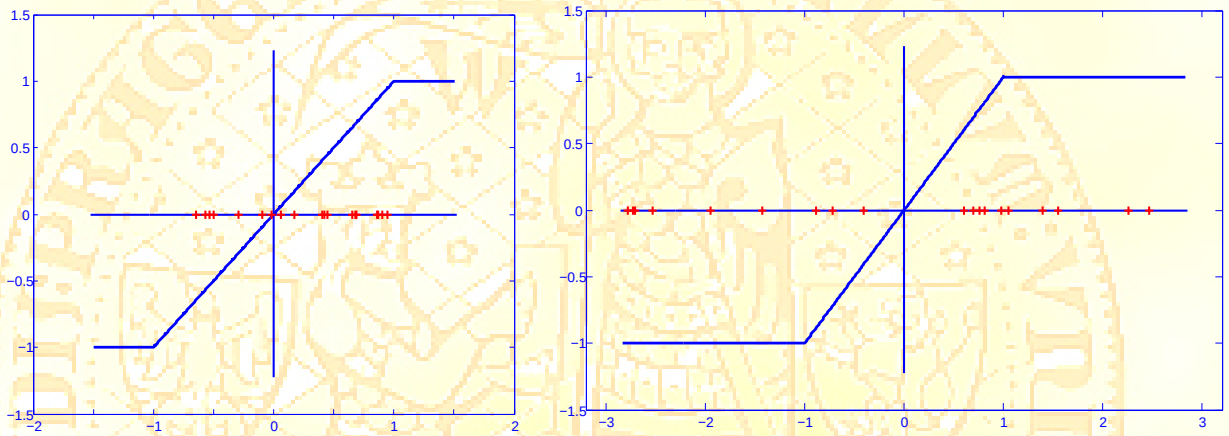


M-estimators - general parameter.

And now, let's consider these data:



M-estimators - general parameter.



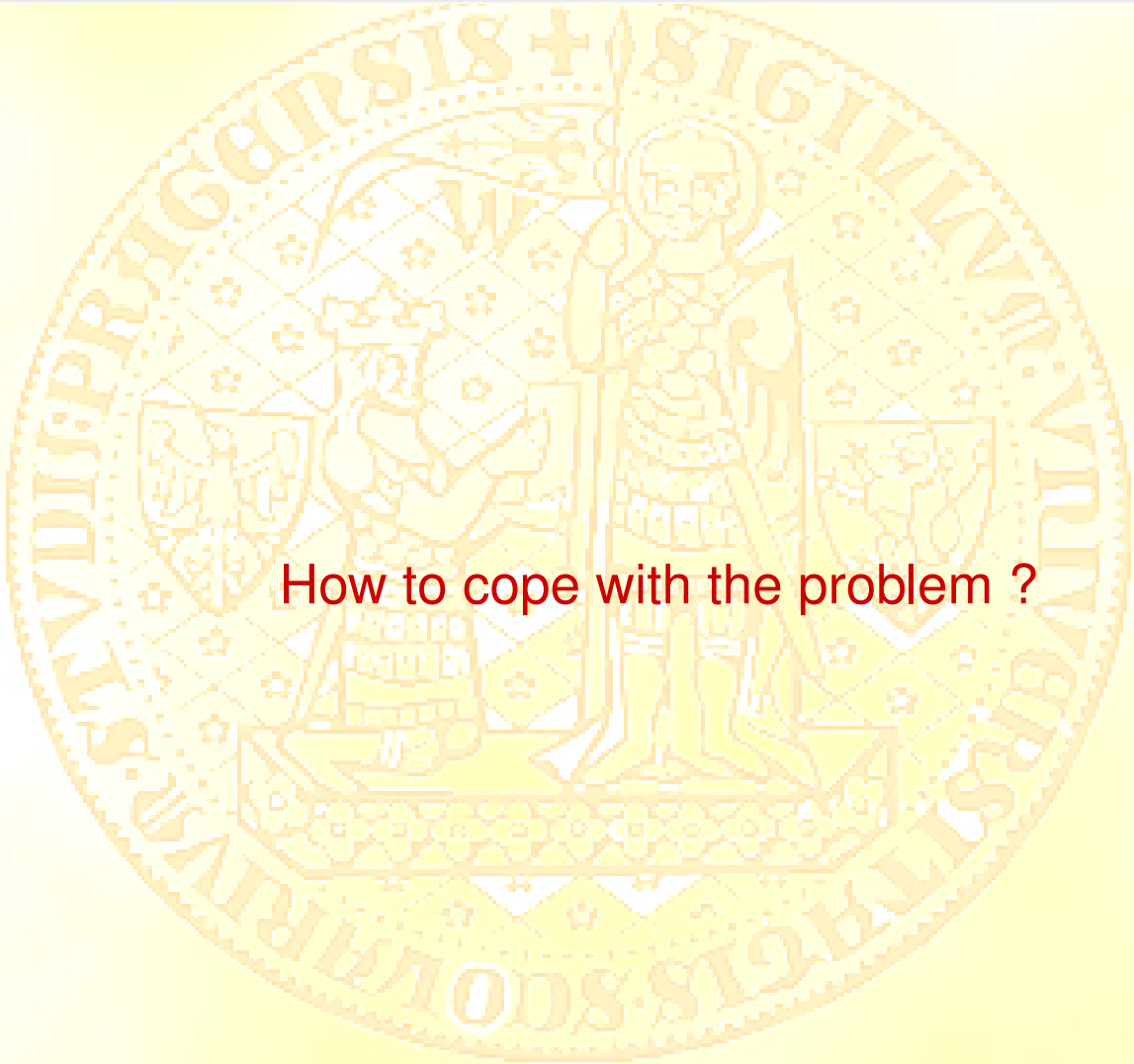
Clearly:

The solution $\hat{\theta}^{(M,n)}$ of the normal equation

$$\sum_{i=1}^{n} \psi\left(x_i, \theta\right) = 0$$

is not scale-equivariant.

M-estimators - general parameter.



How to cope with the problem ?

M-estimators - general parameter.

Let $\hat{\sigma}$ be a (highly) robust estimator
of the standard deviation of data x_i's and solve:

$$\sum_{i=1}^{n} \psi\left(x_i/\hat{\sigma}, \theta\right) = 0.$$

The solution $\hat{\theta}^{(M,n)}$ is then scale-equivariant.

M-estimators - general parameter.

Let $\hat{\sigma}$ be a (highly) robust estimator
of the standard deviation of data x_i's and solve:

$$\sum_{i=1}^{n} \psi\left(x_i/\hat{\sigma}, \theta\right) = 0.$$

The solution $\hat{\theta}^{(M,n)}$ is then scale-equivariant.

An example of such estimator is

$$\hat{\sigma}_{MAD} = 1.483 \operatorname{med}_i \left\{ \left| x_i - \operatorname{med}_j(x_j) \right| \right\}.$$

M-estimators - general parameter.

Let $\hat{\sigma}$ be a (highly) robust estimator
of the standard deviation of data x_i's and solve:

$$\sum_{i=1}^{n} \psi\left(x_i/\hat{\sigma}, \theta\right) = 0.$$

The solution $\hat{\theta}^{(M,n)}$ is then scale-equivariant.

An example of such estimator is

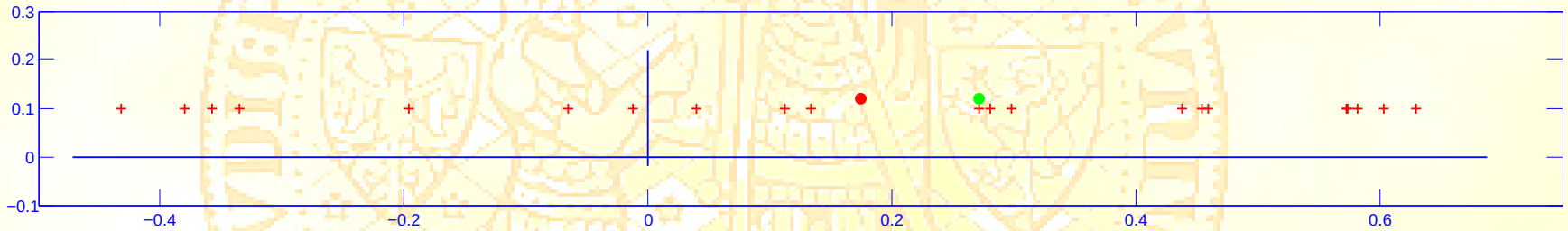
$$\hat{\sigma}_{MAD} = 1.483 \operatorname{med}_i \left\{ \left| x_i - \operatorname{med}_j(x_j) \right| \right\}.$$

(A comparison of $1.483 * MAD$ and s_n is on the next slide.)

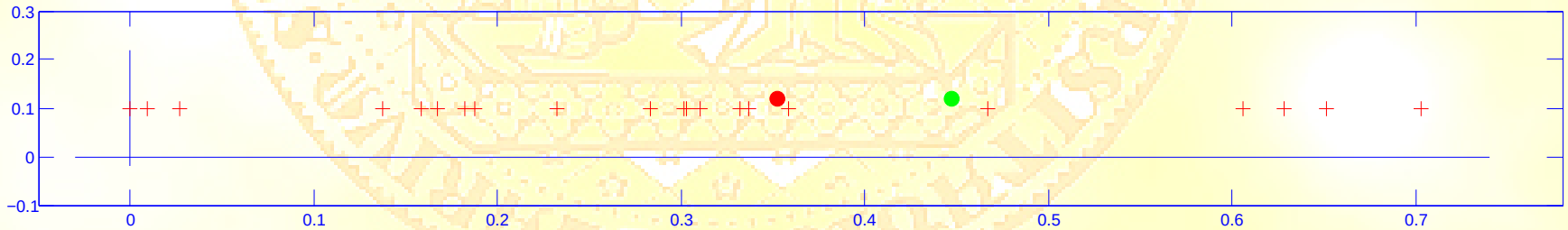
Demonstrating abilities of MAD

Observe the mean • and the median •
and standard deviation s_n • and $\hat{\sigma}_{MAD}$ •.

Non-contaminated data - normal d. f. $\mu = 0$ and $\sigma^2 = \frac{1}{9}$



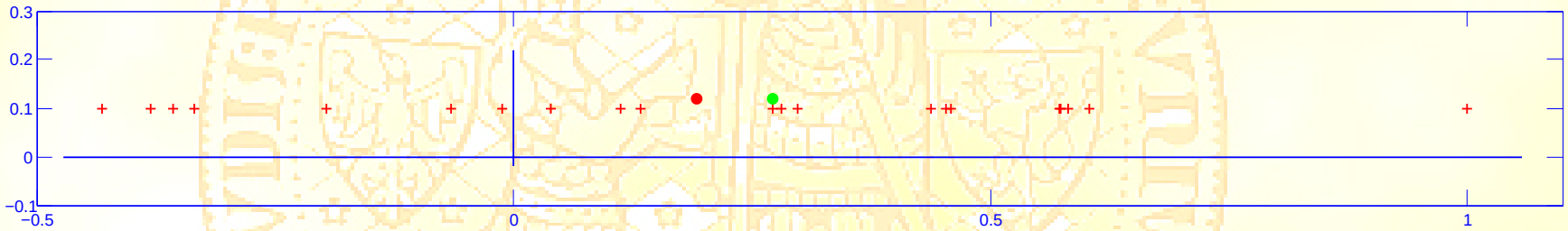
Absolute values of data



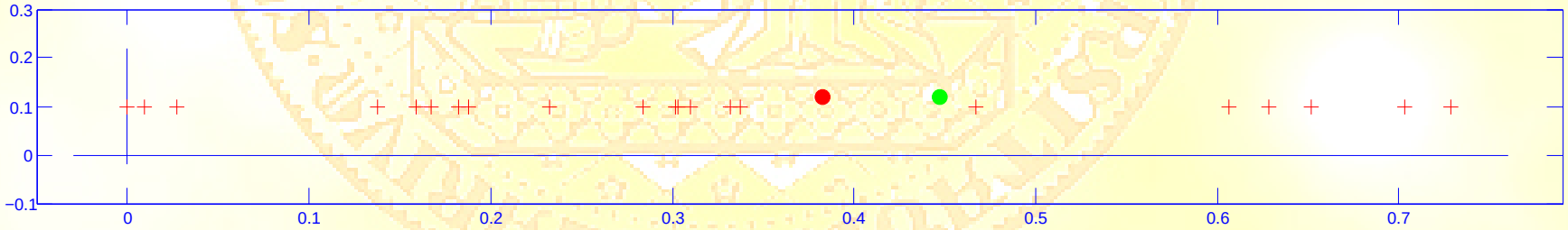
Demonstrating abilities of MAD

Observe the mean • and the median •
and standard deviation s_n • and $\hat{\sigma}_{MAD}$ •.

Contamination at point 1



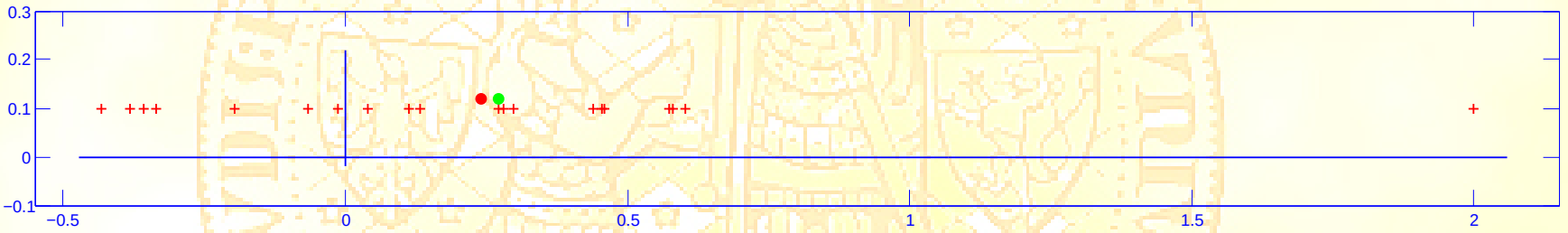
Absolute values of data



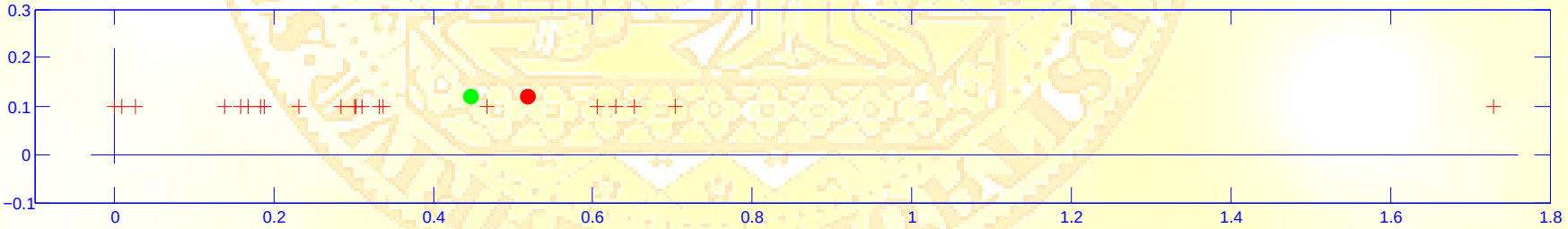
Demonstrating abilities of MAD

Observe the mean • and the median •
and standard deviation s_n • and $\hat{\sigma}_{MAD}$ •.

Contamination at point 2



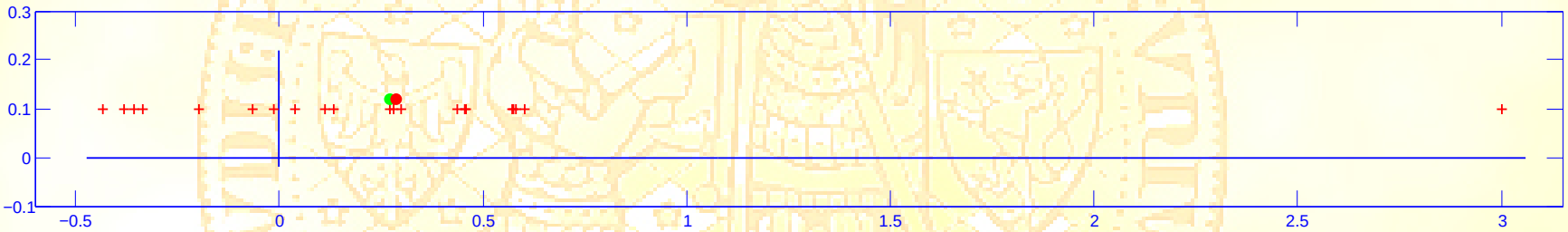
Absolute values of data



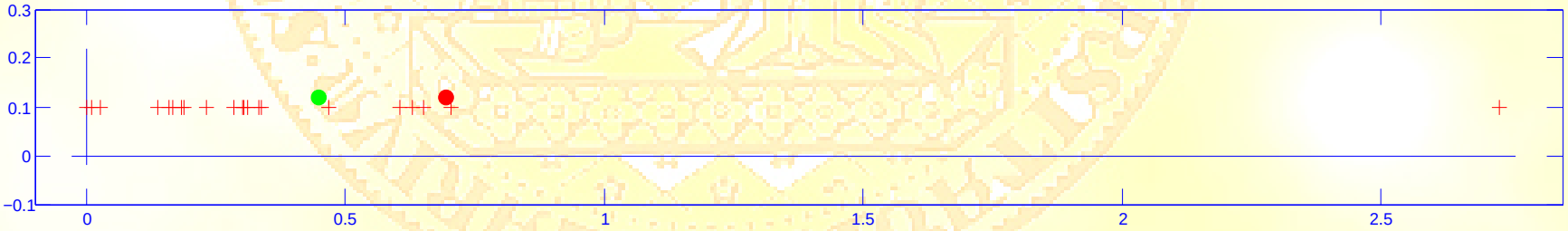
Demonstrating abilities of MAD

Observe the mean • and the median •
and standard deviation s_n • and $\hat{\sigma}_{MAD}$ •.

Contamination at point 3



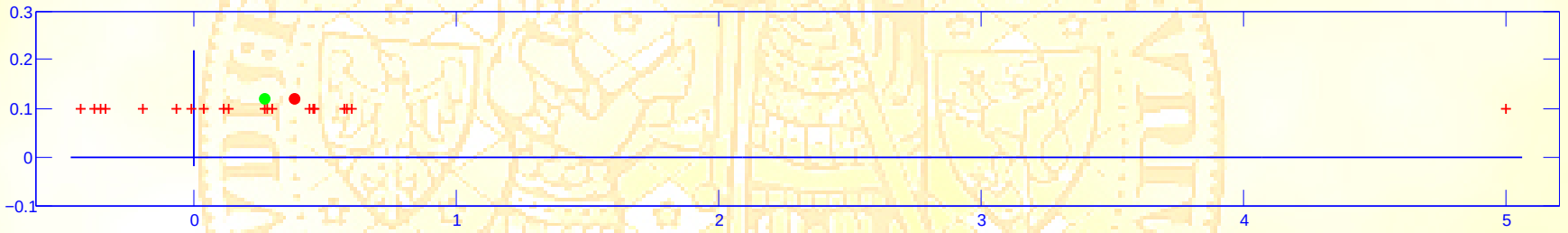
Absolute values of data



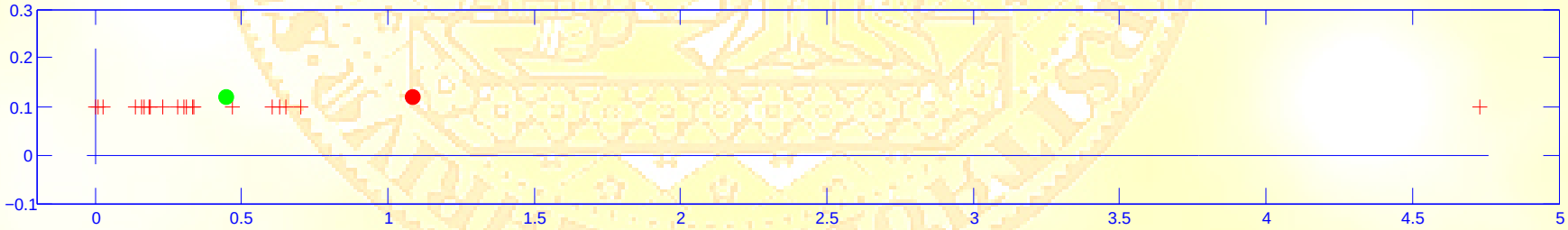
Demonstrating abilities of MAD

Observe the mean • and the median •
and standard deviation s_n • and $\hat{\sigma}_{MAD}$ •.

Contamination at point 5



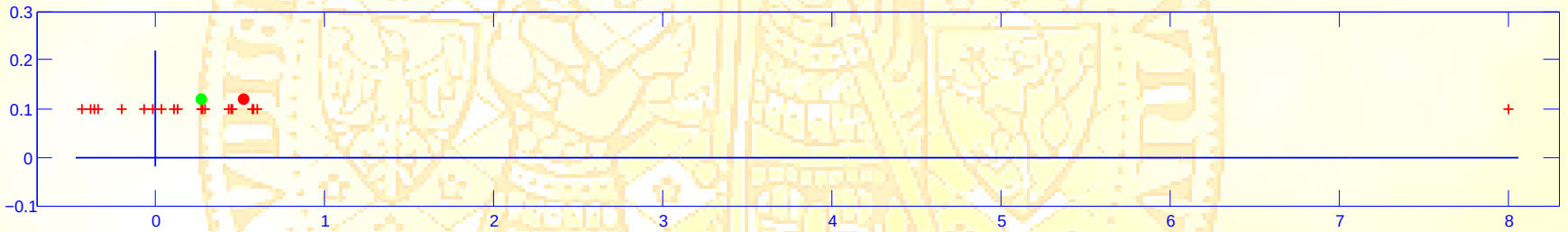
Absolute values of data



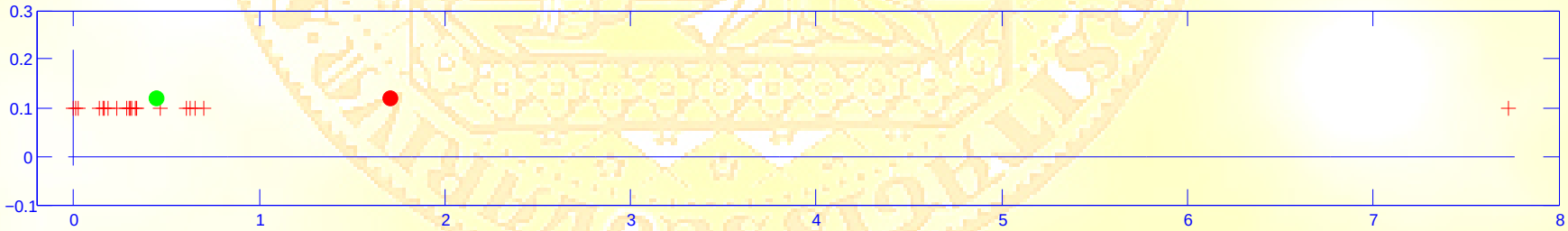
Demonstrating abilities of MAD

Observe the mean • and the median •
and standard deviation s_n • and $\hat{\sigma}_{MAD}$ •.

Contamination at point 8



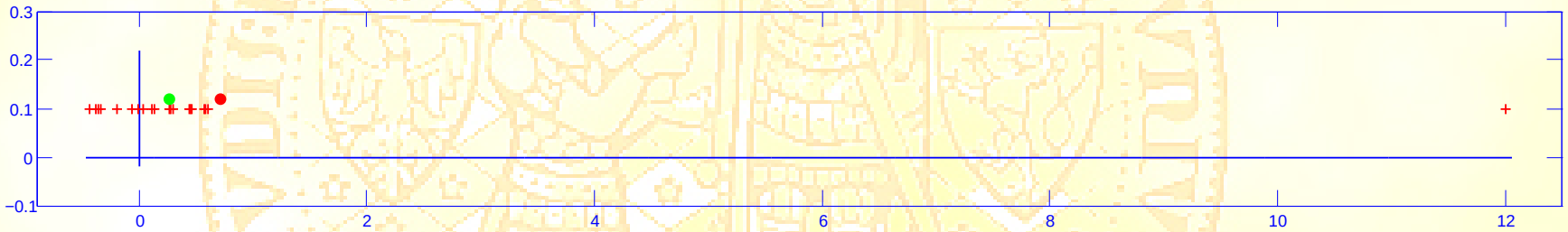
Absolute values of data



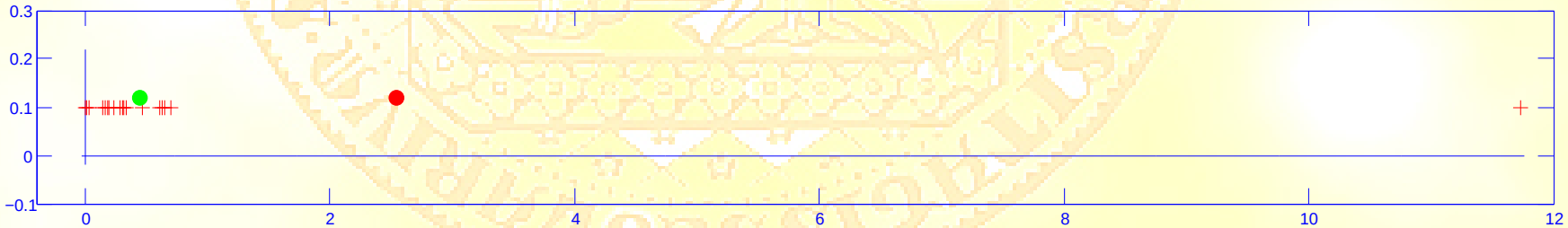
Demonstrating abilities of MAD

Observe the mean • and the median •
and standard deviation s_n • and $\hat{\sigma}_{MAD}$ •.

Contamination at point 12



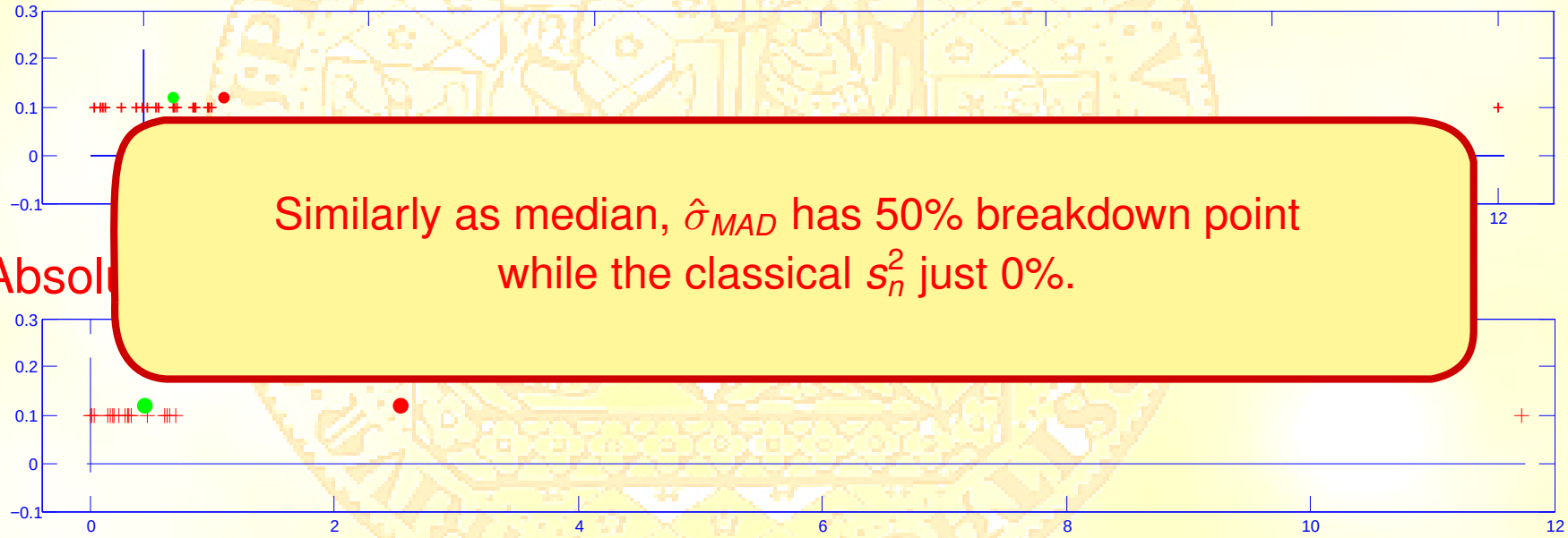
Absolute values of data



Demonstrating abilities of MAD

Observe the mean • and the median •
and standard deviation s_n • and $\hat{\sigma}_{MAD}$ •.

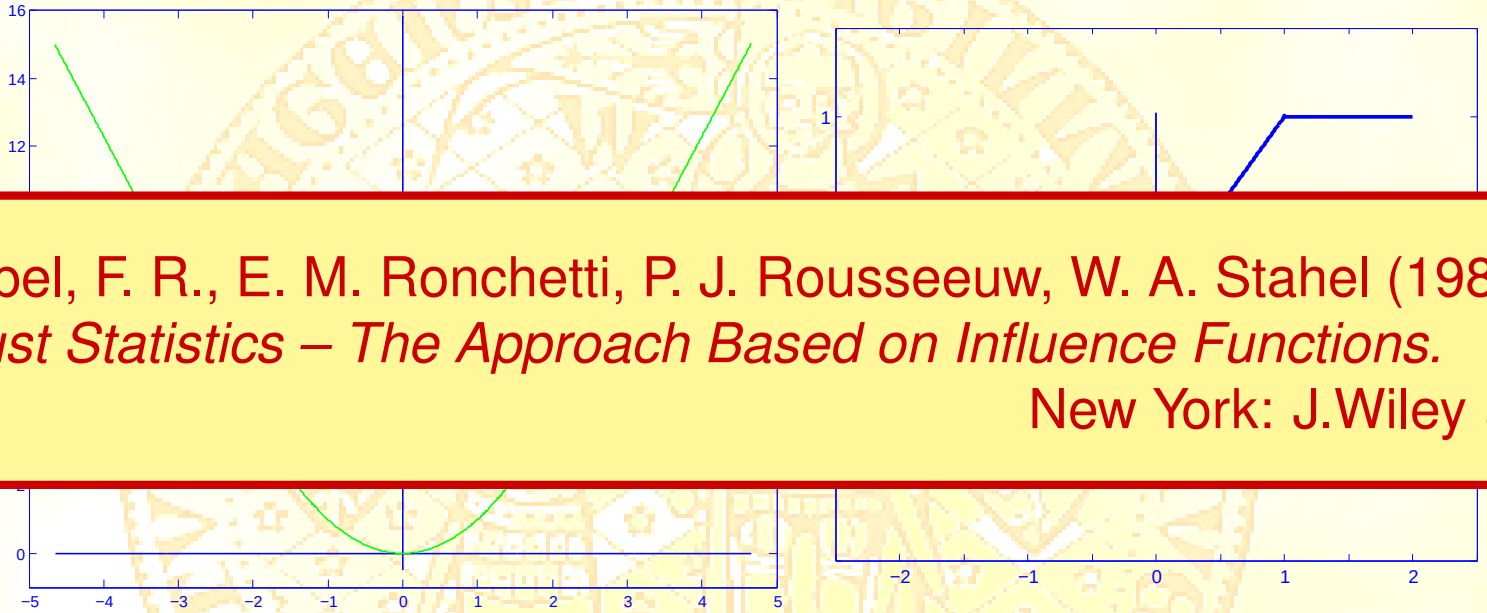
Contamination at point 12



Similarly as median, $\hat{\sigma}_{MAD}$ has 50% breakdown point
while the classical s_n^2 just 0%.

Reviewing the basic families of robust estimators - general parameter.

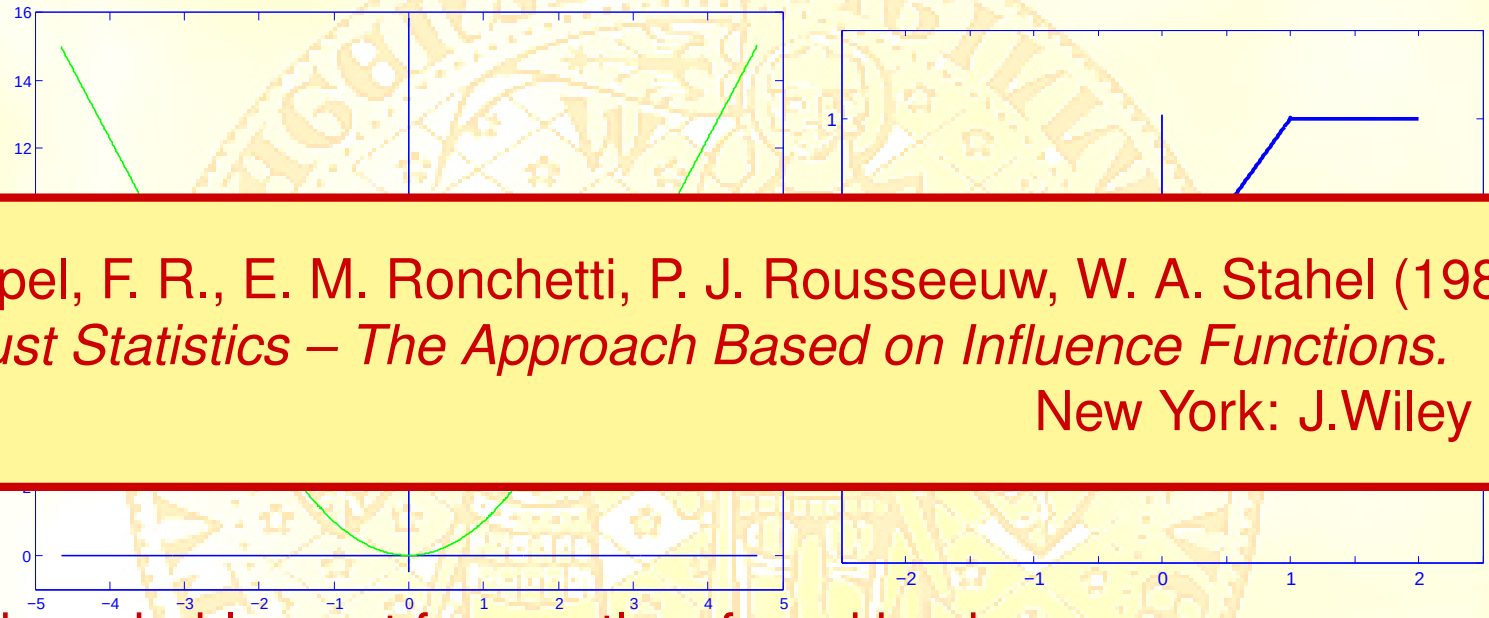
For the nearly exhaustive explanation see:



Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley *&* Son.

Reviewing the basic families of robust estimators - general parameter.

For the nearly exhaustive explanation see:

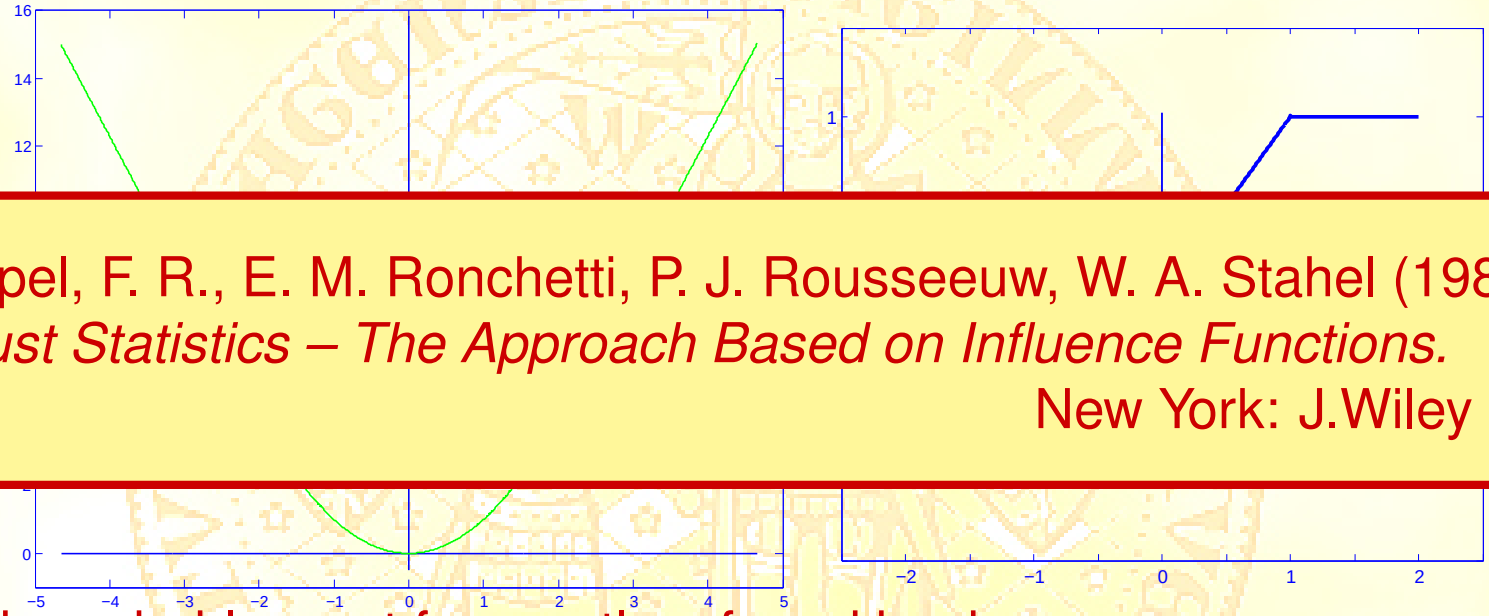


Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley *&* Son.

This is probably most frequently referred book
having an extremely interesting first chapter -
which is without mathematics and can be read as a detective story.

Reviewing the basic families of robust estimators - general parameter.

For the nearly exhaustive explanation see:



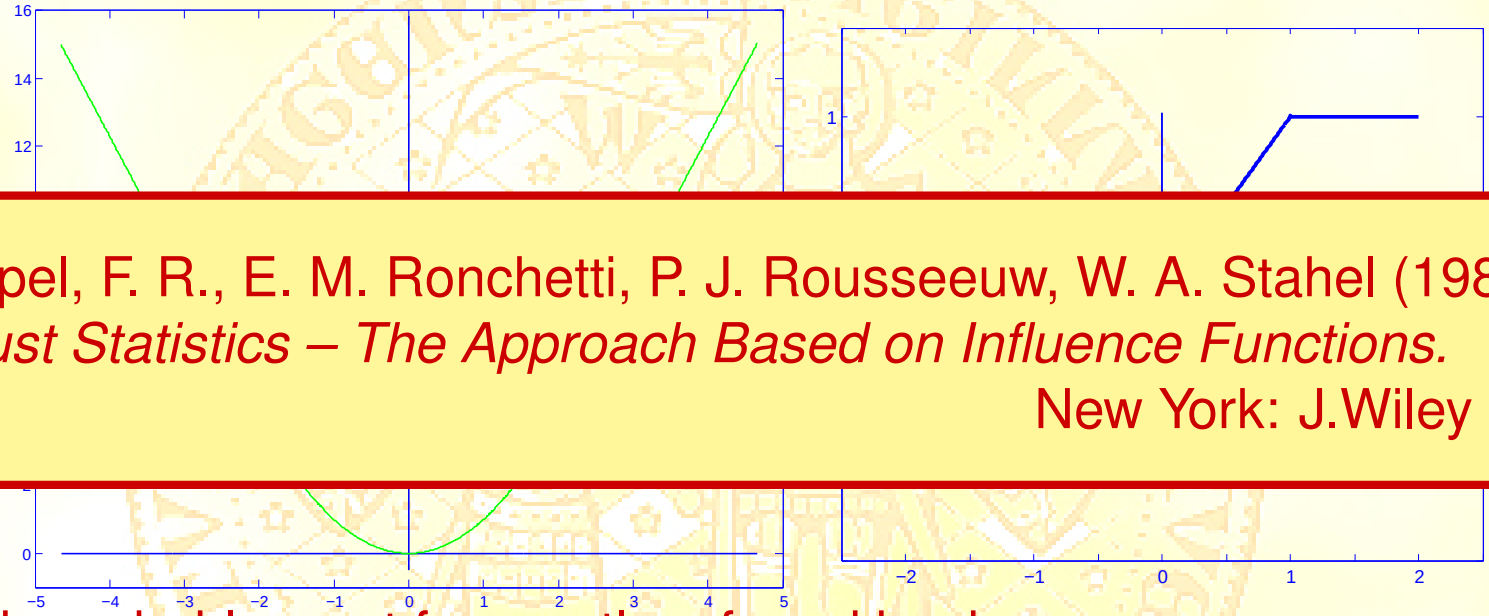
Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley *&* Son.

This is probably most frequently referred book
having an extremely interesting first chapter -
which is without mathematics and can be read as a detective story.

The rest of book is mostly beyond the scope of this basic lecture
but we shall (without the proofs) quote some results from it.

Reviewing the basic families of robust estimators - general parameter.

For the nearly exhaustive explanation see:



Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley *&* Son.

This is probably most frequently referred book
having an extremely interesting first chapter -
which is without mathematics and can be read as a detective story.

The rest of book is mostly beyond the scope of this basic lecture
but we shall (without the proofs) quote some results from it.
(Let's give only one example.)

Example of searching for an optimal M-estimator of location.

Assume the underlying parent d. f. $F(x)$

with differentiable density $f(x)$ which is symmetric

and ask for the M-estimator solving the location problem

and having following properties:

Example of searching for an optimal M-estimator of location.

Assume the underlying parent d. f. $F(x)$

with differentiable density $f(x)$ which is symmetric

and ask for the M-estimator solving the location problem

and having following properties:

1. The efficiency as high as possible,

Example of searching for an optimal M-estimator of location.

Assume the underlying parent d. f. $F(x)$

with differentiable density $f(x)$ which is symmetric

and ask for the M-estimator solving the location problem

and having following properties:

1. The efficiency as high as possible,
2. a priori given gross-error sensitivity.

Example of searching for an optimal M-estimator of location.

Assume the underlying parent d. f. $F(x)$

with differentiable density $f(x)$ which is symmetric

and ask for the M-estimator solving the location problem

and having following properties:

1. The efficiency as high as possible,
2. a priori given gross-error sensitivity.

Example of searching for an optimal M-estimator of location.

Assume the underlying parent d. f. $F(x)$

with differentiable density $f(x)$ which is symmetric

and ask for the M-estimator solving the location problem

and having following properties:

1. The efficiency as high as possible,
2. a priori given gross-error sensitivity.

The solution is given by

$$\psi(x) = \max\{-b, \min\{b, -f'(x)/f(x)\}\}.$$

An example of the likelihood function $f'(x)/f(x)$

Let's consider the standard normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{x^2}{2}\right\},$$

An example of the likelihood function $f'(x)/f(x)$

Let's consider the standard normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{x^2}{2}\right\},$$

i. e.

$$f'(x) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{x^2}{2}\right\} \cdot \{-x\}, = f(x) \cdot \{-x\},$$

An example of the likelihood function $f'(x)/f(x)$

Let's consider the standard normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{x^2}{2}\right\},$$

i. e.

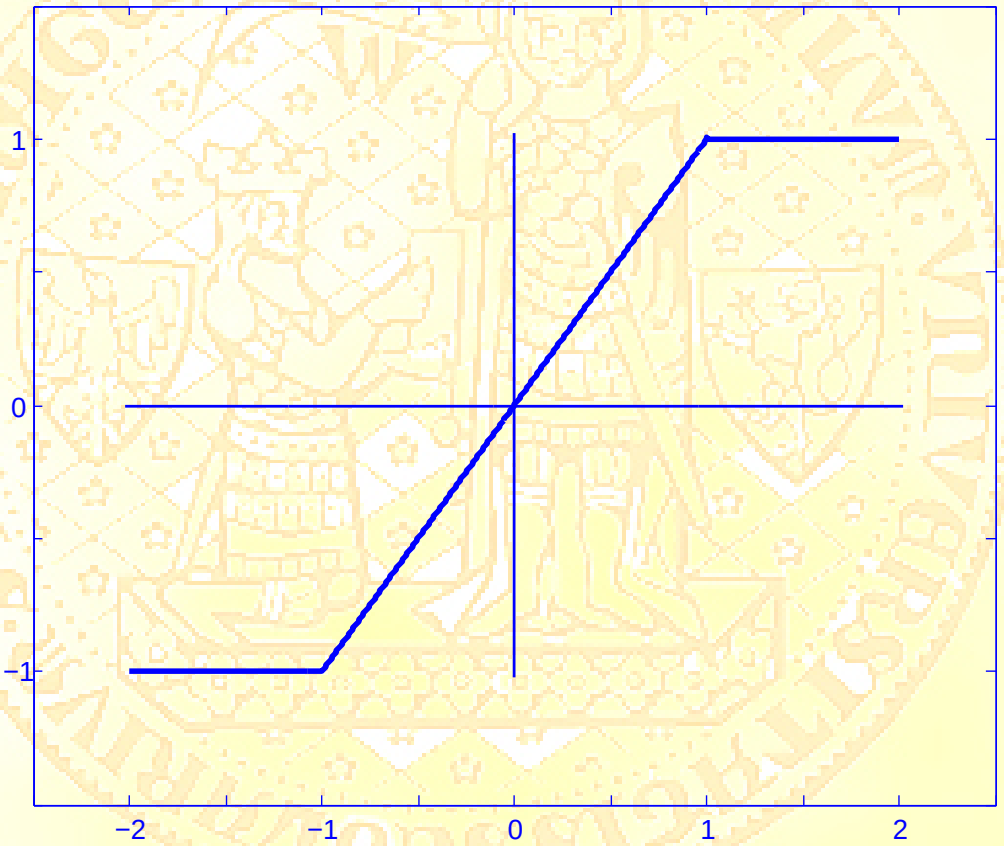
$$f'(x) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{x^2}{2}\right\} \cdot \{-x\}, = f(x) \cdot \{-x\},$$

hence

$$-\frac{f'(x)}{f(x)} = x.$$

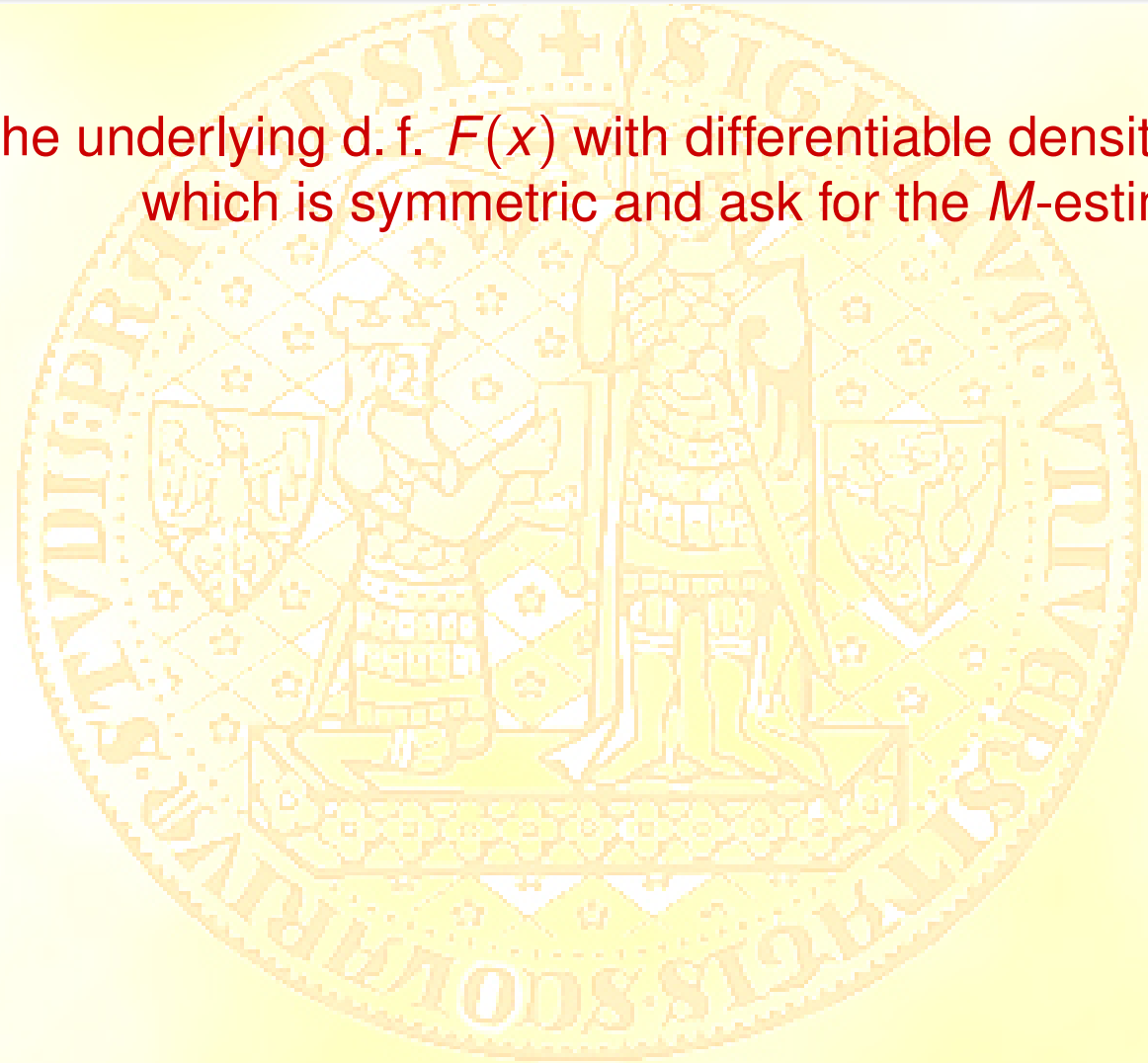
Example of searching for an optimal M-estimator of location.

Specifying $F(x) = \Phi(x)$, we obtain



Example of searching for an optimal M-estimator of location.

Assume the underlying d. f. $F(x)$ with differentiable density $f(x)$ which is symmetric and ask for the M-estimator having:



Example of searching for an optimal M-estimator of location.

Assume the underlying d. f. $F(x)$ with differentiable density $f(x)$ which is symmetric and ask for the M-estimator having:

1. The efficiency as high as possible,

Example of searching for an optimal M-estimator of location.

Assume the underlying d. f. $F(x)$ with differentiable density $f(x)$ which is symmetric and ask for the M-estimator having:

1. The efficiency as high as possible,
2. a priori given gross-error sensitivity,

Example of searching for an optimal M-estimator of location.

Assume the underlying d. f. $F(x)$ with differentiable density $f(x)$ which is symmetric and ask for the M-estimator having:

1. The efficiency as high as possible,
2. a priori given gross-error sensitivity,
3. a priori given rejection point c.

Example of searching for an optimal M-estimator of location.

Assume the underlying d. f. $F(x)$ with differentiable density $f(x)$ which is symmetric and ask for the M-estimator having:

1. The efficiency as high as possible,
2. a priori given gross-error sensitivity,
3. a priori given rejection point c.

Example of searching for an optimal *M*-estimator of location.

Assume the underlying d. f. $F(x)$ with differentiable density $f(x)$ which is symmetric and ask for the *M*-estimator having:

1. The efficiency as high as possible,
2. a priori given gross-error sensitivity,
3. a priori given rejection point c.

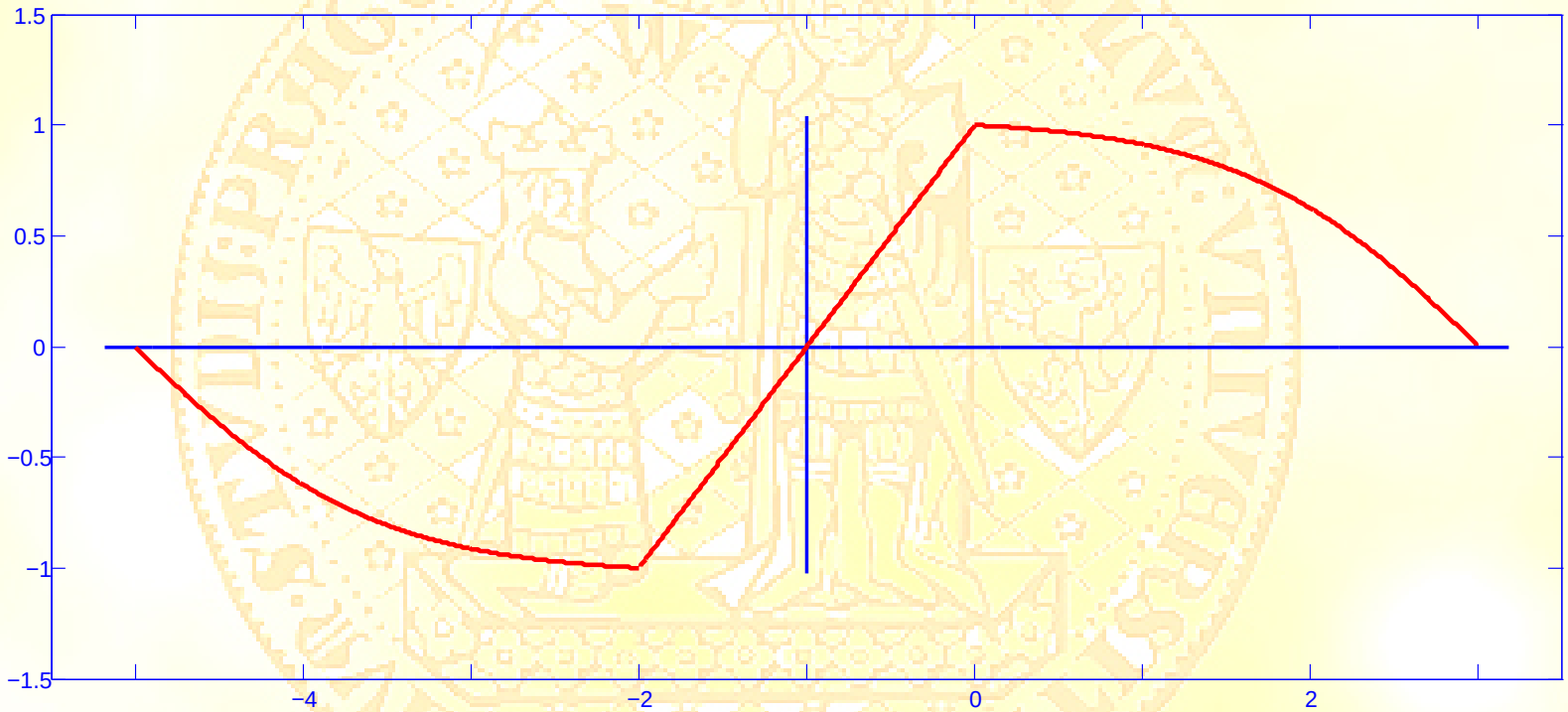
The solution is given by

$$\psi(x) = \max\{-h(x), \min\{h(x), f'(x)/f(x)\}\}$$

where the shape of the function $h(x)$ is given by employment of *tangh*(x) - see next slide.

Example of searching for an optimal *M*-estimator of location.

Specifying $F(x) = \Phi(x)$, we obtain



Other types of estimators

Estimators based on linear (hence the name) combination of order statistics - *L*-estimators

Estimating the location

Observations $z_1, z_2, \ldots, z_n \quad \Rightarrow \quad \underbrace{z_{(1)} \leq z_{(2)} \leq \ldots \leq z_{(n)}}$

These statistics are called order statistics

$$\hat{\mu}^{(L,n)} = \sum_{i=1}^{n} a_i \cdot z_{(i)}$$

where a_i's are a priori selected weights.

Other types of estimators

Estimators based on linear (hence the name) combination of order statistics - *L*-estimators

Estimating the location

Observations $z_1, z_2, ..., z_n \quad \Rightarrow \quad \underbrace{z_{(1)} \leq z_{(2)} \leq ... \leq z_{(n)}}$

These statistics are called order statistics

$$\hat{\mu}^{(L,n)} = \sum_{i=1}^{n} a_i \cdot z_{(i)}$$

where a_i's are a priori selected weights.

Estimating the scale

Put $r_i = \left| z_i - \hat{\mu}^{(L,n)} \right| \quad \Rightarrow \quad r_{(1)} \leq r_{(2)} \leq ... \leq r_{(n)}$

$$\hat{\sigma}^{(L,n)} = \sum_{i=1}^{n} b_i \cdot r_{(i)}$$

where b_i's are again a priori selected weights.

Other types of estimators

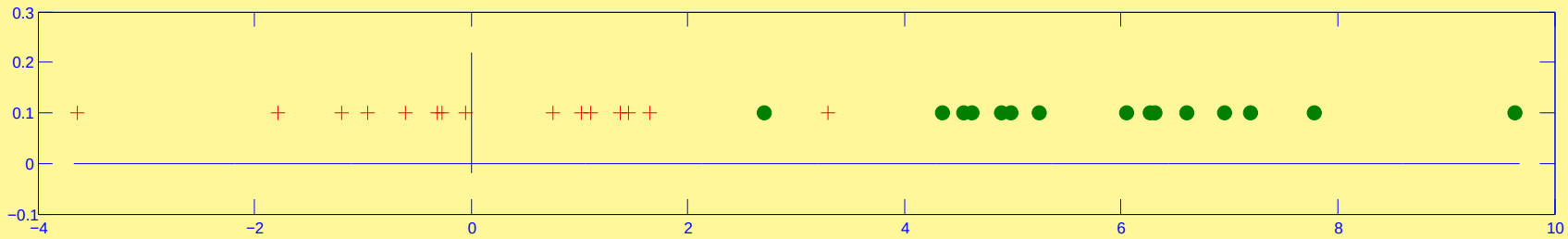
Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations, $\Delta \in R$ and consider data

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n.$$

The situation can looks like this for $\Delta = 3$



Other types of estimators

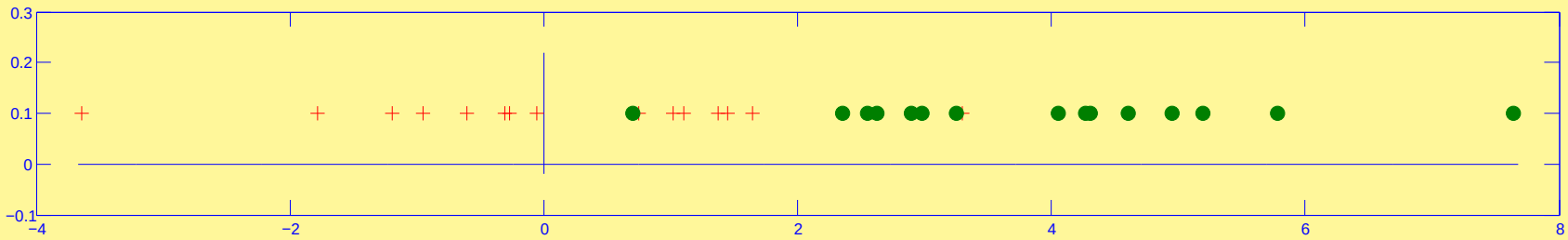
Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations, $\Delta \in R$ and consider data

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n.$$

The situation can looks like this for $\Delta = 2$



Other types of estimators

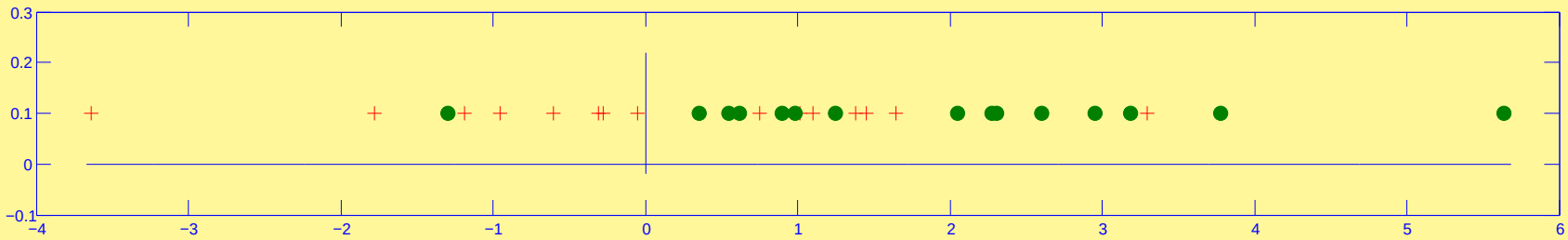
Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations, $\Delta \in R$ and consider data

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n.$$

The situation can looks like this for $\Delta = 1$



Other types of estimators

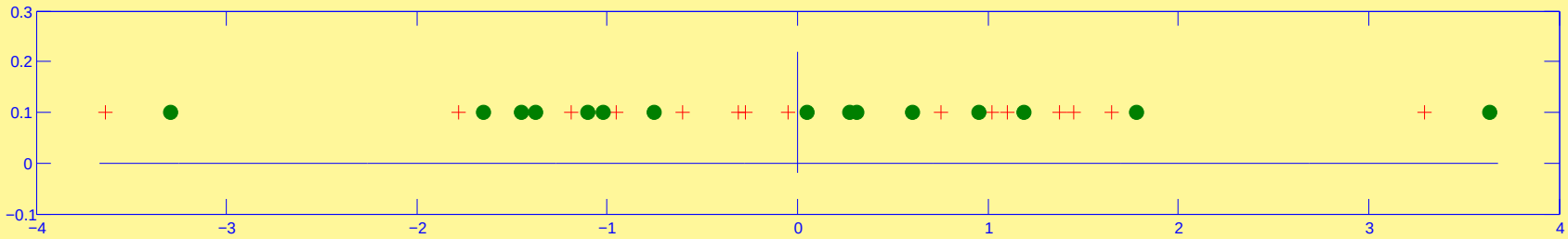
Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations, $\Delta \in R$ and consider data

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n.$$

The situation can looks like this for $\Delta = 0$



Other types of estimators

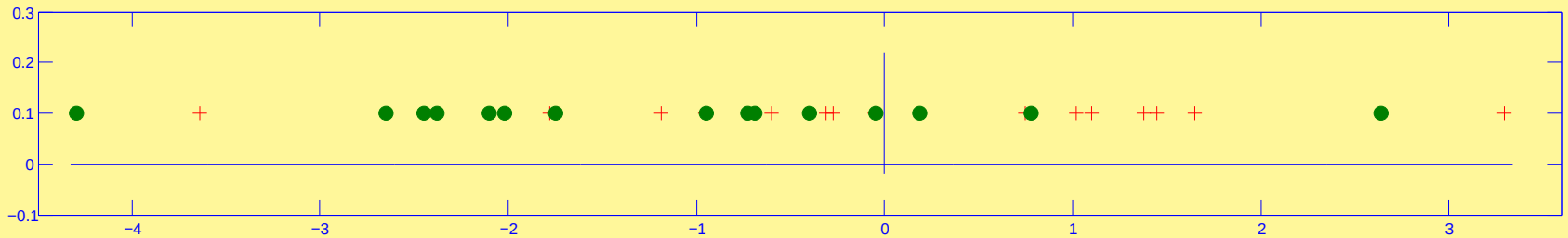
Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations, $\Delta \in R$ and consider data

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n.$$

The situation can looks like this for $\Delta = -0.5$



Other types of estimators

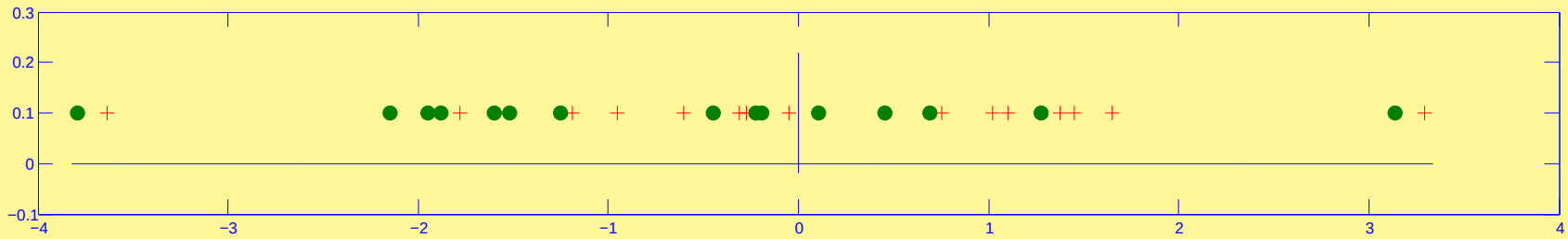
Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations, $\Delta \in R$ and consider data

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n.$$

The situation can looks like this for $\Delta = -0.25$



Other types of estimators

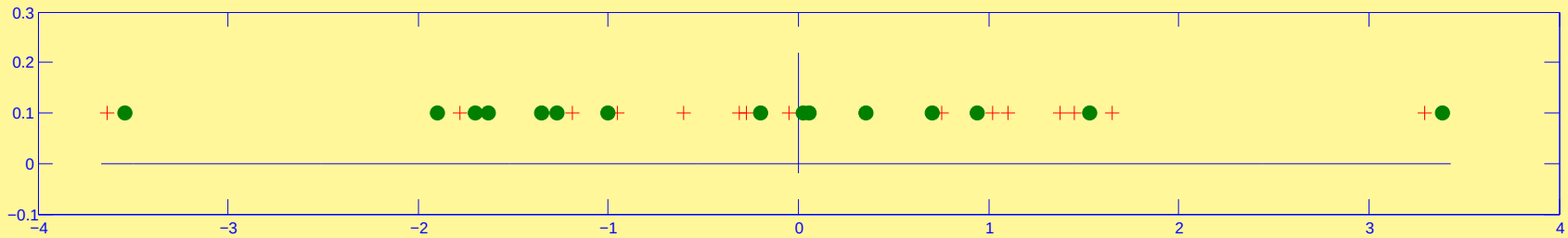
Estimators based on rank statistics (hence the name)
***R*-estimators**

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations, $\Delta \in R$ and consider data

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n.$$

The situation can looks like this for $\Delta = -0.125$



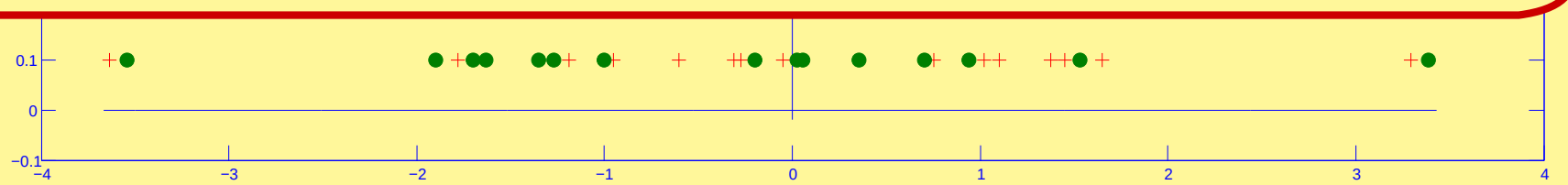
Other types of estimators

Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations, $\Delta \in R$ and consider data

Under assumption that data were generated by a symmetric density we can prove that Δ minimizing distance between $x_1, x_2, ..., x_n$ and $2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n$ is a consistent estimator of location.



Other types of estimators

Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations and $\Delta \in R$.

Let R_i be the rank of the i-th observations in the pooled sample

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n$$

and put $$S_n(\Delta) = \frac{1}{n} \sum_{i=1}^{n} a_n(R_i)$$

where $a_n(R) = n \int_{\frac{R-1}{n}}^{\frac{R}{n}} \Psi(u)\mathrm{d}u$ with $\Psi(u) = \Psi(1-u)$ $(\rightarrow \int_0^1 \Psi(u)\mathrm{d}u = 0)$.

Other types of estimators

Estimators based on rank statistics (hence the name)
R-estimators

Estimating the location

Let $x_1, x_2, ..., x_n$ be observations and $\Delta \in R$.

Let R_i be the rank of the i-th observations in the pooled sample

$$x_1, x_2, ..., x_n, 2\Delta - x_1, 2\Delta - x_2, ..., 2\Delta - x_n$$

and put $$S_n(\Delta) = \frac{1}{n} \sum_{i=1}^{n} a_n(R_i)$$

where $a_n(R) = n \int_{\frac{R-1}{n}}^{\frac{R}{n}} \Psi(u)\mathrm{d}u$ with $\Psi(u) = \Psi(1-u)$ $(\rightarrow \int_0^1 \Psi(u)\mathrm{d}u = 0)$.

Then put

$$\hat{\mu}^{(R,n)} = \underset{\Delta \in R}{\arg\min}\ S_n(\Delta).$$

Other types of estimators

Minimal distance estimators
Estimating a general parameter

Let $\{F_\theta(x)\}_{\theta\in\Theta}$ $\quad x_1, x_2, ..., x_n \rightarrow F^{(n)}(x)$ empirical d. f.

$\mathcal{D}(F, G)$ a distance on the space of all d. f.'s,
e. g. Prokhorov metric π or some I-divergence

Other types of estimators

Minimal distance estimators
Estimating a general parameter

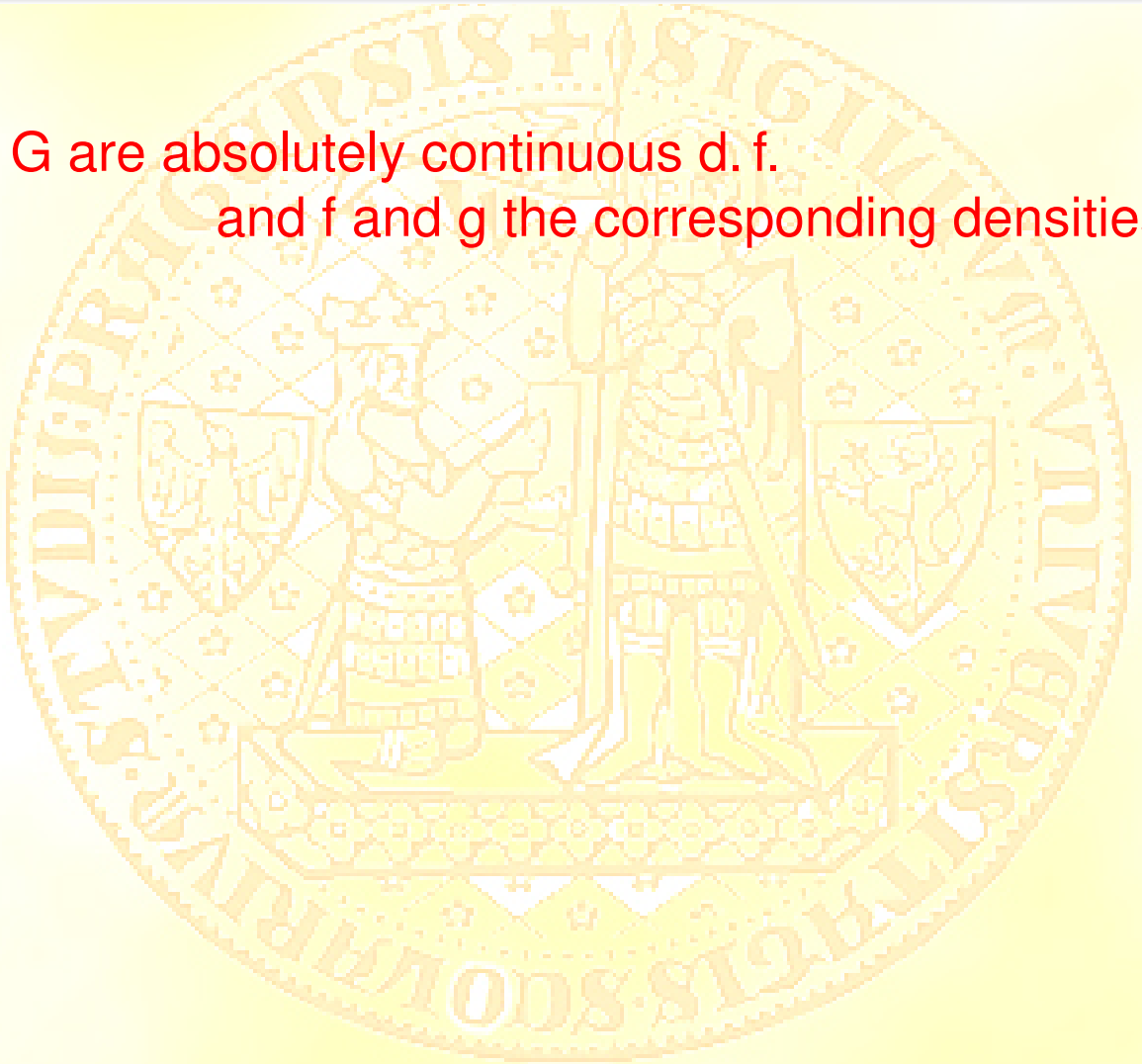
Let $\{F_\theta(x)\}_{\theta\in\Theta}$ $\quad x_1, x_2, ..., x_n \rightarrow F^{(n)}(x)$ empirical d. f.

$\mathcal{D}(F, G)$ a distance on the space of all d. f.'s,
e. g. Prokhorov metric π or some I-divergence

$$\hat{\theta}^{(MD,n)} = \underset{\theta\in\Theta}{\arg\min}\ \mathcal{D}(F_\theta, F^{(n)})$$

Kullbac-Leibler divergence

Let F and G are absolutely continuous d. f.
and f and g the corresponding densities, respectively.



Kullbac-Leibler divergence

Let F and G are absolutely continuous d. f.
and f and g the corresponding densities, respectively.

Then

$$KL(F,G) = \int log\left(\frac{g(x)}{f(x)}\right) \cdot g(x)\mathrm{d}x$$

is called *Kullbac-Leibler divergence*.

Kullbac-Leibler divergence

Let F and G are absolutely continuous d. f.
and f and g the corresponding densities, respectively.

Then

$$KL(F,G) = \int log\left(\frac{g(x)}{f(x)}\right) \cdot g(x)\mathrm{d}x$$

is called *Kullbac-Leibler divergence*.

By Jensen's inequality we easy prove that

$$KL(F,G) \geq 0.$$

Kullbac-Leibler divergence

Let F and G are absolutely continuous d. f.
and f and g the corresponding densities, respectively.

Then

$$KL(F,G) = \int log\left(\frac{g(x)}{f(x)}\right) \cdot g(x)\mathrm{d}x$$

is called *Kullbac-Leibler divergence*.

By Jensen's inequality we easy prove that

$$KL(F,G) \geq 0.$$

The problem with orthogonality - Igor Vajda.

Jensen's inequality

Let $h(x)$ be convex and X a random variable having the mean value $\mathbb{E}X$. Then

$$\mathbb{E}\{h(X)\} \geq h(\mathbb{E}X).$$

Jensen's inequality

Let $h(x)$ be convex and X a random variable having the mean value $\mathbb{E}X$. Then

$$\mathbb{E}\{h(X)\} \geq h(\mathbb{E}X).$$

Proof: As (see the next slide)

$$h(x) \geq h(\mathbb{E}X) + b \cdot (X - \mathbb{E}X),$$

Jensen's inequality

Let $h(x)$ be convex and X a random variable having the mean value $\mathbb{E}X$. Then

$$\mathbb{E}\{h(X)\} \geq h(\mathbb{E}X).$$

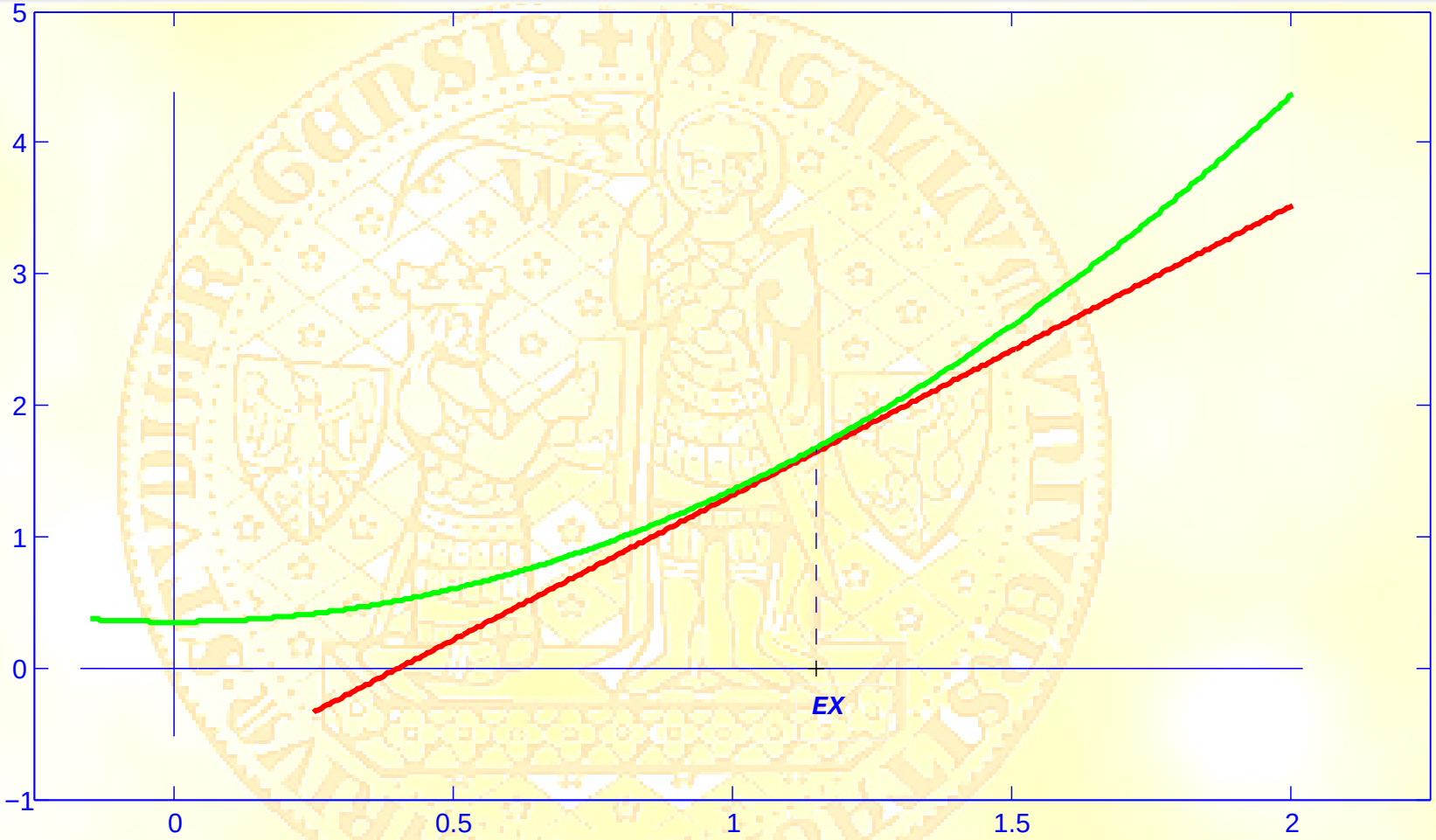
Proof: As (see the next slide)

$$h(x) \geq h(\mathbb{E}X) + b \cdot (X - \mathbb{E}X),$$

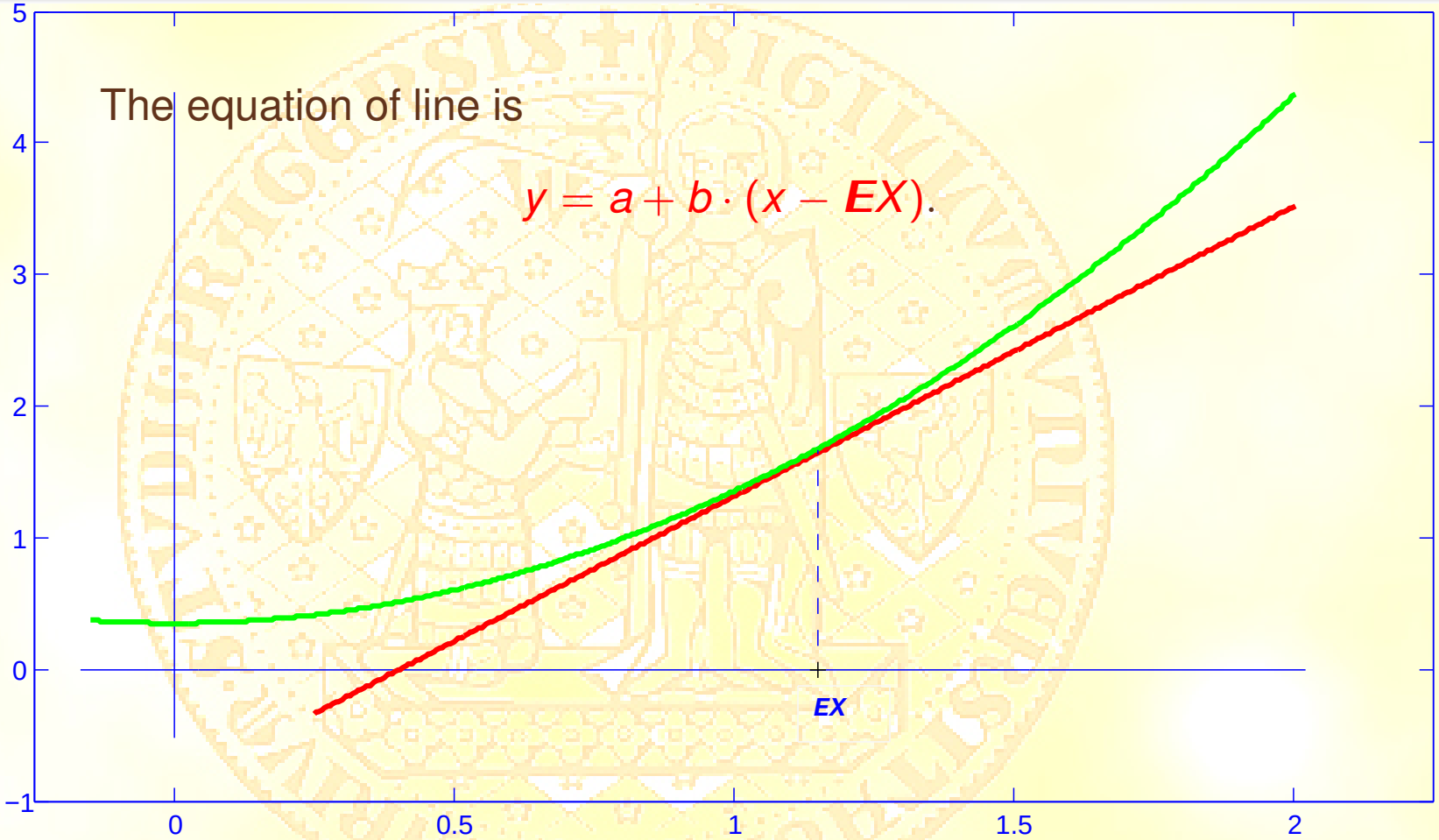
we have

$$\mathbb{E}\{h(X)\} \geq h(\mathbb{E}X) + b \cdot \mathbb{E}(X - \mathbb{E}X) = h(\mathbb{E}X).$$

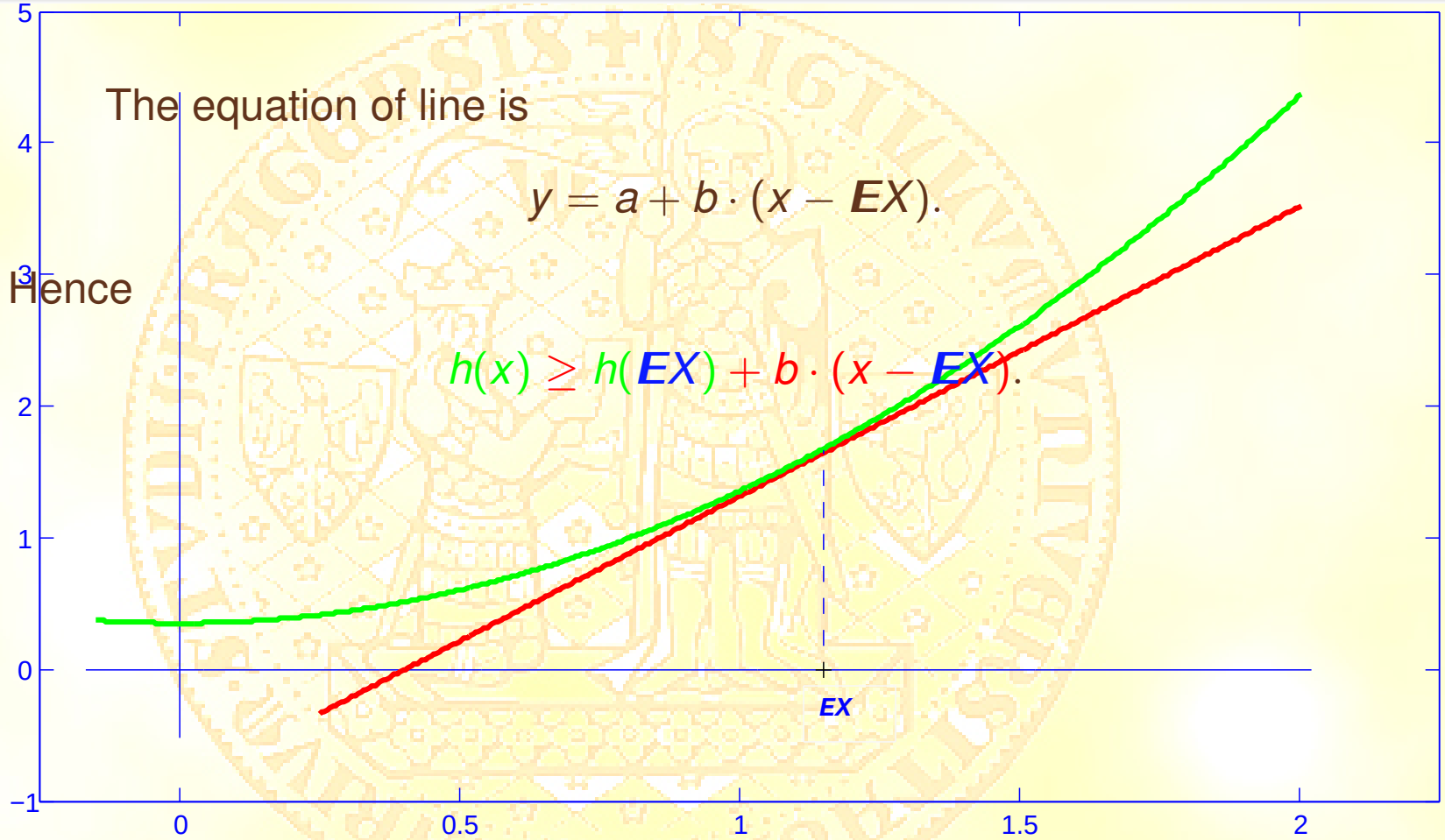
Jensen's inequality



Jensen's inequality



Jensen's inequality



Kullbac-Leibler divergence

By Jensen's inequality we easy prove that

$$KL(F,G) = \int log\left(\frac{g(x)}{f(x)}\right) \cdot g(x)\mathrm{d}x = \mathbb{E}_G log\left(\frac{g(x)}{f(x)}\right) = -\mathbb{E}_G log\left(\frac{f(x)}{g(x)}\right)$$

Kullbac-Leibler divergence

By Jensen's inequality we easy prove that

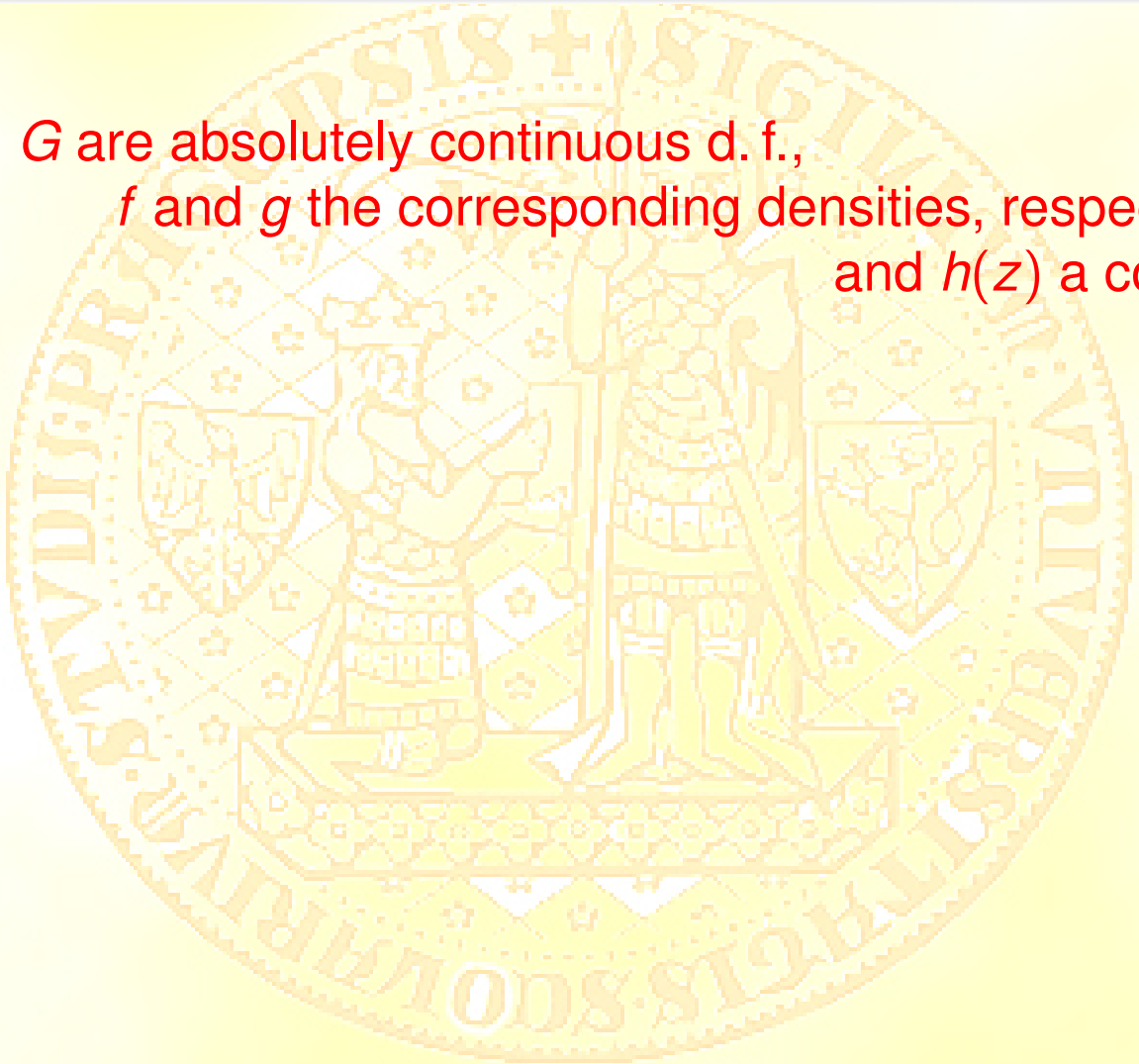
$$KL(F,G) = \int log\left(\frac{g(x)}{f(x)}\right) \cdot g(x)\mathrm{d}x = I\!E_G log\left(\frac{g(x)}{f(x)}\right) = -I\!E_G log\left(\frac{f(x)}{g(x)}\right)$$

As $-log(z)$ is convex function, we have

$$KL(F,G) = -I\!E_G log\left(\frac{f(x)}{g(x)}\right) \geq log\left(\int \frac{f(x)}{g(x)} g(x)\mathrm{dx}\right) = 0.$$

I-divergence

Let F and G are absolutely continuous d. f.,
f and g the corresponding densities, respectively,
and $h(z)$ a convex function.



I-divergence

Let F and G are absolutely continuous d. f.,
f and g the corresponding densities, respectively,
and $h(z)$ a convex function.

Then

$$I(F,G) = \int h\left(\frac{g(x)}{f(x)}\right) \cdot g(x)\mathrm{d}x$$

is called *I-divergence*.

I-divergence

Let F and G are absolutely continuous d. f.,
f and g the corresponding densities, respectively,
and $h(z)$ a convex function.

Then

$$I(F,G) = \int h\left(\frac{g(x)}{f(x)}\right) \cdot g(x)\mathrm{d}x$$

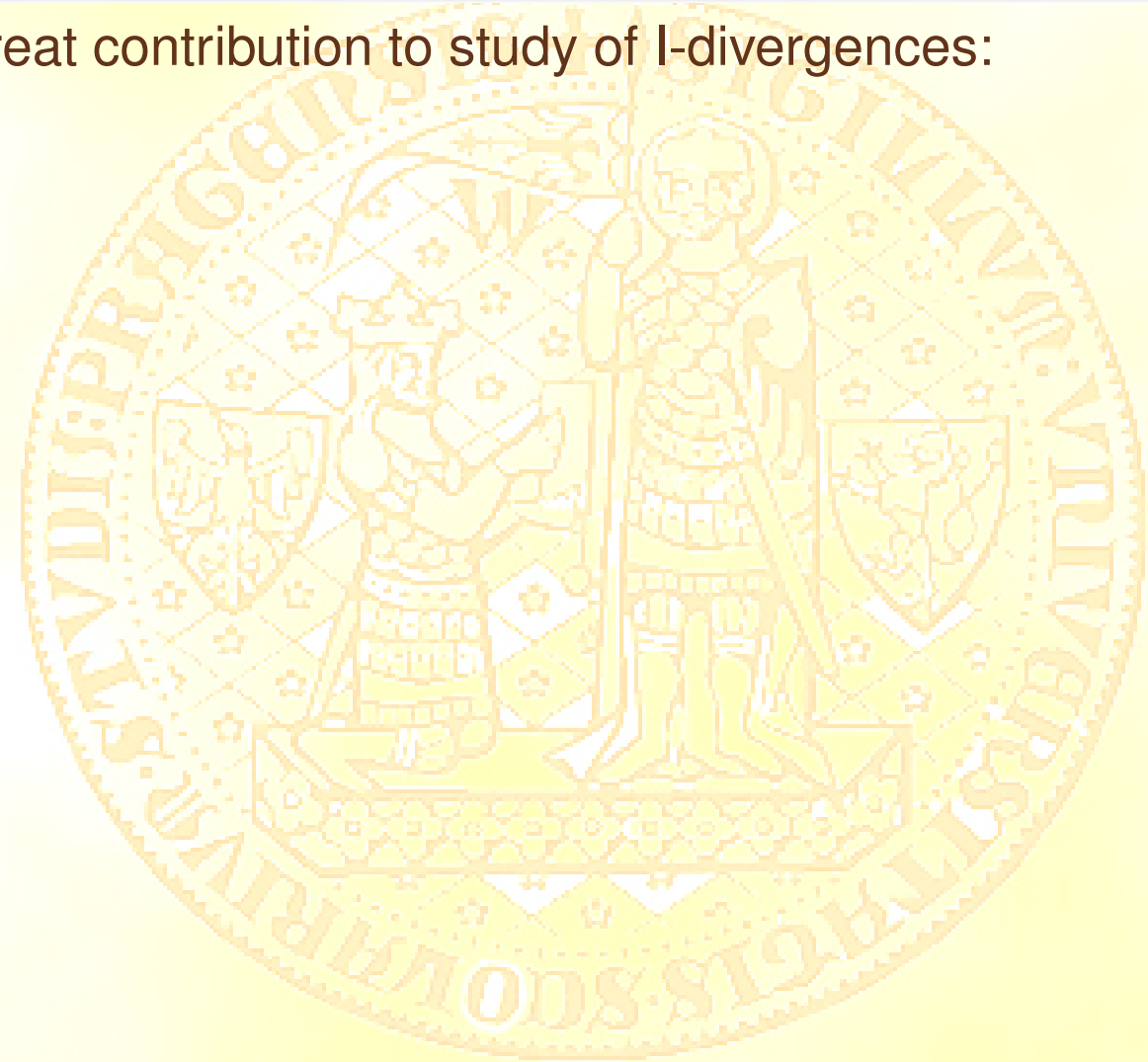
is called *I-divergence*.

By Jensen's inequality we again easy prove that

$$I(F,G) \geq 0.$$

Frequently used divergences

A great contribution to study of I-divergences:



Frequently used divergences

A great contribution to study of I-divergences:

Csiszár, I. (1975): I-divergence geometry of probability distributions and minimization problems.
Ann. Probab. 3, 146-158.

Frequently used divergences

A great contribution to study of I-divergences:

Csiszár, I. (1975): I-divergence geometry of probability distributions and minimization problems.
Ann. Probab. 3, 146-158.

One of the most frequently employed function $h(z)$

$$h(z) = \frac{z^{\alpha} - 1}{\alpha}, \qquad \alpha \in (0, 1].$$

Box, G. E. P., D. R. Cox (1964): An analysis of transformations.
Journal of the Royal Statistical Society, Series B, 26, 211 - 243.

Frequently used divergences

A great contribution to study of I-divergences:

Csiszár, I. (1975): I-divergence geometry of probability distributions and minimization problems.
Ann. Probab. 3, 146-158.

One of the most frequently employed function $h(z)$

$$h(z) = \frac{z^{\alpha} - 1}{\alpha}, \qquad \alpha \in (0, 1].$$

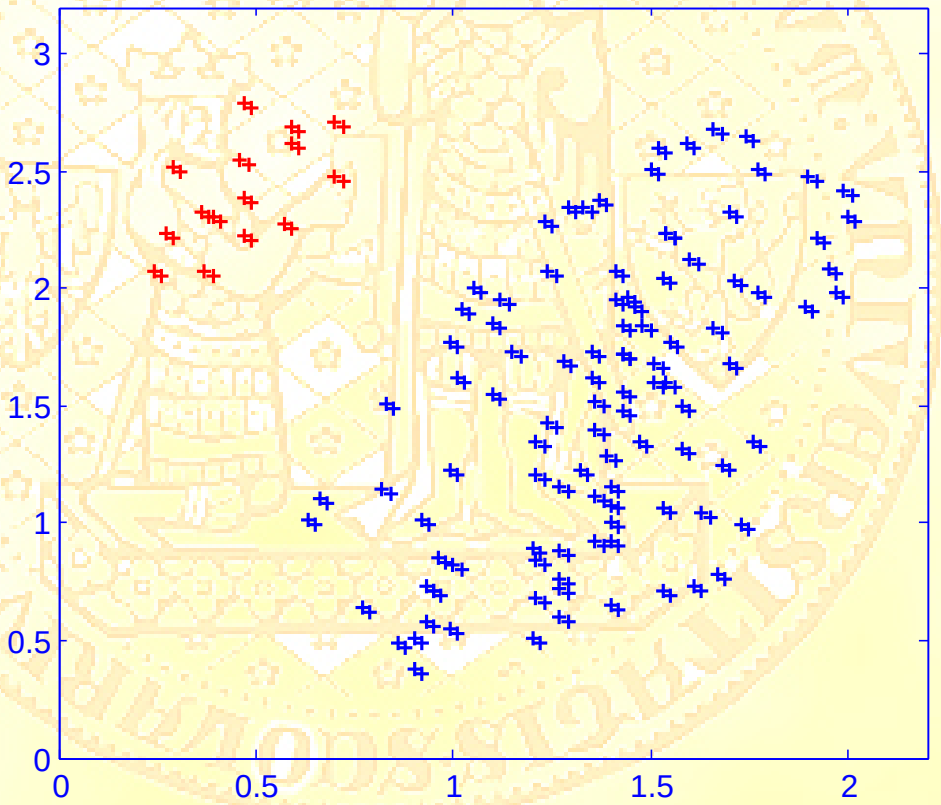
Box, G. E. P., D. R. Cox (1964): An analysis of transformations.
Journal of the Royal Statistical Society, Series B, 26, 211 - 243.

The *I-divergence* is then called α-divergence.

Other types of estimators

Minimal volume estimator

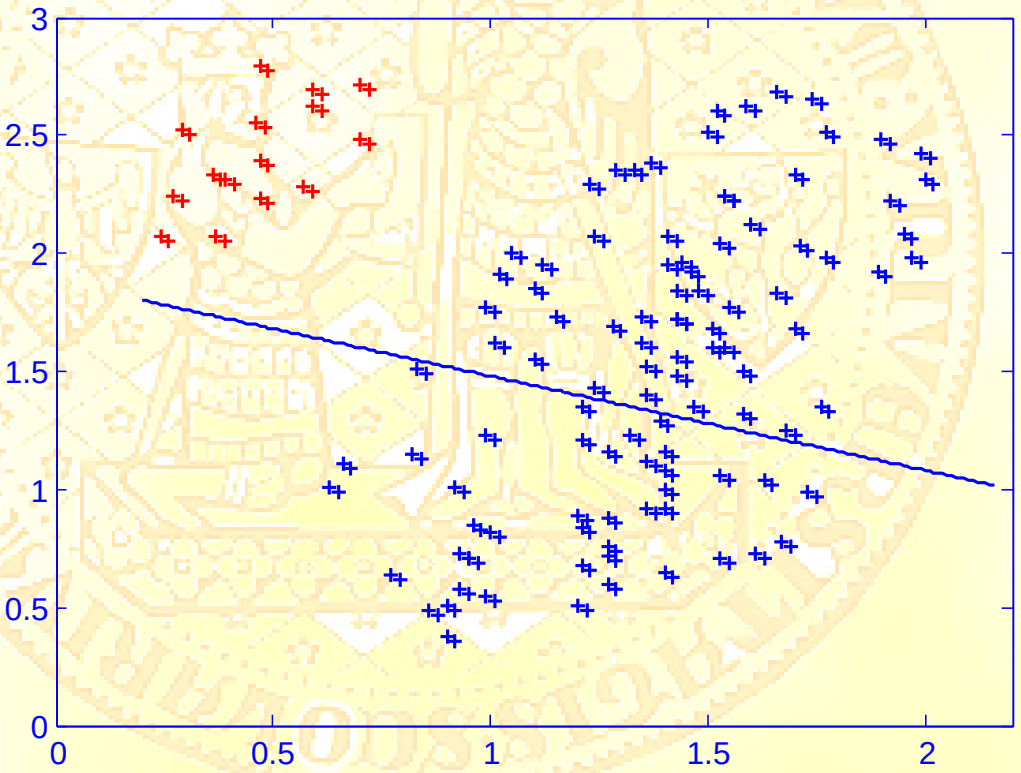
Estimating a regression model



Other types of estimators

By the way, the Ordinary Least Squares gives

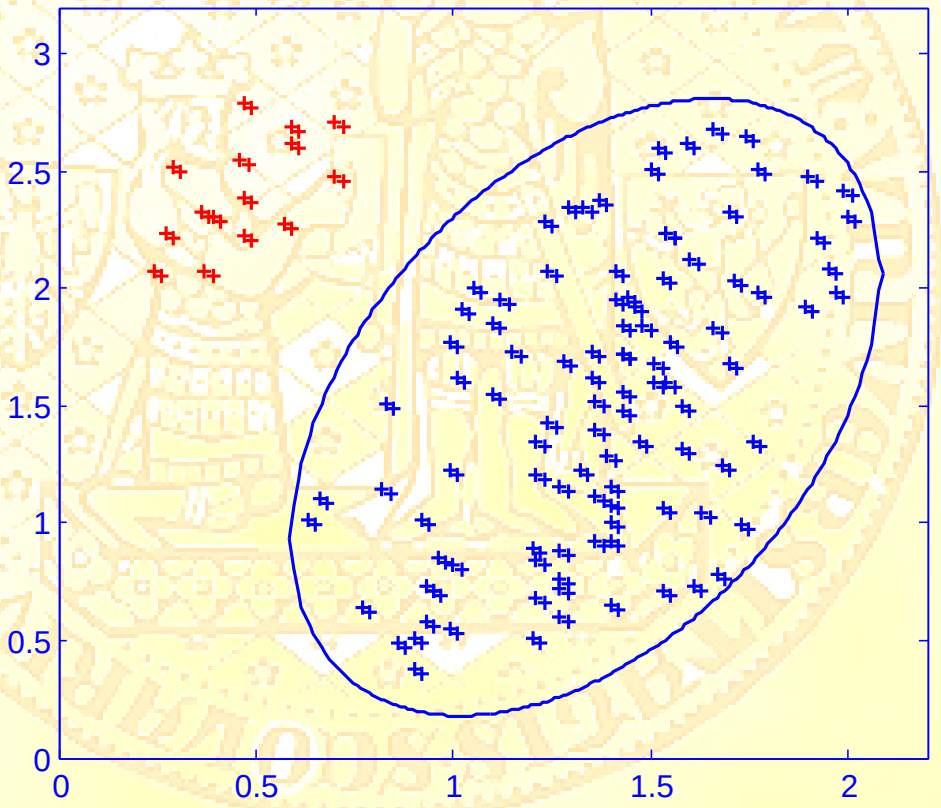
Estimating a regression model



Other types of estimators

Minimal volume estimator

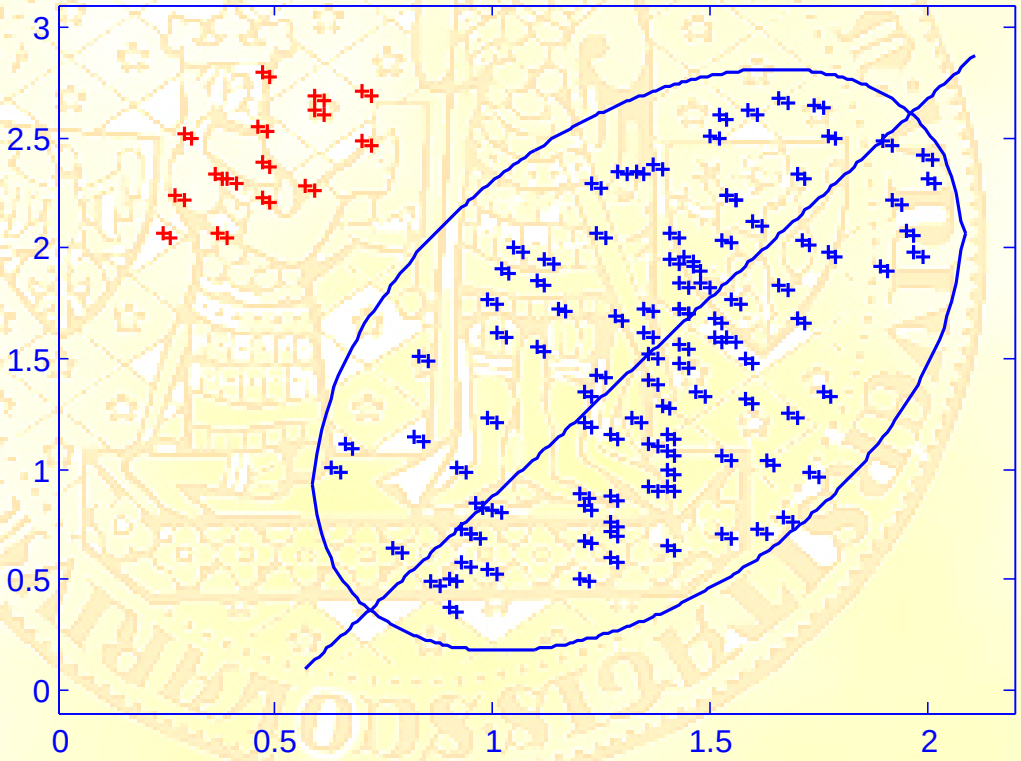
Estimating a general parameter



Other types of estimators

So, it seems we have nearly unmistakeable tool

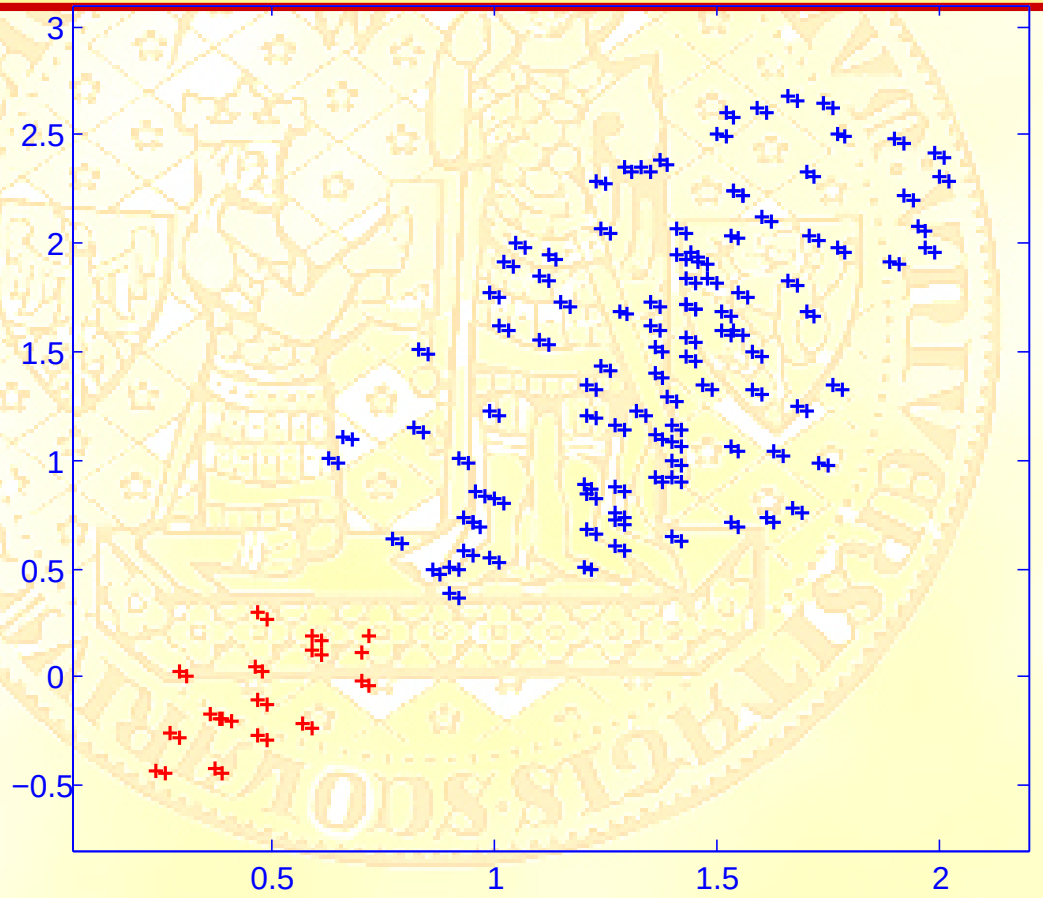
Estimating a regression model



Other types of estimators

But what about such a situation

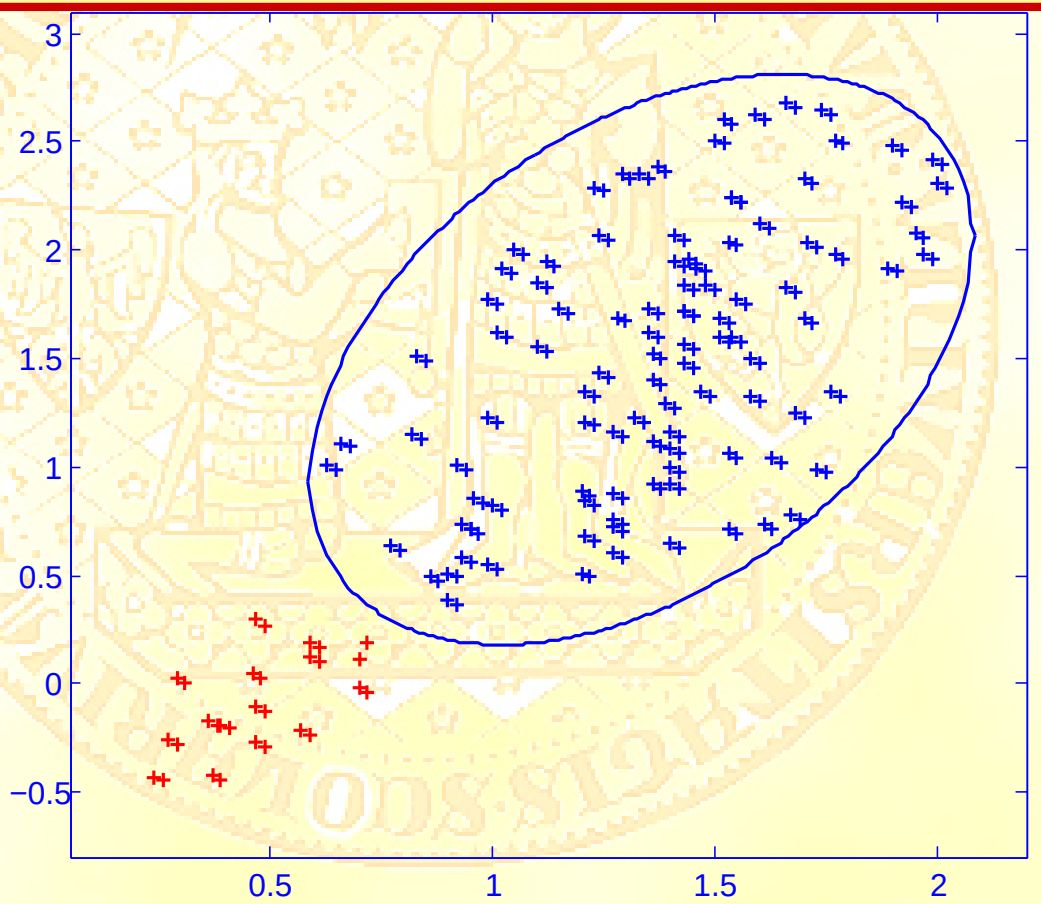
Estimating a regression model



Other types of estimators

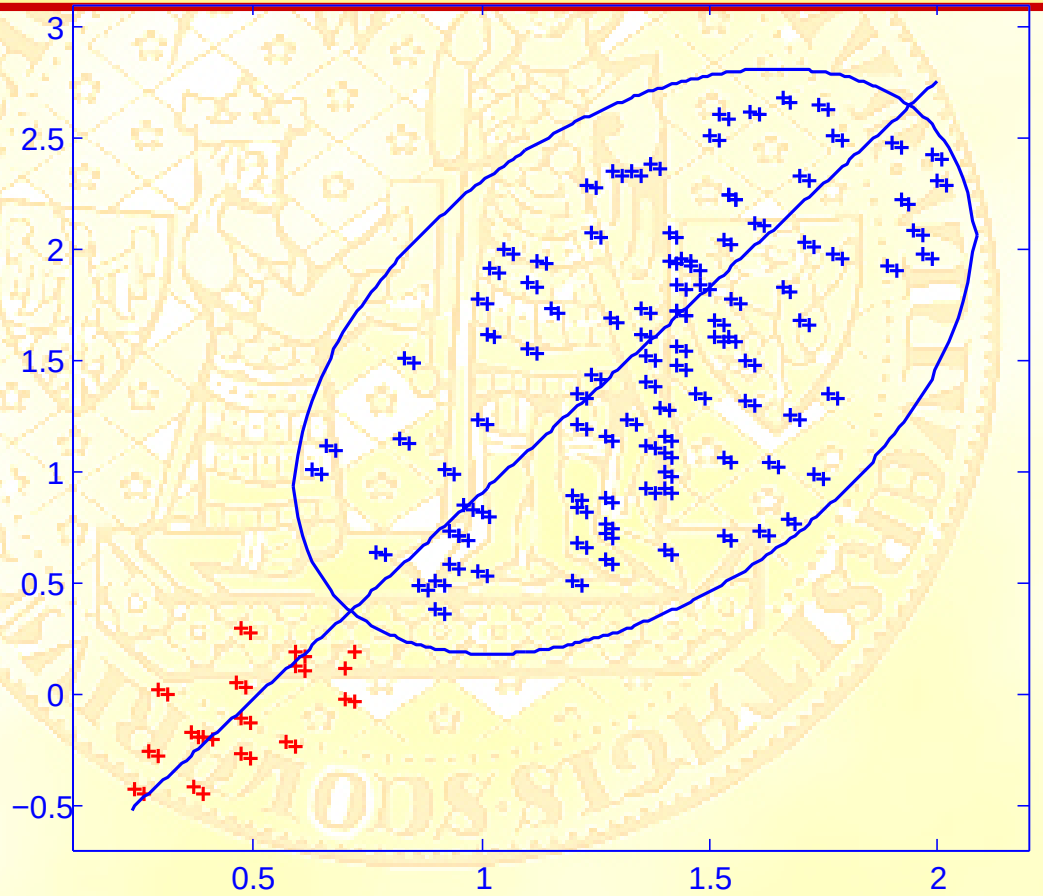
We can proceed as in previous case

Estimating a regression model



Other types of estimators

And the model is reasonable
but we lose idly some information



THANKS FOR ATTENTION