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 1

First of all - the (technical) framework of the course

First of all - the (technical) framework of the course

Technicalities:

1. The text of lecture (both - the full lecture and the handout) will be available on web (at least) from Monday (preceding the lecture), 2 p.m.

First of all - the (technical) framework of the course

Technicalities:

1. The text of lecture (both - the full lecture and the handout) will be available on web (at least) from Monday (preceding the lecture), 2 p.m.
2. An active presence on the seminars will be a necessary and (nearly) sufficient condition for passing the course - - the test at the end will be oriented on "ideas" not on "formulas".

First of all - the (technical) framework of the course

Technicalities:

1. The text of lecture (both - the full lecture and the handout) will be available on web (at least) from Monday (preceding the lecture), 2 p.m.
2. An active presence on the seminars will be a necessary and (nearly) sufficient condition for passing the course - - the test at the end will be oriented on "ideas" not on "formulas".
3. Any comment will be welcomed (to content, on an explanation, correction of ..., etc..)

First of all - the (technical) framework of the course

Technicalities:

1. The text of lecture (both - the full lecture and the handout) will be available on web (at least) from Monday (preceding the lecture), 2 p.m.
2. An active presence on the seminars will be a necessary and (nearly) sufficient condition for passing the course - - the test at the end will be oriented on "ideas" not on "formulas".
3. Any comment will be welcomed (to content, on an explanation, correction of ..., etc..)
4. I encourage You to be active also on lecture - - don't let me escape from any topic without understanding it.

The character of lectures and of seminars

The character of lectures and of seminars

1. The lectures will be oriented on ideas -
   - there will be pictures explainig them or creating an inspiration.

The character of lectures and of seminars

1. The lectures will be oriented on ideas -
   - there will be pictures explainig them or creating an inspiration.
2. Not to disappoint (completely) those
   who came for an exact mathematics
   - some pattern of mathematics will be included.

The character of lectures and of seminars

1. The lectures will be oriented on ideas -
   - there will be pictures explainig them or creating an inspiration.
2. Not to disappoint (completely) those
   who came for an exact mathematics
   - some pattern of mathematics will be included.
3. Some excursions to mathematics will be of general interest -
   - e. g. You can learn how it is with infinity,
   what is countable and uncountable.

The character of lectures and of seminars

1. The lectures will be oriented on ideas -
   - there will be pictures explainig them or creating an inspiration.
2. Not to disappoint (completely) those
   who came for an exact mathematics
   - some pattern of mathematics will be included.
3. Some excursions to mathematics will be of general interest -
   - e. g. You can learn how it is with infinity,
   what is countable and uncountable.
4. There will be some quick reminder(s)
   of something from statistics and econometrics, of history, etc.
   and at the end of term - if we will have some time
   - a drop of philosophy.

The character of lectures and of seminars

1. The lectures will be oriented on ideas -
   - there will be pictures explainig them or creating an inspiration.
2. Not to disappoint (completely) those
   who came for an exact mathematics
   - some pattern of mathematics will be included.
3. Some excursions to mathematics will be of general interest -
   - e. g. You can learn how it is with infinity,
   what is countable and uncountable.
4. There will be some quick reminder(s)
   of something from statistics and econometrics, of history, etc.
   and at the end of term - if we will have some time
   - a drop of philosophy.
5. On seminars - which will be completely under governance
   of Tomáš Křehlík, we assume mainly some exercises with software
   but also Your active role with creating them
   - just to fulfill the word "seminar".

Content of lecture

1. A very first insight into robustness
   - An atractive application as foreword

Content of lecture

1. A very first insight into robustness
   - An atractive application as foreword
2. A bit more modest motivation

Content of lecture

1. A very first insight into robustness
   - An atractive application as foreword
2. A bit more modest motivation
3. Let's start more serious discussion
   - Vulnerability of classical procedures to contamination
   - The classical requirements on estimators

Content

Content

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*

has found:

MAIN SUBGROUP

with number of industries 54 and the 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

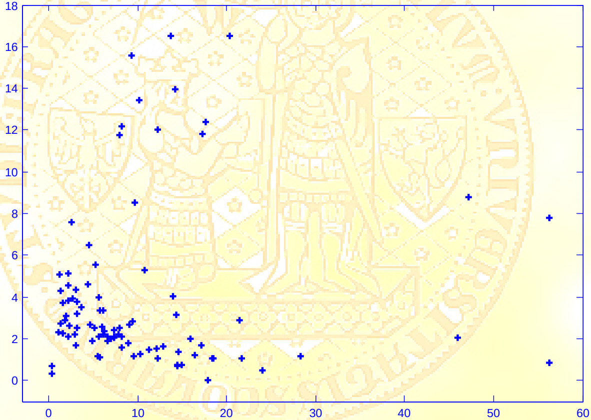
with number of industries 33 and the model

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

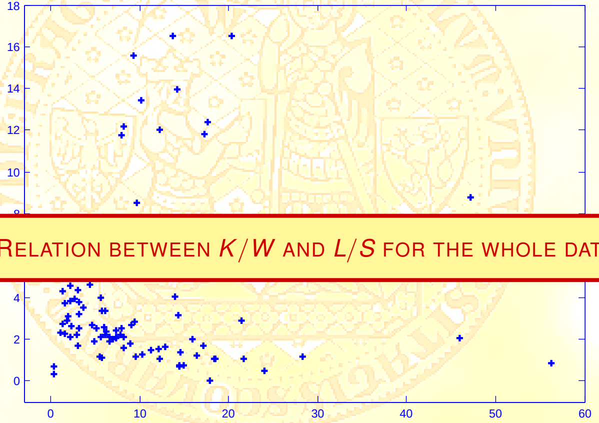
X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest 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*.

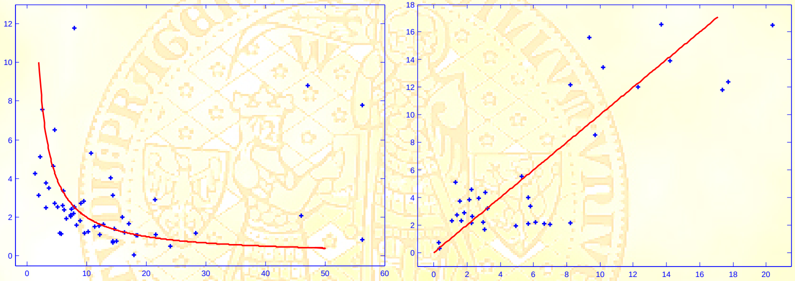


Analysis of the export from the Czech republic to EU in 1994 by means of the *Least Trimmed Squares*.



Relation between K/W and L/S for the whole data.

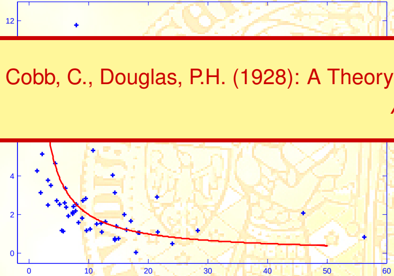
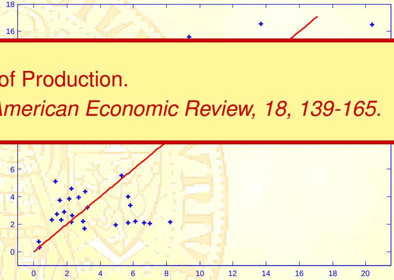
ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *Least Trimmed Squares*.



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).

It seems we have at hand a miraculous method

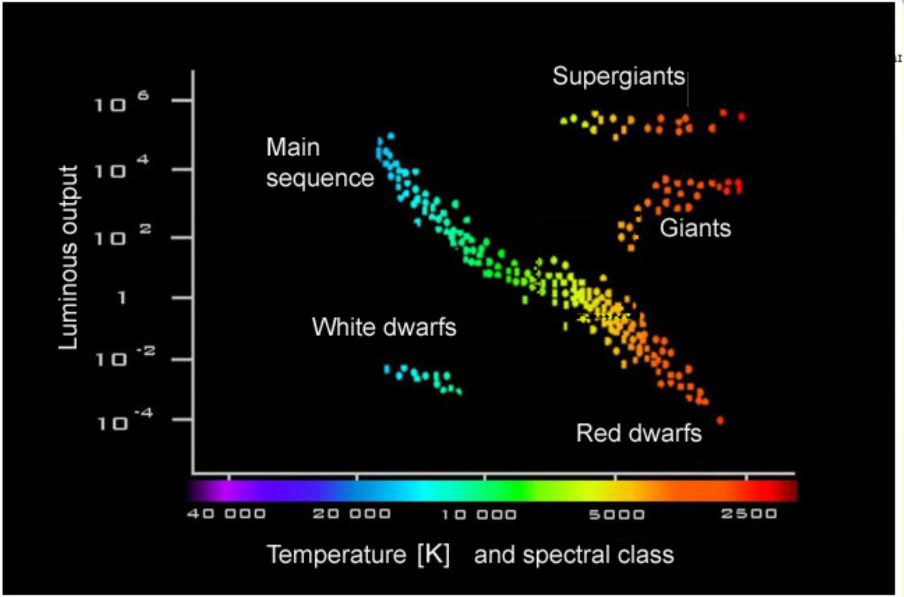
WARNING !!!

We haven't reached something which is

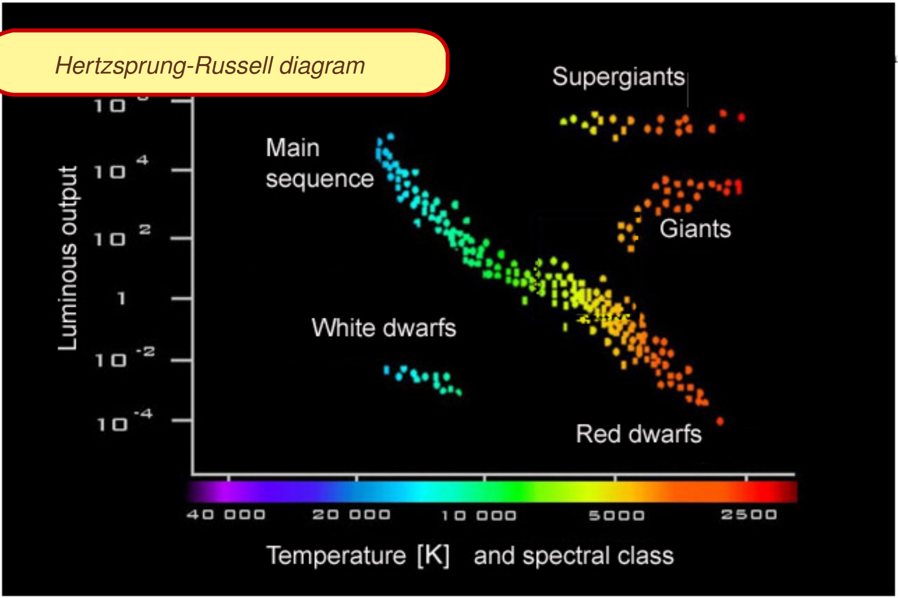
"BOMB und IDIOTEN SICHER"

but which is the powerful tool, if used with a care.

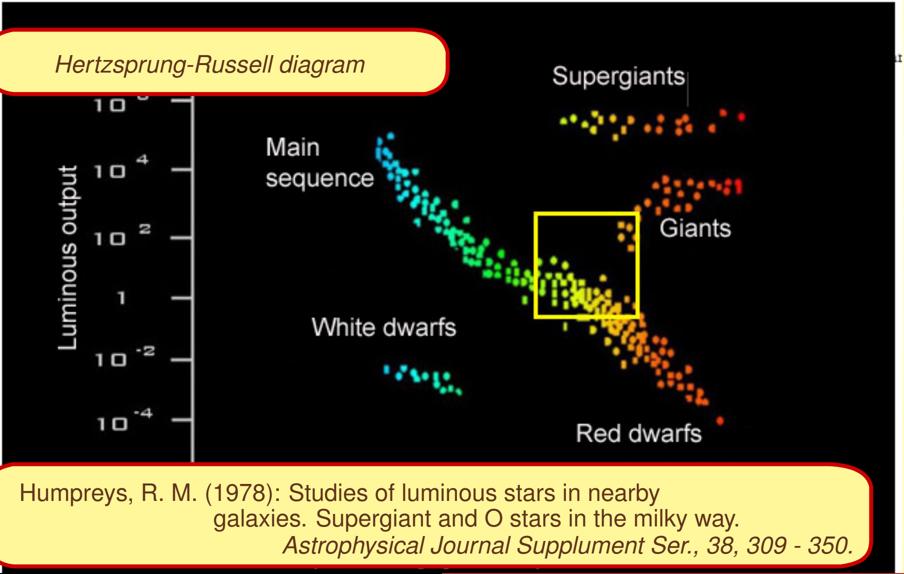
Content


Supergiants
Main sequence
Giants
White dwarfs
Red dwarfs
Luminous output
10^6
10^4
10^2
1
10^{-2}
10^{-4}
40 000
20 000
10 000
5000
2500
Temperature [K] and spectral class

Hertzsprung-Russell diagram

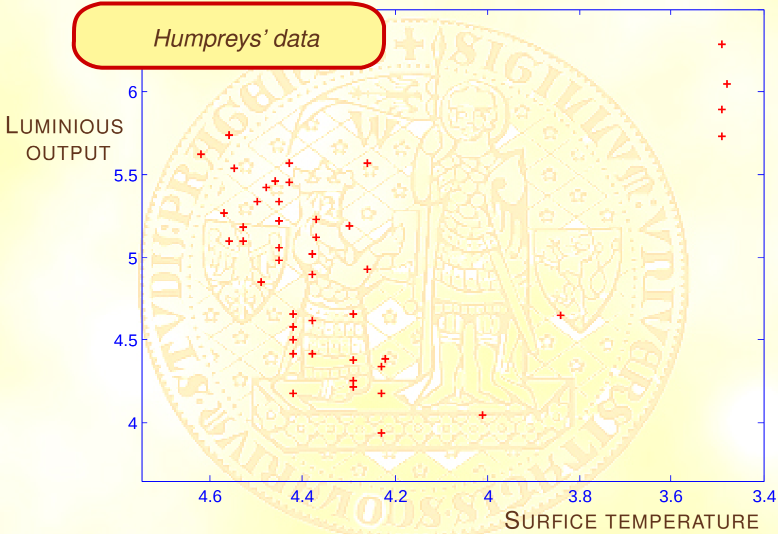


Hertzsprung-Russell diagram

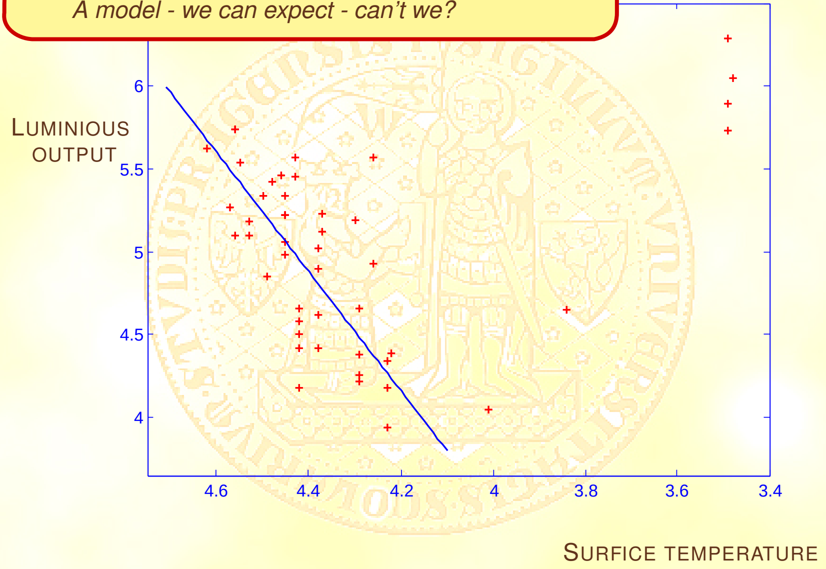


Humpreys, R. M. (1978): Studies of luminous stars in nearby galaxies. Supergiant and O stars in the milky way. *Astrophysical Journal Supplument Ser., 38, 309 - 350.*

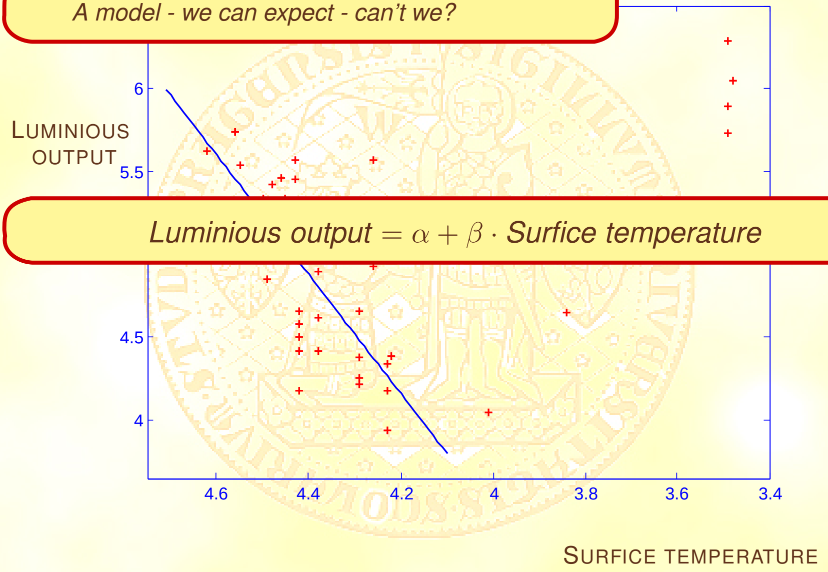
Humpreys' data



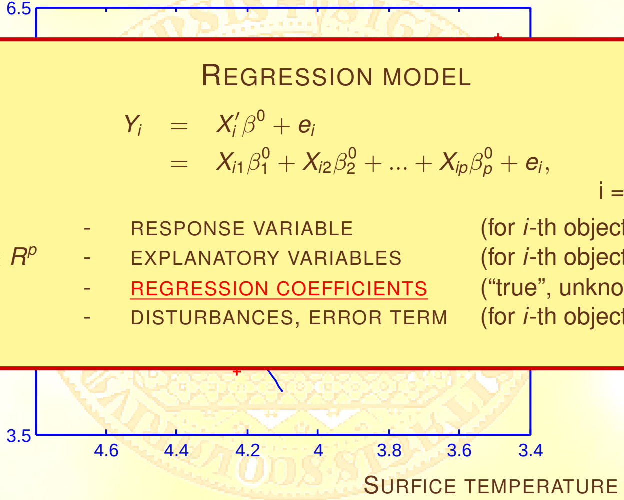
A model - we can expect - can't we?



A model - we can expect - can't we?



Luminious output $= \alpha + \beta \cdot$ *Surfice temperature*

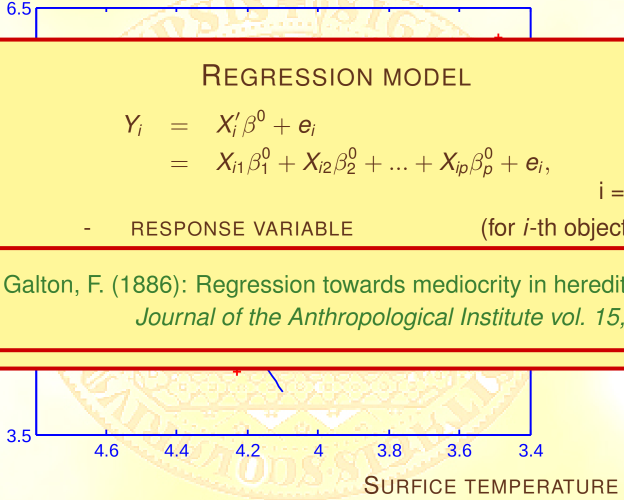


REGRESSION MODEL

$$
\begin{aligned}
Y_i &= X_i'\beta^0 + e_i \\
&= X_{i1}\beta_1^0 + X_{i2}\beta_2^0 + \ldots + X_{ip}\beta_p^0 + e_i,
\end{aligned}
$$

i = 1, 2, ...,n

Y_i	-	RESPONSE VARIABLE	(for i-th object, known)
$X_i \in R^p$	-	EXPLANATORY VARIABLES	(for i-th object, known)
β^0	-	REGRESSION COEFFICIENTS	("true", unknown)
e_i	-	DISTURBANCES, ERROR TERM	(for i-th object, unknown)



REGRESSION MODEL

$$
\begin{aligned}
Y_i &= X_i' \beta^0 + e_i \\
&= X_{i1}\beta_1^0 + X_{i2}\beta_2^0 + \ldots + X_{ip}\beta_p^0 + e_i,
\end{aligned}
$$

i = 1, 2, ...,n

Y_i - RESPONSE VARIABLE (for i-th object, known)

Galton, F. (1886): Regression towards mediocrity in hereditary stature.
Journal of the Anthropological Institute vol. 15,. 246–263.

Regression model

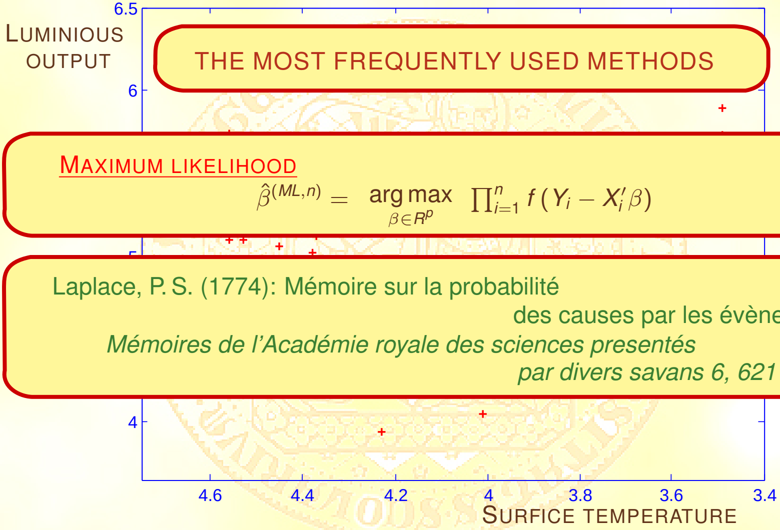
$$Y_i = X_i'\beta^0 + e_i$$
$$= X_{i1}\beta_1^0 + X_{i2}\beta_2^0 + \ldots + X_{ip}\beta_p^0 + e_i,$$

i = 1, 2, ...,n

Y_i - RESPONSE VARIABLE (for i-th object, known)

Galton, F. (1886): Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute vol. 15,. 246–263.*

THE TASK IS TO ESTIMATE UNKNOWN REGRESSION COEFFICIENTS

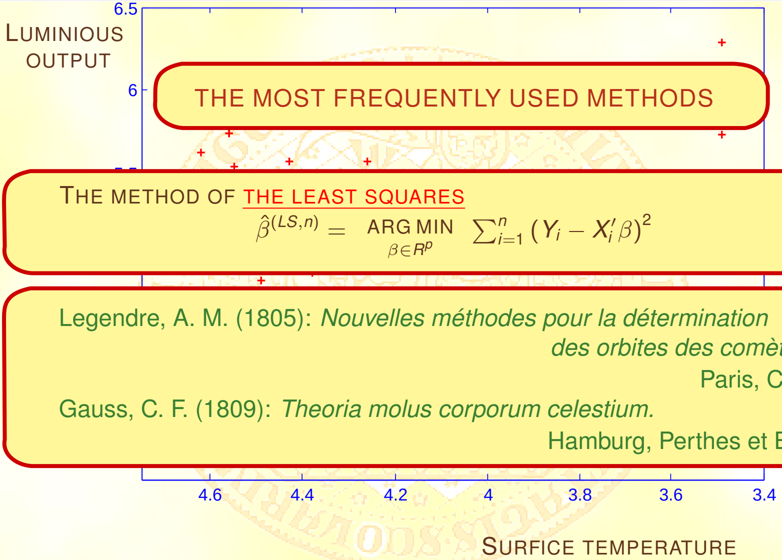


THE MOST FREQUENTLY USED METHODS

<u>Maximum likelihood</u>

$$\hat{\beta}^{(ML,n)} = \underset{\beta \in R^p}{\arg\max} \prod_{i=1}^{n} f\,(Y_i - X_i'\beta)$$

Laplace, P. S. (1774): Mémoire sur la probabilité des causes par les évènemens. *Mémoires de l'Académie royale des sciences presentés par divers savans 6, 621 – 656.*



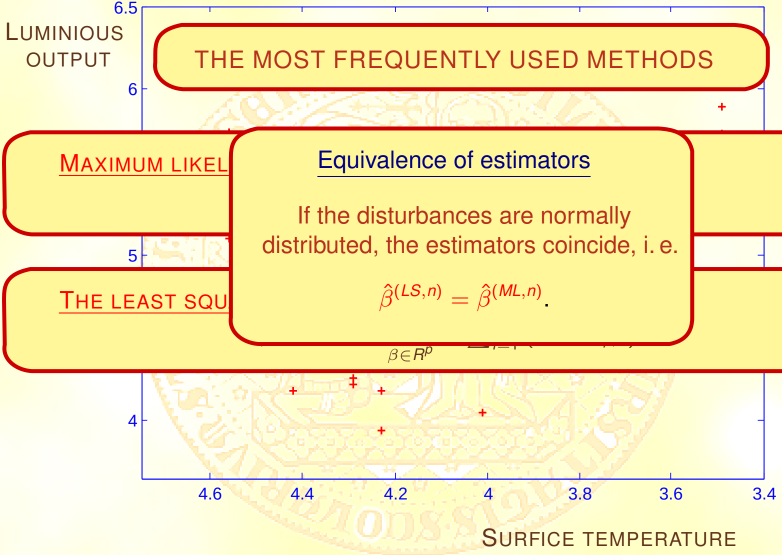
THE MOST FREQUENTLY USED METHODS

THE METHOD OF THE LEAST SQUARES

$$\hat{\beta}^{(LS,n)} = \underset{\beta \in R^p}{\text{ARG MIN}} \sum_{i=1}^{n} (Y_i - X_i'\beta)^2$$

Legendre, A. M. (1805): *Nouvelles méthodes pour la détermination des orbites des comètes.* Paris, Courcier.

Gauss, C. F. (1809): *Theoria molus corporum celestium.* Hamburg, Perthes et Besser.



THE MOST FREQUENTLY USED METHODS

MAXIMUM LIKEL

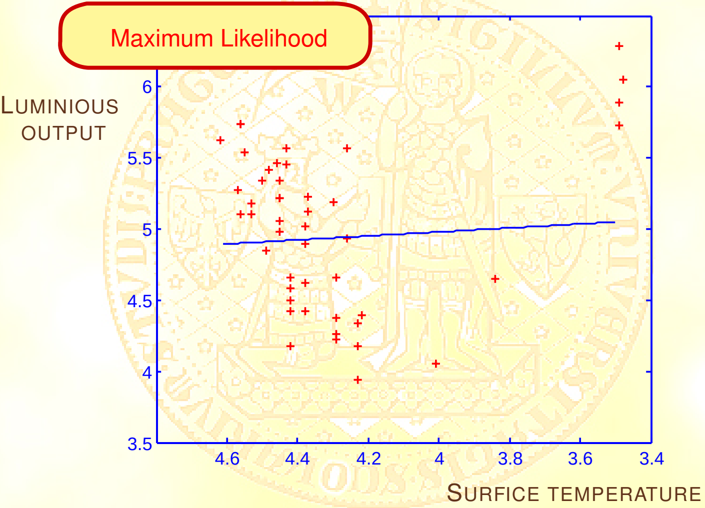
THE LEAST SQU

$\beta \in R^p$

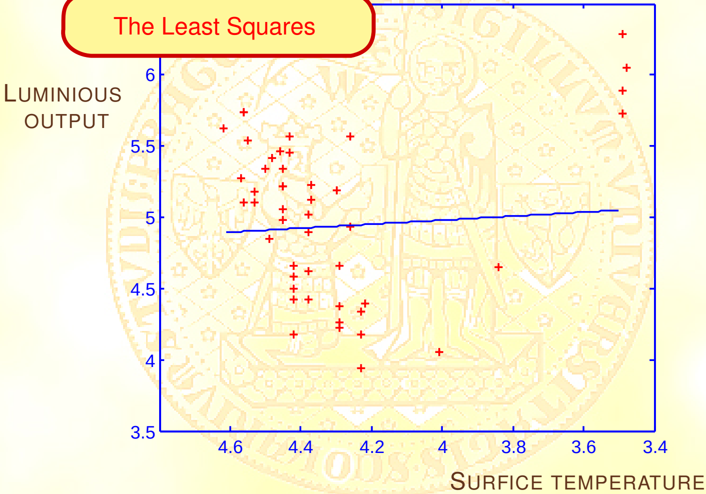
Equivalence of estimators

If the disturbances are normally distributed, the estimators coincide, i. e.

$$\hat{\beta}^{(LS,n)} = \hat{\beta}^{(ML,n)}.$$

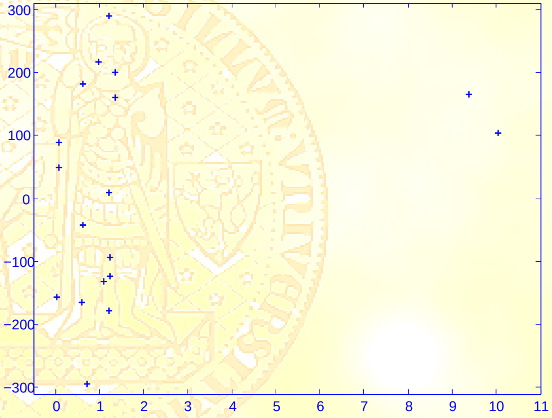

Maximum Likelihood
LUMINIOUS OUTPUT
6
5.5
5
4.5
4
3.5
4.6
4.4
4.2
4
3.8
3.6
3.4
SURFICE TEMPERATURE

The Least Squares



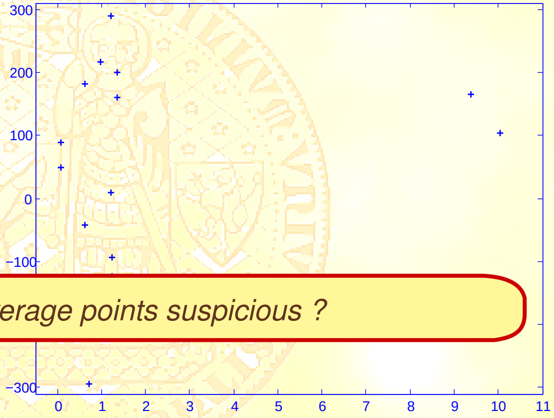
Let's turn to economic data - investments in various industries and their profits

industry	*investment*	*profit*
1	9.39	165.2
2	1.22	9.4
3	0.62	-42.2
4	1.22	289.4
5	1.35	200.4
6	1.21	-179.0
7	1.35	160.4
8	0.07	49.0
9	0.03	-156.1
10	0.07	88.0
11	10.03	103.9
12	0.62	181.7
13	1.23	-93.7
14	1.23	-123.7
15	0.73	-294.4
16	1.10	-131.5
17	0.98	216.3
18	0.61	-165.7



Let’s turn to economic data - investments in various industries and their profits

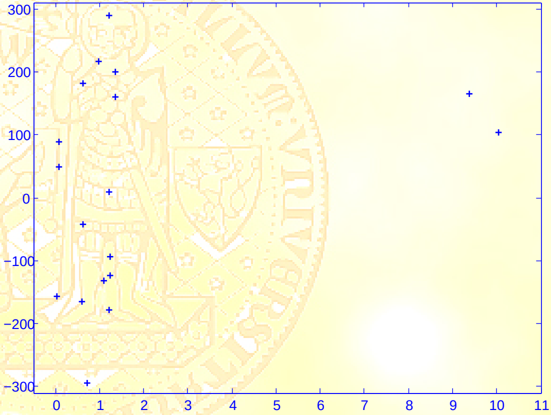
industry	*investment*	*profit*
1	9.39	165.2
2	1.22	9.4
3	0.62	-42.2
4	1.22	289.4
5	1.35	200.4
6	1.21	-179.0
7	1.35	160.4
8	0.07	49.0
9	0.03	-156.1
10	0.07	88.0
11	10.03	103.9
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
14	1.23	-123.7
15	0.73	-294.4
16	1.10	-131.5
17	0.98	216.3
18	0.61	-165.7



Are two leverage points suspicious ?

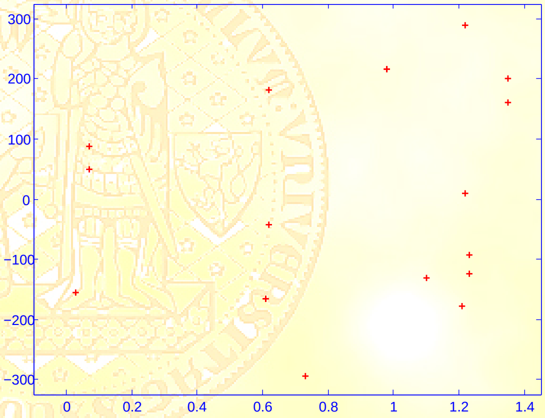
Let's turn to economic data

industry	*investment*	*profit*
1	$\boxed{9.39}$	165.2
2	1.22	9.4
3	0.62	-42.2
4	1.22	289.4
5	1.35	200.4
6	1.21	-179.0
7	1.35	160.4
8	0.07	49.0
9	0.03	-156.1
10	0.07	88.0
11	$\boxed{10.03}$	103.9
12	0.62	181.7
13	1.23	-93.7
14	1.23	-123.7
15	0.73	-294.4
16	1.10	-131.5
17	0.98	216.3
18	0.61	-165.7



Let's turn to economic data - data without leverage points

industry	*investment*	*profit*
2	1.22	9.4
3	0.62	-42.2
4	1.22	289.4
5	1.35	200.4
6	1.21	-179.0
7	1.35	160.4
8	0.07	49.0
9	0.03	-156.1
10	0.07	88.0
12	0.62	181.7
13	1.23	-93.7
14	1.23	-123.7
15	0.73	-294.4
16	1.10	-131.5
17	0.98	216.3
18	0.61	-165.7

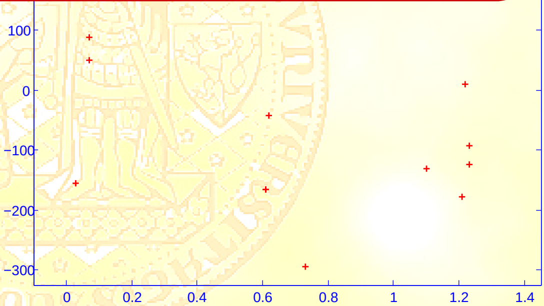


Let's turn to economic data - data without leverage points

A philosophical question$_{(\textit{put by statistical fundamentalist})}$:

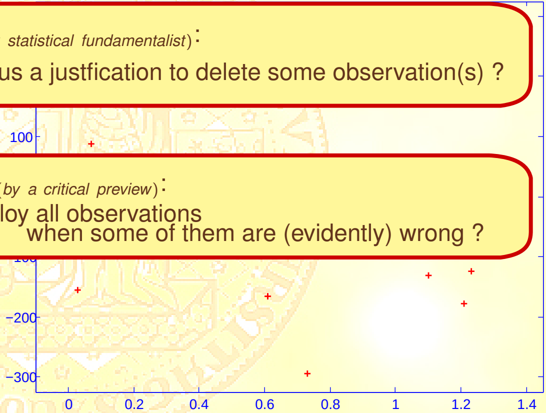
Who gave us a justfication to delete some observation(s) ?

industry	*investment*	*profit*
6	1.21	-179.0
7	1.35	160.4
8	0.07	49.0
9	0.03	-156.1
10	0.07	88.0
12	0.62	181.7
13	1.23	-93.7
14	1.23	-123.7
15	0.73	-294.4
16	1.10	-131.5
17	0.98	216.3
18	0.61	-165.7



Let's turn to economic data - data without leverage points

industry	*investment*	*profit*
6	1.21	-179.0
7	1.35	160.4
13	1.23	-93.7
14	1.23	-123.7
15	0.73	-294.4
16	1.10	-131.5
17	0.98	216.3
18	0.61	-165.7



A philosophical question$_{(\textit{put by statistical fundamentalist})}$:

Who gave us a justfication to delete some observation(s) ?

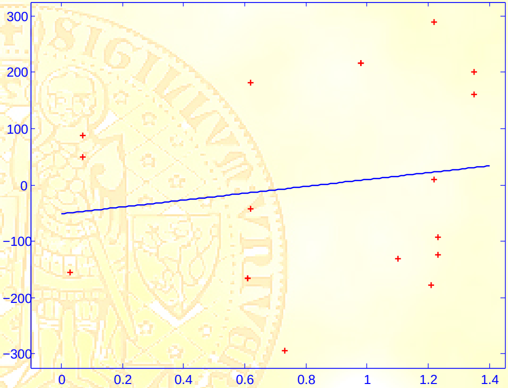
The question can be inverted$_{(\textit{by a critical preview})}$:

Who can force us to employ all observations when some of them are (evidently) wrong ?

The method of the least squares is seen to be our best course when we have thrown overboard a certain portion of our data - a sort of sacrifice which has often to be made by those who sail the stormy seas od Probability.

Francis Ysidro Edgeworth (1887)

industry	*investment*	*profit*
2	1.22	9.4
3	0.62	-42.2
4	1.22	289.4
5	1.35	200.4
6	1.21	-179.0
7	1.35	160.4
8	0.07	49.0
9	0.03	-156.1
10	0.07	88.0
12	0.62	181.7
13	1.23	-93.7
14	1.23	-123.7
15	0.73	-294.4
16	1.10	-131.5
17	0.98	216.3
18	0.61	-165.7



response variable = profit, explanatory variable = investment

$$y_i = \beta_0 + \beta_1 \cdot x_i + u_i \quad i = industry = 2, 3, \ldots, 10, 12, \ldots, 18$$

Graphical analysis

Drawing the data on the screen can help a lot
- but it has one, very significant restriction (limitation).

Graphical analysis

Drawing the data on the screen can help a lot
- but it has one, very significant restriction (limitation).

Could You guess which one it is ?

Graphical analysis

Drawing the data on the screen can help a lot
- but it has one, very significant restriction (limitation).

Could You guess which one it is ?

If no idea,

THE ANSWER WILL BE CLEAR AFTER TRYING TO EMPLOY IT.

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 the 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

How is it with the influence of the individual explanatotory var? Positive sign $\Longrightarrow$ positive influence?

Main subgroup

with number of industries 54 and the 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

Content

Content

Recalling the classical approach to point estimation

Maximum likelihood - solving an extremal problem

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg\max} \prod_{i=1}^{n} f(x_i, \theta)$$

Recalling the classical approach to point estimation

Maximum likelihood - solving an extremal problem

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta\in\Theta} \prod_{i=1}^{n} f(x_i,\theta)$$

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta\in\Theta} log\left\{\prod_{i=1}^{n} f(x_i,\theta)\right\}$$

Recalling the classical approach to point estimation

Maximum likelihood - solving an extremal problem

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} f(x_i, \theta)$$

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \mathit{log} \left\{ \prod_{i=1}^{n} f(x_i, \theta) \right\}$$

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \mathit{log}\left(f(x_i, \theta) \right)$$

Recalling the classical approach to point estimation

Maximum likelihood - solving an extremal problem

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} f(x_i, \theta)$$

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \log \left\{ \prod_{i=1}^{n} f(x_i, \theta) \right\}$$

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log\left(f(x_i, \theta)\right)$$

$$\hat{\theta}^{(ML,n)} = \arg_{\theta \in \Theta} \left\{ \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial \theta} = 0 \right\}$$

Recalling the classical approach to point estimation

Let e. g. $f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial \theta} = 0$$

Recalling the classical approach to point estimation

Let e. g. $f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

$$\hat{\theta}^{(ML,n)} = \arg_{\theta\in\Theta} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial\theta} = 0$$

$\theta = (\mu, \sigma)$

$$\frac{\partial f(x_i,\theta)}{\partial\mu} = 2 \cdot f(x_i, \mu, \sigma^2) \cdot \frac{(x_i-\mu)}{2\sigma^2} \quad \text{and} \quad \frac{\partial f(x_i,\theta)}{\partial\sigma} = -f(x_i, \mu, \sigma^2)\left\{\frac{1}{\sigma} - \frac{(x_i-\mu)^2}{\sigma^3}\right\}$$

Recalling the classical approach to point estimation

Let e. g. $f(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

$$\hat{\theta}^{(ML,n)} = \underset{\theta\in\Theta}{\arg} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial\theta} = 0$$

$\theta = (\mu,\sigma)$

$$\frac{\partial f(x_i,\theta)}{\partial\mu} = 2 \cdot f(x_i,\mu,\sigma^2) \cdot \frac{(x_i-\mu)}{2\sigma^2} \quad \text{and} \quad \frac{\partial f(x_i,\theta)}{\partial\sigma} = -f(x_i,\mu,\sigma^2)\left\{\frac{1}{\sigma} - \frac{(x_i-\mu)^2}{\sigma^3}\right\}$$

$$\hat{\theta}^{(ML,n)} = \underset{\theta\in\Theta}{\arg} \left\{\sum_{i=1}^{n} \frac{(x_i-\mu)}{2\sigma^2} = 0 \ \text{ and } \ \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^3} = \frac{n}{\sigma}\right\}$$

Recalling the classical approach to point estimation

Let e. g. $f(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

$$\hat{\theta}^{(ML,n)} = \underset{\theta\in\Theta}{\arg} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial\theta} = 0$$

$\theta = (\mu,\sigma)$

$$\frac{\partial f(x_i,\theta)}{\partial\mu} = 2\cdot f(x_i,\mu,\sigma^2)\cdot\frac{(x_i-\mu)}{2\sigma^2} \quad \text{and} \quad \frac{\partial f(x_i,\theta)}{\partial\sigma} = -f(x_i,\mu,\sigma^2)\left\{\frac{1}{\sigma} - \frac{(x_i-\mu)^2}{\sigma^3}\right\}$$

$$\hat{\theta}^{(ML,n)} = \underset{\theta\in\Theta}{\arg}\left\{\sum_{i=1}^{n}\frac{(x_i-\mu)}{2\sigma^2} = 0 \text{ and } \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^3} = \frac{n}{\sigma}\right\}$$

$$\hat{\theta}^{(ML,n)} = \left(\hat{\mu}^{(ML,n)} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma}^{(ML,n)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta}^{(ML,n)})^2}\right)$$

Recalling the classical approach to point estimation

Let e. g. $f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial \theta} = 0$$

$\theta = (\mu, \sigma)$

$$\frac{\partial f(x_i,\theta)}{\partial \mu} = 2 \cdot f(x_i, \mu, \sigma^2) \cdot \frac{(x_i-\mu)}{2\sigma^2} \quad \text{and} \quad \frac{\partial f(x_i,\theta)}{\partial \sigma} = -f(x_i, \mu, \sigma^2)\left\{\frac{1}{\sigma} - \frac{(x_i-\mu)^2}{\sigma^3}\right\}$$

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg} \left\{\sum_{i=1}^{n} \frac{(x_i-\mu)}{2\sigma^2} = 0 \ \text{ and } \ \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^3} = \frac{n}{\sigma}\right\}$$

$$\hat{\theta}^{(ML,n)} = \left(\hat{\mu}^{(ML,n)} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma}^{(ML,n)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta}^{(ML,n)})^2}\right)$$

$\hat{\mu}^{(ML,n)} = \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ unbiased, consistent

Recalling the classical approach to point estimation

Let e. g. $f(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

$$\hat{\theta}^{(ML,n)} = \underset{\theta\in\Theta}{\arg} \sum_{i=1}^n \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial\theta} = 0$$

$\theta = (\mu,\sigma)$

$$\frac{\partial f(x_i,\theta)}{\partial\mu} = 2\cdot f(x_i,\mu,\sigma^2)\cdot\frac{(x_i-\mu)}{2\sigma^2} \quad \text{and} \quad \frac{\partial f(x_i,\theta)}{\partial\sigma} = -f(x_i,\mu,\sigma^2)\left\{\frac{1}{\sigma}-\frac{(x_i-\mu)^2}{\sigma^3}\right\}$$

$$\hat{\theta}^{(ML,n)} = \underset{\theta\in\Theta}{\arg}\left\{\sum_{i=1}^n \frac{(x_i-\mu)}{2\sigma^2} = 0 \text{ and } \sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^3} = \frac{n}{\sigma}\right\}$$

$$\hat{\theta}^{(ML,n)} = \left(\hat{\mu}^{(ML,n)} = \frac{1}{n}\sum_{i=1}^n x_i \quad \text{and} \quad \hat{\sigma}^{(ML,n)} = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \hat{\theta}^{(ML,n)})^2}\right)$$

$\hat{\mu}^{(ML,n)} = \bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ unbiased, consistent

$\hat{\sigma}^{(ML,n)} = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i-\bar{x}_n)^2}$ biased, consistent

Recalling the classical approach to point estimation

Let e. g. $f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial \theta} = 0$$

$\theta = (\mu, \sigma)$

$$\frac{\partial f(x_i,\theta)}{\partial \mu} = 2 \cdot f(x_i, \mu, \sigma^2) \cdot \frac{(x_i-\mu)}{2\sigma^2} \quad \text{and} \quad \frac{\partial f(x_i,\theta)}{\partial \sigma} = -f(x_i, \mu, \sigma^2) \left\{ \frac{1}{\sigma} - \frac{(x_i-\mu)^2}{\sigma^3} \right\}$$

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg} \left\{ \sum_{i=1}^{n} \frac{(x_i-\mu)}{2\sigma^2} = 0 \ \text{ and } \ \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^3} = \frac{n}{\sigma} \right\}$$

$$\hat{\theta}^{(ML,n)} = \left(\hat{\mu}^{(ML,n)} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma}^{(ML,n)} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}^{(ML,n)})^2} \right)$$

$\hat{\mu}^{(ML,n)} = \bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$ unbiased, consistent

$\hat{\sigma}^{(ML,n)} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2}$ biased, consistent

$\rightarrow$ $s_n^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$ unbiased, consistent

What we have observed on the previous slide ?

Typical features of the classical estimators

What we have observed on the previous slide ?

Typical features of the classical estimators

Let's consider only estimators which are as $\hat{\mu}^{(ML,n)}$, then:

What we have observed on the previous slide ?

Typical features of the classical estimators

Let's consider only estimators which are as $\hat{\mu}^{(ML,n)}$, then:

Pros:

1. The estimators are defined as solution of extremal problem.

What we have observed on the previous slide ?

Typical features of the classical estimators

Let's consider only estimators which are as $\hat{\mu}^{(ML,n)}$, then:

Pros:

1. The estimators are defined as solution of extremal problem.
2. The extremal problem is (usually) invertible,
   i. e. we have a formula for the estimator,
   hence we can (more or less) easy implement it.

What we have observed on the previous slide ?

Typical features of the classical estimators

Let's consider only estimators which are as $\hat{\mu}^{(ML,n)}$, then:

Pros:

1. The estimators are defined as solution of extremal problem.
2. The extremal problem is (usually) invertible,
   i. e. we have a formula for the estimator,
   hence we can (more or less) easy implement it.
3. They are (mostly) unbiased, consistent, asymptotically normal, etc.

What we have observed on the previous slide ?

Typical features of the classical estimators

Let's consider only estimators which are as $\hat{\mu}^{(ML,n)}$, then:

Pros:

1. The estimators are defined as solution of extremal problem.
2. The extremal problem is (usually) invertible,
   i. e. we have a formula for the estimator,
   hence we can (more or less) easy implement it.
3. They are (mostly) unbiased, consistent, asymptotically normal, etc.
4. If exponential family, usually efficient.

What we have observed on the previous slide ?

Typical features of the classical estimators

Let's consider only estimators which are as $\hat{\mu}^{(ML,n)}$, then:

Pros:

1. The estimators are defined as solution of extremal problem.
2. The extremal problem is (usually) invertible,
   i. e. we have a formula for the estimator,
   hence we can (more or less) easy implement it.
3. They are (mostly) unbiased, consistent, asymptotically normal, etc.
4. If exponential family, usually efficient.

What we have observed on the previous slide ?

Typical features of the classical estimators

Let's consider only estimators which are as $\hat{\mu}^{(ML,n)}$, then:

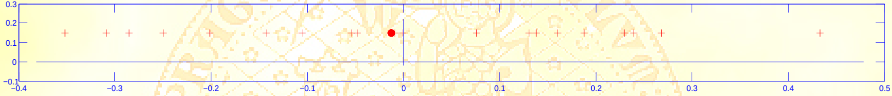
Pros:

1. The estimators are defined as solution of extremal problem.
2. The extremal problem is (usually) invertible,
   i. e. we have a formula for the estimator,
   hence we can (more or less) easy implement it.
3. They are (mostly) unbiased, consistent, asymptotically normal, etc.
4. If exponential family, usually efficient.

Cons: ? ? ? see the next slide !!!

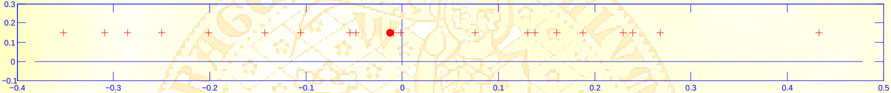
Notice the location of mean

The data generated as standard normal, mean denoted by $\bullet$.

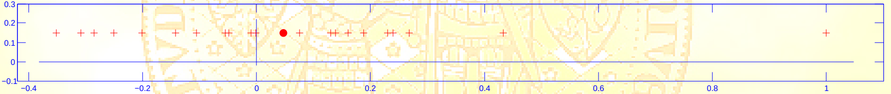


Notice the location of mean

The data generated as standard normal, mean denoted by •.

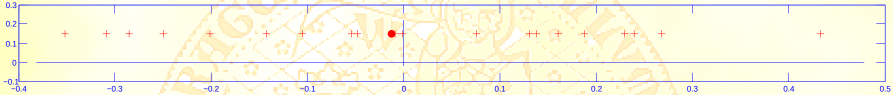


Contamination at 1.

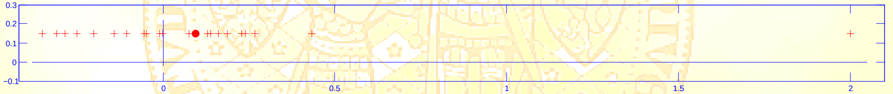


Notice the location of mean

The data generated as standard normal, mean denoted by •.

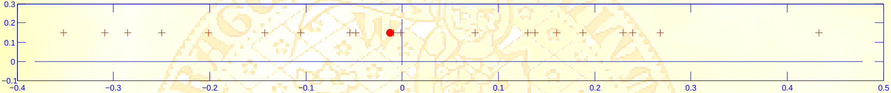


Contamination at 2.

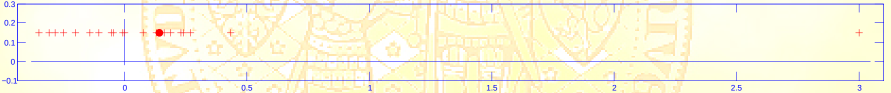


Notice the location of mean

The data generated as standard normal, mean denoted by •.

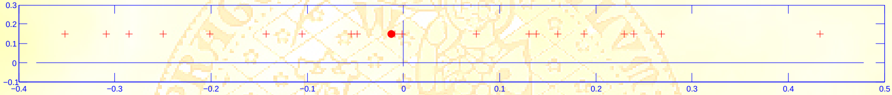


Contamination at 3.

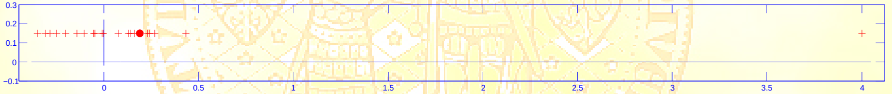


Notice the location of mean

The data generated as standard normal, mean denoted by $\bullet$.

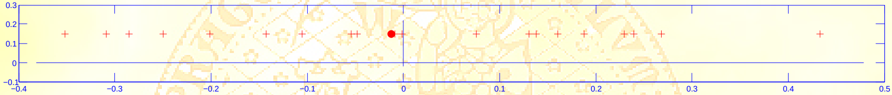


Contamination at 4.

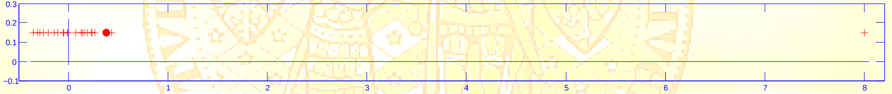


Notice the location of mean

The data generated as standard normal, mean denoted by ●.

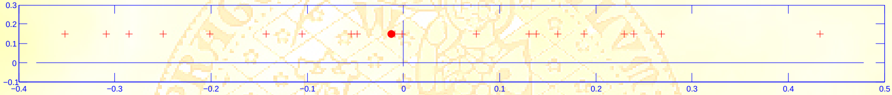


Contamination at 8.

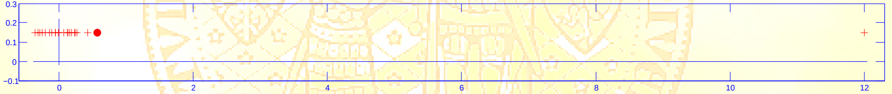


Notice the location of mean

The data generated as standard normal, mean denoted by •.

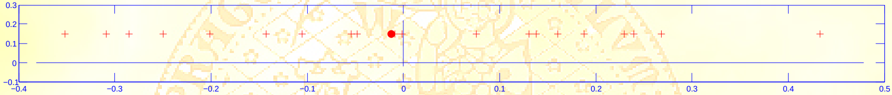


Contamination at 12.

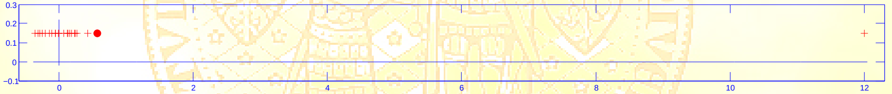


Notice the location of mean

The data generated as standard normal, mean denoted by •.



Contamination at 12.



Conclusion - the classical estimators are (frequently) vulnerable to contamination.

Let's study general reasons causing it - returning a few slides back.

Maximum likelihood - solving an extremal problem

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg\max} \prod_{i=1}^{n} f(x_i, \theta)$$

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg\max} \sum_{i=1}^{n} log(f(x_i, \theta))$$

Let's study general reasons causing it - returning a few slides back.

Maximum likelihood - solving an extremal problem

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} f(x_i, \theta)$$

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} log\left(f(x_i, \theta)\right)$$

Let again $f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}$ and consider only μ

$$\Rightarrow \quad \hat{\mu}^{(ML,n)} = \arg\min_{\mu \in R} \left\{\sum_{i=1}^{n} (x_i - \mu)^2\right\}$$

Let's study general reasons causing it - returning a few slides back.

Maximum likelihood - solving an extremal problem

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} f(x_i, \theta)$$

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} log(f(x_i, \theta))$$

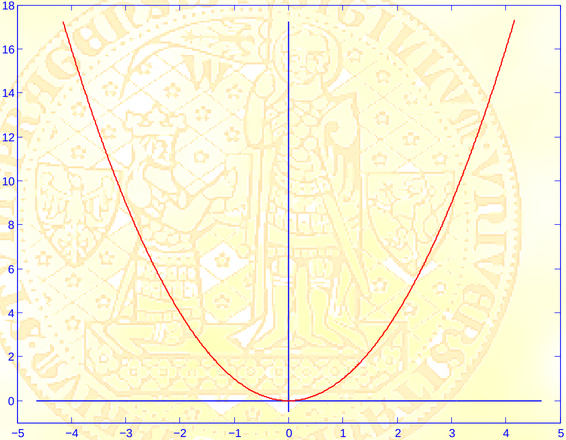
Let again $f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}$ and consider only μ

$$\Rightarrow \quad \hat{\mu}^{(ML,n)} = \arg\min_{\mu \in R} \left\{\sum_{i=1}^{n} (x_i - \mu)^2\right\}$$

The observations with large $(x_i - \mu)^2$ have a large influence on solution.

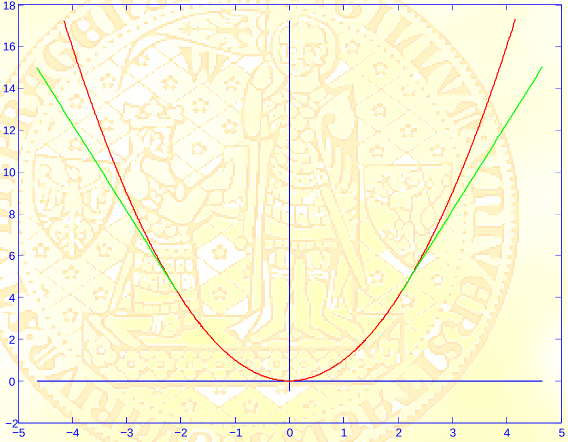
Evidently, low robustness is consequence of quadratic objective function

We have such objective function.



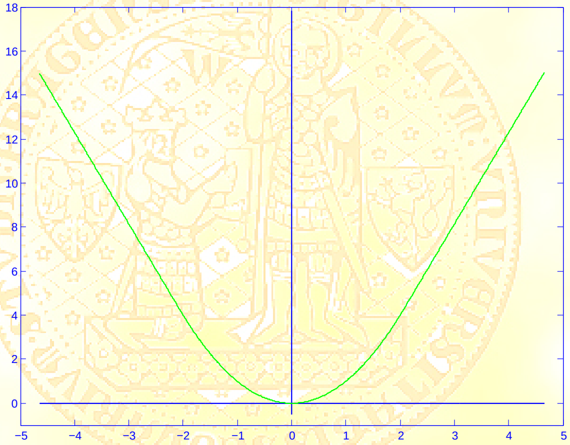
Evidently, low robustness is consequence of quadratic objective function

We should depress influence of large residuals.

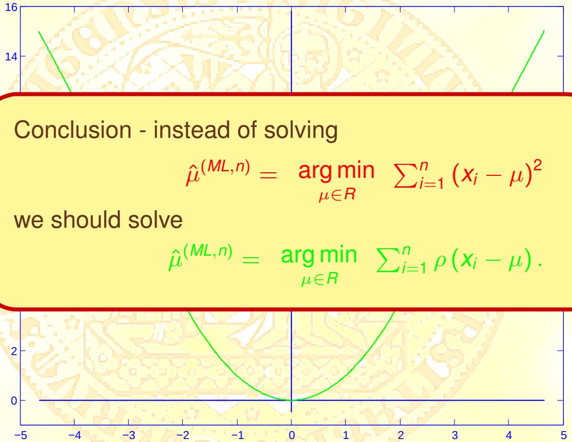


Evidently, low robustness is consequence of quadratic objective function

We should employ such objective function.



Evidently, low robustness is consequence of quadratic objective function



Conclusion - instead of solving

$$\hat{\mu}^{(ML,n)} = \arg\min_{\mu \in R} \sum_{i=1}^{n} (x_i - \mu)^2$$

we should solve

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

Let's study general reasons causing it - an alternative way.

Maximum likelihood - solving the normal equations

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg\max} \prod_{i=1}^{n} f(x_i, \theta) = \underset{\theta \in \Theta}{\arg\max} \sum_{i=1}^{n} log\left(f(x_i, \theta)\right)$$

$$\hat{\theta}^{(ML,n)} = \underset{\theta \in \Theta}{\arg} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial \theta} = 0$$

Let's study general reasons causing it - an alternative way.

Maximum likelihood - solving the normal equations

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta\in\Theta} \prod_{i=1}^{n} f\left(x_i,\theta\right) = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log\left(f\left(x_i,\theta\right)\right)$$

$$\hat{\theta}^{(ML,n)} = \arg_{\theta\in\Theta} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial\theta} = 0$$

Let again $f(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}$, i. e. $\frac{\partial f(x_i,\theta)}{\partial\mu} = f(x_i,\mu,\sigma^2)\cdot\frac{(x_i-\mu)}{\sigma^2}$

and consider only μ $\Rightarrow$ $\hat{\mu}^{(ML,n)} = \arg_{\mu\in R} \left\{\sum_{i=1}^{n}(x_i-\mu) = 0\right\}$

Let's study general reasons causing it - an alternative way.

Maximum likelihood - solving the normal equations

$$\hat{\theta}^{(ML,n)} = \arg\max_{\theta\in\Theta} \prod_{i=1}^{n} f(x_i,\theta) = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} log\left(f(x_i,\theta)\right)$$

$$\hat{\theta}^{(ML,n)} = \arg_{\theta\in\Theta} \sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial\theta} = 0$$

Let again $f(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}$, i. e. $\frac{\partial f(x_i,\theta)}{\partial\mu} = f(x_i,\mu,\sigma^2) \cdot \frac{(x_i-\mu)}{\sigma^2}$

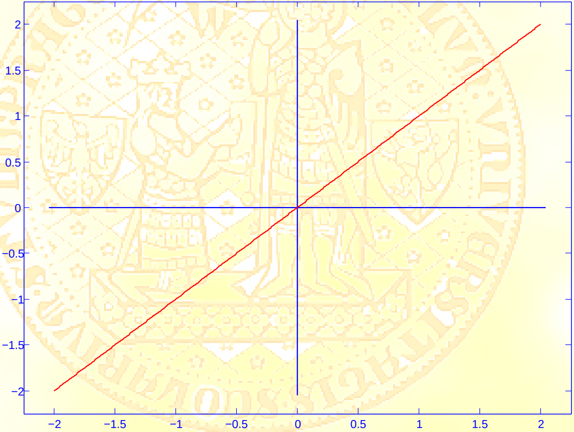
and consider only μ $\Rightarrow$ $\hat{\mu}^{(ML,n)} = \arg_{\mu\in R} \left\{\sum_{i=1}^{n}(x_i-\mu) = 0\right\}$

The same conclusion:

The observations with large $|x_i-\mu|$ have a large influence on solution.

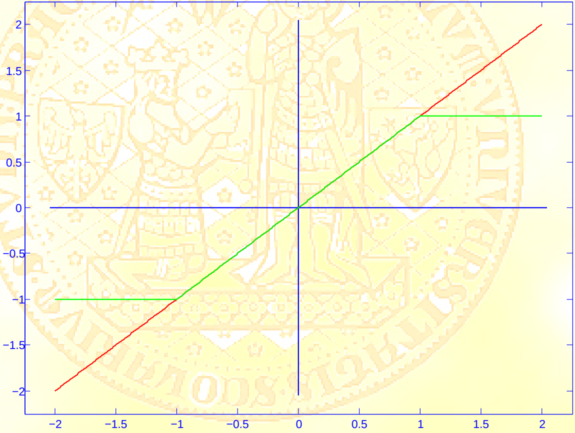
Equivalently, low robustness is consequence of identity in normal equations

We have such influence function.



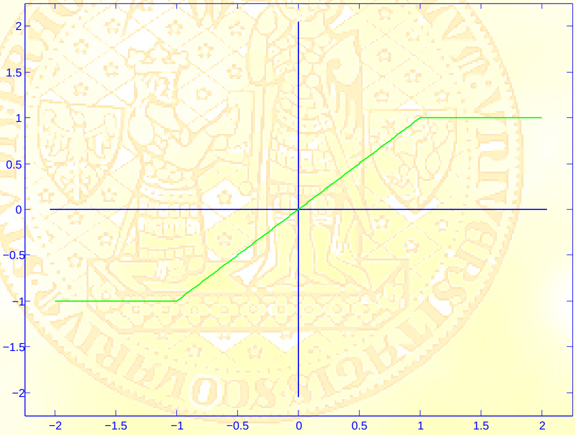
Equivalently, low robustness is consequence of identity in normal equations

We should depress influence of large residuals.

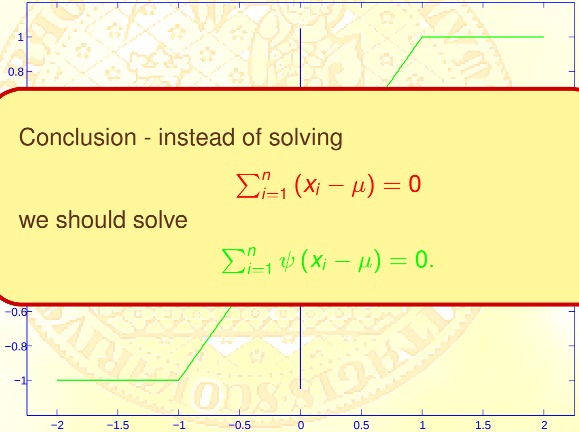


Equivalently, low robustness is consequence of identity in normal equations

We should employ such influence function.



Equivalently, low robustness is consequence of identity in normal equations



Conclusion - instead of solving

$$\sum_{i=1}^{n} (x_i - \mu) = 0$$

we should solve

$$\sum_{i=1}^{n} \psi\,(x_i - \mu) = 0.$$

Content

Recalling the classical requirements on estimators

1. Unbiasedness

Recalling the classical requirements on estimators

1. Unbiasedness
2. Consistency (weak, strong)

Recalling the classical requirements on estimators

1. Unbiasedness
2. Consistency (weak, strong)
3. $\sqrt{n}$-consistency (root-n-consistency)

Recalling the classical requirements on estimators

1. Unbiasedness
2. Consistency (weak, strong)
3. $\sqrt{n}$-consistency (root-n-consistency)
4. Asymptotic normality

Recalling the classical requirements on estimators

1. Unbiasedness
2. Consistency (weak, strong)
3. $\sqrt{n}$-consistency (root-n-consistency)
4. Asymptotic normality
5. Efficiency

Recalling the classical requirements on estimators

1. Unbiasedness
2. Consistency (weak, strong)
3. $\sqrt{n}$-consistency (root-n-consistency)
4. Asymptotic normality
5. Efficiency
6. Scale- and regression-equivariance

Recalling the classical requirements on estimators

1. Unbiasedness
2. Consistency (weak, strong)
3. $\sqrt{n}$-consistency (root-n-consistency)
4. Asymptotic normality
5. Efficiency
6. Scale- and regression-equivariance
7. Admissibility

Recalling the classical requirements on estimators

1. Unbiasedness
2. Consistency (weak, strong)
3. $\sqrt{n}$-consistency (root-n-consistency)
4. Asymptotic normality
5. Efficiency
6. Scale- and regression-equivariance
7. Admissibility

Recalling the classical requirements on estimators

1. Unbiasedness
2. Consistency (weak, strong)
3. $\sqrt{n}$-consistency (root-n-consistency)
4. Let's discuss them one by one.
5. Efficiency
6. Scale- and regression-equivariance
7. Admissibility

Unbiasedness

$$\boldsymbol{E}_\theta\left[\hat{\theta}^n(x_1, x_2, \ldots, x_n)\right] = \int_{\mathcal{X}^n} \hat{\theta}^n(x_1, x_2, \ldots, x_n) f_\theta(x_1, x_2, \ldots, x_n)\mathrm{d}x_1 \cdot \mathrm{d}x_2 \cdot \ldots \cdot \mathrm{d}x_n = \theta$$

Unbiasedness

$$\boldsymbol{E}_\theta\left[\hat{\theta}^n(x_1, x_2, \ldots, x_n)\right] = \int_{\mathcal{X}^n} \hat{\theta}^n(x_1, x_2, \ldots, x_n) f_\theta(x_1, x_2, \ldots, x_n) \mathrm{d}x_1 \cdot \mathrm{d}x_2 \cdot \ldots \cdot \mathrm{d}x_n = \theta$$

Is this requirement justifiable everytime ?

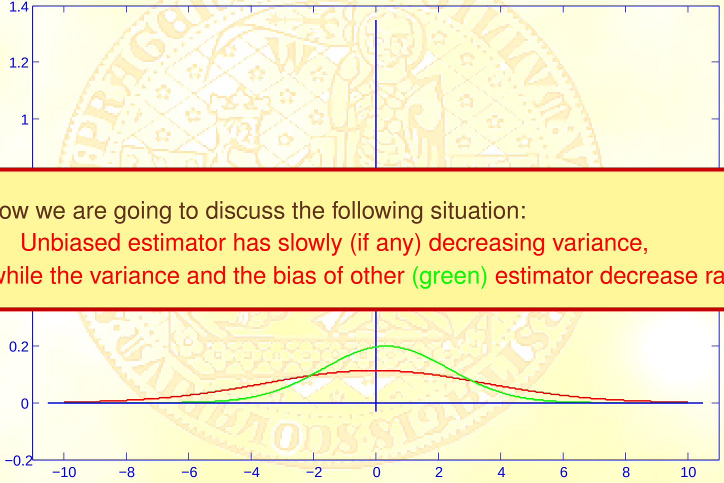
Unbiasedness

$$\boldsymbol{E}_{\theta}\left[\hat{\theta}^{n}(x_1, x_2, ..., x_n)\right] = \int_{\mathcal{X}^n} \hat{\theta}^{n}(x_1, x_2, ..., x_n) f_{\theta}(x_1, x_2, ..., x_n) \mathrm{d}x_1 \cdot \mathrm{d}x_2 \cdot ... \cdot \mathrm{d}x_n = \theta$$

Hoerl, A. E., R. W. Kennard (1970): Ridge regression:
Biased estimation for nonorthogonal problems.
Technometrics 12, 55 - 68.

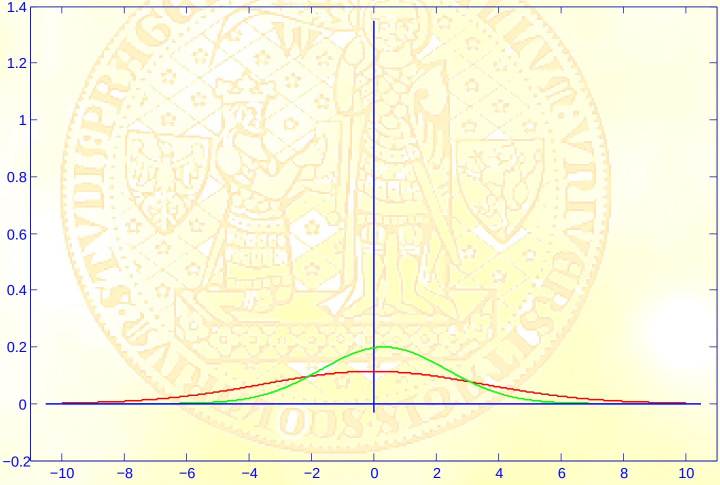
$$\hat{\beta}^{(R,n)} = \left(X'X + \delta \cdot \boldsymbol{I}\right)^{-1} X'Y$$

Possible density of unbiased and biased estimator

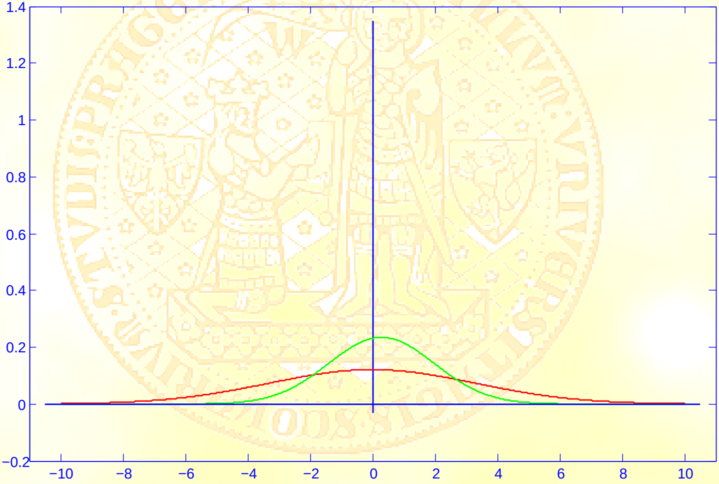


Now we are going to discuss the following situation:
Unbiased estimator has slowly (if any) decreasing variance,
while the variance and the bias of other (green) estimator decrease rapidly.

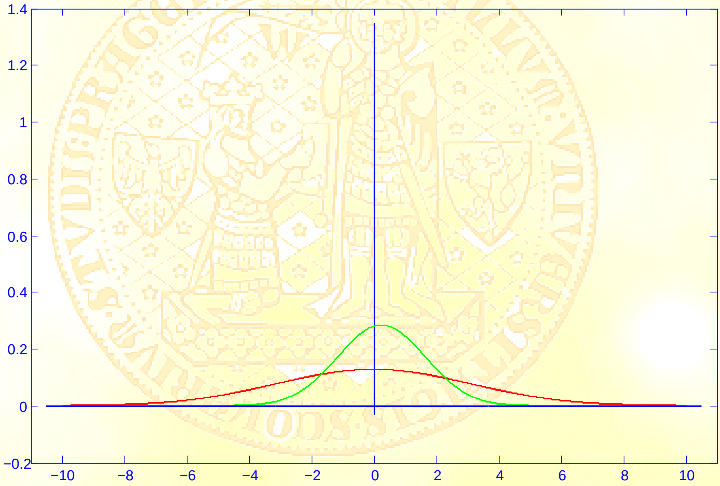
Notice decreasing variance and bias



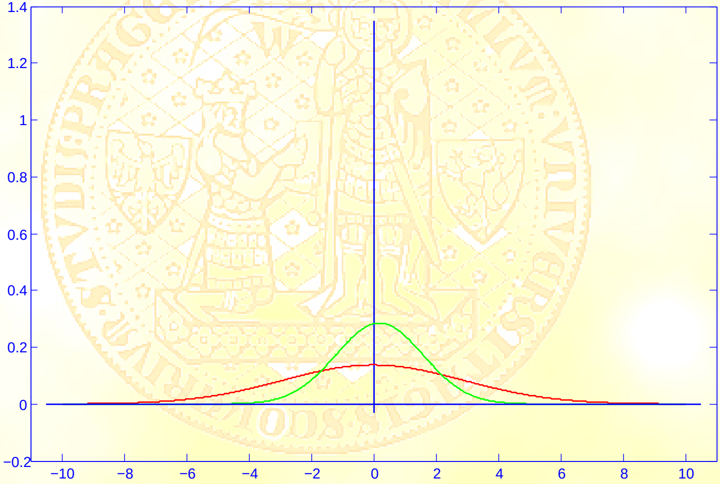
Notice decreasing variance and bias



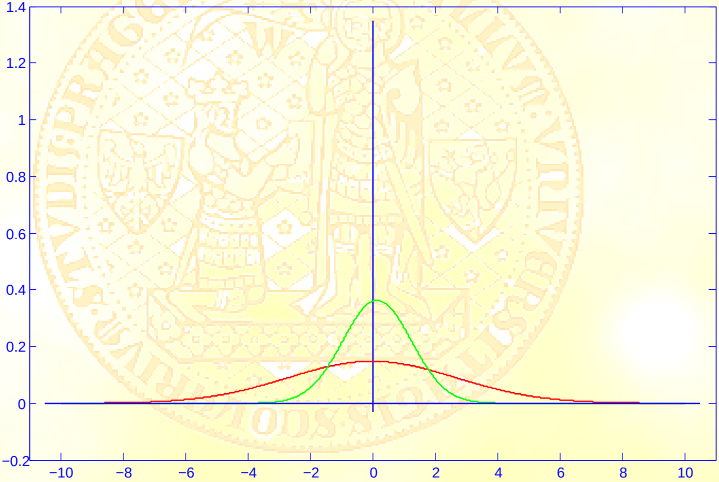
Notice decreasing variance and bias



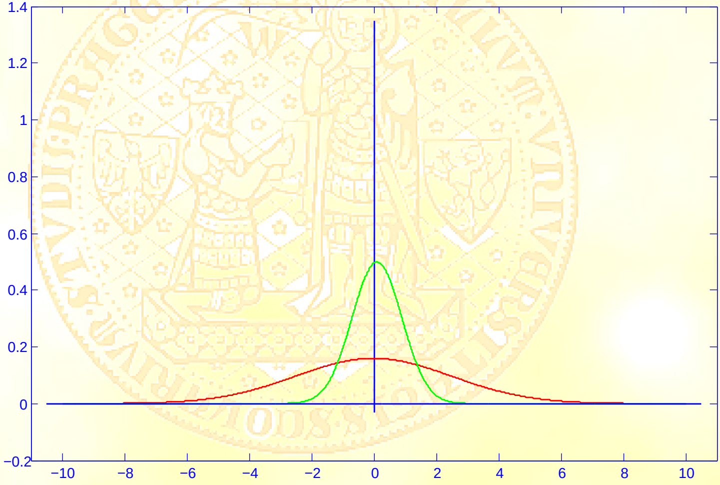
Notice decreasing variance and bias



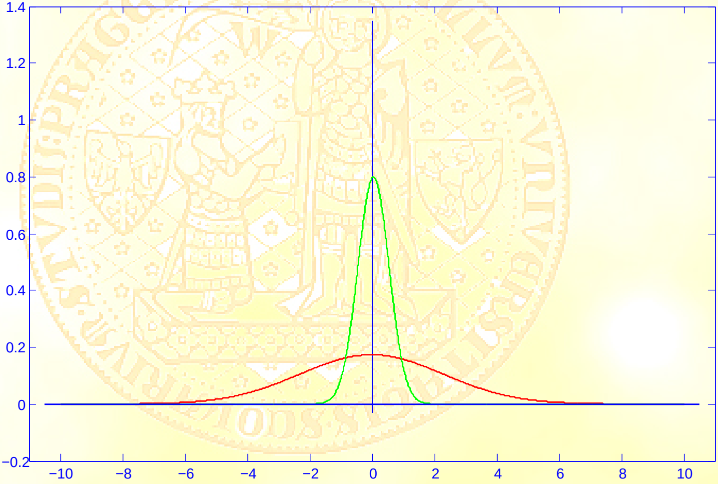
Notice decreasing variance and bias



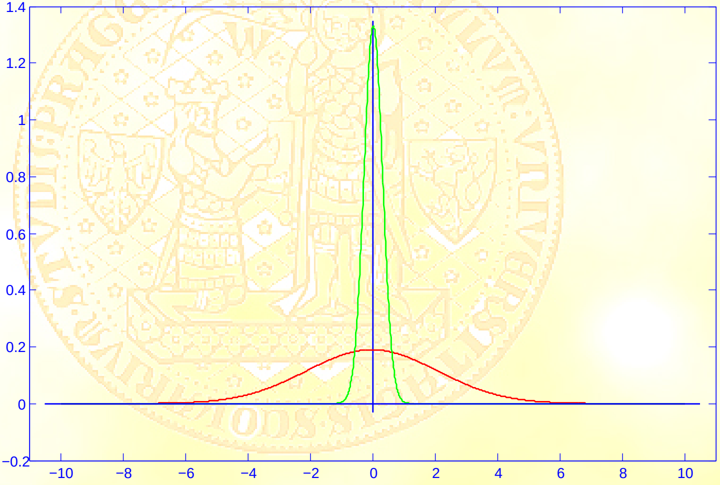
Notice decreasing variance and bias



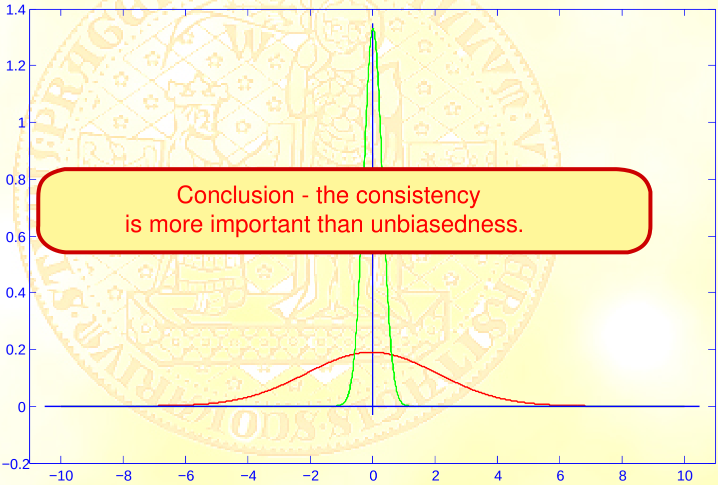
Notice decreasing variance and bias



Notice decreasing variance and bias

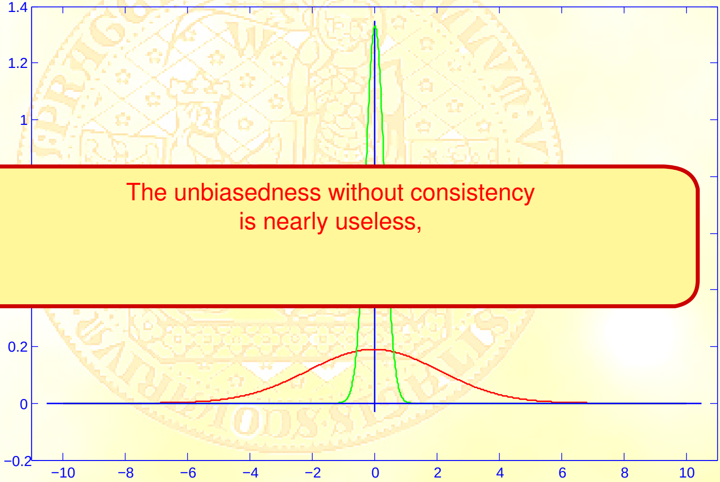


Notice decreasing variance and bias



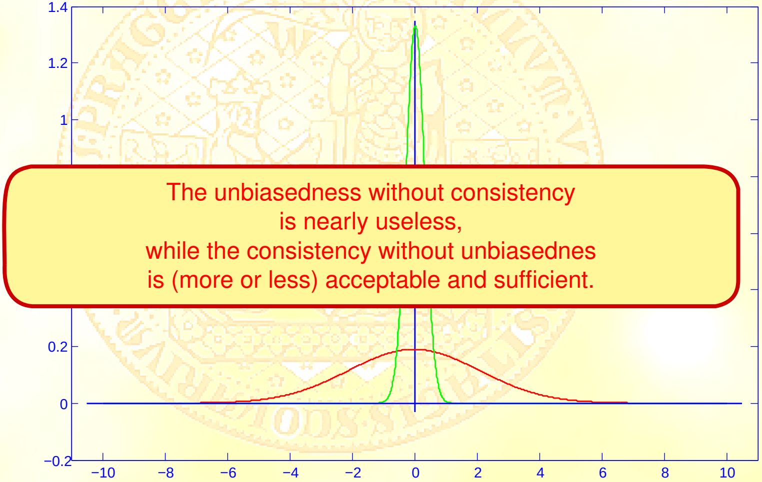
Conclusion - the consistency
is more important than unbiasedness.

Notice decreasing variance and bias



The unbiasedness without consistency
is nearly useless,

Notice decreasing variance and bias



Consistency

1. (Weak) consistency - convergence in probability

Consistency

1. (Weak) consistency - convergence in probability
2. Strong consistency - convergence almost surely

Consistency

1. (Weak) consistency - convergence in probability
2. Strong consistency - convergence almost surely
3. $\sqrt{n}$-consistency (root-n-consistency)

Convergence in probability (weak convergence)

Let X and $\{X_n\}_{n=1}^{\infty}$ be random variable (r. v.) and a sequence of r.v.'s, respectively.

Convergence in probability (weak convergence)

Let X and $\{X_n\}_{n=1}^{\infty}$ be random variable (r. v.) and a sequence of r.v.'s, respectively.

We say that the sequence

$$\{X_n\}_{n=1}^{\infty} \text{ \underline{converge in probability} (\underline{weakly}) to } X$$

if:

$$\forall\,(\varepsilon > 0, \delta > 0) \quad \exists\,(n_{\varepsilon,\delta} \in \mathcal{N}) \quad \forall\,(n \geq n_{\varepsilon,\delta})$$

$$P\left(\{\omega \in \Omega \;:\; |X_n(\omega) - X(\omega)| > \delta\}\right) < \varepsilon$$

Convergence in probability (weak convergence)

Let X and $\{X_n\}_{n=1}^{\infty}$ be random variable (r. v.) and a sequence of r.v.'s, respectively.
We say that the sequence

$$\{X_n\}_{n=1}^{\infty} \; \underline{\textit{converge in probability}} \; (\underline{\textit{weakly}}) \text{ to } X$$

if:

$$\forall(\varepsilon > 0, \delta > 0) \quad \exists(n_{\varepsilon,\delta} \in \mathcal{N}) \quad \forall(n \geq n_{\varepsilon,\delta})$$

$$P(\{\omega \in \Omega \; : \; |X_n(\omega) - X(\omega)| > \delta\}) < \varepsilon$$

or alternatively

$$P(\{\omega \in \Omega \; : \; |X_n(\omega) - X(\omega)| < \delta\}) > 1 - \varepsilon.$$

(Weak) consistency - convergence in probability

In the case that we speak about an estimator of "true" β^0, we say that $\hat{\beta}^{(method,n)}$ is (weakly) consistent if:

$$\forall(\varepsilon > 0, \delta > 0) \quad \exists(n_{\varepsilon,\delta} \in \mathcal{N}) \quad \forall(n \geq n_{\varepsilon,\delta})$$

$$P\left(\left\{\omega \in \Omega : \left\|\hat{\beta}^{(method,n)} - \beta^0\right\| > \delta\right\}\right) < \varepsilon$$

(Weak) consistency - convergence in probability

In the case that we speak about an estimator of "true" β^0,
we say that $\hat{\beta}^{(method,n)}$ is (weakly) consistent if:

$$\forall(\varepsilon > 0, \delta > 0) \quad \exists(n_{\varepsilon,\delta} \in \mathcal{N}) \quad \forall(n \geq n_{\varepsilon,\delta})$$

$$P\left(\left\{\omega \in \Omega : \left\|\hat{\beta}^{(method,n)} - \beta^0\right\| > \delta\right\}\right) < \varepsilon$$

or alternatively

$$P\left(\left\{\omega \in \Omega : \left\|\hat{\beta}^{(method,n)} - \beta^0\right\| < \delta\right\}\right) > 1 - \varepsilon.$$

Convergence almost surely (strong convergence)

Let X and $\{X_n\}_{n=1}^{\infty}$ be r.v. and a sequence of r.v.'s, respectively.

Convergence almost surely (strong convergence)

Let X and $\{X_n\}_{n=1}^{\infty}$ be r.v. and a sequence of r.v.'s, respectively. We say that the sequence

$\{X_n\}_{n=1}^{\infty}$ <u>*converges almost surely*</u> (<u>*strongly*</u>) to X

if:

$$\exists\,(A \in \mathcal{A}, P(A) = 1) \quad \forall\,(\varepsilon > 0, \omega_0 \in A) \quad \exists\,(n_{\varepsilon,\omega_0} \in \mathcal{N}) \quad \forall\,(n \geq n_{\varepsilon,\omega_0})$$

$$|X_n(\omega_0) - X(\omega_0)| < \varepsilon.$$

Strong consistency - convergence almost surely

In the case that we speak about an estimator of "true" β^0,
we say that $\hat{\beta}^{(method,n)}$ is strongly consistent if:

$$\exists\,(A \in \mathcal{A}, P(A) = 1) \quad \forall\,(\varepsilon > 0, \omega_0 \in A) \quad \exists\,(n_{\varepsilon,\omega_0} \in \mathcal{N}) \quad \forall\,(n \geq n_{\varepsilon,\omega_0})$$

$$\left\| \hat{\beta}^{(method,n)}(\omega_0) - \beta^0 \right\| < \varepsilon.$$

$\sqrt{n}$-consistency (root of n consistency)

In this case we typically speak about an estimator of "true" β^0.

$\sqrt{n}$-consistency (root of n consistency)

In this case we typically speak about an estimator of "true" β^0.

Then we say that $\hat{\beta}^{(method,n)}$ is $\sqrt{n}$-consistent if:

$$\forall\,(\varepsilon > 0) \quad \exists\,(K_\varepsilon < \infty \ \text{and} \ n_{\varepsilon,K_\varepsilon} \in \mathcal{N}) \quad \forall\,(n \geq n_{\varepsilon,K_\varepsilon})$$

$$P\left(\left\{\omega \in \Omega : \sqrt{n}\left\|\hat{\beta}^{(method,n)} - \beta^0\right\| > K_\varepsilon\right\}\right) < \varepsilon.$$

$\sqrt{n}$-consistency (root of n consistency)

In this case we typically speak about an estimator of "true" β^0.

Then we say that $\hat{\beta}^{(method,n)}$ is $\sqrt{n}$-consistent if:

$$\forall\,(\varepsilon > 0) \quad \exists\,(K_\varepsilon < \infty \ \text{and} \ n_{\varepsilon,K_\varepsilon} \in \mathcal{N}) \quad \forall\,(n \geq n_{\varepsilon,K_\varepsilon})$$

$$P\left(\left\{\omega \in \Omega \,:\, \sqrt{n}\left\|\hat{\beta}^{(method,n)} - \beta^0\right\| > K_\varepsilon\right\}\right) < \varepsilon.$$

or alternatively

$$P\left(\left\{\omega \in \Omega \,:\, \sqrt{n}\left\|\hat{\beta}^{(method,n)} - \beta^0\right\| \leq K_\varepsilon\right\}\right) > 1 - \varepsilon.$$

Asymptotic normality

It is again case when we speak about an estimator of "true" β^0.

Asymptotic normality

It is again case when we speak about an estimator of "true" β^0.

Then we say that $\hat{\beta}^{(method,n)}$ is asymptotically normal if :

$$\mathcal{L}\left(n^a\left(\hat{\beta}^{(method,n)} - \beta^0\right)\right) \rightarrow \mathcal{N}(0, \Sigma)$$

where $a \leq \frac{1}{2}$ and usually reaches the upper bound, i. e. usually $a = \frac{1}{2}$.

Asymptotic normality

It is again case when we speak about an estimator of "true" β^0.

Then we say that $\hat{\beta}^{(method,n)}$ is asymptotically normal if :

$$\mathcal{L}\left(n^a\left(\hat{\beta}^{(method,n)} - \beta^0\right)\right) \rightarrow \mathcal{N}(0, \Sigma)$$

where $a \leq \frac{1}{2}$ and usually reaches the upper bound, i. e. usually $a = \frac{1}{2}$.

Asymptotic normality is (was? - in the case of OLS, ML, etc.) employed:

1. for constructing (asymptotic) confidence interval

and

Asymptotic normality

It is again case when we speak about an estimator of "true" β^0.

Then we say that $\hat{\beta}^{(method,n)}$ is asymptotically normal if :

$$\mathcal{L}\left(n^a\left(\hat{\beta}^{(method,n)} - \beta^0\right)\right) \rightarrow \mathcal{N}(0, \Sigma)$$

where $a \leq \frac{1}{2}$ and usually reaches the upper bound, i. e. usually $a = \frac{1}{2}$.

Asymptotic normality is (was? - in the case of OLS, ML, etc.) employed:

1. for constructing (asymptotic) confidence interval

   and

2. for verification of $\sqrt{n}$-consistency.

Efficiency

We usually say that $\hat{\beta}^{(method,n)}$ is (asymptotically) efficient, if its covariance matrix reaches (asymptotically)

lower Rao-Cramer bound

in the sense of ordering the matrices by positive semidefiniteness.

Efficiency

We usually say that $\hat{\beta}^{(method,n)}$ is (asymptotically) efficient, if its covariance matrix reaches (asymptotically)

lower Rao-Cramer bound

in the sense of ordering the matrices by positive semidefiniteness.

Sometimes, we say that $\hat{\beta}^{(method,n)}$ is (asymptotically) efficient, if its covariance matrix reaches (asymptotically)

the minimal possible value

in given family of estimators - again in the sense of ordering the matrices by positive semidefiniteness.

Efficiency

Efficiency is:

1. important notion from the pedagogical point view,

Efficiency

Efficiency is:

1. important notion from the pedagogical point view,
2. important from abstract theoretical background of statistics
   - how much we could reach if data would be “clear”,

Efficiency

Efficiency is:

1. important notion from the pedagogical point view,
2. important from abstract theoretical background of statistics
   - how much we could reach if data would be "clear",
3. it can be destroyed by a small deviation
   from the exponential family - Huber's example

Efficiency

Efficiency is:

1. important notion from the pedagogical point view,
2. important from abstract theoretical background of statistics
   - how much we could reach if data would be “clear”,
3. it can be destroyed by a small deviation
   from the exponential family - Huber’s example

Efficiency

Efficiency is:

1. important notion from the pedagogical point view,
2. important from abstract theoretical background of statistics
   - how much we could reach if data would be "clear",
3. it can be destroyed by a small deviation
   from the exponential family - Huber's example

and

4. need not imply too much - Fisher's example.

Small deviation from exact model can cause ...

Huber, P. J. (1980): *Robust Statistics.* New York: J.Wiley and Sons.

$$s_n = \left[\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \right]^{\frac{1}{2}}$$

Small deviation from exact model can cause ...

Huber, P. J. (1980): *Robust Statistics.*
New York: J.Wiley and Sons.

$$s_n = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right]^{\frac{1}{2}} \qquad d_n = \frac{\pi}{2n}\sum_{i=1}^{n}|x_i - \bar{x}_n|$$

Small deviation from exact model can cause ...

Huber, P. J. (1980): *Robust Statistics.* New York: J.Wiley and Sons.

$$s_n = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right]^{\frac{1}{2}} \qquad d_n = \frac{\pi}{2n}\sum_{i=1}^{n}|x_i - \bar{x}_n|$$

$$F(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(\frac{x}{3})$$

Small deviation from exact model can cause ...

Huber, P. J. (1980): *Robust Statistics.*
New York: J.Wiley and Sons.

$$s_n = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right]^{\frac{1}{2}} \qquad d_n = \frac{\pi}{2n}\sum_{i=1}^{n}|x_i - \bar{x}_n|$$

$$F(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(\frac{x}{3})$$

$$ARE_F(\varepsilon) = \lim_{n\to\infty} \frac{\mathrm{var}_F s_n / \mathbb{E}_F^2 s_n}{\mathrm{var}_F d_n / \mathbb{E}_F^2 d_n}$$

Small deviation from exact model can cause ...

ε	0	0.001	0.002	0.05
$ARE(\varepsilon)$	0.876	0.948	1.016	2.035

Small deviation from exact model can cause ...

ε	0	0.001	0.002	0.05
$ARE(\varepsilon)$	0.876	0.948	1.016	2.035

So, 5% of contamination $\rightarrow$ d_n is two times better than s_n.

Is 5% contamination too much or too little?

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel. (1986):
Robust Statistic - The Approach Based on Influence Curve.
New York: J.Wiley and Sons.

E. g. Switzerland has 6% of errors in mortality tables.

Is the efficiency really important or a bit misleading?

Fisher, R. A. (1922): On the mathematical foundation of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A 222, 309 - 368.*

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(\bar{x}_n)}{\mathrm{var}_{t(\nu)}(\bar{x}_n)} = 1 - \frac{6}{\nu(\nu+1)}$$

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(s_n^2)}{\mathrm{var}_{t(\nu)}(s_n^2)} = 1 - \frac{12}{\nu(\nu+1)}$$

Is the efficiency really important or a bit misleading?

Fisher, R. A. (1922): On the mathematical foundation of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A 222, 309 - 368.*

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(\bar{x}_n)}{\mathrm{var}_{t(\nu)}(\bar{x}_n)} = 1 - \frac{6}{\nu(\nu+1)}$$

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(s_n^2)}{\mathrm{var}_{t(\nu)}(s_n^2)} = 1 - \frac{12}{\nu(\nu+1)}$$

Is the efficiency really important or a bit misleading?

Fisher, R. A. (1922): On the mathematical foundation of theoretical statistics.
Philos. Trans. Roy. Soc. London Ser. A 222, 309 - 368.

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(\bar{x}_n)}{\mathrm{var}_{t(\nu)}(\bar{x}_n)} = 1 - \frac{6}{\nu(\nu+1)}$$

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(s_n^2)}{\mathrm{var}_{t(\nu)}(s_n^2)} = 1 - \frac{12}{\nu(\nu+1)}$$

Is the efficiency really important or a bit misleading?

Fisher, R. A. (1922): On the mathematical foundation of theoretical statistics.
Philos. Trans. Roy. Soc. London Ser. A 222, 309 - 368.

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(\bar{x}_n)}{\mathrm{var}_{t(\nu)}(\bar{x}_n)} = 1 - \frac{6}{\nu(\nu+1)}$$

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(s_n^2)}{\mathrm{var}_{t(\nu)}(s_n^2)} = 1 - \frac{12}{\nu(\nu+1)}$$

Is the efficiency really important or a bit misleading?

Fisher, R. A. (1922): On the mathematical foundation of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A 222, 309 - 368.*

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(\bar{x}_n)}{\mathrm{var}_{t(\nu)}(\bar{x}_n)} = 1 - \frac{6}{\nu(\nu+1)}$$

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(s_n^2)}{\mathrm{var}_{t(\nu)}(s_n^2)} = 1 - \frac{12}{\nu(\nu+1)}$$

Is the efficiency really important or a bit misleading?

Fisher, R. A. (1922): On the mathematical foundation of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A 222, 309 - 368.*

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(\bar{x}_n)}{\mathrm{var}_{t(\nu)}(\bar{x}_n)} = 1 - \frac{6}{\nu(\nu+1)}$$

$$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(s_n^2)}{\mathrm{var}_{t(\nu)}(s_n^2)} = 1 - \frac{12}{\nu(\nu+1)}$$

Is the efficiency really important or a bit misleading?

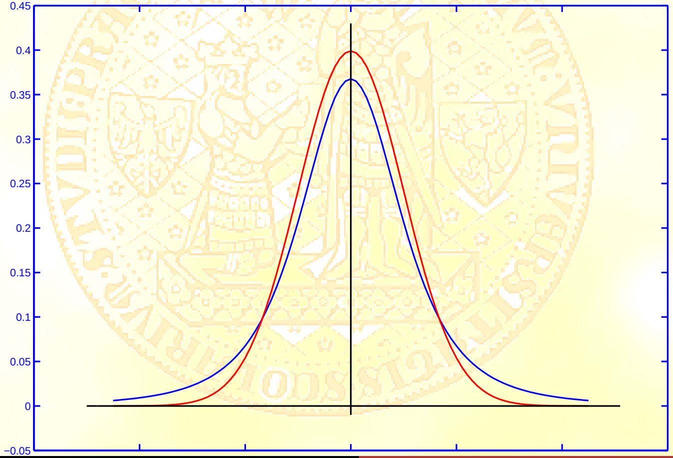
$\lim_{n\to\infty} \frac{\text{var}_{N(0,1)}(T_n)}{\text{var}_{t(\nu)}(T_n)}$	t_9	t_5	t_3
$\bar{x}_n$	0.93	0.80	0.50
s_n^2	0.83	0.40	0!

Is the efficiency really important or a bit misleading?

$\lim_{n\to\infty} \frac{\mathrm{var}_{N(0,1)}(T_n)}{\mathrm{var}_{t(\nu)}(T_n)}$	t_9	t_5	t_3
$\bar{x}_n$	0.93	0.80	0.50
s_n^2	0.83	0.40	$\boxed{0!}$

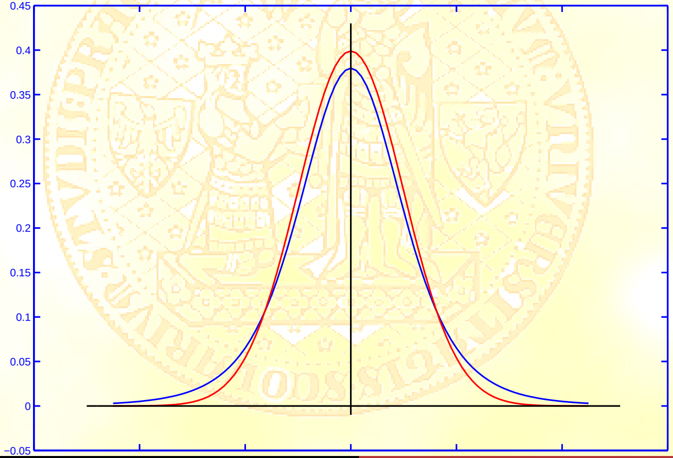
How far is Student density from the normal one ?

THE BLUE CURVE IS STANDARD NORMAL WHILE THE RED ONE IS THE STUDENT'S WITH 3 DEGREES OF FREEDOM.



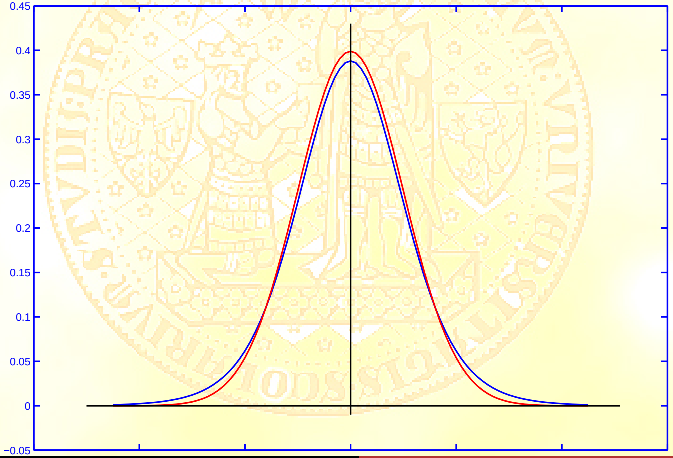
How far is Student density from the normal one ?

THE BLUE CURVE IS STANDARD NORMAL WHILE THE RED ONE IS THE STUDENT'S WITH 5 DEGREES OF FREEDOM.



How far is Student density from the normal one ?

THE BLUE CURVE IS STANDARD NORMAL WHILE THE RED ONE IS THE STUDENT'S WITH 9 DEGREES OF FREEDOM.



The scale- and regression equivariance
and the admissibility will be discussed later on.

THANKS FOR ATTENTION