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 3

Content of lecture

Content

A problem of the classical statistics and econometrics

A tacit hope in ingnoring deviations from ideal models was that they would not matter; that statistical procedures which were optimal under strict model would still be approximately optimal under the approximate model. Unfortunately, it turned out that this hope was often drastically wrong; even mild deviations often have much larger effects than were anticipated by most statisticians.

John W. Tukey (1960)

A problem of the classical statistics and econometrics

A tacit hope in ingnoring deviations from ideal models was that they would not matter; that statistical procedures which were optimal under strict model would still be approximately optimal under the approximate model. Unfortunately, it turned out that this hope was often drastically wrong; even mild deviations often have much larger effects than were anticipated by most statisticians.

John W. Tukey (1960)

(And we gave two examples - Ronald Aylmer Fisher and Peter Huber.)

The main goals of robust statistics

1. To describe the structure best fitting the bulk of data.
2. To identify deviating data points (outliers) or deviating substructures for further treatment, if desired.
3. To identify and give a warning about highly influential data points (leverage points).
4. To deal with unsuspected serial correlation, or more generally, with deviations from the assumed correlation structures.

The four main types of deviations from the strict parametric model

1. The occurence of gross errors.
2. Rounding and grouping.
3. The model may have been conceived as an approximation anyway, e.g., by virtue of CLT.
4. Apart of distributional assumptions, the assumption of independence (or of some specific correlation structure) may only be approximately fulfilled.

Three approaches:

1. Huber's alternative to classical point estimation via neighbourhoods.
2. Huber's alternative to classical testing hypotheses via capacities.
3. Hampel's infinitesimal approach via Prokhorov metric and influence function.

Content

The Hampel's approach is based on two basic ideas and a nice fact:

1. The first idea - any estimator can be interpreted as a function T (say) from the space of all distribution functions $\mathcal{H}$ to the parameter space Θ (say).
2. The second idea - the function T can be studied by an infinitesimal calculus of limits, derivaties, integrals, etc.
3. A nice fact - the Kolmogorov-Smirnov result - the empirical d.f. converge uniformly to the "true" underlying one.

The Hampel's approach is based on two basic ideas and a nice fact:

1. The first idea - any estimator can be interpreted as a function T (say) from the space of all distribution functions $\mathcal{H}$ to the parameter space Θ (say).
2. The second idea - the function T can be studied by an infinitesimal calculus of limits, derivaties, integrals, etc.
3. A nice fact - the Kolmogorov-Smirnov result - the empirical d.f. converge uniformly to the "true" underlying one.

Recalling Taylor's expansion for a real function of real variable

1. The real function of one real variable $f(x)$

   $\rightarrow$ Taylor's expansion of $f(x) = f(x^0) + f'(x^0) \cdot (x - x^0) + \ldots.$

Recalling Taylor's expansion for a real function of real variable

1. The real function of one real variable $f(x)$

   $\rightarrow$ Taylor's expansion of $f(x) = f(x^0) + f'(x^0) \cdot (x - x^0) + \ldots$.

2. Let's recall the derivative of the function $f(x)$ at a given point x_0,

$$f'(x^0) = \lim_{\delta \to 0} \frac{f(x^0 + \delta) - f(x^0)}{\delta}$$

$\rightarrow$ the derivative offers an information about the behaviour of the function in a neighbourhood of x_0.

Recalling Taylor's expansion for a real function of finitely-dimensional variable

1. The real function of several real variables $f(x_1, x_2, ..., x_p)$

   $\rightarrow$ Taylor's expansion of $f(x) = f(x^0) + \sum_{j=1}^{p} \frac{\partial f(x^0)}{\partial x_j} \cdot (x_j - x_j^0) + ...$

Recalling Taylor's expansion for a real function of finitely-dimensional variable

1. The real function of several real variables $f(x_1, x_2, ..., x_p)$

   $\rightarrow$ Taylor's expansion of $f(x) = f(x^0) + \sum_{j=1}^{p} \frac{\partial f(x^0)}{\partial x_j} \cdot (x_j - x_j^0) + ...$

2. Let's recall again the partial derivative of the function $f(x)$ at the point x^0 along the j-th coordinate, i e.

$$\frac{\partial f(x^0)}{\partial x_j} = \lim_{\delta_j \to 0} \frac{f(x^0 + \Delta_j) - f(x^0)}{\delta_j}$$

where $\Delta_j = (0, 0, ..., \delta_j, ..., 0)'$.

Recalling Taylor's expansion for a real function of finitely-dimensional variable

1. The real function of several real variables $f(x_1, x_2, ..., x_p)$

   $\rightarrow$ Taylor's expansion of $f(x) = f(x^0) + \sum_{j=1}^{p} \frac{\partial f(x^0)}{\partial x_j} \cdot (x_j - x_j^0) + ...$

2. Let's recall again the partial derivative of the function $f(x)$ at the point x^0 along the j-th coordinate, i e.

$$\frac{\partial f(x^0)}{\partial x_j} = \lim_{\delta_j \to 0} \frac{f(x^0 + \Delta_j) - f(x^0)}{\delta_j}$$

   where $\Delta_j = (0, 0, ..., \delta_j, ..., 0)'$.

3. Realize that $\max_{j=1,2,...p} \left| \frac{\partial f(x^0)}{\partial x_j} \right|$ is a hint about the behaviour of the function in a neigbourhood of x^0.

Recalling Taylor's expansion for a real function of finitely-dimensional variable

1. The real function of several real variables $f(x_1, x_2, ..., x_p)$

   $\rightarrow$ Taylor's expansion of $f(x) = f(x^0) + \sum_{j=1}^{p} \frac{\partial f(x^0)}{\partial x_j} \cdot (x_j - x_j^0) + ...$

2. Let's recall again the partial derivative of the function $f(x)$ at the point x^0 along the j-th coordinate, i e.

$$\frac{\partial f(x^0)}{\partial x_j} = \lim_{\delta_j \to 0} \frac{f(x^0 + \Delta_j) - f(x^0)}{\delta_j}$$

   where $\Delta_j = (0, 0, ..., \delta_j, ..., 0)'$.

3. Realize that $\max_{j=1,2,...p} \left| \frac{\partial f(x^0)}{\partial x_j} \right|$ is a hint about the behaviour of the function in a neigbourhood of x^0.

Recalling Taylor's expansion for a real function of finitely-dimensional variable

1. The real function of several real variables $f(x_1, x_2, ..., x_p)$

   $\rightarrow$ Taylor's expansion of $f(x) = f(x^0) + \sum_{j=1}^{p} \frac{\partial f(x^0)}{\partial x_j} \cdot (x_j - x_j^0) + ...$

2. Let's recall again the partial derivative of the function $f(x)$ at the point x^0 along the j-th coordinate, i e.

Realize that when computing $\frac{\partial f(x^0)}{\partial x_j}$, we change only one coordinate,

3. Realize that $\max_{j=1,2,...p} \left| \frac{\partial f(x^0)}{\partial x_j} \right|$ is a hint about the behaviour of the function in a neigbourhood of x^0.

Recalling Taylor's expansion for a real function of finitely-dimensional variable

1. The real function of several real variables $f(x_1, x_2, ..., x_p)$

   $\rightarrow$ Taylor's expansion of $f(x) = f(x^0) + \sum_{j=1}^{p} \frac{\partial f(x^0)}{\partial x_j} \cdot (x_j - x_j^0) + ...$

2. Let's recall again the partial derivative of the function $f(x)$ at the point x^0 along the j-th coordinate, i e.

Realize that when computing $\frac{\partial f(x^0)}{\partial x_j}$, we change only one coordinate, i. e. we compute the derivative in one direction.

3. Realize that $\max_{j=1,2,...p} \left| \frac{\partial f(x^0)}{\partial x_j} \right|$ is a hint about <u>the behaviour of the function in a neigbourhood of x^0.</u>

Let's think about the partial derivative once again - a bit alternative approach.

Consider the partial derivative of
the function $f(x)$ at the point x^0 along the j-th coordinate, i e.

$$\frac{\partial f(x^0)}{\partial x_j} = \lim_{\delta \to 0} \frac{f(x^0 + \delta \cdot \Delta_j) - f(x^0)}{\delta}$$

where $\Delta_j = (0, 0, ..., 1, ..., 0)'$ - the unit is on the j-th position.

Generalizing Taylor's expansion for a real function of uncountably-dimensional variable

1. Denote a degenerated (at the point x) d. f. Δ_x
$\rightarrow \quad \Delta_x(v) = 0$ for $v \leq x$, $\Delta_x(v) = 1$ otherwise.

Generalizing Taylor's expansion for a real function of uncountably-dimensional variable

1. Denote a degenerated (at the point x) d. f. Δ_x
$\rightarrow \quad \Delta_x(v) = 0$ for $v \leq x$, $\Delta_x(v) = 1$ otherwise.

2. Fix a functional $T : \mathcal{H} \rightarrow R$ and consider the partial derivative of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

Generalizing Taylor's expansion for a real function of uncountably-dimensional variable

1. Denote a degenerated (at the point x) d. f. Δ_x
$\rightarrow \quad \Delta_x(v) = 0$ for $v \leq x$, $\Delta_x(v) = 1$ otherwise.
2. Fix a functional $T : \mathcal{H} \rightarrow R$ and consider the partial derivative of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

3. At those x's at which $IF(x, T, F)$ exists, call it *influence function* of the functional T at the d. f. F.

Generalizing Taylor's expansion for a real function of uncountably-dimensional variable

1. Denote a degenerated (at the point x) d. f. Δ_x
   $\rightarrow \quad \Delta_x(v) = 0$ for $v \leq x$, $\Delta_x(v) = 1$ otherwise.
2. Fix a functional $T : \mathcal{H} \rightarrow R$ and consider the partial derivative of the functional T at the point F along the x-th coordinate, i. e.
$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\Big((1-\delta)F(.) + \delta \cdot \Delta_x\Big) - T\Big(F(.)\Big)}{\delta}.$$
3. At those x's at which $IF(x, T, F)$ exists, call it *influence function* of the functional T at the d. f. F.

Generalizing Taylor's expansion for a real function of uncountably-dimensional variable

1. Denote a degenerated (at the point x) d. f. Δ_x

Notice that the influence function $IF(x, T, F)$ has three arguments:

1. the point x at which the contamination is assumed,

of the functional T at the d. f. F.

Generalizing Taylor's expansion for a real function of uncountably-dimensional variable

1. Denote a degenerated (at the point x) d. f. Δ_x

Notice that the influence function $IF(x, T, F)$ has three arguments:

1. the point x at which the contamination is assumed,
2. the functional T in question

of the functional T at the d. f. F.

Generalizing Taylor's expansion for a real function of uncountably-dimensional variable

1. Denote a degenerated (at the point x) d. f. Δ_x

Notice that the influence function $IF(x, T, F)$ has three arguments:

1. the point x at which the contamination is assumed,
2. the functional T in question

of the functional T at the d. f. F.

Generalizing Taylor's expansion for a real function of uncountably-dimensional variable

1. Denote a degenerated (at the point x) d. f. Δ_x

Notice that the influence function $IF(x, T, F)$ has three arguments:

1. the point x at which the contamination is assumed,
2. the functional T in question

and finally

3. the d. f. F, as the point of space $\mathcal{H}$.

of the functional T at the d. f. F.

What is the influence function good for?

1. Under some technical conditions

$$T(F_n) \cong T(F) + \int IF(x, F, T)\mathrm{d}F_n(x) + remainder_1,$$

i. e

$$T(F_n) \cong T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_1$$

What is the influence function good for?

1. Under some technical conditions

$$T(F_n) \cong T(F) + \int IF(x, F, T)\mathrm{d}F_n(x) + remainder_1,$$

i. e

$$T(F_n) \cong T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_1$$

2. It means that if we add new observation, say x_{n+1}, the value of estimator changes approximately from

$$T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) \quad \text{to} \quad T(F) + \frac{1}{n+1}\sum_{i=1}^{n+1} IF(x_i, F, T).$$

Asymptotic normality of estimator follows from ... .

1. We had, under some technical conditions

$$T(F_n) \cong T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_1$$

or equivalently

$$\sqrt{n}\,(T(F_n) - T(F)) \cong \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_2. \qquad (1)$$

Asymptotic normality of estimator follows from ... .

1. We had, under some technical conditions

$$T(F_n) \cong T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_1$$

or equivalently

$$\sqrt{n}(T(F_n) - T(F)) \cong \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_2. \qquad (1)$$

2. So, under the conditions allowing to apply CLT, the asymptotic normality of $T(F_n)$ follows directly from (**??**).

Asymptotic normality of estimator follows from ... .

1. We had, under some technical conditions

$$T(F_n) \cong T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_1$$

or equivalently

$$\sqrt{n}(T(F_n) - T(F)) \cong \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_2. \qquad (1)$$

2. So, under the conditions allowing to apply CLT,
the asymptotic normality of $T(F_n)$ follows directly from (**??**).

Asymptotic normality of estimator follows from ... .

1. We had, under some technical conditions

$$T(F_n) \cong T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_1$$

or equivalently

$$\sqrt{n}\,(T(F_n) - T(F)) \cong \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF(x_i, F, T) + remainder_2. \qquad (1)$$

2. So, under the conditions allowing to apply CLT, the asymptotic normality of $T(F_n)$ follows directly from (**??**).

Please, keep the last two slides in mind for a moment.

So, $\frac{1}{n+1} IF(x_{n+1}, F, T)$ represents a contribution

of the observation x_{n+1} to the functional $T(F_n)$.

So, $\frac{1}{n+1} IF(x_{n+1}, F, T)$ represents a contribution of the observation x_{n+1} to the functional $T(F_n)$.

Conclusion: The influence function $IF(x, F, T)$ (IF) predetermines or predestinates (many) properties of estimator.

So, $\frac{1}{n+1} IF(x_{n+1}, F, T)$ represents a contribution of the observation x_{n+1} to the functional $T(F_n)$.

Conclusion: The influence function $IF(x, F, T)$ (IF) predetermines or predestinates (many) properties of estimator.

That is why the characteristics of the estimator will be defined by means of IF.

So, $\frac{1}{n+1} IF(x_{n+1}, F, T)$ represents a contribution of the observation x_{n+1} to the functional $T(F_n)$.

Conclusion: The influence function $IF(x, F, T)$ (IF) predetermines or predestinates (many) properties of estimator.

That is why the characteristics of the estimator will be defined by means of IF.

And that is what we'll discuss today:

We are going to discuss:

1. Examples of the influence function

We are going to discuss:

1. Examples of the influence function
2. Recalling the classical requirements on the point estimator

We are going to discuss:

1. Examples of the influence function
2. Recalling the classical requirements on the point estimator
3. Adding some new ones defined by means of the influence function

We are going to discuss:

1. Examples of the influence function
2. Recalling the classical requirements on the point estimator
3. Adding some new ones defined by means of the influence function

We are going to discuss:

1. Examples of the influence function
2. Recalling the classical requirements on the point estimator
3. Adding some new ones defined by means of the influence function

Prior to it, let's recall the idea of interpreting the point estimator
as a function (functional) of empirical distribution function.

We had at the second lecture:

Estimator as a function of distribution function

1. Consider e. g. $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Estimator as a function of distribution function

1. Consider e. g. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
2. Let $F_n(.) \in \mathcal{H}$ be an empirical d. f. corresponding to the observations $x_1, x_2, ..., x_n$, then $T(F_n) = \int x \mathrm{d}F_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$
(because $F_n(x)$has positive $\mathrm{d}F_n(x)$ of size $\frac{1}{n}$ just at the points $x_1, x_2, ..., x_n$).

Estimator as a function of distribution function

1. Consider e. g. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
2. Let $F_n(.) \in \mathcal{H}$ be an empirical d. f. corresponding to the observations $x_1, x_2, ..., x_n$, then $T(F_n) = \int x \mathrm{d}F_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$
   (because $F_n(x)$ has positive $\mathrm{d}F_n(x)$ of size $\frac{1}{n}$ just at the points $x_1, x_2, ..., x_n$).
3. If we plug-in instead of empirical d. f. the underlying d. f. F, we obtain a function(al) $T : \mathcal{H} \rightarrow R^k$ $T(F) = \int x \mathrm{d}F(x) = \mathbb{E}X$ which is a theoretical counterpart to the estimator.

Estimator as a function of distribution function

1. Consider e. g. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
2. Let $F_n(.) \in \mathcal{H}$ be an empirical d. f. corresponding to the observations $x_1, x_2, ..., x_n$, then $T(F_n) = \int x \mathrm{d}F_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$
   (because $F_n(x)$has positive $\mathrm{d}F_n(x)$ of size $\frac{1}{n}$ just at the points $x_1, x_2, ..., x_n$).
3. If we plug-in instead of empirical d. f. the underlying d. f. F, we obtain a function(al) $T : \mathcal{H} \rightarrow R^k$ $T(F) = \int x \mathrm{d}F(x) = \mathbb{E}X$
   which is a theoretical counterpart to the estimator.
4. Typically, for any estimator we have a theoretical counterpart so that we can write $\hat{\theta}^{(n)} = T_n(F_n)$ and $\theta = T(F)$, where F_n is the empirical d. f. corresponding to the underlying d. f..

Estimator as a function of distribution function

1. Consider e. g. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
2. Let $F_n(.) \in \mathcal{H}$ be an empirical d. f. corresponding to the observations $x_1, x_2, ..., x_n$, then $T(F_n) = \int x \mathrm{d}F_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$
   (because $F_n(x)$has positive $\mathrm{d}F_n(x)$ of size $\frac{1}{n}$ just at the points $x_1, x_2, ..., x_n$).
3. If we plug-in instead of empirical d. f. the underlying d. f. F, we obtain a function(al) $T : \mathcal{H} \rightarrow R^k$ $T(F) = \int x \mathrm{d}F(x) = I\!\!E X$
   which is a theoretical counterpart to the estimator.
4. Typically, for any estimator we have a theoretical counterpart so that we can write $\hat{\theta}^{(n)} = T_n(F_n)$ and $\theta = T(F)$, where F_n is the empirical d. f. corresponding to the underlying d. f..

Estimator as a function of distribution function

1. Consider e. g. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
2. Let $F_n(.) \in \mathcal{H}$ be an empirical d. f. corresponding to the observations x

   Do you remember?

   $x_1, x_2, ..., x_n)$.
3. If we plug-in instead of empirical d. f. the underlying d. f. F, we obtain a function(al) $T : \mathcal{H} \rightarrow R^k$ $T(F) = \int x \mathrm{d}F(x) = \mathbb{E}X$ which is a theoretical counterpart to the estimator.
4. Typically, for any estimator we have a theoretical counterpart so that we can write $\hat{\theta}^{(n)} = T_n(F_n)$ and $\theta = T(F)$, where F_n is the empirical d. f. corresponding to the underlying d. f..

Estimator as a function of distribution function - another example

1. We could consider also $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
$= \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2.$

Estimator as a function of distribution function - another example

1. We could consider also $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$
$$= \frac{1}{n-1}\sum_{i=1}^n x_i^2 - \frac{n}{n-1}\bar{x}^2.$$
2. Denote the funcional from previous slide by $T^{(1)}(F_n) = \bar{x}$.

Estimator as a function of distribution function - another example

1. We could consider also $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
$$= \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2.$$
2. Denote the funcional from previous slide by $T^{(1)}(F_n) = \bar{x}$.
3. Let $F_n(.) \in \mathcal{H}$ be an empirical d. f. corresponding to the observations $x_1, x_2, ..., x_n$ and put $T^{(2)}(F_n) = \frac{n}{n-1}\int x^2 \mathrm{d}F_n(x) = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2$.

Estimator as a function of distribution function - another example

1. We could consider also $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
$$= \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2.$$
2. Denote the funcional from previous slide by $T^{(1)}(F_n) = \bar{x}$.
3. Let $F_n(.) \in \mathcal{H}$ be an empirical d. f. corresponding to the observations $x_1, x_2, ..., x_n$ and put $T^{(2)}(F_n) = \frac{n}{n-1}\int x^2 \mathrm{d}F_n(x) = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2$.
4. Then we have $s_n^2 = T^{(2)}(F_n) - \frac{n}{n-1}\left[T^{(1)}(F_n)\right]^2$.

Estimator as a function of distribution function - another example

1. We could consider also $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$
$= \frac{1}{n-1}\sum_{i=1}^n x_i^2 - \frac{n}{n-1}\bar{x}^2.$
2. Denote the funcional from previous slide by $T^{(1)}(F_n) = \bar{x}$.
3. Let $F_n(.) \in \mathcal{H}$ be an empirical d. f. corresponding to the observations $x_1, x_2, ..., x_n$ and put $T^{(2)}(F_n) = \frac{n}{n-1}\int x^2 \mathrm{d}F_n(x) = \frac{1}{n-1}\sum_{i=1}^n x_i^2$.
4. Then we have $s_n^2 = T^{(2)}(F_n) - \frac{n}{n-1}\left[T^{(1)}(F_n)\right]^2$.
5. If we plug-in instead of empirical d. f. the underlying d. f. F, we obtain a theoretical counterpart to the estimator.

Usually we can find a function $h(x)$ (say) so that we have

$$T(F_n) = \int h(x) \mathrm{d}F_n(x).$$

Usually we can find a function $h(x)$ (say) so that we have

$$T(F_n) = \int h(x) \mathrm{d}F_n(x).$$

In previous examples we had

1. for $T^{(1)}(F_n)$ the function $h^{(1)}(x) = x$

Usually we can find a function $h(x)$ (say) so that we have

$$T(F_n) = \int h(x) \mathrm{d}F_n(x).$$

In previous examples we had

1. for $T^{(1)}(F_n)$ the function $h^{(1)}(x) = x$

Usually we can find a function $h(x)$ (say) so that we have

$$T(F_n) = \int h(x)\mathrm{d}F_n(x).$$

In previous examples we had

1. for $T^{(1)}(F_n)$ the function $h^{(1)}(x) = x$

   and

2. for $T^{(2)}(F_n)$ the function $h^{(2)}(x) = \frac{n}{n-1}x^2$.

Content

Let's recall that for the standard normal distribution we use usually Φ and for its density ϕ.

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \rightarrow 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi) = \mathbb{E}_\Phi(X) = \int X d\Phi = \int z \cdot \phi(z) dz$.

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi) = \mathbb{E}_\Phi(X) = \int X \mathrm{d}\Phi = \int z \cdot \phi(z) \mathrm{d}z$.
2. $T\left(\Phi(.)\right) = \frac{1}{\sqrt{2\pi}} \int z \cdot \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z = 0$

Returning to the definition of influence function

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi) = E_\Phi(X) = \int X \mathrm{d}\Phi = \int z \cdot \phi(z) \mathrm{d}z$.
2. $T\left(\Phi(.)\right) = \frac{1}{\sqrt{2\pi}} \int z \cdot \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z = 0$
3. $T\left((1-\delta)\Phi(.) + \delta \cdot \Delta_x\right)$
$= \int z \left\{(1-\delta)\frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{z^2}{2}\right\} + \delta \cdot \Delta_x\right\} \mathrm{d}z = (1-\delta) \cdot 0 + \delta \cdot x.$

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x$) and consider the partial derivative
 of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi) = E_\Phi(X) = \int X \mathrm{d}\Phi = \int z \cdot \phi(z)\mathrm{d}z$.
2. $T\left(\Phi(.)\right) = \frac{1}{\sqrt{2\pi}} \int z \cdot \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z = 0$
3. $T\left((1-\delta)\Phi(.) + \delta \cdot \Delta_x\right)$
   $= \int z \left\{(1-\delta)\frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{z^2}{2}\right\} + \delta \cdot \Delta_x\right\} \mathrm{d}z = (1-\delta) \cdot 0 + \delta \cdot x.$
4. Finally, $IF(x, T, \Phi) = \lim_{\delta \to 0} \frac{\delta \cdot x}{\delta} = x.$

Let's recall once again that for the standard normal distribution we use usually Φ and for its density ϕ.

Let's recall once again that for the standard normal distribution we use usually Φ and for its density ϕ.

Let's generalize it a bit so that Φ_{μ,σ^2} and ϕ_{μ,σ^2} will denote normal d. f. and the normal density with mean μ and variance σ^2, respectively.

Let's recall once again that for the standard normal distribution we use usually Φ and for its density ϕ.

Let's generalize it a bit so that Φ_{μ,σ^2} and ϕ_{μ,σ^2} will denote normal d. f. and the normal density with mean μ and variance σ^2, respectively.

And compute the $IF(x, T, \Phi_{\mu,\sigma^2})$.

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x^2$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x^2$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x,T,F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi_{\mu,\sigma^2}) = \mathbb{E}_{\Phi_{\mu,\sigma^2}}(X) = \int X \mathrm{d}\Phi_{\mu,\sigma^2} = \int z \cdot \phi_{\mu,\sigma^2}(z)\mathrm{d}z$.

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x^2$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi_{\mu,\sigma^2}) = \mathbb{E}_{\Phi_{\mu,\sigma^2}}(X) = \int X \mathrm{d}\Phi_{\mu,\sigma^2} = \int z \cdot \phi_{\mu,\sigma^2}(z)\mathrm{d}z$.
2. $T\left(\Phi_{\mu,\sigma^2}(.)\right) = \frac{1}{\sqrt{2\pi\sigma}} \int z \cdot \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\} \mathrm{d}z = \mu$

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x^2$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi_{\mu,\sigma^2}) = \mathbb{E}_{\Phi_{\mu,\sigma^2}}(X) = \int X \mathrm{d}\Phi_{\mu,\sigma^2} = \int z \cdot \phi_{\mu,\sigma^2}(z)\mathrm{d}z$.
2. $T\left(\Phi_{\mu,\sigma^2}(.)\right) = \frac{1}{\sqrt{2\pi\sigma}} \int z \cdot \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\} \mathrm{d}z = \mu$
3. $T\left((1-\delta)\Phi_{\mu,\sigma^2}(.) + \delta \cdot \Delta_x\right)$
$= \int z \left\{(1-\delta)\frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\} + \delta \cdot \Delta_x\right\} \mathrm{d}z = (1-\delta) \cdot \mu + \delta \cdot x.$

Fix a functional $T : \mathcal{H} \rightarrow R$ (now T is given by $h(x) = x^2$) and consider the partial derivative
of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi_{\mu,\sigma^2}) = \mathbb{E}_{\Phi_{\mu,\sigma^2}}(X) = \int X \mathrm{d}\Phi_{\mu,\sigma^2} = \int z \cdot \phi_{\mu,\sigma^2}(z)\mathrm{d}z$.
2. $T\left(\Phi_{\mu,\sigma^2}(.)\right) = \frac{1}{\sqrt{2\pi\sigma}} \int z \cdot \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\} \mathrm{d}z = \mu$
3. $T\left((1-\delta)\Phi_{\mu,\sigma^2}(.) + \delta \cdot \Delta_x\right)$
$= \int z \left\{(1-\delta)\frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\} + \delta \cdot \Delta_x\right\} \mathrm{d}z = (1-\delta) \cdot \mu + \delta \cdot x.$
4. Finally, $IF(x, T, \Phi) = \lim_{\delta \to 0} \frac{\delta \cdot (\mu + x)}{\delta} = \mu + x.$

Returning once again to the definition of influence function

Fix a functional $T : \mathcal{H} \rightarrow R$ and consider the partial derivative of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

Fix a functional $T : \mathcal{H} \rightarrow R$ and consider the partial derivative of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi) = \mathbb{E}_{\Phi}(X^2) = \int X^2 \mathrm{d}\Phi = \int z^2 \cdot \phi(z)\mathrm{d}z$.

Fix a functional $T : \mathcal{H} \rightarrow R$ and consider the partial derivative of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x,T,F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi) = \mathbb{E}_\Phi(X^2) = \int X^2 \mathrm{d}\Phi = \int z^2 \cdot \phi(z)\mathrm{d}z$.
2. $T\left(\Phi(.)\right) = \frac{1}{\sqrt{2\pi}} \int z^2 \cdot \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z = 1$

Returning once again to the definition of influence function

Fix a functional $T : \mathcal{H} \rightarrow R$ and consider the partial derivative of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x, T, F) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right) - T\left(F(.)\right)}{\delta}.$$

1. Fix $T(\Phi) = I\!\!E_\Phi(X^2) = \int X^2 \mathrm{d}\Phi = \int z^2 \cdot \phi(z)\mathrm{d}z$.
2. $T\left(\Phi(.)\right) = \frac{1}{\sqrt{2\pi}} \int z^2 \cdot \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z = 1$
3. $T\left((1-\delta)\Phi(.) + \delta \cdot \Delta_x\right)$

$$= \int z^2 \left\{(1-\delta)\frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{z^2}{2}\right\} + \delta \cdot \Delta_x\right\} \mathrm{d}z = (1-\delta) \cdot 1 + \delta \cdot x^2.$$

Fix a functional $T : \mathcal{H} \rightarrow R$ and consider the partial derivative of the functional T at the point F along the x-th coordinate, i. e.

$$IF(x,T,F) = \lim_{\delta\to 0} \frac{T\Big((1-\delta)F(.) + \delta\cdot\Delta_x\Big) - T\Big(F(.)\Big)}{\delta}.$$

1. Fix $T(\Phi) = \mathbb{E}_\Phi(X^2) = \int X^2 \mathrm{d}\Phi = \int z^2 \cdot \phi(z)\mathrm{d}z$.
2. $T\Big(\Phi(.)\Big) = \frac{1}{\sqrt{2\pi}} \int z^2 \cdot \exp\left\{-\frac{z^2}{2}\right\}\mathrm{d}z = 1$
3. $T\Big((1-\delta)\Phi(.) + \delta\cdot\Delta_x\Big)$
   $= \int z^2 \left\{(1-\delta)\frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{z^2}{2}\right\} + \delta\cdot\Delta_x\right\}\mathrm{d}z = (1-\delta)\cdot 1 + \delta\cdot x^2.$
4. Finally, $IF(x,T,\Phi) = \lim_{\delta\to 0} \frac{(1-\delta)\cdot 1 + \delta\cdot x^2 - 1}{\delta} = -1 + x^2.$

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)

Let's discuss them from the point of view of robust procedures -
- we know already enough about it to be able to do it.

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)

Let's start with admissibility, recalling its definition.

5. Efficiency
6. Scale- and regression-equivariance
7. Admissibility

Definition of Mean Square Error (MSE):

Let $\hat{\theta}$ be an estimator, then

$$MSE(\hat{\theta}, \theta) = \mathbb{E}_{\theta} \left(\hat{\theta} - \theta \right)^2.$$

Definition of Mean Square Error (MSE):

Let $\hat{\theta}$ be an estimator, then

$$MSE(\hat{\theta}, \theta) = \mathbb{E}_\theta \left(\hat{\theta} - \theta\right)^2.$$

We say that the point estimator $\hat{\theta}^{(1)}$ is better than $\hat{\theta}^{(2)}$ if

1. $\forall(\theta \in \Theta) \quad MSE(\hat{\theta}^{(1)}, \theta) \leq MSE(\hat{\theta}^{(2)}, \theta)$,

Definition of Mean Square Error (MSE):

Let $\hat{\theta}$ be an estimator, then

$$MSE(\hat{\theta}, \theta) = \mathbb{E}_\theta \left(\hat{\theta} - \theta\right)^2.$$

We say that the point estimator $\hat{\theta}^{(1)}$ is better than $\hat{\theta}^{(2)}$ if

1. $\forall(\theta \in \Theta) \quad MSE(\hat{\theta}^{(1)}, \theta) \leq MSE(\hat{\theta}^{(2)}, \theta)$,
2. $\exists(\theta_0 \in \Theta) \quad MSE(\hat{\theta}^{(1)}, \theta_0) < MSE(\hat{\theta}^{(2)}, \theta_0)$.

Definition of Mean Square Error (MSE):

Let $\hat{\theta}$ be an estimator, then

$$MSE(\hat{\theta}, \theta) = \mathbb{E}_\theta \left(\hat{\theta} - \theta\right)^2.$$

We say that the point estimator $\hat{\theta}^{(1)}$ is better than $\hat{\theta}^{(2)}$ if

1. $\forall(\theta \in \Theta) \quad MSE(\hat{\theta}^{(1)}, \theta) \leq MSE(\hat{\theta}^{(2)}, \theta)$,
2. $\exists(\theta_0 \in \Theta) \quad MSE(\hat{\theta}^{(1)}, \theta_0) < MSE(\hat{\theta}^{(2)}, \theta_0)$.

Definition of Mean Square Error (MSE):

Let $\hat{\theta}$ be an estimator, then

$$MSE(\hat{\theta}, \theta) = \mathbb{E}_\theta \left(\hat{\theta} - \theta\right)^2.$$

We say that the point estimator $\hat{\theta}^{(1)}$ is better than $\hat{\theta}^{(2)}$ if

1. $\forall(\theta \in \Theta) \quad MSE(\hat{\theta}^{(1)}, \theta) \leq MSE(\hat{\theta}^{(2)}, \theta)$,
2. $\exists(\theta_0 \in \Theta) \quad MSE(\hat{\theta}^{(1)}, \theta_0) < MSE(\hat{\theta}^{(2)}, \theta_0)$.

Definition of admissible estimator:

Let $\hat{\theta}$ be an estimator, then we say that $\hat{\theta}$ is admissible if there is not an estimator better than $\hat{\theta}$.

(And we assume that it holds independently on number of observations.)

We don't require (study?) admissibility for robust estimators because:

1. It is much more important to compare them on the base of other properties as the level of robustness, a loss of efficiency, etc.

We don't require (study?) admissibility for robust estimators because:

1. It is much more important to compare them on the base of other properties as the level of robustness, a loss of efficiency, etc.
2. Moreover, frankly speaking, we are not able to compute (exactly) nearly any finite-sample characteristic of these estimators. And hence also not the $MSE(\hat{\theta}^{(n)})$.

We don't require (study?) admissibility for robust estimators because:

1. It is much more important to compare them on the base of other properties as the level of robustness, a loss of efficiency, etc.
2. Moreover, frankly speaking, we are not able to compute (exactly) nearly any finite-sample characteristic of these estimators.

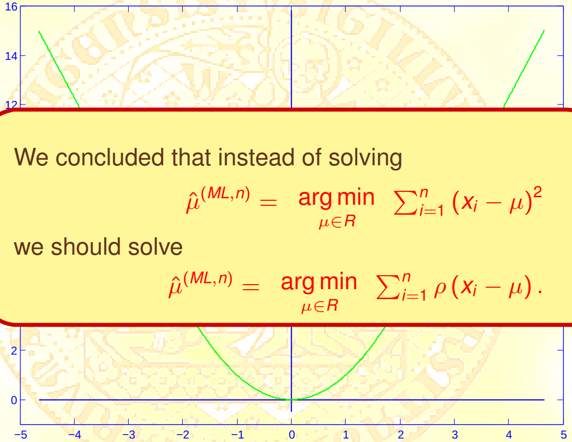
   And hence also not the $MSE(\hat{\theta}^{(n)})$.

We don't require (study?) admissibility for robust estimators because:

1. It is much more important to compare them on the base of other properties as the level of robustness, a loss of efficiency, etc.
2. Moreover, frankly speaking, we are not able to compute (exactly) nearly any finite-sample characteristic of these estimators.
   And hence also not the $MSE(\hat{\theta}^{(n)})$.

Now, let's return to the first lecture and discuss the unbiasedness.

Recalling that we found on the first lecture

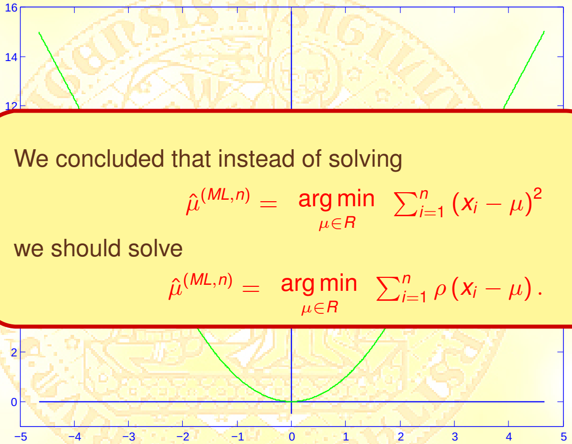


We concluded that instead of solving

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

we should solve

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



We concluded that instead of solving

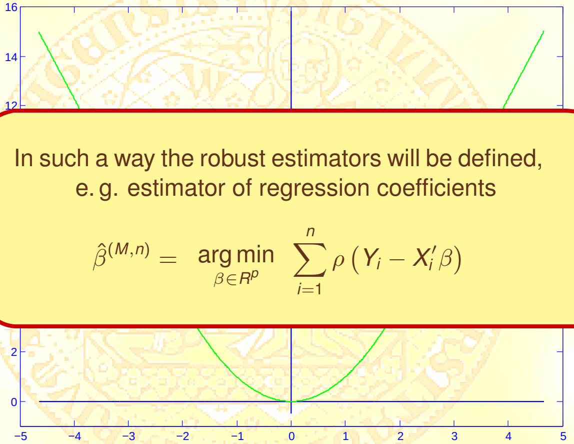
$$\hat{\mu}^{(ML,n)} = \underset{\mu \in R}{\arg\min} \sum_{i=1}^{n} (x_i - \mu)^2$$

we should solve

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

Notice the bottom line in the frame !!

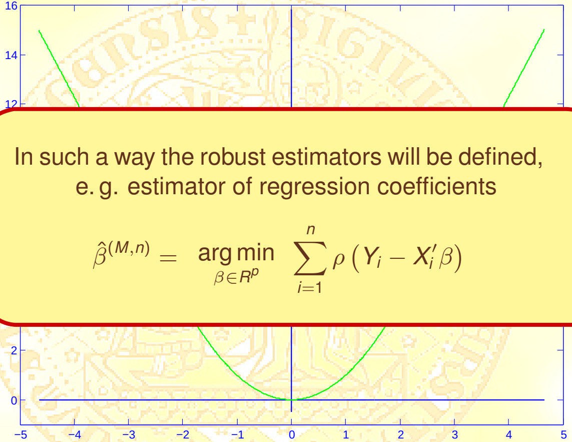
Recalling that we found on the first lecture



In such a way the robust estimators will be defined, e. g. estimator of regression coefficients

$$\hat{\beta}^{(M,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} \rho\left(Y_i - X_i'\beta\right)$$

Recalling that we found on the first lecture

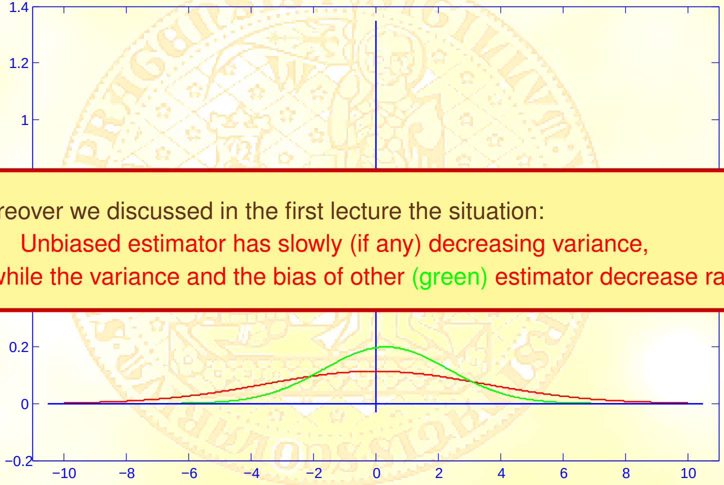


In such a way the robust estimators will be defined, e. g. estimator of regression coefficients

$$\hat{\beta}^{(M,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} \rho\left(Y_i - X_i'\beta\right)$$

But then we cannot (typically) find a formula for (robust) estimators and hence we cannot prove (compute ?!?) unbiasedness.

Possible density of unbiased and biased estimator

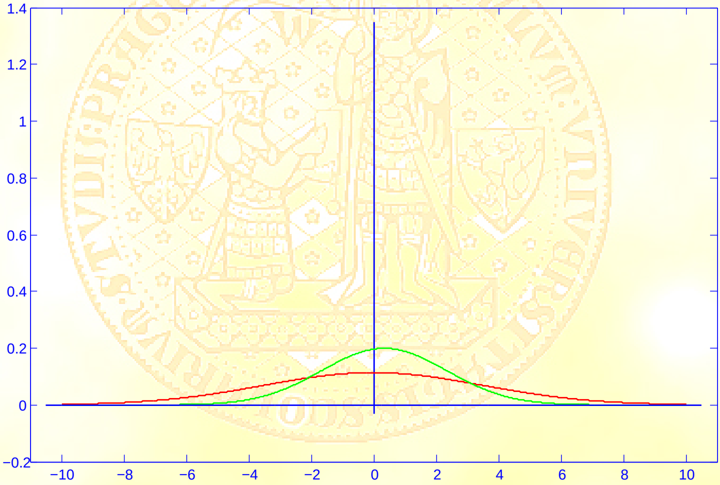


Moreover we discussed in the first lecture the situation:

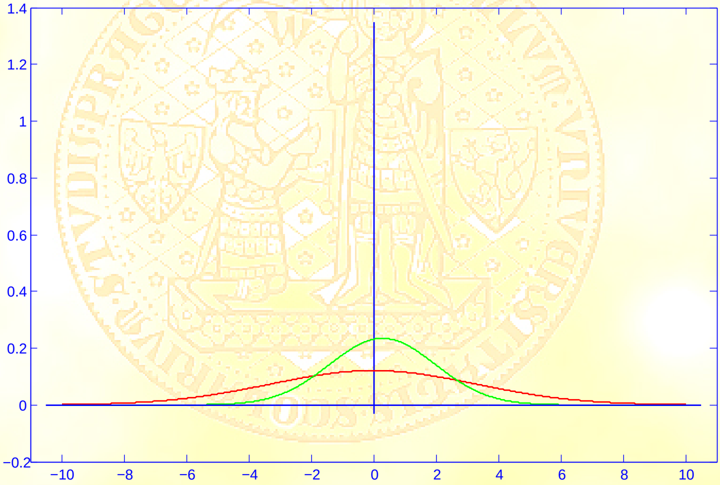
Unbiased estimator has slowly (if any) decreasing variance,

while the variance and the bias of other (green) estimator decrease rapidly.

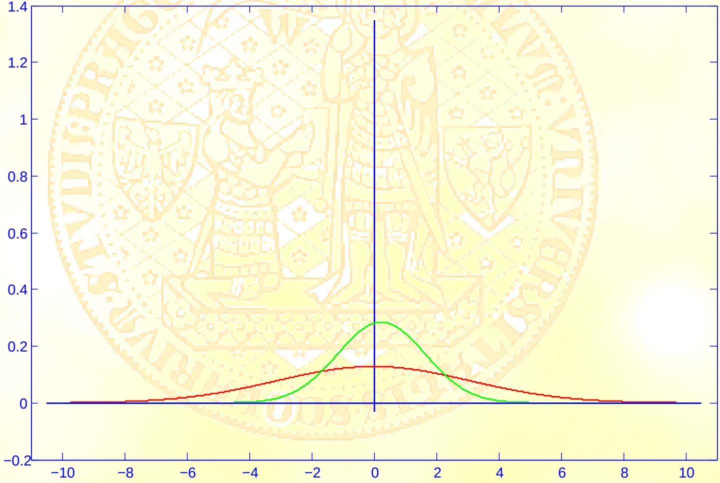
Notice decreasing variance and bias



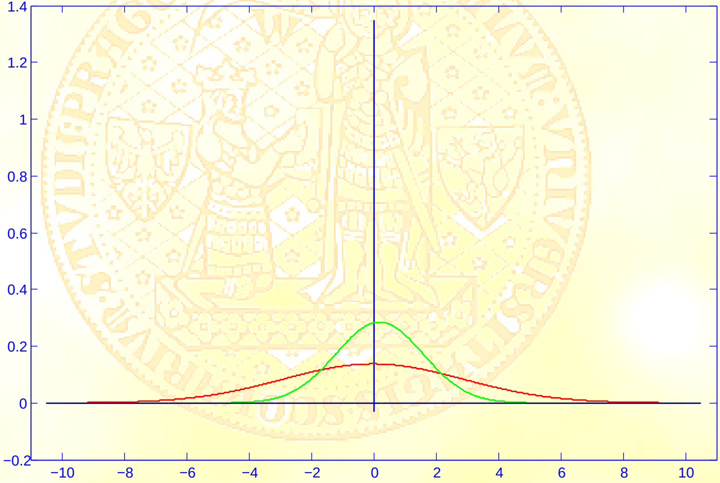
Notice decreasing variance and bias



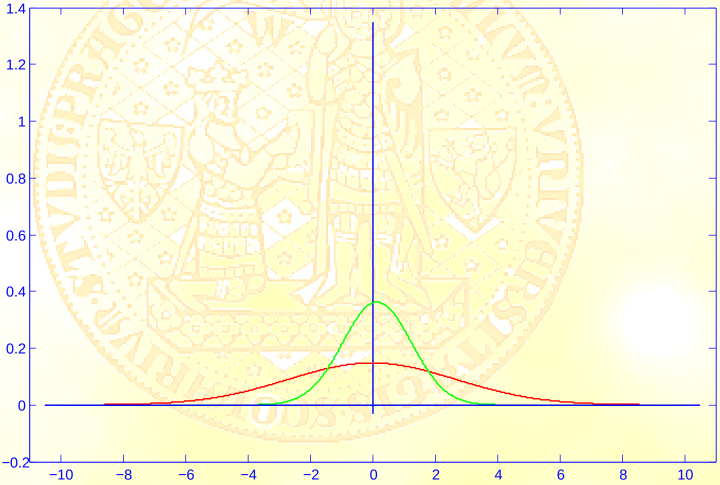
Notice decreasing variance and bias



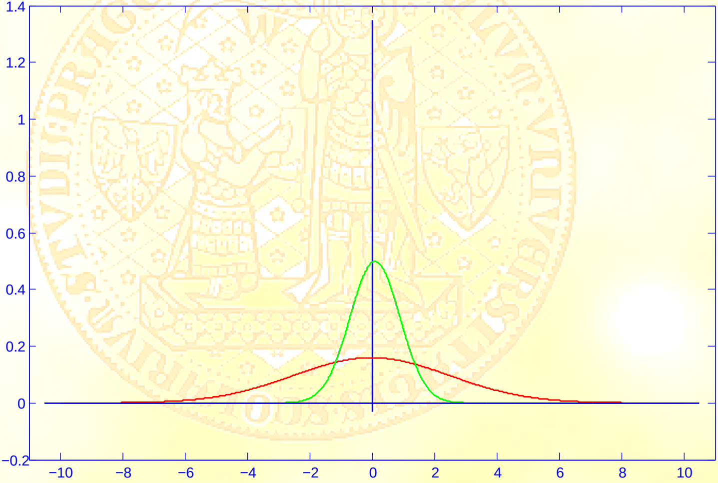
Notice decreasing variance and bias



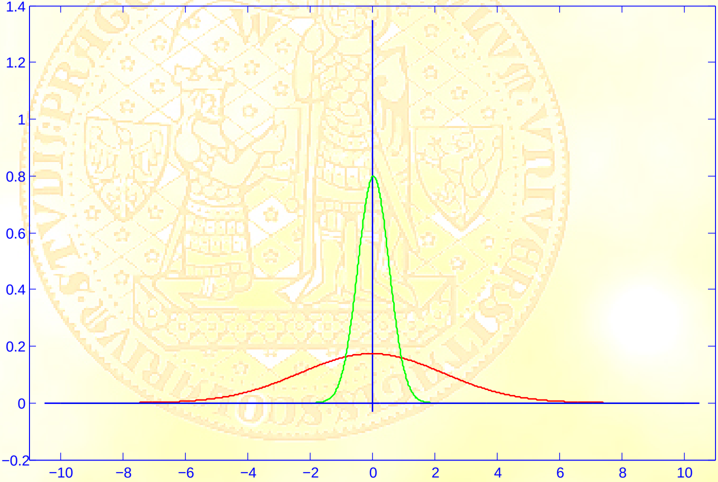
Notice decreasing variance and bias



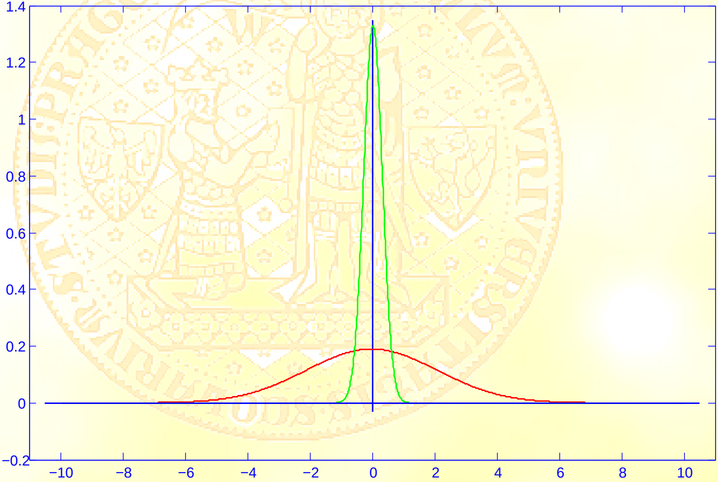
Notice decreasing variance and bias

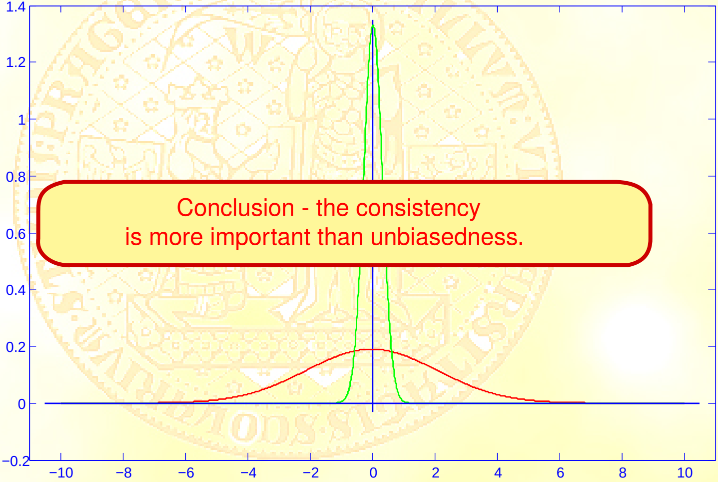


Notice decreasing variance and bias



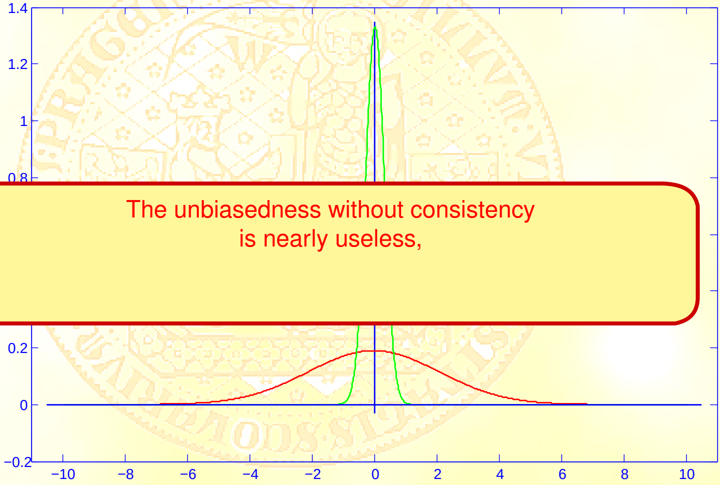
Notice decreasing variance and bias





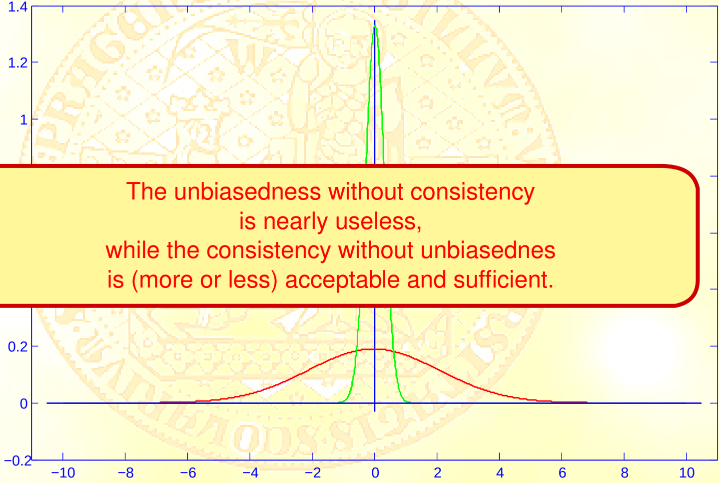
Conclusion - the consistency
is more important than unbiasedness.

Notice decreasing variance and bias



The unbiasedness without consistency
is nearly useless,

Notice decreasing variance and bias



The unbiasedness without consistency
is nearly useless,
while the consistency without unbiasednes
is (more or less) acceptable and sufficient.

Nearly concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)

Nearly concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)
2. $\sqrt{n}$-consistency (root-n-consistency)

Nearly concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)
2. $\sqrt{n}$-consistency (root-n-consistency)
3. Asymptotic normality

Nearly concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)
2. $\sqrt{n}$-consistency (root-n-consistency)
3. Asymptotic normality
4. Loss of efficiency as small as possible

Nearly concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)
2. $\sqrt{n}$-consistency (root-n-consistency)
3. Asymptotic normality
4. Loss of efficiency as small as possible
5. Scale- and regression-equivariance

Nearly concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)

We still didn't discuss scale- and regression-equivariance

4. Loss of efficiency as small as possible
5. Scale- and regression-equivariance

Nearly concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)

We still didn't discuss scale- and regression-equivariance
- so let's do it.

4. Loss of efficiency as small as possible
5. Scale- and regression-equivariance

Framework: $Y_i = X_i'\beta^0 + e_i$

$i = 1, 2, \ldots, n$

Equivariance of $\hat{\beta}^{(n)}$

$$\hat{\beta}(Y, X) : M(n, p+1) \to R^p$$

scale-equivariant : $\forall c \in R^+ \quad \hat{\beta}(cY, X) = c\hat{\beta}(Y, X)$

regression-equivariant : $\forall b \in R^p \quad \hat{\beta}(Y + Xb, X) = \hat{\beta}(Y, X) + b$

Framework: $Y_i = X_i'\beta^0 + e_i$

$i = 1, 2, \ldots, n$

Equivariance of $\hat{\beta}^{(n)}$

$$\hat{\beta}(Y, X) : M(n, p+1) \to R^p$$

scale-equivariant : $\forall c \in R^+ \quad \hat{\beta}(cY, X) = c\hat{\beta}(Y, X)$

regression-equivariant : $\forall b \in R^p \quad \hat{\beta}(Y + Xb, X) = \hat{\beta}(Y, X) + b$

Examples : $\hat{\beta}^{(OLS,n)} = (X'X)^{-1} X'Y$

Framework: $Y_i = X_i'\beta^0 + e_i$

$i = 1, 2, \ldots, n$

Equivariance of $\hat{\beta}^{(n)}$

$$\hat{\beta}(Y, X) : M(n, p+1) \rightarrow R^p$$

scale-equivariant : $\forall c \in R^+ \quad \hat{\beta}(cY, X) = c\hat{\beta}(Y, X)$

regression-equivariant : $\forall b \in R^p \quad \hat{\beta}(Y + Xb, X) = \hat{\beta}(Y, X) + b$

Examples : $\hat{\beta}^{(OLS,n)} = (X'X)^{-1} X'Y$

$$\hat{\beta}^{(L_1,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} |Y_i - X_i'\beta|$$

Framework: $Y_i = X_i'\beta^0 + e_i$

$i = 1, 2, \ldots, n$

Equivariance - invariance of $\hat{\sigma}^2$

$$\hat{\sigma}^2(Y, X) : M(n, p+1) \to R^+$$

scale-equivariant : $\forall c \in R^+ \quad \hat{\sigma}^2(cY, X) = c^2\hat{\sigma}^2(Y, X)$

regression-invariant : $\forall b \in R^p \quad \hat{\sigma}^2(Y + Xb, X) = \hat{\sigma}^2(Y, X)$

Framework: $Y_i = X_i' \beta^0 + e_i$

$i = 1, 2, \ldots, n$

Equivalariance - invariance of $\hat{\sigma}^2$

$$\hat{\sigma}^2(Y, X) : M(n, p+1) \rightarrow R^+$$

scale-equivariant : $\forall c \in R^+ \quad \hat{\sigma}^2(cY, X) = c^2 \hat{\sigma}^2(Y, X)$

regression-invariant : $\forall b \in R^p \quad \hat{\sigma}^2(Y + Xb, X) = \hat{\sigma}^2(Y, X)$

Examples : $s_n^2 = \frac{1}{n-p} \sum_{i=1}^{n} r_i^2(\hat{\beta}^{(OLS,n)})$

What is the equivariance of $\hat{\beta}^{(n)}$ good for ?

1. When the units of measurement have been changed,
   we don't need to recalculate the estimator
   - we just shift the decimal point

   (we are used to it from classical statistics).

What is the equivariance of $\hat{\beta}^{(n)}$ good for ?

1. When the units of measurement have been changed,
   we don't need to recalculate the estimator
   - we just shift the decimal point
   (we are used to it from classical statistics).
2. The requirement of invariance and equivariance
   removed superefficiency.

Finally, concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)
2. $\sqrt{n}$-consistency (root-n-consistency)
3. Asymptotic normality
4. Loss of efficiency as small as possible
5. Scale- and regression-equivariance

Finally, concluding:

The requirements overtaken from the classical statistics

1. Consistency (typically weak, i. e. in probability)

And now we add some others which correspond to the spirit of the discussion we have passed up to this moment.

4. Loss of efficiency as small as possible
5. Scale- and regression-equivariance

Returning to IF once again

Let's recall that if we add new observation, say x_{n+1},
the value of estimator changes from

$$T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) \quad \text{to} \quad T(F) + \frac{1}{n+1}\sum_{i=1}^{n+1} IF(x_i, F, T).$$

Returning to IF once again

Let's recall that if we add new observation, say x_{n+1},
the value of estimator changes from

$$T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) \quad \text{to} \quad T(F) + \frac{1}{n+1}\sum_{i=1}^{n+1} IF(x_i, F, T).$$

So, $IF(x_{n+1}, F, T)$ represents
a contribution of the observation x_{n+1} to the functional $T(F_n)$.

Returning to IF once again

Let's recall that if we add new observation, say x_{n+1},
the value of estimator changes from

$$T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) \quad \text{to} \quad T(F) + \frac{1}{n+1}\sum_{i=1}^{n+1} IF(x_i, F, T).$$

So, $IF(x_{n+1}, F, T)$ represents
a contribution of the observation x_{n+1} to the functional $T(F_n)$.

Influence function $IF(x, F, T)$ (IF) predetermines or predestinates (many) properties of estimator.

Returning to IF once again

Let's recall that if we add new observation, say x_{n+1},
the value of estimator changes from

$$T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(x_i, F, T) \quad \text{to} \quad T(F) + \frac{1}{n+1}\sum_{i=1}^{n+1} IF(x_i, F, T).$$

So, $IF(x_{n+1}, F, T)$ represents
a contribution of the observation x_{n+1} to the functional $T(F_n)$.

Influence function $IF(x, F, T)$ (IF) predetermines or predestinates (many) properties of estimator.

So, let's define a couple of new requirements by it.

Hampel's approach - characteristics of the functional T at the d. f. F

- Clearly,

$$\gamma^* = \sup_{x \in R} |IF(x, T, F)|$$

represents a maximal possible contribution of observation x to the value of the functional T provided the d. f. which generated data was F.

Hampel's approach - characteristics of the functional T at the d. f. F

- Clearly,

$$\gamma^* = \sup_{x \in R} |IF(x, T, F)|$$

  represents a maximal possible contribution of observation x to the value of the functional T provided the d. f. which generated data was F.

- γ^* is called *gross-error sensitivity*.

We should depress the influence of large number of small shifts

Hampel's approach - characteristics of the functional T at the d. f. F

- Similarly, the maximal Lipschitz ratio

$$\lambda^* = \sup_{x,y \in R} \left| \frac{IF(x,T,F) - IF(y,T,F)}{x - y} \right|$$

represents a maximal possible contribution to the value of the functional T provided the d. f. which generated data was F by a rounding observation x.

Hampel's approach - characteristics of the functional T at the d. f. F

- Similarly, the maximal Lipschitz ratio

$$\lambda^* = \sup_{x,y \in R} \left| \frac{IF(x,T,F) - IF(y,T,F)}{x - y} \right|$$

  represents a maximal possible contribution to the value of the functional T provided the d. f. which generated data was F by a rounding observation x.

- λ^* is called *local-shift sensitivity*.

Hampel's approach - characteristics of the functional T at the d. f. F

- Finally,

$$\rho^* = \inf \left\{ r \in R^+ \; : \; IF(x, T, F) = 0, \; |x| > r \right\}$$

represents a value such that any observation which is in absolute value larger then ρ^* brings no contribution to the value of the functional T provided the d. f. which generated data was F.

Hampel's approach - characteristics of the functional T at the d. f. F

- Finally,

$$\rho^* = \inf \left\{ r \in R^+ \; : \; IF(x,T,F) = 0, \; |x| > r \right\}$$

represents a value such that any observation which is in absolute value larger then ρ^* brings no contribution to the value of the functional T provided the d. f. which generated data was F.

- ρ^* is called *rejection point*.

There is still one theoretical requirement on the robust estimator which seemed to be

- at the early days of robust methods -

the most important - the breakdown point.

There is still one theoretical requirement on the robust estimator which seemed to be

- at the early days of robust methods -

the most important - the breakdown point.

It became - for some time - even an obsession of some statisticians and econometricians.

There is still one theoretical requirement on the robust estimator which seemed to be

- at the early days of robust methods -

the most important - the breakdown point.

It became - for some time - even an obsession of some statisticians and econometricians.

Finally, there are also a couple of practical requirements. All these topics will be discussed in the next lectures.

THANKS FOR ATTENTION