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 4

## Content of lecture

1. Repetiton is mother of wisdom(Jan mos Komensk)
2. The breakdown point
3. Specification of robustness characteristics for classical estimators

## A brief repetition of some points from previous lectures

We have concluded:

**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

## A brief repetition of some points from previous lectures

Then we added:

**The requirements enlarging the classical paradigma**

1. Gross-error sensitivity
$$\gamma^* = \sup_{x \in R} |IF(x, T, F)|$$
2. Local-shift sensitivity
$$\lambda^* = \sup_{x,y \in R} \left| \frac{IF(x, T, F) - IF(y, T, F)}{x - y} \right|$$
3. Rejection point
$$\rho^* = \inf \left\{ r \in R^+ \; : \; IF(x, T, F) = 0, \;\; |x| > r \right\}$$
4. Breakdown point - will be discused as the first topic today

## A brief repetition of some points from previous lectures

Then we added$_{(continued)}$:

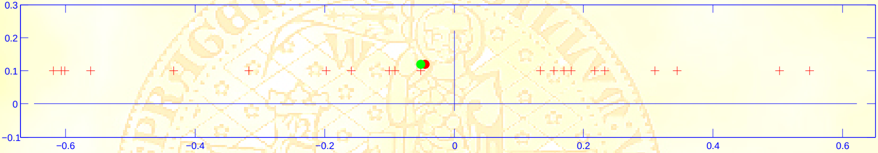
**The requirements enlarging the classical paradigma**

1. Tight algorithm and reliable implementation
   - invention and verification
2. Good heuristic - to convince people to employ it

First of

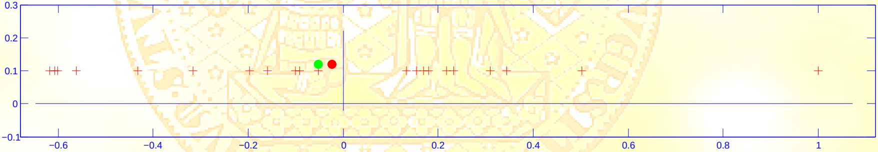
Observe the mean • and the median •

Non-contaminated data



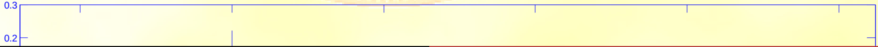
Co

at point 1



Co

at point 2



Then there is of course a question:

Why we use more frequently mean than median?

$$\lim_{n\to\infty} \frac{\mathrm{var}_\Phi\left(\bar{x}^{(n)}\right)}{\mathrm{var}_\Phi\left(median^{(n)}\right)} = 0.6....$$

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

- An overall characteristic of the functional (the estimator) is

$$\varepsilon^* = \sup\Big\{\varepsilon \le 1 : \exists K_\varepsilon \subset \Theta, \quad K_\varepsilon \quad \textit{compact}$$

$$\pi(F,G) < \varepsilon \;\Rightarrow\; G(\{T_n \in K_\varepsilon\}) \underset{n\to\infty}{\longrightarrow} 1\Big\}$$

where $\pi(F,G)$ is the *Prokhorov metric* of $F(x)$ and $G(x)$
and $T_n$ is an empirical counterpart to the functional $T$.

- $\varepsilon^*$ is called *breakdown point*

(explanation of Prokhorov metric is on one of the next slides,
then finite sample breakdown point).

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

- An overall characteristic of the functional (the estimator) is

$\varepsilon^*$

It is not trivial to understand this definition -
- we shall try after some preparation.

where $\pi(F, G)$ is the *Prokhorov metric* of $F(x)$ and $G(x)$
and $T_n$ is an empirical counterpart to the functional $T$.

- $\varepsilon^*$ is called *breakdown point*.

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

An overall characteristic of the functional (the estimator) is

The definition contains new notions:

compact set and Prokhorov metric.

1. Firstly, what is a compact set? And what is good for ?
2. Secondly, an inspiration why we need Prokhorov metric.
3. Thirdly, an explanation what Prokhorov metric is.

$\varepsilon^*$ is called *breakdown point*.

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

- An overall characteristic of the functional (the estimator) is

Then we return to the definiton of breakdown point.

1. Firstly, we explain the sense of it.
2. Then we try to read the mathematics.

where $\pi(F, G)$ is the *Prokhorov metric* of $F(x)$ and $G(x)$
and $T_n$ is an empirical counterpart to the functional $T$.

- $\varepsilon^*$ is called *breakdown point*.

So, let us start with compact set

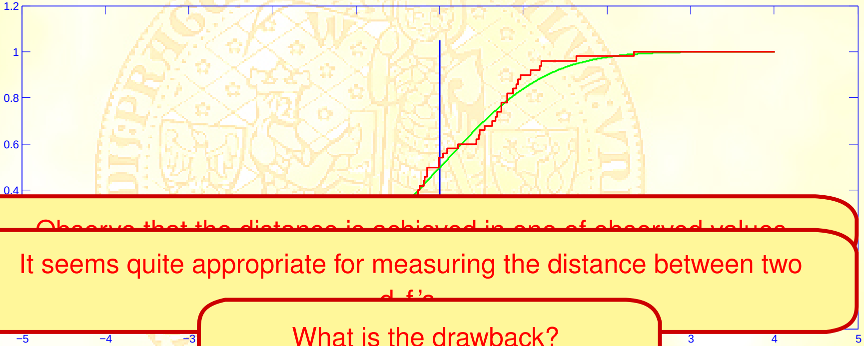
- Open and closed sets
  $C$ is closed set if for any sequence $\{x_n\}_{n=1}^{\infty} \subset C$ such that

$$\exists \left( \lim_{n\to\infty} x_n = x_0 \right) \quad \Rightarrow \quad x_0 \in C.$$

- $C$ is compact if it is closed and $\forall \ (x \in C) \quad \|x\| < K < \infty$.
- The sense of compact sets and the use of compactness:
  Any open cover contains a finite subcover.

## Recalling Kolmogorov-Smirnov metric

$$D(F,G) = \sup_{-\infty < x < \infty} |F(x) - G(x)|$$




Observe that the distance is achieved in one of observed values

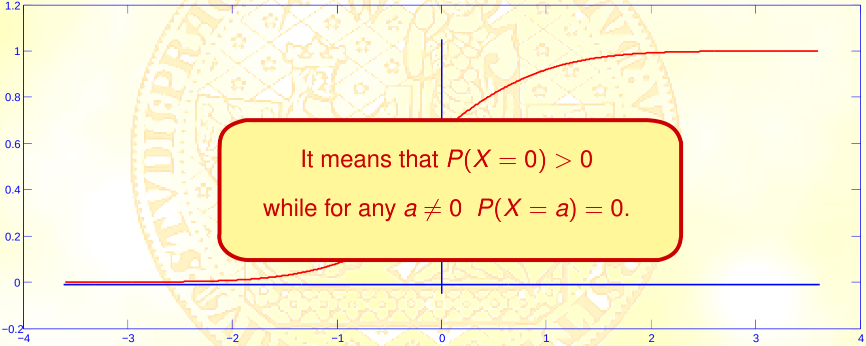
It seems quite appropriate for measuring the distance between two d.f.'s

What is the drawback?

(An answer on the next slides)

# Intuitive convergence of d. f.'s - but the Kolmogorov-Smirnov distance

Distribution function with a jump in the origine



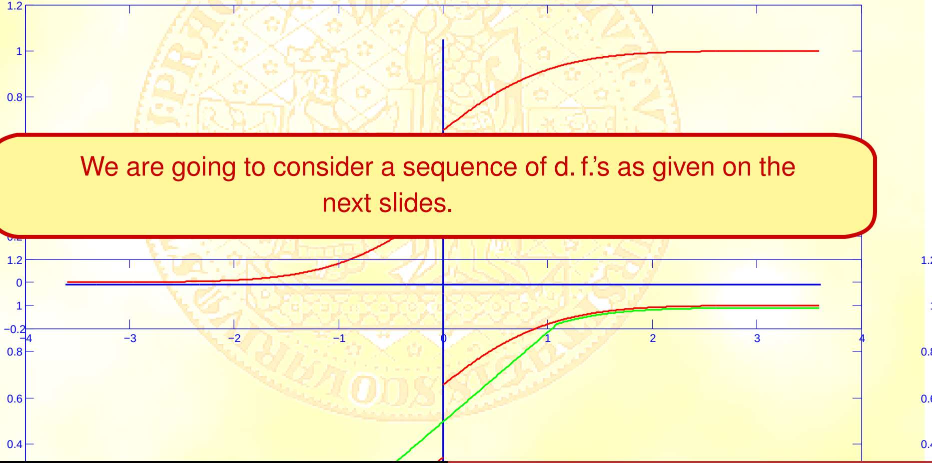

It means that $P(X = 0) > 0$

while for any $a \neq 0$ $P(X = a) = 0$.

# Intuitive convergence of d. f.'s - but the Kolmogorov-Smirnov distance

D. f. with a jump in the origineWidth of linear part 2.21.8

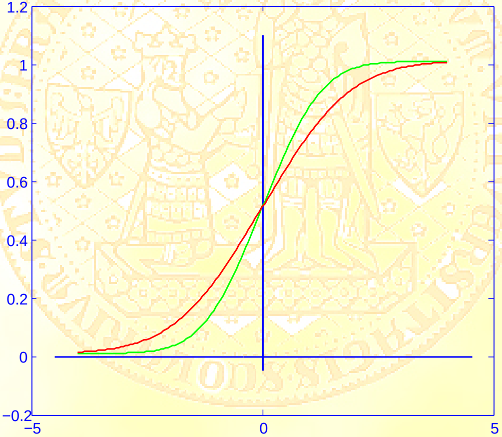
1.51.51.10.80.40.30.20.10.080.06.




We are going to consider a sequence of d. f.'s as given on the next slides.

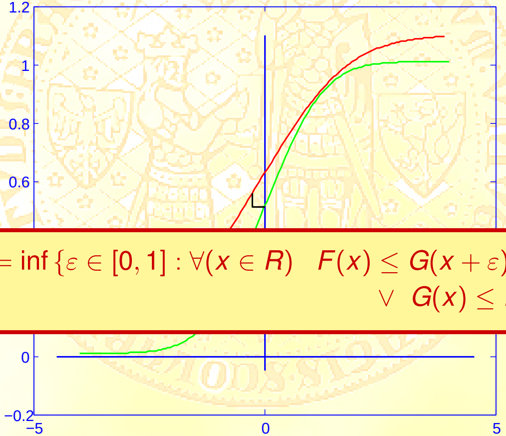
# Explaining Prohorov distance (notice different transcription)

Coinsider two d. f.'s



## Explaining Prokhorov distance

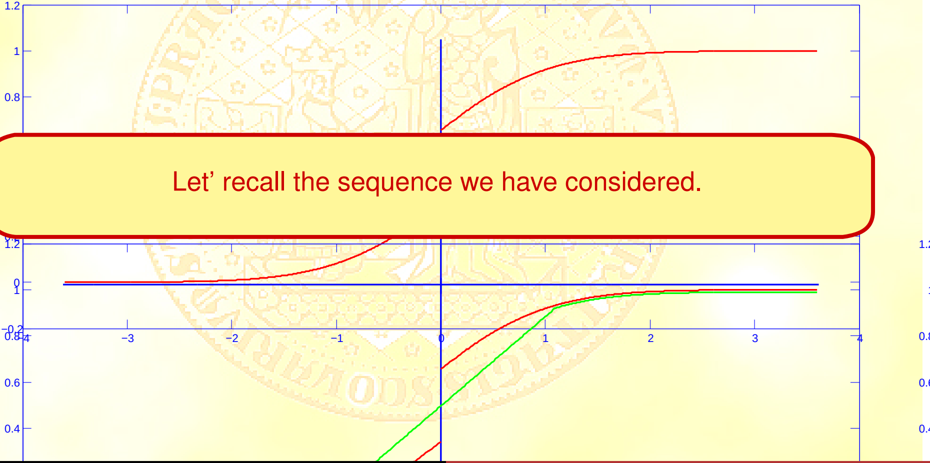
### Prokhorov metric




$$\pi(F,G) = \inf\{\varepsilon \in [0,1] : \forall(x \in R) \quad F(x) \leq G(x+\varepsilon)+\varepsilon \ \vee \ \vee \ G(x) \leq F(x+\varepsilon)+\varepsilon\}$$

## Comparing Kolmogorov-Smirnov and Prokhorov metrics

D. f. with a jump in the origineWidth of linear part 2.21.8 1.51.10.80.40.30.20.10.080.06.

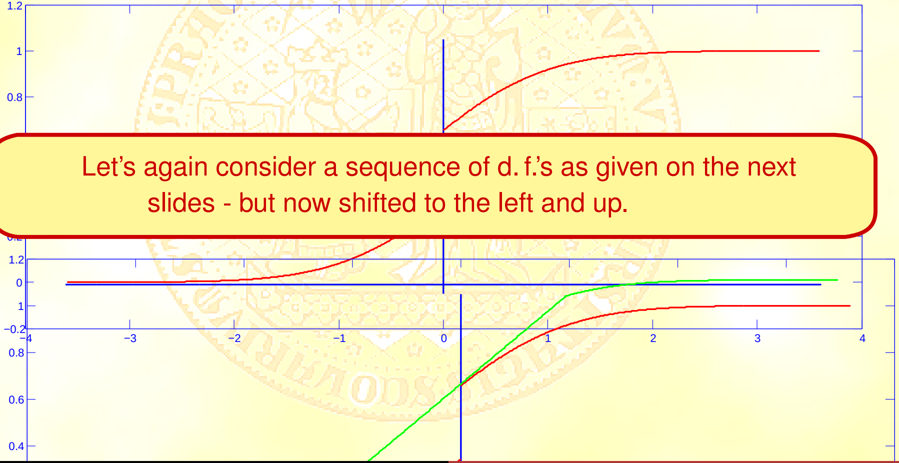



Let' recall the sequence we have considered.

# The convergence in Prokhorov metrics corresponds with intuitive idea of it.

D. f. with a jump in the origineWidth of linear part 2.21.8

1.51.10.80.40.30.20.10.080.06.




Let's again consider a sequence of d. f.'s as given on the next slides - but now shifted to the left and up.

## The global empirical characteristc of estimator.

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

- *Breakdown point* - "finite" sample version

$$x_1, x_2, ..., x_n \quad \Rightarrow \quad T_n(x_1, x_2, ..., x_n)$$

- Find maximal $m_n$ such that for any

$$y_1, y_2, ..., y_{m_n} \quad \Rightarrow \quad |T_n(x_1, x_2, ..., x_{n-m_n}, y_1, y_2, ..., y_{m_n})| < \infty$$

( $0 < T_n(x_1, x_2, ..., x_{n-m_n}, y_1, y_2, ..., y_{m_n}) < \infty$ - for scale ).

- Put

$$\varepsilon^* = \lim_{n\to\infty} \frac{m_n}{n}$$

(we'll return to it later on,

now let's return to the exact definition of the breakdown point).

## Let's read the definiton of breakdown point

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

- An overall characteristic of the functional (the estimator) is

$$\varepsilon^* = \sup\Big\{\varepsilon \leq 1 : \exists K_\varepsilon \subset \Theta, \quad K_\varepsilon \quad \textit{compact}$$

$$\pi(F, G) < \varepsilon \;\Rightarrow\; G(\{T_n \in K_\varepsilon\}) \underset{n\to\infty}{\longrightarrow} 1\Big\}$$

where $\pi(F, G)$ is the *Prokhorov metric* of $F(x)$ and $G(x)$
and $T_n$ is an empirical counterpart to the functional $T$.

- $\varepsilon^*$ is called *breakdown point*.

Let's rewrite the mathematical part of definition on the next slide.

## Let's read the definiton of breakdown point

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

$$\varepsilon^* = \sup\left\{\varepsilon \le 1 : \exists K_\varepsilon \subset \Theta, \quad K_\varepsilon \;\; \textit{compact} \quad \pi(F,G) < \varepsilon \;\Rightarrow\; G(\{T_n \in K_\varepsilon\}) \underset{n\to\infty}{\longrightarrow} 1\right\}$$

1. Let's assume that $T \in R$, so it is connected with a parameter of $F$.
2. $K_\varepsilon$ could be sufficiently wide interval - for this moment.
3. The definition says that $T_n$ is so "good" for estimating "$F$" that whenever $G$ is sufficiently close to $F$, $T_n$ converges in probability with respect to $G$ to something finite.

That's all.

## Specifying the “robustness” characteristics for the location parameter

At the beginning of the third lecture
we have computed influence function for $T(\Phi) = \boldsymbol{E}_\Phi Z$:

Recalling definition of influence function is :
Fix a functional $T : \mathcal{H} \to R$ .....

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

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

## Specifying the “robustness” characteristics for the location parameter

We easy verify that the same computation can be done whenever r. v. $Z$ has finite mean value $T(F) = \boldsymbol{E}_F Z = \mu \in R$:

Fix a functional $T : \mathcal{H} \rightarrow R$ ...

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

1. Fix $T(F) = \boldsymbol{E}_F(Z) = \int Z \mathrm{d}\Phi = \int z \cdot f(z) \mathrm{d}z$ .
2. $T\left(F(.)\right) = \int z \cdot f(z) \mathrm{d}z = \mu$
3. $T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right)$ $= \int z \left\{(1-\delta)f(x) + \delta \cdot \Delta_x\right\} \mathrm{d}z = (1-\delta) \cdot \mu + \delta \cdot x.$
4. Finally, $IF(x, T, F) = \lim_{\delta \to 0} \frac{\delta \cdot (-\mu + x)}{\delta} = -\mu + x.$

## Specifying the “robustness” characteristics for the location parameter

Hence the “robustness” characteristics of $T(F) = \boldsymbol{E}_F(X)$ are:

1. The gross error sensitivity $\gamma^* = \sup_{x \in R} |IF(x, T, F)| = \infty$.
2. The local-shift sensitivity $\lambda^* = \sup_{x,y \in R} \left| \frac{IF(x,T,F) - IF(y,T,F)}{x-y} \right| = 1$.
3. The rejection point $\rho^* = \inf \left\{ r \in R^+ \ : \ IF(x, T, F) = 0, \ \ |x| > r \right\} = \infty$.
4. The breakdown point $\varepsilon^* = 0$

(the last characteristic is “derived heuristically”
from the finite version of breakdown point).

## Specifying the “robustness” characteristics for the scale parameter

At the beginning of the third lecture
we have also computed influence function for $T(\Phi) = \boldsymbol{E}_\Phi X^2$:

The ~~Recalling again definition of influence function is~~ :
Fix a functional $T : \mathcal{H} \to R$ .....

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

1. Fix $T(\Phi) = \boldsymbol{E}_\Phi(X^2) = \int X^2 \mathrm{d}\Phi = \int z^2 \cdot \phi(z)\mathrm{d}z$ .
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, $IF(x, T, \Phi) = \lim_{\delta \to 0} \frac{(1-\delta)\cdot 1 + \delta \cdot x^2 - 1}{\delta} = -1 + x^2$.

## Specifying the "robustness" characteristics for the scale parameter

We easy verify that the same computation can be done whenever r. v. $Z$ has finite variance $T(F) = \boldsymbol{E}_F (Z - \boldsymbol{E}Z)^2 = \sigma^2 \in R^+$:

Fix a functional $T : \mathcal{H} \to R$ ...

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

---

1. Fix $T(F) = \boldsymbol{E}_F(Z - \boldsymbol{E}Z)^2 = \int (Z - \boldsymbol{E}Z)^2 \, \mathrm{d}F = \int (z - \boldsymbol{E}Z)^2 \cdot f(z)\mathrm{d}z$.
2. $T\left(F(.)\right) = \int (z - \boldsymbol{E}Z)^2 \cdot f(z)\mathrm{d}z = \sigma^2$
3. $T\left((1-\delta)F(.) + \delta \cdot \Delta_x\right)$
$= \int (z - \boldsymbol{E}Z)^2 \{(1-\delta)f(z) + \delta \cdot \Delta_x\} \mathrm{d}z = (1-\delta)\cdot\sigma^2 + \delta \cdot (x - \boldsymbol{E}Z)^2$.
4. Finally, $IF(x,T,F) = \lim_{\delta \to 0} \frac{\delta \cdot \left(-\sigma^2 + (x - \boldsymbol{E}Z)^2\right)}{\delta} = -\sigma^2 + (x - \boldsymbol{E}Z)^2$.

## Specifying the “robustness” characteristics for the scale parameter

Hence the “robustness” characteristics of $T(F) = \boldsymbol{E}_F(Z - \boldsymbol{E}Z)^2$ are:

1. The gross error sensitivity $\gamma^* = \sup_{x \in R} |IF(x, T, F)| = \infty$.
2. The local-shift sensitivity $\lambda^* = \sup_{x,y \in R} \left| \frac{IF(x,T,F) - IF(y,T,F)}{x-y} \right| = \infty$.
3. The rejection point $\rho^* = \inf \left\{ r \in R^+ \ : \ IF(x, T, F) = 0, \ \ |x| > r \right\} = \infty$.
4. The breakdown point $\varepsilon^* = 0$

(the last characteristic is again “derived heuristically”
from the finite version of breakdown point).

*THANKS FOR ATTENTION*