One-Dimensional Robust Estimators

Martin Branda

Stochastic modelling for economy and finance

12th October 2009
Obsah

1. The influence function
2. Robustness measures
3. Empirical influence function
4. The breakdown point and qualitative robustness
Obsah

1. The influence function
2. Robustness measures
3. Empirical influence function
4. The breakdown point and qualitative robustness
**ASS:** One-dimensional i.i.d. observations $X_1, \ldots, X_n$ from sample space $\mathcal{X} \subseteq \mathbb{R}$. 
**ASS:** One-dimensional i.i.d. observations $X_1, \ldots, X_n$ from sample space $\mathcal{X} \subseteq \mathbb{R}$.

**Parametric model** - family of probability distributions $F_\theta$ on the sample space $\mathcal{X}$, $\theta$ belongs to an *open convex* parameter space $\Theta \subseteq \mathbb{R}$. 
**ASS:** One-dimensional i.i.d. observations $X_1, \ldots, X_n$ from sample space $\mathcal{X} \subseteq \mathbb{R}$.

**Paramateric model** - family of probability distributions $F_\theta$ on the sample space $\mathcal{X}$, $\theta$ belongs to an *open convex* parameter space $\Theta \subseteq \mathbb{R}$.

**EX:** $\mathcal{X} = \{0, \ldots, N\}$, $\Theta = (0, 1)$, and $F_\theta$ is the binomial distribution with probability $\theta$ of success.
**ASS:** One-dimensional i.i.d. observations $X_1, \ldots, X_n$ from sample space $\mathcal{X} \subseteq \mathbb{R}$.

**Paramateric model** - family of probability distributions $F_\theta$ on the sample space $\mathcal{X}$, $\theta$ belongs to an *open convex* parameter space $\Theta \subseteq \mathbb{R}$.

**EX:** $\mathcal{X} = \{0, \ldots, N\}$, $\Theta = (0, 1)$, and $F_\theta$ is the binomial distribution with probability $\theta$ of success.

**ASS:** $\exists$ pdf $f_\theta$ with respect to a $\sigma-$finite measure $\lambda$ on $\mathcal{X}$.
Classical vs. robust statistics

\[ \{ F_\theta, \theta \in \Theta \} \]

We will investigate deviations from assumed distribution (its shape).
Empirical distribution $G_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq x)$ based on sample $X_1, \ldots, X_n$. 
Empirical distribution \( G_n(x) = 1/n \sum_{i=1}^{n} I(x_i \leq x) \) based on sample \( X_1, \ldots, X_n \).

Estimator of \( \theta \)

\[
T_n = T_n(X_1, \ldots, X_n) = T_n(G_n),
\]

i.e. \( \{ T_n, n \geq 1 \} \) sequence of estimators. IDEALLY, the observations are i.i.d. according to a member of the parametric model \( \{ F_\theta, \theta \in \Theta \} \), but the class \( \mathcal{F}(\mathcal{X}) \) of all possible p.d. on \( \mathcal{X} \) is much larger.
The influence function

Assymptotic estimator

**ASS:** $\exists$ P-lim

$$T_n(G_n) n \rightarrow \infty T(G)$$

Then $T(G)$ is the *asymptotic value* of $\{T_n, n \geq 1\}$ at the true distribution $G (X_1, \ldots, X_n \text{G–i.i.d.}):$
The influence function

**Assymptotic estimator**

**ASS:** \( \exists \) \( P \)-lim

\[
T_n(G_n) \xrightarrow{n \to \infty} T(G)
\]

Then \( T(G) \) is the *asymptotic value* of \( \{T_n, n \geq 1\} \) at the true distribution \( G(X_1, \ldots, X_n \text{ i.i.d.}) \):

\[
T : \text{domain } (T) \to \mathbb{R},
\]

\( \text{domain } (T) \subseteq \mathcal{F}(\mathcal{X}) \) for which the estimator is well defined.
The influence function

Assymptotic estimator

**ASS:** ∃ P-lim

\[ T_n(G_n) \xrightarrow{n \to \infty} T(G) \]

Then \( T(G) \) is the *asymptotic value* of \( \{ T_n, n \geq 1 \} \) at the true distribution \( G(X_1, \ldots, X_n, G \text{–i.i.d.}): \)

\[ T : \text{domain } (T) \to \mathbb{R}, \]

domain \( (T) \subseteq \mathcal{F}(\mathcal{X}) \) for which the estimator is well defined.

**ASS:** Asymptotic normality

\[ \mathcal{L}_G(\sqrt{n}[T_n - T(G)]) \xrightarrow{w} N(0, V(T, G)), \]

where \( V(T, G) \) is the asymptotic variance of \( \{ T_n, n \geq 1 \} \) at \( G \).
Fisher consistency

**ASS:** Fisher consistency

\[ T(F_\theta) = \theta, \quad \forall \theta \in \Theta. \]

\( \approx \) the model estimator asymptotically measures the right quantity.
The influence function

Fisher consistency

**ASS: Fisher consistency**

\[ T(F_\theta) = \theta, \quad \forall \theta \in \Theta. \]

\[ \approx \text{the model estimator asymptotically measures the right quantity.} \]

Consistency: \( P - \lim_{n \to \infty} T_n = \theta, \)

Asymptotic unbiasedness: \( \lim_{n \to \infty} E T_n = \theta. \)
Stochastic programming

\[ T(F_n) = \min_{x \in X} \frac{1}{n} \sum_{i=1}^{n} f(x, X_i), \]  
\[ T(F) = \min_{x \in X} \mathbb{E}_F[f(x, \omega)] \]

\( \emptyset \neq X \subseteq \mathbb{R}^n \text{ closed, } f : \mathbb{R}^n \times X \to \mathbb{R}. \)
The influence function

Contaminated distribution

\[ F, G \in \mathcal{F}(\mathcal{X}) \]

\[ (1 - t)F + tG, \quad t \in [(0, 1)]. \]  \hfill (3)
Gateaux (Frechet) differentiability

**ASS:** domain $(T)$ convex subset of the set of all finite signed measures on $\mathcal{X}$ containing more than one element.
The influence function

Gateaux (Frechet) differentiability

**ASS:** domain \((T)\) convex subset of the set of all finite signed measures on \(\mathcal{X}\) containing more than one element.

Then \(T\) is Gateaux differentiable at the distribution \(F\) in domain \((T)\), if there exists a real function \(a_1\) such that for all \(G \in \text{domain} (T)\) it holds that

\[
\lim_{t \to 0} \frac{T((1 - t)F + tG) - T(F)}{t} = \int_{\mathcal{X}} a_1(x) dG(x) \quad (4)
\]

and we can use the notion

\[
d/dt[T((1 - t)F + tG)]_{t=0}
\]

for the limit.
In the above situation, $T$ is called the *von Mises functional* with first kernel function $a_1$. 
In the above situation, $T$ is called the \textit{von Mises functional} with first kernel function $a_1$.

Setting $G = F$ we obtain

$$\int a_1(x) dF(x) = 0,$$

hence we may write

$$\int a_1(x) dG(x) = \int a_1(x) d(G - F)(x)$$
Influence function

\[ G = \Delta_x \] is the Dirac measure.

**Definition**

The influence function (IF) of \( T \) at \( F \) is given by

\[
IF(x; T, F) = \lim_{t \to 0^+} \frac{T((1 - t)F + t\Delta_x) - T(F)}{t}
\]  (5)

in those \( x \in \mathcal{X} \) where the limit exists.
The influence function

The influence function (IF) of $T$ at $F$ is given by

$$IF(x; T, F) = \lim_{t \to 0^+} \frac{T((1 - t)F + t\Delta_x) - T(F)}{t}$$

(5)

in those $x \in \mathcal{X}$ where the limit exists.

- IF describes the effect of an infinitesimal contamination at the point $x$ on the estimate $T$, standardized by the mass of the contamination.
- IF measures the asymptotic bias caused by contamination in the observations.
First-order von Mises expansion of $T$ at $F$ ($G$ is near $F$)

$$T(G) = T(F) + \int IF(x; T, F)d(G - F)(x) + \text{remainder.} \quad (6)$$

Using Glivenko-Cantelli $F_n \Rightarrow G$ we obtain the approximation

$$T_n(F_n) \approx T(F) + \int IF(x; T, F)dF_n(x) + \text{remainder,} \quad (7)$$

where we used

$$\int IF(x; T, F)dF(x) = 0. \quad (8)$$
Asymptotic variance

It yields

$$\sqrt{n}(T_n - T(F)) \approx \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(X_i; T, F) + \text{remainder}. \quad (9)$$

Then

$$\mathcal{L}_F(\sqrt{n}[T_n - T(F)]) \xrightarrow{w} N(0, V(T, F)), \quad (10)$$

where

$$V(T, F) = \int IF(x; T, F)^2 dF(x).$$
Asymptotic analysis

Two estimates \( \{ T_n, n \geq 1 \} \), \( \{ S_n, n \geq 1 \} \), the asymptotic relative efficiency

\[
ARE_{T,S} = \frac{V(T, F)}{V(S, F)}.
\]  

(11)

The Fisher information at \( F_{\theta^*} \) for some fixed \( \theta^* \in \Theta \)

\[
J(F_{\theta^*}) = \int \left( \frac{d}{d\theta} \ln f_{\theta}(x) \right)_{\theta=\theta^*}^2 dF_{\theta^*}(x).
\]  

(12)

Let \( 0 < J(F_{\theta^*}) < \infty \), using Fisher consistency we obtain

\[
\frac{d}{d\theta} \left[ \int IF(x; T, F_{\theta^*}) dF_{\theta}(x) \right]_{\theta=\theta^*} = \frac{d}{d\theta} [T(F_{\theta})]_{\theta=\theta^*} = \left[ \frac{d\theta}{d\theta} \right] = 1
\]  

(13)
Changing the order of differentiation and integration

\[
1 = \int IF(x; T, F_{\theta^*}) \frac{d}{d\theta} \left[ f_{\theta}(x) \right]_{\theta=\theta^*} d\lambda(x)
\]

\[
= \int IF(x; T, F_{\theta^*}) \frac{d}{d\theta} \left[ \ln f_{\theta}(x) \right]_{\theta=\theta^*} dF_{\theta^*}(x).
\]

Using Cauchy-Schwartz inequality

\[
(V(T, F_{\theta^*}) = \int IF(x; T, F_{\theta^*})^2 dF_{\theta^*}(x) \geq \frac{1}{J(F_{\theta^*})}.
\]

Equality holds iff \(IF(x; T, F_{\theta^*})\) is proportional to \(d/d\theta \left[ \ln f_{\theta}(x) \right]_{\theta=\theta^*}\) a.e.
Then, the estimator is asymptotically efficient iff

\[
IF(x; T, F_{\theta^*}) = \frac{1}{J(F_{\theta^*})} \cdot \frac{d}{d\theta} \left[ \ln f_{\theta}(x) \right]_{\theta=\theta^*}
\]  

(15)

**Asymptotic efficiency** of an estimator

\[
e = \frac{1}{V(T, F_{\theta^*}) \cdot J(F_{\theta^*})}.
\]  

(16)
Obsah

1. The influence function
2. Robustness measures
3. Empirical influence function
4. The breakdown point and qualitative robustness
Gross-error sensitivity of $T$ at $F$

$$\gamma^*(T, F) = \sup_x |IF(x; T, F)|,$$  \hspace{1cm} (17)

the supremum being taken over all $x$ where $IF(x; T, F)$ exists.
Gross-error sensitivity of $T$ at $F$

$$\gamma^*(T, F) = \sup_x |IF(x; T, F)|,$$  \hspace{1cm} (17)

the supremum being taken over all $x$ where $IF(x; T, F)$ exists.

It measures the worst influence which a small amount of contamination of fixed size can have on the value of the estimator.
B-robustness

The estimator $T$ is said to be **B-robust** iff $\gamma^*(T, F) < \infty$. (In conflict with asymptotic efficiency.)
B-robustness

The estimator $T$ is said to be \textbf{B-robust} iff $\gamma^*(T, F) < \infty$. (In conflict with asymptotic efficiency.)

The estimator $T$ is said to be \textbf{optimal B-robust} iff it can not be improved simultaneously with respect to $V(T, F)$ and $\gamma^*(T, F)$ ("biobjective minimization").
B-robustness

The estimator $T$ is said to be **B-robust** iff $\gamma^*(T, F) < \infty$. (In conflict with asymptotic efficiency.)

The estimator $T$ is said to be **optimal B-robust** iff it can not be improved simultaneously with respect to $V(T, F)$ and $\gamma^*(T, F)$ ("biobjective minimization").

**Most B-robust** estimators iff Fisher consistent and $0 \leq \gamma^*(T, F) < \infty$ (will be).
Local shift sensitivity of $T$ at $F$

\[ \lambda^*(T, F) = \sup_{x \neq y} \frac{|IF(x; T, F) - IF(y; T, F)|}{|x - y|}, \]  

which is the smallest Lipschitz constant the IF obeys.
Local shift sensitivity of $T$ at $F$

$$\lambda^*(T, F) = \sup_{x \neq y} \frac{|IF(x; T, F) - IF(y; T, F)|}{|x - y|}$$ (18)

which is the smallest Lipschitz constant the IF obeys.

The effect of shifting an observation slightly from a point $x$ to a neighboring point $y$. 
Rejection point of $T$ at symmetric $F$

\[
\rho^*(T, F) = \inf\{r > 0 : IF(x; T, F) = 0 \text{ when } |x| > r\}, \quad (19)
\]

taking $\inf\emptyset = \infty$. 

Rejection point of $T$ at symmetric $F$

$$\rho^*(T, F) = \inf\{r > 0 : IF(x; T, F) = 0 \text{ when } |x| > r\},$$  
(19)

taking $\inf\emptyset = \infty$.

IF vanishes outside a certain area, the contamination in those points have no influence on the estimator, hence the points can be rejected.
Example I: arithmetic mean

Standard normal cdf $F_\theta(x) = \Phi(x - \theta)$. The arithmetic (sample) mean $T_n = 1/n \sum_{i=1}^{n} X_i$ and corresponding functional $T(G) = \int udG(u)$ defined for all $G$ for which it is finite. $T$ is Fisher consistent.
Example I: arithmetic mean

Standard normal cdf $F_\theta(x) = \Phi(x - \theta)$. The arithmetic (sample) mean $T_n = 1/n \sum_{i=1}^{n} X_i$ and corresponding functional $T(G) = \int udG(u)$ defined for all $G$ for which it is finite. $T$ is Fisher consistent.

$$IF(x; T, \Phi) = \lim_{t \to 0^+} \frac{(1 - t) \int ud\Phi(u) + t \int ud\Delta_x(u) - \int ud\Phi(u)}{t}$$

$$= \lim_{t \to 0^+} \frac{tx}{t} = x.$$

Clearly

$$\int IF(x; T, \Phi)d\Phi(x) = 0,$$

$$V(T, \Phi) = \int IF(x; T, \Phi)^2d\Phi(x) = 1.$$
$J(\Phi) = 1$, so $\int IF(x; T, \Phi)^2 d\Phi(x) = J(\Phi)^{-1}$. Since it holds

$$IF(x; T, \Phi) = \frac{1}{J(\Phi)} \cdot \frac{d}{d\theta} \left[ \ln \phi(x) \right]_{\theta=\theta^*},$$

(20)

the estimator is asymptotically efficient.
\( J(\Phi) = 1\), so \( \int IF(x; T, \Phi)^2d\Phi(x) = J(\Phi)^{-1} \). Since it holds

\[
IF(x; T, \Phi) = \frac{1}{J(\Phi)} \cdot \frac{d}{d\theta} \left[ \ln \phi(x) \right]_{\theta = \theta^*}, (20)
\]

the estimator is asymptotically efficient.

- Gross-error sensitivity \( \gamma^* = \infty \),
- Local shift sensitivity \( \lambda^* = 1 \),
- Rejection point \( \rho^* = \infty \).

Mean is dangerous in any situation where outliers might occur.
Example II: median

\[ T_n = \begin{cases} 
  X_{\left(\frac{n+1}{2}\right)} & n \text{ even} \\
  \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+1}{2}\right)}}{2} & n \text{ odd}
\end{cases} \]

\[ T(G) = G^{-1}(1/2) \text{ (or midpoint of the interval } \{x, G(x) = 1/2\}). \text{ Fisher consistent.} \]
Example II: median

\[
T_n = \begin{cases} 
X\left(\frac{n+1}{2}\right) & n \text{ even} \\
\frac{X\left(\frac{n}{2}\right) + X\left(\frac{n+1}{2}\right)}{2} & n \text{ odd}
\end{cases}
\]

\[T(G) = G^{-1}(1/2)\] (or midpoint of the interval \(\{x, G(x) = 1/2\}\)). Fisher consistent.

Will be shown that

\[
IF(x; T, Phi) = \frac{\text{sign}(x)}{2\phi(0)}.
\]
Robustness measures

\[ \int IF(x; T, \Phi) d\Phi(x) = 0, \]

\[ V(T, \Phi) = \int IF(x; T, \Phi)^2 d\Phi(x) = (2\phi(0))^{-2} = \frac{\pi}{2}, \]

\[ e = \frac{2}{\pi}. \]
\[
\int IF(x; T, \Phi) d\Phi(x) = 0,
\]
\[
V(T, \Phi) = \int IF(x; T, \Phi)^2 d\Phi(x) = (2\phi(0))^{-2} = \pi/2,
\]
\[e = 2/\pi.\]

- Gross-error sensitivity \( \gamma^* = (2\phi(0))^{-1} \) (B-robustness),
- Local shift sensitivity \( \lambda^* = \infty \) (sensitivity to shifting near the center of symmetry),
- Rejection point \( \rho^* = \infty \).
Example III: Poisson model

$\mathcal{X} = \{0, 1, \ldots \}$, $\lambda$ is the counting measure, $\theta \in \Theta = (0, \infty)$,

$$f_\theta(k) = \frac{\theta^k}{k!} e^{-\theta}.$$ 

The maximum likelihood estimator for $\theta$ is the sample mean, i.e.

$$T_n = \frac{1}{n} \sum_{i=1}^n X_i$$

and corresponding functional

$$T(F) = \int udF(u) = \sum_{k=0}^{\infty} kf(k).$$

Fisher consistency $T(F_\theta) = \theta$. 

$M. \text{ Branda (MFF UK)}$
Let $x \in \{0, 1, \ldots \}$ and $\theta^* \in \Theta$ be fixed

$$
\text{IF}(x; T, F_{\theta^*}) = \lim_{t \to 0^+} \frac{\sum_{k=0}^{\infty} k[(1 - t)f_{\theta^*}(k) + tI_x(k)] - \sum_{k=0}^{\infty} kf_{\theta^*}}{t} \\
= \lim_{t \to 0^+} \frac{t \sum_{k=0}^{\infty} kl_x(k) - t \sum_{k=0}^{\infty} kf_{\theta^*}(k)}{t} \\
= x - \theta^*.
$$
\[
\int IF(x; T, F_{\theta^*})dF_{\theta^*}(x) = \sum_{k=0}^{\infty} (k - \theta^*) f_{\theta^*}(k) = 0,
\]

\[
V(T, F_{\theta^*}) = \int IF(x; T, F_{\theta^*})^2dF_{\theta^*}(x) = \sum_{k=0}^{\infty} (k - \theta^*)^2 f_{\theta^*}(k) = \theta^*,
\]

Since \( J(F_{\theta^*}) = (\theta^*)^{-1} \), the asymptotic efficiency is \( e = 1 \).
\[
\int IF(x; T, F_{\theta^*}) dF_{\theta^*}(x) = \sum_{k=0}^{\infty} (k - \theta^*) f_{\theta^*}(k) = 0,
\]

\[
V(T, F_{\theta^*}) = \int IF(x; T, F_{\theta^*})^2 dF_{\theta^*}(x) = \sum_{k=0}^{\infty} (k - \theta^*)^2 f_{\theta^*}(k) = \theta^*,
\]

Since \( J(F_{\theta^*}) = (\theta^*)^{-1} \), the asymptotic efficiency is \( e = 1 \).

- Gross-error sensitivity \( \gamma^* = \infty \) (no B-robustness),
- Local shift sensitivity \( \lambda^* = 1 \),
- Rejection point \( \rho^* = ??\infty \) (asymmetric).
Obsah

1. The influence function
2. Robustness measures
3. Empirical influence function
4. The breakdown point and qualitative robustness
Empirical influence function

\{ T_n, n \geq 1 \}, a sample \{ x_1, \ldots, x_{n-1} \}, then

addition version

\[ IF^1(x; T) = T_n(x_1, \ldots, x_{n-1}, x), \]

replacement version

\[ IF^2(x; T) = T_n(x_1, \ldots, x_n) \text{ for } x = x_n. \]
Tukey’s sensitivity curve

Sensitivity curve

\[ SC_n(x) = \frac{T_n(x_1, \ldots, x_{n-1}, x) - T_{n-1}(x_1, \ldots, x_{n-1})}{1/n}. \]

Sensitivity curve for functional estimator based on empirical distribution functions \( \{F_n, n \geq 1\} \) (contamination size \( t = 1/n \))

\[ SC_n(x) = n \left[ T \left( \left( 1 - \frac{1}{n} \right) F_{n-1} + \frac{1}{n} \Delta x \right) - T(F_{n-1}) \right]. \]
Tukey’s sensitivity curve

\[ SC_n(x) = \frac{T_n(x_1, \ldots, x_{n-1}, x) - T_{n-1}(x_1, \ldots, x_{n-1})}{1/n} . \]

**Sensitivity curve** for functional estimator based on empirical distribution functions \( \{ F_n, n \geq 1 \} \) (contamination size \( t = 1/n \))

\[ SC_n(x) = n \left[ T \left( \left( 1 - \frac{1}{n} \right) F_{n-1} + \frac{1}{n} \Delta_x \right) - T(F_{n-1}) \right] . \]

Under some assumptions \( SC_n(x) \to IF(x; T, F) \) as \( n \to \infty \).
Jacknife estimator

The \( i \)th jacknife pseudo-value is defined by

\[
T_{ni}^* = nT_n(x_1, \ldots, x_n) - (n - 1)T_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)
\]

E.g. for arithmetic mean \( T_{ni}^* = X_i \).
Jacknife estimator

The $i$th jacknife pseudovalue is defined by

$$T^*_{n_i} = nT_n(x_1, \ldots, x_n) - (n - 1)T_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

E.g. for arithmetic mean $T^*_{n_i} = X_i$.

We get a pseudosample $(T^*_{n_1}, \ldots, T^*_{nn})$, which we use to compute a corrected estimate

$$T^*_n = \sum_{i=1}^{n} T^*_{n_i}$$

(often less biased than $T_n$).
The variance

\[ V_n = \frac{1}{n - 1} \sum_{i=1}^{n} (T_{ni}^* - T_n^*)^2 \approx V(T, F). \]
Obsah

1. The influence function
2. Robustness measures
3. Empirical influence function
4. The breakdown point and qualitative robustness
Previous concept - *local stability*. Below we will study **GLOBAL reliability** of an estimator.
Previous concept - *local stability*. Below we will study **GLOBAL reliability** of an estimator.

Three main reasons why parametric model does not hold exactly:

- Rounding of the observations.
- Occurrence of gross errors (i.e. fraction of the data may be very differently distributed).
- Our model is only "idealized" approximation of reality.
The breakdown point and qualitative robustness

Prohorov distance

**Prohorov distance** of two probability distributions $F$ and $G$ in $\mathcal{F}(\mathcal{X})$ ($P_F$ denotes the probability measure that corresponds to the distribution function $F \in \mathcal{F}(\mathcal{X})$):

$$\pi(P_F, P_G) = \pi(F, G) = \inf\{\varepsilon : P_F(A) \leq P_G(A^\varepsilon) + \varepsilon \text{ for all events } A\},$$

where $A^\varepsilon = \{x \in \mathbb{R} : d(A, x) < \varepsilon\}$. 
Prohorov distance of two probability distributions $F$ and $G$ in $\mathcal{F}(\mathcal{X})$ ($P_F$ denotes the probability measure that corresponds to the distribution function $F \in \mathcal{F}(\mathcal{X})$):

$$\pi(P_F, P_G) = \pi(F, G) = \inf\{\varepsilon : P_F(A) \leq P_G(A^\varepsilon) + \varepsilon \text{ for all events } A\},$$

where $A^\varepsilon = \{x \in \mathbb{R} : d(A, x) < \varepsilon\}$.

The distance formalizes previous "three points" (round error, different distribution, weak convergence).
Breakdown point

Definition

The breakdown point $\varepsilon^*$ of the sequence of estimators $\{T_n, n \geq 1\}$ at $F$ is defined by $\varepsilon^* = \varepsilon^*(\{T_n, n \geq 1\}, F) = \sup\{\varepsilon \leq 1 : \text{there is a compact set } K_\varepsilon \subset \Theta \text{ such that } \pi(P_F, P_G) = \pi(F, G) < \varepsilon \text{ implies } P_G(T_n \in K_\varepsilon) \to 1 \text{ as } n \to \infty\}$.

It provides some guidance up to what distance from the model the linear approximation provided by IF can be used (will be more).
Other distances

The **gross-error breakdown point**: instead of \( \pi(F, G) < \varepsilon \) we consider the family of contaminated distributions

\[
G \in \{(1 - \varepsilon)F + \varepsilon H : H \in \mathcal{F}(X)\}.
\]
Finite-sample breakdown point

Definition

The finite-sample breakdown point $\varepsilon^*$ of the sequence of the estimators $T_n$ at the sample $(x_1, \ldots, x_n)$ is given by

$$
\varepsilon_n^* = \varepsilon^*(T_n, x_1, \ldots, x_n) = \frac{1}{n} \max \{ m : \\
\max \sup_{\{i_1, \ldots, i_{n-m}\} \subset \{1, \ldots, n\}} |T_n(z_1, \ldots, z_n)| < \infty \},
$$

where the sample $(z_1, \ldots, z_n)$ is obtained by replacing $m$ data points by the arbitrary values $y_1, \ldots, y_m (\in \mathcal{X})$, i.e.

$$(z_1, \ldots, z_n) = (x_{i_1}, \ldots, x_{i_{n-m}}, y_1, \ldots, y_m).$$
Remarks

- Usually it does not depend on \((x_1, \ldots, x_n)\) and \(\varepsilon_n^* \to \varepsilon^*\).
- For arithmetic mean \(\varepsilon_n^* = 0\).
- For location parameters \(0 < |T_n(z_1, \ldots, z_n)| < \infty\).
Qualitative robustness

Definition

We say that a sequence of estimators \( \{ T_n, n \geq 1 \} \) is **qualitatively robust** at \( F \) if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that for all \( G \in \mathcal{F}(X) \) and for all \( n \)

\[
\pi(F, G) < \delta \Rightarrow \pi(\mathcal{L}_F(T_n), \mathcal{L}_G(T_n)) < \varepsilon.
\]
Continuity

Definition

We say that a sequence of estimators \( \{ T_n, n \geq 1 \} \) is \textbf{continuous} at \( F \) if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) and \( n_0 \) such that for all \( m, n \geq n_0 \) and for all empirical cdf \( F_n, F_m \)

\[
\left( \pi(F, F_n) < \delta \  \& \ \pi(F, F_m) < \delta \right) \Rightarrow |T_n(F_n) - T_m(F_m)| < \varepsilon.
\]
Theorem

A sequence of estimators \( \{T_n, n \geq 1\} \) which is continuous at \( F \) and for which all \( T_n \) are continuous functions of the observation, is qualitative robust.
Relations

Theorem

In case the estimators are generated by a functional $T$, i.e. $T_n(F_n) = T(F_n)$, then $T$ is continuous with respect to the Prohorov distance at all $F$ if and only if $\{T_n, n \geq 1\}$ is qualitatively robust at all $F$ and satisfies $\pi(L_F(T_n), \Delta_{T(F)}) \to 0$ as $n \to \infty$ for all $F$. 
The breakdown point and qualitative robustness

Relations

Example

- The arithmetic mean is nowhere qualitatively robust and nowhere continuous, with $\varepsilon^* = 0$.
- The median is QR and C at $F$ if $F^{-1}(1/2)$ contains only one point, always $\varepsilon^* = 1/2$. 

Reference


Thank you for your attention.