# 3: Measuring Robustness

March 15, 2022 8:30 PM

- Sensitivity curve
  $$\widehat{\mu}(x_1, \ldots, x_n, x_0) - \widehat{\mu}(x_1, \ldots, x_n)$$

**3.1 Influence Function**

- Influence function

  Approximation to the behavior of estimator $\widehat{\theta}$ when the sample contains a small fraction of $\epsilon$ outliers

  $$\mathrm{IF}_{\widehat{\theta}}(x_0, F) = \lim_{\varepsilon \downarrow 0} \frac{\widehat{\theta}_\infty((1-\varepsilon)F + \varepsilon\delta_{x_0}) - \widehat{\theta}_\infty(F)}{\varepsilon}$$
  $$= \frac{\partial}{\partial\varepsilon} \widehat{\theta}_\infty((1-\varepsilon)F + \varepsilon\delta_0)\Big|_{\varepsilon\downarrow 0},$$

  where $\delta_{x_0}$ is the point-mass at $x_0$ and "↓" stands for "limit from the right"

  where
  - $\widehat{\theta}_\infty((1-\varepsilon)F + \varepsilon\delta_{x_0})$

    is the asymptotic value of the estimator when the underlying distribution is F and a fraction $\epsilon$ of outliers is equal to $x_0$

    $$\widehat{\theta}_\infty((1-\varepsilon)F + \varepsilon\delta_{x_0}) \approx \widehat{\theta}_\infty(F) + \varepsilon \mathrm{IF}_{\widehat{\theta}}(x_0, F)$$

- Standardized sensitivity curve
  $$\mathrm{SC}_n(x_0) = \frac{\widehat{\theta}_{n+1}(x_1, \ldots, x_n, x_0) - \widehat{\theta}_n(x_1, \ldots, x_n)}{1/(n+1)},$$
  $$= (n+1)\left(\widehat{\theta}_{n+1}(x_1, \ldots, x_n, x_0) - \widehat{\theta}_n(x_1, \ldots, x_n)\right)$$
  - just replace epsilon with 1/(n+1)
- Result: 1/(n+1) -> 0
  $$\mathrm{SC}_n(x_0) \to_{a.s.} \mathrm{IF}_{\widehat{\theta}}(x_0, F),$$
- M-estimator for location influence function
  $$\mathrm{IF}_{\widehat{\mu}}(x_0, F) = \frac{\psi(x_0 - \widehat{\mu}_\infty)}{\mathrm{E}\psi'(x - \widehat{\mu}_\infty)}$$
- M-estimator for scale incfluence function
  $$\mathrm{IF}_{\widehat{\sigma}}(x_0, F) = \widehat{\sigma}_\infty \frac{\rho(x_0/\widehat{\sigma}_\infty) - \delta}{\mathrm{E}(x/\widehat{\sigma}_\infty)\rho'(x/\widehat{\sigma}_\infty)}.$$
- General M-estimator theta hat is the solution of
  $$\sum^n \Psi(x_i, \widehat{\theta}) = 0$$
- General M-estimator influence function

$$\mathrm{IF}_{\hat{\theta}}(x_0, F) = -\frac{\Psi(x_0, \hat{\theta}_\infty)}{B(\hat{\theta}_\infty, \Psi)}$$

$$B(\theta, \Psi) = \frac{\partial}{\partial\theta}\mathrm{E}\Psi(x, \theta),$$

or

$$\mathrm{IF}(x; T, F_\theta) = \frac{\Psi(x, \theta)}{\int \frac{\partial}{\partial\theta}\Psi(y, \theta) dF_\theta(y)}.$$

**3.2 Breakdown Point**

- Definition

> **Definition 3.1** *The asymptotic contamination BP of the estimator $\hat{\theta}$ at $F$, denoted by $\varepsilon^*(\hat{\theta}, F)$, is the largest $\varepsilon^* \in (0, 1)$ such that for $\varepsilon < \varepsilon^*$, $\hat{\theta}_\infty((1-\varepsilon)F + \varepsilon G)$ remains bounded away from the boundary of $\Theta$ for all $G$.*

- Intuition: the breakdown point (BP) of an estimator $\hat{\theta}$ of the parameter $\theta$ is the largest amount of contamination (proportion of atypical points) that the data may contain such that $\hat{\theta}$ still gives some information about $\theta$

1. Location M-estimators

    For monotonic $\psi$, let

    $k_1 = -\psi(-\infty),\ k_2 = \psi(\infty)$

    Then BP is $\epsilon_1^*$ and $\epsilon_2^*$for $+\infty$ and $-\infty$ correspondingly

    $$\varepsilon_j^* = \frac{k_j}{k_1 + k_2}\ (j = 1, 2)$$

    For redescending

2. Scale and dispersion estimators
    - BP for
        - SDL: 0
        - MAD: 1/2
        - IQR: 1/4

**3.3 Maximum Asymptotic Bias**

- Maximum Bias (MS)

    $$\mathrm{MB}_{\hat{\theta}}(\varepsilon, \theta) = \max\{|\mathrm{b}_{\hat{\theta}}(F, \theta)| : F \in \mathcal{F}_{\varepsilon,\theta}\}$$

    where b() is bias

- Contamination sensitivity of $\hat{\theta}$

    $$\gamma_c(\hat{\theta}, \theta) = \left[\frac{d}{d\varepsilon}\mathrm{MB}_{\hat{\theta}}(\varepsilon, \theta)\right]_{\varepsilon=0}$$

**3.4 Balancing robustness and efficiency**

- we recommend when estimating location the bisquare M-estimator with previously computed MAD

**3.5 Optimal robustness**

1. Bias and variance optimality of location estimators
    - Minimax bias: median has smallest maximum bias among all shift equivariant estimators
    - Minimax variance
2. Bias optimality of scale and dispersion estimators
    - For simultaneous estimation of location and scale/dispersion with the monotone $\psi$-

function, the minimax bias estimator is well approximated by the MAD for all $\varepsilon < 0.5$
    - For M-estimators of scale with a general location estimator that includes location M-estimators with redescending $\psi$-functions, the minimax bias estimator is well approximated by the Shorth dispersion estimator (the shortest half of the data, see Problem 2.16b) for a wide range of $\varepsilon < 0.5$
3. Infinitesimal approach
4. Hampel approach
5. Balancing bias and variance

**3.6 Multidimensional parameters**

- Let $\theta = (\mu, \sigma)$ then

$$\Psi_1(x, \boldsymbol{\theta}) = \psi\left(\frac{x-\mu}{\sigma}\right) \text{ and } \Psi_2(x, \boldsymbol{\theta}) = \rho_{\text{scale}}\left(\frac{x-\mu}{\sigma}\right) - \delta.$$

and the estimators satisfy

$$\sum_{i=1}^{n} \boldsymbol{\Psi}(x_i, \widehat{\boldsymbol{\theta}}) = \mathbf{0},$$

- Generalized IF of $\widehat{\theta}$

$$\mathrm{IF}_{\widehat{\theta}}(x_0, F) = -\mathbf{B}^{-1}\boldsymbol{\Psi}(x_0, \widehat{\boldsymbol{\theta}}_\infty),$$

where

$$B_{jk} = \mathrm{E}\left\{ \left. \frac{\partial \Psi_j(x, \boldsymbol{\theta})}{\partial \theta_k} \right|_{\theta=\theta_F} \right\}.$$

- Multidimensional M-estimators are asymptotically normal with covariance

$$\mathbf{V} = \mathbf{B}^{-1}(\mathrm{E}\boldsymbol{\Psi}(x_0, \boldsymbol{\theta})\boldsymbol{\Psi}(x_0, \boldsymbol{\theta})')\mathbf{B}^{-1'},$$

**3.7 Estimators as Functionals**

- A "function" T whose argument is a distribution (a *functional)* is

$$T(F) = \mathrm{E}_F x = \int x dF(x).$$

- Mean

  It follows that $T(\widehat{F}_n) = \bar{x}$. If $\mathbf{x}$ is an i.i.d. sample from $F$, the law of large numbers implies that $T(\widehat{F}_n) \to_p T(F)$ when $n \to \infty$.
- M-estimators

  $\mathrm{T}(F)$is the solution $\theta$ of

$$\mathrm{E}_F \Psi(x, \theta) = 0.$$

  Then $T(\widehat{F}_n)$(empirical) is a solution of

$$\mathrm{E}_{\widehat{F}_n} \Psi(x, \theta) = \frac{1}{n} \sum_{i=1}^{n} \Psi(x_i, \theta) = 0.$$

- Qualitative robustness: an estimator corresponding to a functional T is said to be qualitatively robust at F if T is continuous at F according to the metric d; that is, for all $\varepsilon$ there exists $\delta$ such that d(F, G) < $\delta$ implies |T(F) − T(G)| < $\varepsilon$.