Robustness of Statistical Estimators:
Influence Functions, Sensitivity, and Efficiency

1 Robustness: What and Why

Classical estimators (mean, least squares, MLE) are often optimal under idealized models but can behave poorly under small deviations such as outliers or heavy tails.

Definition 1.1 (Robustness)**.** Robustness refers to the stability of a statistical procedure under small departures from modeling assumptions.

We ask:

> How much can a single observation, or a small fraction of the data, distort an estimator?

2 Sensitivity Curve: R Examples

The sensitivity curve is a finite-sample diagnostic that approximates the influence function by replacing one observation with an outlier.

Here, we compare three different estimates of the 'location' of the data. Thus, our interest is to know the location of the data, but we statistically quantify it using three different estimators, namely, mean, median, and Hodges-Legmann (HL).

2.1 Sensitivity curve for the mean

Consider a clean sample from a standard normal distribution. We replace one observation by a value z and track the effect on the sample mean.

```
set.seed(1)
n <- 100
x <- rnorm(n)

z.grid <- seq(-10, 10, length = 200)
sc.mean <- sapply(z.grid, function(z) {
  x.new <- x
  x.new[1] <- z
  n * (mean(x.new) - mean(x))
})

plot(z.grid, sc.mean, type = "l",
     main = "Sensitivity␣Curve:␣Sample␣Mean",
     xlab = "Contamination␣point␣z",
     ylab = "SC_n(z)")
abline(h = 0, lty = 2)
```

The linear growth confirms that the mean has an *unbounded* influence function.

2.2 Sensitivity curve for the median

```
sc.median <- sapply(z.grid, function(z) {
  x.new <- x
  x.new[1] <- z
  n * (median(x.new) - median(x))
})

plot(z.grid, sc.median, type = "l",
     main = "Sensitivity␣Curve:␣Sample␣Median",
     xlab = "Contamination␣point␣z",
     ylab = "SC_n(z)")
abline(h = 0, lty = 2)
```

The bounded curve reflects the robustness of the median.

Hodges–Lehmann Estimator Another method to compute the location of the data is the Hodges-Lehmann (HL).

Motivation: Rank-based tests such as the Wilcoxon signed-rank test and the Wilcoxon rank-sum (Mann–Whitney) test are widely used for inference on location parameters under minimal distributional assumptions. The Hodges–Lehmann (HL) estimator provides a *point estimator* of the underlying location shift that is naturally associated with these rank-based tests.

Key idea: Replace the mean (which is sensitive to outliers) by a *median of pairwise averages or differences*, yielding a robust, interpretable estimator of location.

One-sample Hodges–Lehmann estimator

Let $X_1, \ldots, X_n$ be i.i.d. observations from a continuous distribution symmetric about a location parameter θ. The one-sample HL estimator of θ is defined as

$$\hat{\theta}_{\mathrm{HL}} = \operatorname{median}\left\{ \frac{X_i + X_j}{2} : 1 \le i \le j \le n \right\}.$$

Remark 2.1. The set includes all pairwise averages, including X_i itself (when $i = j$). Taking the median rather than the mean ensures robustness to extreme observations.

Example (Sensitivity curve for HL)

```
HL_z_n_plus_1 = apply(z_n_plus_1_df, 1 , function(x){
  x = c(z_n, x)
  hl.loc(x)
})
sensitivity_HL = sensitivity(HL_z_n_plus_1,
                             mean(z_n), length(z_n))
```

Example (Sensitivity curve plot)

```
library(tidyr)
library(ggplot2)

df = data.frame(z_n_plus_1 = z_n_plus_1_df$z_n_plus_1,
                sensitivity_mean = sensitivity_mean,
                sensitivity_median = sensitivity_median,
```

```
               sensitivity_HL = sensitivity_HL)

df_long = gather(df, key = "estimator", value = "value", -z_n_plus_1)
df_long$estimator = factor(df_long$estimator)

ggplot(data = df_long) +
  geom_line(aes(x = z_n_plus_1, y = value,
               group = estimator, color = estimator)) +
  xlab(bquote(z[n+1])) +
  scale_color_discrete(name = "Estimator",
                      labels = c("HL", "Mean", "Median")) +
  ylab(bquote(S(z[n+1])))
```

3 Statistical Functionals

An estimator can be viewed as a functional of the data-generating distribution.

Definition 3.1 (Statistical Functional)**.** A statistical functional is a mapping

$$T : \mathcal{F} \to \mathbb{R},$$

where $\mathcal{F}$ is a class of probability distributions.

Examples:

- Mean: $T(F) = \int x \, dF(x)$
- Median: $T(F) = F^{-1}(1/2)$
- Variance: $T(F) = \int (x - \mu)^2 \, dF(x)$
- Regression coefficient as a functional of joint (X, Y) distribution

Given data $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$, the empirical distribution

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

induces the plug-in estimator $T(\hat{F}_n)$.

Example: If T is mean, $T(\hat{F}_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$.

4 Contamination Models

To formalize robustness, we perturb F.

Definition 4.1 (Gross-Error Contamination)**.** For $0 < \varepsilon \ll 1$ and arbitrary G,

$$F_\varepsilon = (1 - \varepsilon)F + \varepsilon G.$$

A fundamental special case is point-mass contamination.

Definition 4.2 (Point-Mass Contamination)**.** For $z \in \mathbb{R}$,

$$F_\varepsilon = (1 - \varepsilon)F + \varepsilon \delta_z.$$

This models the effect of inserting an infinitesimal outlier at z.

5 Sensitivity Curve

Definition 5.1 (Sensitivity Curve)**.** The sensitivity curve of T at sample size n is

$$\mathrm{SC}_n(z) = n\big(T(\hat{F}_{n,z}) - T(\hat{F}_n)\big),$$

where $\hat{F}_{n,z}$ is the empirical distribution with one observation replaced by z.

Remark 5.1. The sensitivity curve is the finite-sample analogue of the influence function.

As $n \to \infty$,

$$\mathrm{SC}_n(z) \longrightarrow \mathrm{IF}(z; T, F).$$

6 Influence Function

Definition 6.1 (Influence Function)**.** The influence function of T at F is

$$\mathrm{IF}(z; T, F) = \lim_{\varepsilon \to 0} \frac{T((1-\varepsilon)F + \varepsilon\delta_z) - T(F)}{\varepsilon} == \lim_{\varepsilon \to 0} \frac{T(F + \varepsilon(\delta_z - F)) - T(F)}{\varepsilon}.$$

This is technically valid only for discrete data. But we can always pretend that the observed dataset is the support (or the possible values of the random variable) and hence discrete.

Remark 6.1. It is the Gateaux derivative of T at F in the direction $\delta_z - F$.

7 Examples of Influence Functions

7.1 Mean

Here $T(F) = \int x dF(x)$.

$$\mathrm{IF}(z; \mathrm{mean}, F) = z - \mu.$$

Unbounded $\Rightarrow$ non-robust.

7.2 Median

If $f(\theta) > 0$,

$$\mathrm{IF}(z; \mathrm{median}, F) = \frac{1}{2f(\theta)} \left(\overset{\mathrm{ind}}{\sim} (z \leq \theta) - \frac{1}{2} \right).$$

Bounded $\Rightarrow$ robust.

7.3 M-estimators

Let θ solve $\mathbb{E}[\psi(X - \theta)] = 0$. Then

$$\mathrm{IF}(z) = \frac{\psi(z - \theta)}{\mathbb{E}[\psi'(X - \theta)]}.$$

Bounded ψ implies bounded influence.

8 Influence Functions and Plug-in Standard Errors

Theorem 8.1 (Asymptotic Linearization)**.** *If T is smooth,*

$$T(\hat{F}_n) - T(F) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(X_i; T, F) + o_p(n^{-1/2}).$$

Hence,

$$\mathrm{Var}(T(\hat{F}_n)) \approx \frac{1}{n}\mathbb{E}[\mathrm{IF}(X; T, F)^2].$$

Remark 8.1. This yields a *plug-in standard error* estimator:

$$\widehat{\mathrm{SE}}^2 = \frac{1}{n^2}\sum_{i=1}^{n} \widehat{\mathrm{IF}}(X_i)^2.$$

9 Breakdown Point

Definition 9.1 (Finite-Sample Breakdown Point)**.** The breakdown point is the smallest fraction of contamination that can cause the estimator to take arbitrarily large values.

Remark 9.1. In finite samples, the breakdown point depends on n. For example, the finite-sample breakdown point of the sample median is $\lfloor (n-1)/2 \rfloor / n$, which converges to 0.5 as $n \to \infty$, whereas the finite-sample breakdown point of the sample mean is $1/n$, which converges to 0.

Remark 9.2. The following are *asymptotic breakdown points*, i.e., limits of the finite-sample breakdown point as the sample size $n \to \infty$:

- Sample mean: asymptotic breakdown point = 0
- Sample median: asymptotic breakdown point = 0.5
- Least squares estimator: asymptotic breakdown point = 0

Remark 9.3. Breakdown point measures *global robustness*, while influence functions measure *local robustness*.

Breakdown point is a measure of *global robustness* that quantifies the smallest amount of contamination capable of causing an estimator to fail catastrophically.

9.1 Finite-sample breakdown point

Let $T_n = T(x_1, \ldots, x_n)$ be an estimator based on a sample $\boldsymbol{x} = (x_1, \ldots, x_n)$.

Definition 9.2 (Finite-sample breakdown point)**.** The finite-sample breakdown point of T_n at $\boldsymbol{x}$ is defined as

$$\varepsilon_n^*(T_n, \boldsymbol{x}) = \frac{1}{n}\min\left\{ m : \sup_{\boldsymbol{z}} \left\| T(x_1', \ldots, x_n') \right\| = \infty \right\},$$

where $(x_1', \ldots, x_n')$ is obtained from $(x_1, \ldots, x_n)$ by replacing *any* m observations by arbitrary values $\boldsymbol{z} = (z_1, \ldots, z_m)$.

Remark 9.4. The norm $\|\cdot\|$ reflects divergence of the estimator (e.g., absolute value for location parameters or Euclidean norm for vector-valued estimators).

9.2 Asymptotic breakdown point

Because ε_n^* depends on both n and the realized sample, an asymptotic notion is often used.

Definition 9.3 (Asymptotic breakdown point)**.** The asymptotic breakdown point of an estimator sequence $\{T_n\}$ is

$$\varepsilon^*(T) = \liminf_{n\to\infty} \inf_{\boldsymbol{x}} \varepsilon_n^*(T_n, \boldsymbol{x}),$$

where the infimum is taken over all samples $\boldsymbol{x}$ of size n.

Remark 9.5. The asymptotic breakdown point represents the largest fraction of contamination that an estimator can tolerate in the worst case as the sample size grows.

9.3 Examples

- **Sample mean:** Replacing a single observation by an arbitrarily large value causes divergence, so
$$\varepsilon_n^* = \frac{1}{n}, \qquad \varepsilon^* = 0.$$
- **Sample median:** At least $\lfloor (n-1)/2 \rfloor$ observations must be contaminated to move the median arbitrarily, giving
$$\varepsilon_n^* = \frac{\lfloor (n-1)/2 \rfloor}{n}, \qquad \varepsilon^* = 0.5.$$
- **Least squares estimator:** A single high-leverage outlier can cause divergence, hence
$$\varepsilon^* = 0.$$

Remark 9.6. Breakdown point captures *global robustness* and complements the influence function, which measures *local robustness*.

10 Efficiency and Cramér–Rao Bounds

10.1 Parametric Cramér–Rao Lower Bound

For parametric model p_θ,

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{nI(\theta)}.$$

Efficient estimators achieve equality.

10.2 Nonparametric Efficiency Bound

In semiparametric models,

$$\mathrm{Var}(\hat{\psi}) \geq \mathrm{Var}(\mathrm{EIF}),$$

where EIF is the efficient influence function.

11 Efficient Influence Function (EIF)

Definition 11.1 (Efficient Influence Function)**.** Let $\psi = T(F)$ be a target parameter defined as a functional of the data-generating distribution F, and let $\mathcal{T}$ denote the tangent space of the statistical model at F. The *efficient influence function* (EIF) for ψ is the unique function $\phi^\star \in \mathcal{T}$ such that:

- $\mathbb{E}_F[\phi^\star(Z)] = 0$,
- $\phi^\star$ is an influence function for ψ (i.e., it represents a valid first-order perturbation of T),
- $\mathrm{Var}_F(\phi^\star(Z))$ is minimized among all influence functions in $\mathcal{T}$.

Remark 11.1. Equivalently, the EIF is the orthogonal projection of any valid influence function onto the tangent space of the model. It achieves the *semiparametric efficiency bound*, meaning that no regular, asymptotically linear estimator of ψ can have smaller asymptotic variance.

Remark 11.2. The EIF plays three simultaneous roles:

- it characterizes the lowest possible asymptotic variance for estimating ψ,
- it provides the asymptotic linear expansion of efficient estimators,
- it directly yields variance estimators and confidence intervals.

Thus, deriving the EIF is equivalent to solving the optimal inference problem for ψ under the assumed model.

12 EIF Examples

12.1 Mean

Let $Z \sim F$ with $\mu = \mathbb{E}(Z)$. The parameter of interest is

$$\psi = \mu.$$

The EIF is

$$\mathrm{EIF}(z) = z - \mu.$$

Remark 12.1. In this case, the model is fully nonparametric, the tangent space consists of all mean-zero square-integrable functions, and the sample mean attains the efficiency bound. Thus, the classical sample mean is semiparametrically efficient.

12.2 Average Treatment Effect (ATE)

Let $Z = (Y, A, X)$, where:

- $A \in \{0, 1\}$ is a treatment indicator,
- Y is the observed outcome,
- X is a vector of covariates.

The target parameter is the average treatment effect

$$\psi = \mathbb{E}\{Y(1) - Y(0)\}.$$

Define:

$$\mu_a(X) = \mathbb{E}(Y \mid A = a, X), \qquad e(X) = \mathbb{P}(A = 1 \mid X).$$

Under standard identification assumptions (consistency, ignorability, positivity), the EIF for ψ in the nonparametric model is

$$\mathrm{EIF}(Z) = \frac{A}{e(X)}\big(Y - \mu_1(X)\big) - \frac{1-A}{1-e(X)}\big(Y - \mu_0(X)\big) + \mu_1(X) - \mu_0(X) - \psi.$$

Remark 12.2. This EIF decomposes naturally into three components:

- an outcome residual weighted by inverse propensity for treated units,
- an outcome residual weighted by inverse propensity for control units,
- a plug-in bias correction term $\mu_1(X) - \mu_0(X) - \psi$.

Remark 12.3. Estimators constructed by solving the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{EIF}(Z_i) = 0$$

are *doubly robust*: they remain consistent if either the outcome regression $\mu_a(X)$ or the propensity score $e(X)$ is correctly specified. When both are correctly specified, such estimators achieve the semiparametric efficiency bound.

Remark 12.4. The EIF provides a direct recipe for practical inference:

- point estimation via estimating equations,
- standard error estimation via the empirical variance of the EIF,
- construction of Wald-type confidence intervals.

This is why EIFs lie at the core of modern causal inference and semiparametric theory.