Robustness

Definitions: A sequence $T_n$ of estimators is **qualitatively robust** at $F = F_0$ if the sequence of mappings taking $F$ into the distribution of $T_n$ under $F$ is equicontinuous with respect to a suitable metric.

The **influence curve** $IC(x) = d_1 T(F; \varepsilon_x - F)$ measures the influence of an additional observation $x$ on the estimator $T(F_n)$, given a large sample from $F$.

The **breakdown point** of $T$ is the smallest fraction of gross errors that carries the estimator beyond all bounds.

The **gross error sensitivity** (GES) of $T$ is $\sup_x |IC(x)|$.

An estimator is **B-robust** if it has bounded GES.

The **efficiency** of $T$ at the distribution $F$ is $I_1(F) = \int IC(x)^2 \, dF(x)$ where $I_1(F)$ is the Fisher information in a single observation from $F$.

An **M-estimator** is the solution to $\int \psi(x; \theta) dF_n(x) = 0$. A **location M-estimator** has $\psi(x; \theta) = \psi(x - \theta)$.

**Theorem:** The breakdown point of a location M-estimator is $1/(1+\eta)$ where $\eta = \max \left( \frac{\psi(\infty)}{\psi(-\infty)} - \frac{\psi(-\infty)}{\psi(\infty)} \right)$.

Let $s(x; \theta_0)$ be the score function for a nominal distribution $F_0 \equiv F_{\theta_0}$. Define $\tilde{\psi}(x) = [s(x; \theta_0) - a]_b$, where

$$[h(x)]^u_\ell = \begin{cases} y & h(x) < y \\ h(x) & y \leq h(x) \leq z \\ z & h(x) > z \end{cases}$$

and $a$ is chosen to make $\tilde{\psi}$ integrate to zero with respect to $dF_0$.

**Theorem:** Among all location M-estimators with $\int \psi dF_0 = 0$, $\int \psi(x) s(x; \theta_0) dF_0(x) \neq 0$, and $\sup_x \left| \int \psi(x) s(x; \theta_0) dF_0(x) \right| \leq c$, that based on $\tilde{\psi}$ minimizes $\int \psi^2 dF_0 / \left( \int \psi s dF_0 \right)^2$. It is, up to a multiplicative constant, the unique such estimator, the most efficient location M-estimator with this bound on its GES.

**Theorem:** The most B-robust estimator (i.e., having the smallest GES) is the median.