Influence Functions

Empirical Distribution Function

Definition 1.1.1. Suppose $(X_i) \overset{\text{IID}}{\sim} F_X$. The **empirical (cumulative) distribution function** $\mathbb{F}_n$ is defined as:

$$\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x).$$

■

Definition 1.1.2. The **Kolmogorov-Smirnov** distance between two distributions functions F and G, mapping $\mathbb{R} \to \mathbb{R}$, is defined as:

$$d(F, G) = \sup_{x\in\mathbb{R}} \big|F(x) - G(x)\big|. \tag{1.1.1}$$

■

Theorem 1.1.1 (Glivenko-Cantelli). Suppose $(X_i) \overset{\text{IID}}{\sim} F_X$. The empirical distribution function $\mathbb{F}_n$ converges *uniformly almost surely* to F_X:

$$\sup_{x\in\mathbb{R}} \big|\mathbb{F}_n(x) - F_X(x)\big| \xrightarrow{as} 0.$$

□

Remark 1.1.1. See Serfling (1980), theorem 2.1.4. ♦

Statistical Functionals

2.1 Definition

Definition 1.2.1. A **functional** T is a mapping from a function space $\mathscr{F}$ into $\mathbb{R}$. Typically, $\mathscr{F}$ is a linear function (vector) space. ■

Example 1.2.1. Let $\{F, F_0\}$ denote distribution functions. Examples of *statistical functionals* include:

- The expectation of a specified function $\psi : \mathbb{R} \to \mathbb{R}$:

$$T(F) = \int \psi(x) dF(x). \tag{1.2.2}$$

  Functionals of the form (1.2.2) are **linear**.

- The kth central moment:

$$T(F) = \int \left\{ x - \int \xi dF(\xi) \right\}^k dF(x).$$

- The *Cramer von Mises* distance:

$$T(F) = \int \left\{ F(x) - F_0(x) \right\}^2 dF_0(x).$$

♠

Example 1.2.2. Let H_x denote the Heaviside function:

$$H_x(t) = \int_{-\infty}^{t} \delta_x(u) du = I(t \geq x),$$

where $\delta_x(u)$ is a Dirac spike localized to x. The empirical distribution function is expressible as:

$$\mathbb{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} H_{X_i}(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq t).$$

The evaluation of a linear functional at $\mathbb{F}_n$ is:

$$T(\mathbb{F}_n) = \int \psi(x) d \left\{ \frac{1}{n} \sum_{i=1}^{n} H_{X_i}(x) \right\} = \frac{1}{n} \sum_{i=1}^{n} \int \psi(x) \delta_{X_i}(x) dx = \frac{1}{n} \sum_{i=1}^{n} \psi(X_i).$$

♠

2.2 Continuity

Definition 1.2.2 (Lehmann 1999, definition 6.2.1)**.** Let d denote the KS metric (1.1.1). A statistical functional T is **continuous** if for any sequence $G_n \to F$,

$$\lim_{n\to\infty} d(G_n, F) = 0 \implies \lim_{n\to\infty} T(G_n) = T(F). \tag{1.2.3}$$

■

Proposition 1.2.1 (Functional Consistency). Suppose $(X_i) \overset{\text{IID}}{\sim} F_X$, and T is a continuous functional (1.2.3), then $T(\mathbb{F}_n)$ is consistent for $T(F)$, i.e. $T(\mathbb{F}_n) \xrightarrow{p} T(F)$. ♦

Proposition 1.2.2 (Huber 2009, lemma 2.1)**.** A *linear functional* L defined on the space $\mathscr{F}$ of probability measures on $(\mathbb{R}, \mathscr{B})$ is **continuous** if and only if it is expressible as:

$$L(F) = \int \psi(x) dF(x),$$

for some *bounded, continuous* function $\psi : \mathbb{R} \to \mathbb{R}$. ♦

2.3 Differentiability

Definition 1.2.3. Let F and G denote distribution functions. Define the **ϵ-contaminated** distribution:

$$F_{G,\epsilon} = (1 - \epsilon)F + \epsilon G. \tag{1.2.4}$$

The **Gateaux differential** $\partial T(F; G - F)$ of the functional T at F in the direction of $G - F$ is defined by:

$$\partial T(F; G - F) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon}\big\{T(F_{G,\epsilon}) - T(F)\big\}. \tag{1.2.5}$$

In general, the kth Gateaux differential is obtained via:

$$\partial^k T(F; G - F) = \lim_{\epsilon \downarrow 0} \left\{\frac{\partial^k T(F_{G,\epsilon})}{\partial \epsilon^k}\right\}.$$

■

Example 1.2.3. Consider the Cramer von Mises functional:

$$T(F) = \int \big(F - F_0\big)^2 dF_0.$$

The evaluation at the ϵ-contaminated distribution is:

$$T(F_{G,\epsilon}) = \int \big(F + \epsilon(G - F) - F_0\big)^2 dF_0.$$

The derivative with respect to ϵ is:

$$\frac{\partial}{\partial \epsilon} T(F_{G,\epsilon}) = 2\int \big(F + \epsilon(G - F) - F_0\big)(G - F) dF_0$$

Taking the limit as $\epsilon \downarrow 0$ gives:

$$\partial T(F; G - F) = \lim_{\epsilon \downarrow 0} \frac{\partial}{\partial \epsilon} T(F_{G,\epsilon}) = 2\int (F - F_0)(G - F) dF_0.$$

♠

Discussion 1.2.1. A functional T is **Frechet differentiable** if there exists a functional $\partial T(F;G-F)$, *linear in* $G-F$, such that:

$$\big|T(G_n) - T(F) - \partial T(F;G_n - F)\big| = o\big\{d(F,G_n)\big\}$$

for all sequences (G_n) such that $d(G_n,F) \to 0$ as $n \to \infty$.

If T is continuous in a neighborhood of F and *Frechet differentiable* at F, then the Frechet derivative $\partial T(F;G-F)$ is *continuous* at F (Huber 2009, proposition 2.19). By proposition (1.2.2), there exists a bounded, continuous function ψ_F such that:

$$\partial T(F;G-F) = \int \psi_F(x)d\big\{G(x) - F(x)\big\} = \int \left\{\psi_F(x) - \int \psi_F(\xi)dF(\xi)\right\} dG(x).$$

Moreover, if the Frechet derivative exists, then the Gateaux derivative exists, and the two are equal (Serfling 1980, section 6.2). Thus, continuity and Frechet differentiability at F suffice to represent the Gateaux derivative $\partial T(F;G-F)$ as:

$$\partial T(F;G-F) = \int \varphi_F(x)dG(x),$$

where φ_F bounded, continuous, and mean-zero. When Frechet differentiability cannot be established, it is often *assumed* that the Gateaux derivative is expressible as:

$$\partial h(F;G-F) = \int \psi_F(x)d\big\{G(x) - F(x)\big\} = \int \varphi_F(x)dG(x).$$

for some measurable function ψ_F. Bounding and continuity of $\varphi_F(x)$ are not assumed.

♠

Influence Functions

3.1 Definition

Example 1.3.1. Suppose T is a statistical functional whose Gateaux derivative admits the representation:

$$\partial T(F;G-F) = \int \varphi_F(x)dG(x).$$

To isolate the function $\varphi_F(x)$, set $G(t) = H_x(t)$, then:

$$\partial T(F;H_x - F) = \int \varphi_F(t)dH_x(t) = \int \varphi_F(t)\delta_x(t)dt = \varphi_F(x).$$

Therefore, for any other G, the Gateaux derivative is expressible as:

$$\partial T(F;G-F) = \int \varphi_F(x)dG(x) = \int \big\{\partial T(F;H_x - F)\big\}dG(x).$$

$\varphi_F = \partial T(F;G-F)$ is described as the *influence function*. ♠

Definition 1.3.1. Suppose $(X_i) \overset{\text{IID}}{\sim} F$, and T is a statistical functional. Recall that $H_x(t) = I(x \leq t)$ denotes the Heaviside function. The **influence function** of h is the Gateaux derivative of T at F in the direction of $H_x - F$:

$$\partial T(F; H_x - F) = \lim_{\epsilon \downarrow 0} \left[\frac{\partial}{\partial \epsilon} T\big\{F + \epsilon(H_x - F)\big\} \right]. \tag{1.3.6}$$

■

3.2 Properties

Proposition 1.3.1. Influence functions inherit the chain, product, and quotient rules from differentiation.

i. (**Linearity**) If $T = \sum_{j=1}^{m} \alpha_j T_j$ is a linear combination of functionals, then:

$$\partial T(F; H_x - F) = \sum_{j=1}^{m} \alpha_j \partial T_j(F; H_x - F).$$

ii. (**Product Rule**) If $T = T_1 \cdot T_2$ is a product of functionals, then:

$$\partial T(F; H_x - F) = \partial T_1(F; H_x - F) \cdot T_2(F) + T_1(F) \cdot \partial T_2(F; H_x - F).$$

iii. (**Quotient Rule**) If $T = T_1 / T_2$ is a ratio of functionals, then:

$$\partial T(F; H_x - F) = \frac{\partial T_1(F; H_x - F) \cdot T_2(F) - T_1(F) \cdot \partial T_2(F; H_x - F)}{T_2^2(F)}.$$

iv. (**Chain Rule**) If $T = U \circ T_1$ is a differentiable function of a functional, then:

$$\partial T(F; H_x - F) = \dot{U}\big\{T_1(F)\big\} \partial T_1(F; H_x - F).$$

♦

Example 1.3.2 (Linear Functional)**.** Consider the linear functional:

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

Evaluating at $F_{x,\epsilon}$:

$$T(F_{x,\epsilon}) = \int \psi(u) d\big\{(1-\epsilon)F(u) + \epsilon H_x(u)\big\} = (1-\epsilon) \int \psi(u) dF(u) + \epsilon \psi(x)$$

Taking the derivative with respect to ϵ:

$$\frac{\partial T(F_{x,\epsilon})}{\partial \epsilon} = -\int \psi(u) dF(u) + \psi(x).$$

Thus, the influence function is:

$$\varphi_F(x) = \psi(x) - E\big\{\psi(x)\big\}.$$

♠

Example 1.3.3 (Sample Variance)**.** The variance functional is expressible as:

$$T(F) = \int \big\{x - T_1(F)\big\}^2 dF(x),$$

where $T_1(F)$ is the mean functional. Expanding the quadratic:

$$T(F) = \int \big\{x^2 + T_1^2(F) - 2xT_1(F)\big\} dF(x) = T_2(F) - T_1^2(F).$$

where T_2 is the second moment functional. Since the kth moment is a linear functional, the influence functions for T_1 and T_2 are:

$$\varphi_1(x) = x - \mu_1, \qquad \varphi_2(x) = x^2 - m_2.$$

By linearity and the chain rule, the influence function of T is:

$$\varphi(x) = \big(x^2 - m_2\big) - 2\mu_1(x - \mu_1).$$

♠

Example 1.3.4 (Third Central Moment)**.** The 3rd central moment functional is:

$$T(F) = \int \big\{x - T_1(F)\big\}^3 dF(x).$$

Upon expansion:

$$T(F) = T_3(F) - 3T_2(F)T_1(F) + 2T_1^3(F)$$

By linearity, the product and chain rules, the influence function is:

$$\varphi(x) = \big(x^3 - \mu_3\big) - 3\Big\{\big(x^2 - m_2\big)\mu_1 + m_2(x - \mu_1)\Big\} + 6\mu_1^2(x - \mu_1).$$

♠

3.3 von Mises Expansion

Definition 1.3.2 (Serfling 1980, section 6.2)**.** Suppose T is a statistical functional. The **von Mises expansion** of order m is:

$$T(\mathbb{F}_n) - T(F) = V_{n,m} + R_{n,m}, \tag{1.3.7}$$

where $V_{n,m}$ is given by:

$$V_{n,m} = \sum_{k=1}^{m} \frac{1}{k!} \partial^k T(F; \mathbb{F}_n - F)$$

The *remainder* is expressible either as $R_{n,m} = \big\{T(\mathbb{F}_n) - T(F)\big\} - V_{n,m}$, or as:

$$R_{n,m} = \frac{1}{(m+1)!} \left[\frac{\partial^{m+1}}{\partial \epsilon^{m+1}} T\big\{F + \epsilon(\mathbb{F}_n - F)\big\} \right]_{\epsilon=\epsilon^*}$$

for $\epsilon^* \in [0,1]$. To validate the von Mises expansion (1.3.7), it is necessary to show that the scaled remainder is asymptotically negligible:

$$n^{m/2} R_{n,m} \xrightarrow{p} 0.$$

■

3.4 Asymptotic Normality

Proposition 1.3.2. Suppose $(X_i) \overset{\text{IID}}{\sim} F$ and T is a statistical functional that admits an *influence function expansion* of the form:

$$\sqrt{n}\big\{T(\mathbb{F}_n) - T(F)\big\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_F(X_i) + \sqrt{n} R_n,$$

where $E\big\{\varphi_F(x)\big\} = 0$ and $\sqrt{n} R_n \xrightarrow{p} 0$. Then:

$$\sqrt{n}\big\{T(\mathbb{F}_n) - T(F)\big\} \xrightarrow{\mathcal{L}} N\big(0, \gamma_F^2\big), \tag{1.3.8}$$

where $\gamma_F^2 = E\big\{\varphi_F^2(X)\big\}$. ♦

Proof. Since T admits an influence function:

$$\begin{aligned} T(\mathbb{F}_n) - T(F) &= \int \varphi_F(x) d\mathbb{F}_n(x) + R_n \\ &= \int \varphi_F(x) d \left\{ \frac{1}{n} \sum_{i=1}^{n} H_{X_i}(x) \right\} + R_n = \frac{1}{n} \sum_{i=1}^{n} \varphi_F(X_i) + R_n. \end{aligned}$$

Rescaling by $\sqrt{n}$ gives:

$$\sqrt{n}\big\{T(\mathbb{F}_n) - T(F)\big\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_F(X_i) + \sqrt{n}R_n.$$

Since $E\big\{\varphi_F(X_i)\big\} = 0$, by the ordinary CLT:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_F(X_i) \xrightarrow{\mathcal{L}} N\big(0, \gamma_F^2\big).$$

By hypothesis, the scaled remainder is asymptotically negligible $\sqrt{n}R_n \xrightarrow{p} 0$. The conclusion follows from Slutsky's theorem. ■

Remark 1.3.1 (Lehmann 1999, section 6.3)**.** Consistency and asymptotic normality of $T(\mathbb{F})$ as an estimator of $T(F)$ may be established via the following steps:

i. Derive the influence function for T:

$$\varphi_F(x) = \lim_{\epsilon\downarrow 0}\left[\frac{\partial}{\partial\epsilon}T\big\{F + \epsilon(H_x - F)\big\}\right].$$

ii. Verify that the influence function is unbiased:

$$\int \varphi_F(x)dF(x) = 0.$$

iii. Verify that the remainder is asymptotically negligible:

$$\sqrt{n}R_n = \sqrt{n}\left\{T(\mathbb{F}_n) - T(F) - \frac{1}{n}\sum_{i=1}^{n}\varphi_F(X_i)\right\} \xrightarrow{p} 0.$$

iv. Determine the asymptotic variance:

$$\gamma_F^2 = E\big\{\varphi_F^2(X)\big\} = \int \varphi_F^2(x)dF(x).$$

♦

3.5 Examples

Example 1.3.5 (Cramer von Mises Distance)**.** Recall from a previous example that the Gateaux derivative of the Cramer von Mises functional is:

$$\partial T(F; G-F) = 2\int (F-F_0)(G-F)dF_0.$$

Therefore, the candidate influence function is:

$$\varphi_F(x) = 2\int (F-F_0)(H_x - F)dF_0.$$

The influence function is unbiased since:

$$E\big\{\varphi_F(X)\big\} = 2\int (F-F_0)\big\{E(H_X)-F\big\}dF_0 = 2\int (F-F_0)(F-F)dF_0 = 0.$$

The remainder takes the form:

$$\begin{aligned}
R_n &= T(\mathbb{F}_n) - T(F) - \frac{2}{n}\sum_{i=1}^{n}\int (F-F_0)(H_x-F)dF_0 \\
&= \int (\mathbb{F}_n - F_0)^2 dF_0 - \int (F-F_0)^2 dF_0 - 2\int (F-F_0)(\mathbb{F}_n - F)dF_0 \\
&= \int \big\{(\mathbb{F}_n - F_0)^2 - (F-F_0)^2 - 2(F-F_0)(\mathbb{F}_n - F)\big\}dF_0 \\
&= \int \big\{\mathbb{F}_n - 2\mathbb{F}_n F + F^2\big\}dF_0 = \int (\mathbb{F}_n - F)^2 dF_0.
\end{aligned}$$

The remainder is bounded above by:

$$|R_n| \leq \sup_{x\in\mathbb{R}} \big\{\mathbb{F}_n(x) - F(x)\big\}^2.$$

By the Dvoretzky, Kiefer, Wolfowitz inequality:

$$\sqrt{n}\left\{\sup_{x\in\mathbb{R}} \big|\mathbb{F}_n(x) - F(x)\big|\right\} = \mathcal{O}_p(1),$$

therefore:

$$\sqrt{n}\big|R_n\big| \leq \frac{1}{\sqrt{n}}\cdot n \cdot \sup_{x\in\mathbb{R}} \big\{\mathbb{F}_n(x) - F(x)\big\}^2 = \mathcal{O}(n^{-1/2}) \xrightarrow{p} 0.$$

This establishes the influence function expansion for the Cramer von Mises functional:

$$\sqrt{n}\big\{T(\mathbb{F}_n) - T(F)\big\} = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\int (F-F_0)(H_{X_i} - F)dF_0 + \mathcal{O}_p(n^{-1/2}).$$

Now,

$$\sqrt{n}\big\{T(\mathbb{F}_n) - T(F)\big\} \xrightarrow{\mathcal{L}} N\big(0, \gamma_F^2\big),$$

with asymptotic variance:

$$\gamma_F^2 = E\{\varphi_F^2(X)\} = 4E_F\left\{\int (F - F_0)(H_X - F)dF_0\right\}^2.$$

♠

Example 1.3.6 (Sample Quantile)**.** Consider the pth sample quantile:

$$t_p = T(F) = F^{-1}(p).$$

Towards obtaining the influence function, let $t_{p,\epsilon} = T(F_{x,\epsilon}) = F_{x,\epsilon}^{-1}(p)$, then:

$$p = F_{x,\epsilon}(t_{p,\epsilon}) = (1-\epsilon)F(t_{p,\epsilon}) + \epsilon H_x(t_{p,\epsilon}).$$

Taking the implicit derivative with respect to ϵ:

$$0 = -F(t_{p,\epsilon}) + (1-\epsilon)f(t_{p,\epsilon})\frac{\partial t_{p,\epsilon}}{\partial \epsilon} + H_x(t_{p,\epsilon}) + \epsilon\delta_x(t_{p,\epsilon})\frac{\partial t_{p,\epsilon}}{\partial \epsilon}$$

In the limit as $\epsilon \downarrow 0$:

$$0 = -F(t_p) + f(t_p)\left\{\frac{\partial t_{p,\epsilon}}{\partial \epsilon}\right\}_{\epsilon=0^+} + H_x(t_p).$$

Identify $\{\partial_\epsilon t_{p,\epsilon}\}_{\epsilon=0^+}$ as the influence function, and rearrange to obtain:

$$\varphi_F(x) = \frac{F(t_p) - H_x(t_p)}{f(t_p)} = \frac{p - I(x \le t_p)}{f(t_p)}.$$

The influence function is unbiased:

$$E\{\varphi_F(X)\} = \frac{p - F(t_p)}{f(t_p)} = \frac{p-p}{f(t_p)} = 0.$$

From Serfling (1980), theorem 2.5.1:

$$\sqrt{n}\left(T(\mathbb{F}_n) - T(F) - \frac{1}{n}\sum_{i=1}^{n}\frac{p - I(x \le t_p)}{f(t_p)}\right) \xrightarrow{p} 0.$$

Therefore:

$$\sqrt{n}\{T(\mathbb{F}_n) - T(F)\} \xrightarrow{\mathcal{L}} N(0, \gamma_F^2),$$

where the asymptotic variance is:

$$\gamma_F^2 = \text{Var}\{\varphi_F(X)\} = \frac{F(t_p)\{1 - F(t_p)\}}{f^2(t_p)} = \frac{p(1-p)}{f^2(t_p)}.$$

♠

Example 1.3.7 (M-Estimation)**.** Suppose $(X_i) \overset{\text{IID}}{\sim} F$, and consider an M-estimator $\hat{\theta}_n = T(\mathbb{F}_n)$, defined implicitly by the relation:

$$0 = \frac{1}{n}\sum_{i=1}^{n}\psi(X_i;\hat{\theta}_n) = \int \psi\big\{u; T(\mathbb{F}_n)\big\}d\mathbb{F}_n(u),$$

where ψ is the *estimating equation*. Towards finding the influence function, consider evaluation of the estimating relation at $F_{x,\epsilon}$:

$$\begin{aligned} 0 &= \int \psi\big\{u; T(F_{x,\epsilon})\big\}dF_{x,\epsilon}(u) = \int \psi\big\{u; T(F_{x,\epsilon})\big\}d\big\{(1-\epsilon)F(u) + \epsilon H_x(u)\big\} \\ &= (1-\epsilon)\int \psi\big\{u; T(F_{x,\epsilon})\big\}dF(u) + \epsilon\psi\big\{u; T(F_{x,\epsilon})\big\} \end{aligned}$$

Taking the implicit derivative with respect to ϵ:

$$\begin{aligned} 0 = &-\int \psi\big\{u; T(F_{x,\epsilon})\big\}dF(u) + (1-\epsilon)\int \dot{\psi}\big\{u; T(F_{x,\epsilon})\big\}\frac{\partial T(F_{x,\epsilon})}{\partial\epsilon}dF(u) \\ &+ \psi\big\{x; T(F_{x,\epsilon})\big\} + \epsilon\dot{\psi}\big\{u; T(F_{x,\epsilon})\big\}\frac{\partial T(F_{x,\epsilon})}{\partial\epsilon}. \end{aligned}$$

In the limit as $\epsilon \downarrow 0$:

$$\begin{aligned} 0 &= -\int \psi\big\{u; T(F)\big\}dF(u) + \int \dot{\psi}\big\{u; T(F_{x,\epsilon})\big\}dF(u)\cdot\left\{\frac{\partial T(F_{x,\epsilon})}{\partial\epsilon}\right\}_{\epsilon=0^+} + \psi\big\{x; T(F)\big\} \\ &= -E_F\big\{\psi(X;\theta)\big\} + E_F\big\{\dot{\psi}(X;\theta)\big\}\cdot\varphi_F(x) + \psi(x;\theta) \end{aligned}$$

Upon rearranging:

$$\varphi_F(x) = -\frac{\psi(x;\theta) - E_F\big\{\psi(X;\theta)\big\}}{E_F\big\{\dot{\psi}(X;\theta)\big\}}.$$

Under the regularity conditions for M-estimation the remainder is asymptotically negligible, and:

$$\sqrt{n}\big\{\hat{\theta}_n - \theta\big\} \xrightarrow{\mathcal{L}} N\big(0, \gamma_F^2\big),$$

with asymptotic variance:

$$\text{Var}\big\{\varphi_F(X)\big\} = \frac{\text{Var}\big\{\psi(X;\theta)\big\}}{E_F^2\big\{\dot{\psi}(X;\theta)\big\}}.$$

In the case of maximum likelihood estimation, $\psi(x;\theta) = \dot{\ell}(x;\theta)$, $\ell(x;\theta) = \ln f(x;\theta)$, is the *score equation*, which is unbiased $E\big\{\psi(X;\theta)\big\} = E\big\{\dot{\ell}(X;\theta)\big\} = 0$.

The negative expectation of $\dot{\psi}(X;\theta)$ is the Fisher information:

$$-E_F\big\{\dot{\psi}(X;\theta)\big\} = -E_F\big\{\ddot{\ell}(X;\theta)\big\} = \mathcal{I}(\theta).$$

The influence function takes the form:

$$\varphi_F(x) = -\mathcal{I}^{-1}(\theta)\dot{\ell}(x;\theta),$$

and the asymptotic variance is:

$$\gamma_F^2 = \frac{\text{Var}\big\{\dot{\ell}(X;\theta)\big\}}{\mathcal{I}^2(\theta)} = \frac{\mathcal{I}(\theta)}{\mathcal{I}^2(\theta)} = \mathcal{I}^{-1}(\theta).$$

♠

Example 1.3.8 (V-Estimation)**.** Consider a V-estimator of order $m = 2$ with symmetric kernel function $h(X_1, X_2)$:

$$T(F) = \int h(x_1, x_2)dF(x_1)dF(x_2).$$

Towards finding the influence function, consider the evaluation at $F_{x,\epsilon}$:

$$\begin{aligned} T(F_{x,\epsilon}) = &\int h(x_1, x_2)d\big\{(1-\epsilon)F(x_1) + \epsilon H_x(x_1)\big\}d\big\{(1-\epsilon)F(x_2) + \epsilon H_x(x_2)\big\} \\ =&(1-\epsilon)^2\int h(x_1,x_2)dF(x_1)dF(x_2) + (1-\epsilon)\epsilon\int h(x_1,x)dF(x_1) \\ &+ \epsilon(1-\epsilon)\int h(x,x_2)dF(x_2) + \epsilon^2 h(x,x). \end{aligned}$$

Recall that $\theta = E_F\big\{h(X_1, X_2)\big\}$, and that $h_1(x) = E\big\{h(X_1, X_2)|X_1 = x\big\}$ denotes the first projection. The functional at $F_{x,\epsilon}$ is expressible as:

$$T(F_{x,\epsilon}) = (1-\epsilon)^2\theta + 2\epsilon(1-\epsilon)h_1(x) + \epsilon^2 h(x,x).$$

Taking the derivative with respect to ϵ:

$$\frac{\partial T(F_{x,\epsilon})}{\partial\epsilon} = -2(1-\epsilon)\theta + 2(1-\epsilon)h_1(x) - 2\epsilon h_1(x) + 2\epsilon h(x,x).$$

In the limit as $\epsilon \downarrow 0$:

$$\varphi_F(x) = 2\big\{h_1(x) - \theta\big\}$$

The remainder takes the form:

$$R_n = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j=1}^{n}\big\{h(X_i, X_j) - \theta\big\} - \frac{2}{n}\sum_{i=1}^{n}\big\{h_1(X_i) - \theta\big\}$$

Supposing $\text{Var}\big\{h(X,X)\big\} < \infty$, it can be shown that $n\text{Var}(R_n) \to 0$ (see Serfling 1980, 6.3.2), which implies $\sqrt{n}R_n = o_p(1)$. Therefore:

$$\sqrt{n}\big\{\hat{\theta}_n - \theta\big\} \xrightarrow{\mathcal{L}} N\big(0, \gamma_F^2\big),$$

where the asymptotic variance is:

$$\text{Var}\big\{\varphi_F(X)\big\} = 4\text{Var}\big\{h_1(X)\big\} = 4\zeta_1^2.$$

♠

Jackknife

4.1 Definition

Definition 1.4.1. Suppose $(X_i) \overset{\text{IID}}{\sim} F$, and let $\hat{\theta}_n = t(x_1, \cdots, x_n)$ denote an estimator. Denote by $\hat{\theta}_{(-i)} = t(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ the estimate obtained by omitting the ith observation. Define the ith **pseudo-value** as:

$$\tilde{\theta}_i = n\hat{\theta}_n - (n-1)\hat{\theta}_{(-i)}.$$

The jackknife estimator of θ is:

$$\hat{\theta}_J = \frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i.$$

■

4.2 Empirical Influence Function

Proposition 1.4.1. Define $\mathbb{F}_{(-i)}$ as the empirical distribution function obtained by omitting the ith observation. $\mathbb{F}_{(-i)}$ is related to the overall distribution function $\mathbb{F}_n$ via:

$$\mathbb{F}_{(-i)} = \mathbb{F}_n + \epsilon_n\big(H_{X_i} - \mathbb{F}_n\big).$$

Thus, $\mathbb{F}_{(-i)}$ may be regarded as an ϵ-contaminated distribution (1.2.4). ♦

Proof. Expressing $\mathbb{F}_n$ in terms of $\mathbb{F}_{(-i)}$:

$$\mathbb{F}_n = \frac{1}{n}X_i + \frac{1}{n}\sum_{j\neq i}^{n} H_{X_j} = \frac{1}{n}H_{y_i} + \frac{n-1}{n}\frac{1}{n-1}\sum_{j\neq i}\delta_{X_j} = \frac{1}{n}H_{X_i} + \frac{n-1}{n}\mathbb{F}_{(-i)}.$$

Solving for $\mathbb{F}_{(-i)}$:

$$\mathbb{F}_{(-i)} = \frac{n}{n-1}\mathbb{F}_n - \frac{1}{n-1}H_{X_i} = \left(1 + \frac{1}{n-1}\right)\mathbb{F}_n - \frac{1}{n-1}H_{X_i}.$$

Let $\epsilon_n = -(n-1)^{-1}$, then:

$$\mathbb{F}_{(-i)} = \mathbb{F}_n - \epsilon_n\mathbb{F}_n + \epsilon_n H_{X_i} = \mathbb{F}_n + \epsilon_n(H_{X_i} - \mathbb{F}_n).$$

■

Proposition 1.4.2. The difference between the ith *pseudo-value* $\tilde{\theta}_i$ and the overall sample estimator $\hat{\theta}_n$ is an empirical approximation to the influence function at X_i:

$$\tilde{\theta}_i - \hat{\theta}_n = \hat{\varphi}_F(X_i).$$

♦

Proof. Consider the ith pseudo-value:

$$\tilde{\theta}_i = n\hat{\theta}_n - (n-1)\hat{\theta}_{(-i)} = \hat{\theta}_n - (n-1)\big(\hat{\theta}_{(-i)} - \hat{\theta}_n\big),$$
$$\tilde{\theta}_i - \hat{\theta}_n = \frac{\hat{\theta}_{(-i)} - \hat{\theta}_n}{-(n-1)^{-1}} = \frac{\hat{\theta}_{(-i)} - \hat{\theta}_n}{\epsilon_n}.$$

Writing $\theta = T(F)$ as a statistical functional:

$$\tilde{\theta}_i - \hat{\theta}_n = \epsilon_n^{-1}\big\{T(\mathbb{F}_{(-i)}) - T(\mathbb{F}_n)\big\} = \epsilon_n^{-1}\Big[T\big\{\mathbb{F}_n + \epsilon_n(H_{X_i} - \mathbb{F}_n)\big\} - T(\mathbb{F}_n)\Big].$$

Comparison to the Gateaux derivative (1.2.5) suggests that for $n \to \infty$:

$$\tilde{\theta}_i - \hat{\theta}_n \approx \varphi_F(X_i).$$

■

Example 1.4.1 (Jackknife Variance Estimator)**.** Suppose $\hat{\theta}_n = T(\mathbb{F}_n)$ admits an influence function expansion:

$$\sqrt{n}\big(\hat{\theta}_n - \theta\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_F(X_i) + o_p(1).$$

The asymptotic variance of $\sqrt{n}\big(\hat{\theta}_n - \theta\big)$ is $\mathrm{Var}\big\{\varphi_F(X)\big\}$, which is estimated by:

$$\hat{V}\big\{\varphi_F(X)\big\} \doteq \frac{1}{n}\sum_{i=1}^{n}\varphi_F^2(X_i).$$

Approximating the true influence function φ_F by the empirical influence of X_i:

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_F^2(X_i) \approx \frac{1}{n}\sum_{i=1}^{n}\hat{\varphi}_F^2(X_i) = \frac{1}{n}\sum_{i=1}^{n}\big(\tilde{\theta}_i - \hat{\theta}_n\big)^2.$$

This approximation motivates the jackknife variance estimator:

$$\hat{V}_J\big\{\sqrt{n}(\hat{\theta}_n - \theta)\big\} = \frac{1}{n-1}\sum_{i=1}^{n}\big(\tilde{\theta}_i - \hat{\theta}_n\big)^2.$$

Note that this is an estimate of $\mathrm{Var}\big\{\sqrt{n}(\hat{\theta}_n - \theta)\big\}$. The jackknife estimate of $\mathrm{Var}(\hat{\theta}_n)$ is:

$$\hat{V}_J(\hat{\theta}_n) = \frac{1}{n}\cdot\frac{1}{n-1}\sum_{i=1}^{n}\big(\tilde{\theta}_i - \hat{\theta}_n\big)^2.$$

Expressed in terms of the leave-1-out estimates:

$$\hat{V}_J(\hat{\theta}_n) = \frac{n-1}{n}\sum_{i=1}^{n}\big(\hat{\theta}_{(-i)} - \hat{\theta}_n\big)^2.$$

♠

References

References for influence functions:

- Serfling. *Approximation Theorems of Mathematical Statistics* (1980); Chapter 6.
- Lehmann. *Elements of Large Sample Theory* (1999); Chapter 6.
- Tsiatis. *Semiparametric Theory and Missing Data* (2006); Chapter 3.
- Huber & Ronchetti. *Robust Statistics* (2009); Chapters 1 & 2.