STATISTICAL FUNCTIONALS

Suppose $X_1, \ldots, X_n$ are iid with probability distribution F.

We consider estimating quantities derived from F by means of a **statistical functional**, that is by a function $T : \mathcal{D} \to \mathbb{R}$, which takes probability distributions in a space $\mathcal{D}$ of distributions and returns real numbers.

We thus wish to estimate

$$\theta_0 = T(F),$$

where $\theta_0 \in \mathbb{R}$ represents some property or feature of F.

We will consider estimating $\theta_0 = T(F)$ with

$$\hat{\theta}_n = T(\hat{F}_n),$$

where $\hat{F}_n$ is the empirical distribution of $X_1, \ldots, X_n$.

So our estimator of $\theta_0 = T(F)$ is obtained by applying the same functional to the empirical distribution of $X_1, \ldots, X_n$.

The empirical distribution is given by

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

where δ_z is the distribution placing unit mass at the point z for any $z \in \mathbb{R}$.

From now on we will use $F, \hat{F}_n$ to represent probability distributions (measures) or the corresponding cdfs interchangeably. So we may write

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \le x),$$

since the cdf of δ_z is given by $\mathbb{1}(x \ge z)$ for $x \in \mathbb{R}$.

We give some examples of statistical functionals and their "plug-in" estimators:

Examples

(i) The mean:

$$\mu = T(F) = \int x \, dF(x) = \begin{cases} \int_{\mathbb{R}} x f(x) dx & \text{if } F \text{ has pdf } f \\ \sum_{x \in \mathcal{X}} x \cdot p_X(x) & \text{if } F \text{ has pmf } p_X \text{ with support on } \mathcal{X} \\ \vdots & \end{cases}$$

Estimator is

$$\hat{\mu} = T(\hat{F}_n) = \int x \, d\hat{F}_n(x)$$
$$= \int x \, d\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}\right)(x)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \int x \, d\delta_{X_i}(x)$$
$$= \frac{1}{n}\sum_{i=1}^{n} X_i \quad .$$

(ii) The variance:

$$\sigma^2 = T(F) = \int \left(x - \int t \, dF(t)\right)^2 dF(x)$$

Estimator is

$$\hat{\sigma}^2 = T(\hat{F}_n) = \int \left(x - \int t \, d\hat{F}_n(t)\right)^2 d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2.$$

(iii) The τ^{th} quantile:

$$\xi_\tau = T(F) = \inf\{x : F(x) \geq \tau\}$$

Estimator is

$$\hat{\xi}_\tau = T(\hat{F}_n) = \inf\{x : \hat{F}_n(x) \geq \tau\} = X_{(\lceil \tau n \rceil)},$$

where $X_{(1)} \leq \dots \leq X_{(n)}$ are the order statistics.

(iv) Shape parameter under Gamma$(\alpha, 1)$:

$\alpha = T(F)$ is the value of t that solves $\displaystyle\int \left(\log x - \frac{\Gamma'(t)}{\Gamma(t)}\right) dF(x) = 0$

Estimator

$\hat{\alpha} = T(\hat{F}_n)$ is the value of t that solves

$$0 = \int \left(\log x - \frac{\Gamma'(t)}{\Gamma(t)}\right) d\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \log X_i - \frac{\Gamma'(t)}{\Gamma(t)}.$$

(v) The α-trimmed mean:

$$\mu_\alpha = T(F) = \frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x \, dF(x)$$

Estimator is

usually put here $\dfrac{1}{n-2\lfloor \alpha n \rfloor}$

$$\hat{\mu}_\alpha = \frac{1}{1-2\alpha}\int_{\hat{F}_n^{-1}(\alpha)}^{\hat{F}_n^{-1}(1-\alpha)} x \cdot d\hat{F}_n(x) = \frac{1}{n(1-2\alpha)} \sum_{i=\lfloor \alpha n \rfloor+1}^{n-\lfloor \alpha n \rfloor} X_{(i)},$$

where $X_{(1)} \le \cdots \le X_{(n)}$ are the order statistics.

von Mises Expansion for statistical functionals

With a view to establishing central limit type results for statistical functionals, which will be of the form

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) \xrightarrow{D} N(0, \vartheta), \text{ for some } \vartheta < \infty,$$

we work towards expressing the difference $T(\hat{F}_n) - T(F)$ as a mean of iid random variables with mean 0 and some variance ϑ.

In particular we wish to write

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) = \underbrace{\sqrt{n}\,\bar{Y}_n}_{\xrightarrow{d} N(0,\sigma^2)} + \underbrace{\sqrt{n}\,R(\hat{F}_n - F)}_{\xrightarrow{p} 0},$$

where

- $\bar{Y}_n$ is the mean of iid rvs $Y_1, \dots, Y_n$ such that $\mathbb{E}Y_1 = 0$, $\mathbb{E}Y_1^2 = \vartheta$.
- $\sqrt{n}\,R(\hat{F}_n - F)$ is a remainder term which can be made to vanish under certain assumptions.

To achieve this, we consider something like a Taylor expansion of T around F, evaluated at $\hat{F}_n$.

We need to define a derivative for statistical functionals.

Von Mises Derivative

The von Mises derivative of T at F in the direction G is defined as

$$T_F^{(1)}(G-F) = \frac{d}{d\varepsilon} T\left(F + \varepsilon(G-F)\right)\Big|_{\varepsilon=0},$$

provided there exists a function φ_F, not depending on G, such that

$$T_F^{(1)}(G-F) = \int \varphi_F(x)\, d(G-F)\, dx$$

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

We will sometimes use the notation $F_\varepsilon = F + \varepsilon(G-F)$, and then write

$$T_F^{(1)}(G-F) = \frac{d}{d\varepsilon} T(F_\varepsilon)\Big|_{\varepsilon=0}.$$

With the von Mises derivative we will write

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) = \sqrt{n}\, T_F^{(1)}(\hat{F}_n - F) + \sqrt{n}\, R(\hat{F}_n - F),$$

when we want $\sqrt{n}\, T_F^{(1)}(\hat{F}_n - F) \xrightarrow{D} N(0, \vartheta)$.

Influence curve:

The function ψ_F is called the <u>influence curve</u> of the functional T at F.

We see that we may obtain $\psi_F(x)$ as

$$\psi_F(x) = T_F^{(1)}(\delta_x - F) = \frac{d}{d\varepsilon} T\left(F + \varepsilon(\delta_x - F)\right)\Big|_{\varepsilon=0},$$

by writing

$$\int \frac{d}{d\varepsilon} T\left(F + \varepsilon(\delta_x - F)\right)\Big|_{\varepsilon=0} d(G-F)(x)$$

$$= \frac{d}{d\varepsilon}\left[\int T(F + \varepsilon(\delta_x - F))\, dG(x) - \int T(F + \varepsilon(\delta_x - F))\, dF(x)\right]\Bigg|_{\varepsilon=0}$$

$$= \frac{d}{d\varepsilon} T\left(F + \varepsilon(G - F)\right)\Big|_{\varepsilon=0} - \frac{d}{d\varepsilon} T\left(F + \varepsilon(F - F)\right)\Big|_{\varepsilon=0}$$

$$= T_F^{(1)}(G - F).$$

The influence curve measures the change in $T(F)$ when F is perturbed by the addition of a point mass at x.

Influence curves play an important role in the study of robust estimation — when one considers the effect of outliers, for example.

Exercise: Show that $T_F^{(1)}(\hat{F}_n - F) = \frac{1}{n}\sum_{i=1}^n \varphi_F(X_i)$

Solution: We have

$$T_F^{(1)}(\hat{F}_n - F) = \int \varphi_F(x)\, d(\hat{F}_n - F)(x)$$

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

$$= \int \varphi_F(x)\, d\hat{F}_n(x)$$

$$= \int \varphi_F(x)\, d\left(\frac{1}{n}\sum_{i=1}^n \delta_{X_i}\right)(x)$$

$$= \frac{1}{n}\sum_{i=1}^n \int \varphi_F(x)\, d\delta_{X_i}(x)$$

$$= \frac{1}{n}\sum_{i=1}^n \varphi_F(X_i).$$

We may now write

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n T_F^{(1)}(\delta_{X_i} - F) + \sqrt{n}\, R(\hat{F}_n - F),$$

noting that by the central limit theorem

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n T_F^{(1)}(\delta_{X_i} - F) \xrightarrow{D} N(0, \vartheta),$$

where

$$\vartheta = \int [\varphi_F(x)]^2\, dF(x) = \mathrm{Var}\left(T_F^{(1)}(\delta_{X_1} - F)\right)$$

provided this is finite. Note that $\mathbb{E}\, T_F^{(1)}(\delta_{X_1} - F) = \int \varphi_F(x)\, dF(x) = 0.$

EXAMPLES OF VON MISES EXPANSIONS FOR STATISTICAL FUNCTIONALS

The mean:

Let $X_1, \dots, X_n \overset{ind}{\sim} F$ with mean μ and variance $\sigma^2 < \infty$.

Consider

$$\mu = T(F) = \int x \, dF(x)$$

$$\hat{\mu} = T(\hat{F}_n) = \int x \, d\hat{F}_n(x) = \bar{X}_n.$$

- The von Mises derivative is

$$\begin{aligned} T_F^{(1)}(G-F) &= \frac{d}{d\varepsilon} T(F_\varepsilon) \Big|_{\varepsilon=0} \\ &= \frac{d}{d\varepsilon} \int x \, F_\varepsilon(x) \Big|_{\varepsilon=0} \\ &= \frac{d}{d\varepsilon} \int x \, d\big(F + \varepsilon(G-F)\big)(x) \Big|_{\varepsilon=0} \\ &= \frac{d}{d\varepsilon} \left[\int x \, dF(x) + \varepsilon \int x \, d(G-F)(x) \right] \Big|_{\varepsilon=0} \\ &= \int x \, d(G-F)(x) \end{aligned}$$

- The influence function at x_0 is

$$T_F^{(1)}(\delta_{x_0} - F) = \int x \, d(\delta_{x_0} - F)(x) = x_0 - \mu.$$

and

$$\mathrm{Var}\left[T_F^{(1)}(\delta_{X_1} - F)\right] = \mathbb{E}(X_1 - \mu)^2 = \sigma^2.$$

- We can write the von Mises expansion

$$\sqrt{n}\left(\bar{X}_n - \mu\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(X_i - \mu\right),$$

noting that the remainder is zero and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(X_i - \mu\right) \xrightarrow{D} N\left(0, \sigma^2\right) \quad \text{as} \quad n \to \infty.$$

Smooth functions of the mean:

Let $X_1, \dots, X_n \overset{ind}{\sim} F$ with mean μ and variance $\sigma^2 < \infty$.

Consider

$$\theta_0 = T(F) = g\left(\int x \, dF(x)\right) = g(\mu) \qquad \text{for some } g: \mathbb{R} \to \mathbb{R}.$$

$$\hat{\theta}_n = T(\hat{F}_n) = g\left(\int x \, d\hat{F}_n(x)\right) = g(\bar{X}_n)$$

- The von Mises derivative is

$$T_F^{(1)}(G-F) = \frac{d}{d\varepsilon} T(F_\varepsilon) \Big|_{\varepsilon=0}$$

$$= \frac{d}{d\varepsilon} \, g\left(\int x \, dF_\varepsilon(x)\right) \Big|_{\varepsilon=0}$$

$$= g'\left(\int x F_\varepsilon(x)\right) \frac{d}{d\varepsilon} \int x \, d\left(F + \varepsilon(G-F)\right)(x) \Big|_{\varepsilon=0}$$

$$= g'\left(\int x \, dF(x)\right) \int x \, d(G-F)(x).$$

- The influence curve at x_0 is

$$T_F^{(1)}(\delta_{x_0} - F) = g'\left(\int x \, dF(x)\right)\left(x_0 - \int x \, dF(x)\right) = g'(\mu)(x_0 - \mu)$$

and

$$\mathrm{Var}\left[T_F^{(1)}(\delta_{X_1} - F)\right] = [g'(\mu)]^2 \, \mathrm{Var}(X_1 - \mu) = [g'(\mu)]^2 \sigma^2.$$

- We can write the von Mises expansion

$$\sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g'(\mu)(X_i - \mu) + \sqrt{n}\, R(\hat{F}_n - F),$$

where

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g'(\mu)(X_i - \mu) \xrightarrow{D} N\left(0, [g'(\mu)]^2 \sigma^2\right) \quad \text{as} \quad n \to \infty$$

and $\sqrt{n}\, R(\hat{F}_n - F)$ depends on the smoothness of g.

Quantiles:

Let $X_1, \dots, X_n \overset{ind}{\sim} F$ with continuous density f

Consider

$$\xi_\tau = T(F) = \inf\{x : F(x) \geq \tau\} = F^{-1}(\tau)$$

$$\hat{\xi}_\tau = T(\hat{F}_n) = \inf\{x : \hat{F}_n(x) \geq \tau\} = X_{(\lceil \tau n \rceil)},$$

where $X_{(1)} \leq \dots \leq X_{(n)}$ are the order statistics.

- The von Mises derivative is found as follows: Write

$$F_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right) = \tau \qquad \Rightarrow \qquad \frac{d}{d\varepsilon} F_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\Big|_{\varepsilon=0} = 0$$

Then we have

$$0 = \frac{d}{d\varepsilon}\left[F\left(F_\varepsilon^{-1}(\tau)\right) + \varepsilon (G-F)\left(F_\varepsilon^{-1}(\tau)\right)\right]\Bigg|_{\varepsilon=0}$$

$$= \Bigg[f\left(F_\varepsilon^{-1}(\tau)\right) \underbrace{\frac{d}{d\varepsilon} F_\varepsilon^{-1}(\tau)} + (G-F)\left(F_\varepsilon^{-1}(\tau)\right) + \varepsilon \frac{d}{d\varepsilon}(G-F)\left(F_\varepsilon^{-1}(\tau)\right)\Bigg]\Bigg|_{\varepsilon=0}$$

$$= f\left(F^{-1}(\tau)\right) T_F^{(1)}(G-F) + (G-F)\left(F^{-1}(\tau)\right).$$

$$= f\left(F^{-1}(\tau)\right) T_F^{(1)}(G-F) + G\left(F^{-1}(\tau)\right) - \tau$$

$\Leftrightarrow$

$$T_F^{(1)}(G-F) = \frac{\tau - G\left(F^{-1}(\tau)\right)}{f\left(F^{-1}(\tau)\right)},$$

provided $f\left(F^{-1}(\tau)\right) > 0$.

- The influence function at x_0 is

$$T_F^{(1)}(\delta_{x_0} - F) = \frac{\tau - \mathbb{1}\left(F^{-1}(\tau) \geq x_0\right)}{f\left(F^{-1}(\tau)\right)} = \frac{\tau - \mathbb{1}\left(x_0 \leq \xi_\tau\right)}{f(\xi_\tau)}$$

and

$$\mathrm{Var}\left[T_F^{(1)}(\delta_{X_1}-F)\right] = \mathrm{Var}\left[\frac{\tau - \mathbb{1}(X_1 \le \xi_\tau)}{f(\xi_\tau)}\right] = \frac{\tau(1-\tau)}{f^2(\xi_\tau)}.$$

- We can write the von Mises expansion

$$\sqrt{n}\left(\hat{\xi}_\tau - \xi_\tau\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\tau - \mathbb{1}(X_i \le \xi_\tau)}{f(\xi_\tau)}\right) + \sqrt{n}\, R(\hat{F}_n - F),$$

where

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\tau - \mathbb{1}(X_i \le \xi_\tau)}{f(\xi_\tau)} \xrightarrow{D} N\left(0, \frac{\tau(1-\tau)}{f^2(\xi_\tau)}\right) \quad \text{as} \quad n \to \infty,$$

and Ghosh (1971) showed $\sqrt{n}\, R(\hat{F}_n - F) \xrightarrow{P} 0$ provided $f(\xi_\tau) > 0$.

L-Estimates:

Let $X_1, \dots, X_n \overset{ind}{\sim} F$ with continuous density f.

Consider

$$\theta_0 = T(F) = \int_0^1 J(u) F^{-1}(u)\, du \quad \text{for some } J:(0,1) \to \mathbb{R}.$$

$$\hat{\theta}_n = T(\hat{F}_n) = \int_0^1 J(u) \hat{F}_n^{-1}(u)\, du = \sum_{i=1}^n u_i \cdot X_{(i)},$$

where $u_i = \int_{\frac{i-1}{n}}^{\frac{i}{n}} J(u)\, du$, $i = 1, \dots, n$. Draw a picture of $\hat{F}_n^{-1}(u) = \inf\{x : \hat{F}_n(x) \ge u\}$.

So L-estimators are linear combinations of order statistics.

- The von Mises derivative is

$$T_F^{(1)}(G-F) = \frac{d}{d\varepsilon} T(F_\varepsilon)\Big|_{\varepsilon=0}$$

$$= \frac{d}{d\varepsilon}\left[\int_0^1 J(u)\, F_\varepsilon^{-1}(u)\, du\right]\Bigg|_{\varepsilon=0}$$

$$= \int_0^1 J(u) \frac{d}{d\varepsilon} F_\varepsilon^{-1}(u)\Big|_{\varepsilon=0} du$$

From our work on quantiles

$$= \int_0^1 J(u)\left[\frac{u - G(F^{-1}(u))}{f(F^{-1}(u))}\right] du$$

$u = F(y)$, $y = F^{-1}(u)$, $\frac{du}{dy} = f(y)$

$$= \int_{-\infty}^{\infty} J(F(y))\,[F(y) - G(y)]\, dy$$

- The influence function at x_0 is

$$T_F^{(1)}(\delta_{x_0} - F) = \int_{-\infty}^{\infty} J(F(y))\,[F(y) - \mathbb{1}(y \geq x_0)]\, dy$$

$$= -\int_{x_0}^{\infty} J(F(y))\, dy + \int_{-\infty}^{\infty} J(F(y))\, F(y)\, dy$$

$$= \int_{-\infty}^{\infty} J(F(y))\, dy - \int_{x_0}^{\infty} J(F(y))\, dy + \int_{-\infty}^{\infty} J(F(y))\,[F(y) - 1]\, dy$$

$$= \int_{-\infty}^{x_0} J(F(y))\, dy + \int_{-\infty}^{\infty} J(F(y))\,[F(y) - 1]\, dy$$

<u>Example of an L-estimator:</u>

The <u>α-trimmed mean</u> is defined via an L-functional as

$$\mu_\alpha = T(F) = \frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x \, dF(x) = \frac{1}{1-2\alpha}\int_\alpha^{1-\alpha} F^{-1}(u)\, du ,$$

corresponding to

$$J(u) = \begin{cases} (1-2\alpha)^{-1} & \alpha < u < 1-\alpha \\ 0 & \text{otherwise.} \end{cases}$$

This functional induces (approximately) the estimator

$$\hat{\mu}_\alpha = T(\hat{F}_n) = \frac{1}{n-2\lfloor \alpha n \rfloor} \sum_{i=\lceil \alpha n \rceil}^{n-\lfloor \alpha n \rfloor} X_{(i)} .$$

- The influence function of the α-trimmed mean at x_0 is found as follows: We consider the different pieces of

$$T_F^{(1)}(\delta_{x_0}-F) = \int_{-\infty}^{x_0} J(F(y))\, dy + \int_{-\infty}^{\infty} J(F(y))\,[F(y)-1]\, dy$$

under the α-trimmed mean choice of $J(u)$. We have

$$\int_{-\infty}^{\infty} J(F(y))\, F(y)\, dy = \frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \underbrace{F(y)}_{v}\, \underbrace{dy}_{du}$$

$$= \frac{1}{1-2\alpha}\left[y\cdot F(y)\Big|_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} - \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} y\cdot f(y)\, dy \right]$$

$$\mu_\alpha = \frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} y\, f(y)\, dy$$

$$= \frac{1}{1-2\alpha}\left[F^{-1}(1-\alpha)(1-\alpha) - F^{-1}(\alpha)\alpha\right] - \mu_\alpha .$$

Next,

$$\int_{-\infty}^{\infty} J(F(y))\,dy = \frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} 1 \quad dy = \frac{1}{1-2\alpha}\left[F^{-1}(1-\alpha) - F^{-1}(\alpha)\right],$$

so that

$$\begin{aligned}\int_{-\infty}^{\infty} J(F(y))\left[F(y)-1\right] dy &= \frac{1}{1-2\alpha}\left[F^{-1}(1-\alpha)(1-\alpha) - F^{-1}(\alpha)\alpha\right] - \mu_\alpha \\ &\quad - \frac{1}{1-2\alpha}\left[F^{-1}(1-\alpha) - F^{-1}(\alpha)\right] \\ &= \frac{1}{1-2\alpha}\left[F^{-1}(\alpha) - \underbrace{\left((1-2\alpha)\mu_\alpha - \alpha\left(F^{-1}(1-\alpha) - F^{-1}(\alpha)\right)\right)}_{=:\,\tilde{\mu}_\alpha.}\right] \\ &= \frac{1}{1-2\alpha}\left[F^{-1}(\alpha) - \tilde{\mu}_\alpha\right],\end{aligned}$$

where

$$\tilde{\mu}_\alpha = (1-2\alpha)\mu_\alpha - \alpha\left(F^{-1}(1-\alpha) - F^{-1}(\alpha)\right).$$

Now we have

$$\int_{-\infty}^{x_0} J(F(y))\,dy = \begin{cases} 0, & x_0 < F^{-1}(\alpha) \\ \dfrac{x_0 - F^{-1}(\alpha)}{1-2\alpha}, & F^{-1}(\alpha) < x_0 < F^{-1}(1-\alpha) \\ \dfrac{F^{-1}(1-\alpha) - F^{-1}(\alpha)}{1-2\alpha}, & F^{-1}(1-\alpha) < x_0. \end{cases}$$

Putting everything together, we have the influence function

$$T_F^{(1)}(\delta_{x_0} - F) = \begin{cases} \frac{1}{1-2\alpha}\left[F^{-1}(\alpha) - \tilde{\mu}_\alpha\right], & x_0 < F^{-1}(\alpha) \\ \frac{1}{1-2\alpha}\left[x_0 - \tilde{\mu}_\alpha\right], & F^{-1}(\alpha) < x_0 < F^{-1}(1-\alpha) \\ \frac{1}{1-2\alpha}\left[F^{-1}(1-\alpha) - \tilde{\mu}_\alpha\right], & F^{-1}(1-\alpha) < x_0, \end{cases}$$

and

$$\begin{aligned} \mathrm{Var}\left(T_F^{(1)}(\delta_{X_1} - F)\right) &= \left(\frac{1}{1-2\alpha}\right)^2 \Bigg\{\alpha\left[F^{-1}(\alpha) - \tilde{\mu}_\alpha\right]^2 + \alpha\left[F^{-1}(1-\alpha) - \tilde{\mu}_\alpha\right]^2 \\ &\qquad + \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} (x - \tilde{\mu}_\alpha)^2 \, dF(x) \Bigg\} \\ &=: \vartheta. \end{aligned}$$

- The von Mises expansion is

$$\sqrt{n}\left(\hat{\mu}_\alpha - \mu_\alpha\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_F^{(1)}(\delta_{X_i} - F) + \sqrt{n}\, R(\hat{F}_n - F),$$

where

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_F^{(1)}(\delta_{X_i} - F) \xrightarrow{D} N(0, \vartheta) \quad \text{as} \quad n \to \infty.$$

M-Estimators:

Let $X_1, \dots, X_n$ be iid with distribution F.

Consider

$$\theta_0 = T(F) = \text{the value of } t \text{ which solves } \int \psi(x,t)\, dF(x) = 0$$

$$\hat{\theta}_n = T(\hat{F}_n) = \text{the value of } t \text{ which solves } \int \psi(x,t)\, d\hat{F}_n(x) = 0$$

- The von Mises derivative is found as follows: Define

$$\int \psi(x, T(F_\varepsilon))\, dF_\varepsilon(x) = 0 .$$

Then

$$0 = \frac{d}{d\varepsilon} \int \psi(x, T(F_\varepsilon))\, dF_\varepsilon(x) \Big|_{\varepsilon=0}$$

$$= \frac{d}{d\varepsilon} \int \psi(x, T(F_\varepsilon))\, d\big(F + \varepsilon(G-F)\big)(x) \Big|_{\varepsilon}$$

$$= \Big[\frac{d}{d\varepsilon} \int \psi(x, T(F_\varepsilon))\, dF(x) + \int \psi(x, T(F_\varepsilon))\, d(G-F)(x)$$

Set $\lambda_F(t) = \int \psi(x,t)\, dF(x)$

$$+ \varepsilon \frac{d}{d\varepsilon} \int \psi(x, T(F_\varepsilon))\, d(G-F)(x) \Big] \Big|_{\varepsilon=0}$$

$$= \lambda_F'(T_F) \underbrace{\frac{d}{d\varepsilon} T(F_\varepsilon) \Big|_{\varepsilon=0}}_{T_F^{(1)}(G-F)} + \underbrace{\int \psi(x, T(F))\, dG(x)}$$

$$= \lambda_F'(T_F)\, T_F^{(1)}(G-F) + \lambda_G(T(F))$$

So we have

$$T_F^{(1)}(G-F) = \frac{-\lambda_G(T(F))}{\lambda'_F(T_F)} .$$

- The influence function at x_0 is

$$T_F^{(1)}(\delta_{x_0}-F) = \frac{-\lambda_{\delta_{x_0}}(T(F))}{\lambda'_F(T_F)} = -\frac{\psi(x_0, T(F))}{\lambda'_F(T_F)} ,$$

and

$$\mathrm{Var}\left(T_F^{(1)}(\delta_{X_1}-F)\right) = \frac{\int \psi^2(x, T(F))\, dF(x)}{\left[\lambda'_F(T_F)\right]^2} .$$

- We can write the von Mises expansion

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi(X_i, T_F)}{\lambda'_F(T_F)} + \sqrt{n}\, R(\hat{F}_n - F)$$

where

$$-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi(X_i, T_F)}{\lambda'_F(T_F)} \xrightarrow{D} N\left(0, \frac{\int \psi^2(x, T(F))\, dF(x)}{\left[\lambda'_F(T(F))\right]^2}\right).$$

The behavior of $\sqrt{n}\, R(\hat{F}_n - F)$ depends on the function ψ.

Example of M-estimators: Maximum likelihood estimators

For $X_1, \dots, X_n$ iid with pdf or pmf $f(x;\theta)$, let

$$\ell_n(\theta; X_1, \dots, X_n) = \sum_{i=1}^{n} \log f(X_i;\theta).$$

Then the MLE $\hat{\theta}_n$ can be written as the solution to

$$0 = \int \psi(x,t)\, d\hat{F}_n(x), \qquad \text{with} \qquad \psi(x,t) = \frac{\partial}{\partial\theta} \log f(x;\theta)\Big|_{\theta=t}.$$

The corresponding "population" quantity θ_0 is the solution to

$$0 = \int \psi(x,t)\, dF(x).$$

According to our previous work, we have the expansion

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} \log f(X_i;\theta)\Big|_{\theta=\theta_0}}{\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(X_1;\theta)\Big|_{\theta=\theta_0}\right]} + \sqrt{n}\, R(\hat{F}_n - F),$$

where

$$-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} \log f(X_i;\theta)\Big|_{\theta=\theta_0}}{\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(X_i;\theta)\Big|_{\theta=\theta_0}\right]} \xrightarrow{D} N(0, v),$$

with

$$v = \frac{\mathbb{E}\left(\frac{\partial}{\partial\theta} \log f(X_1;\theta)\Big|_{\theta=\theta_0}\right)^2}{\left(\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(X_i;\theta)\Big|_{\theta=\theta_0}\right]\right)^2} \overset{\text{pg 146 of Serfling}}{=} \frac{1}{\left(\mathbb{E}\, \frac{\partial}{\partial\theta} \log f(X_1;\theta)\Big|_{\theta=\theta_0}\right)^2}$$

under typical maximum likelihood regularity conditions.

POWERFUL RESULTS

Under a type of differentiability called Hadamard differentiability, which we define later, the remainder term in the von Mises expansion of $\sqrt{n}\,(T(\hat{F}_n) - T(F))$ will vanish, giving the following nice result.

<u>Theorem</u> (CLT for Hadamard differentiable functionals):

If T is Hadamard differentiable then

(i) $$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) \longrightarrow N(0, \vartheta) \text{ in dist. as } n \to \infty$$

with $$\vartheta = \int \left[T_F^{(1)}(\delta_x - F)\right]^2 dF(x).$$

(ii) $$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right)/\sqrt{\hat{\vartheta}} \to N(0,1) \text{ in dist as } n \to \infty,$$

with $$\hat{\vartheta} = \int \left[T_{\hat{F}_n}^{(1)}(\delta_x - \hat{F}_n)\right]^2 d\hat{F}_n(x).$$

Note that result (ii) of the theorem gives that

$$T(\hat{F}_n) \pm z_{\alpha/2}\sqrt{\hat{\vartheta}/n}$$

will have coverage probability converging to $1-\alpha$ as $n \to \infty$.

<u>Examples of $\hat{\vartheta}$:</u>

- <u>The mean:</u>

We have $$T_F^{(1)}(\delta_{x_0} - F) = \int x \, d(\delta_{x_0} - F)(x) = x_0 - \mu$$

$$T_{\hat{F}_n}^{(1)}(\delta_{x_0} - \hat{F}_n) = \int x \, d(\delta_{x_0} - \hat{F}_n)(x) = x_0 - \bar{x}_n,$$

so that

$$\begin{aligned}\hat{\vartheta} &= \int \left[T^{(1)}_{\hat{F}_n}(\delta_x - \hat{F}_n)\right]^2 d\hat{F}_n(x)\\ &= \int (x-\bar{x}_n)^2 d\hat{F}_n(x)\\ &= \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2.\\ &=: \hat{\sigma}_n^2.\end{aligned}$$

- <u>Smooth function of the mean</u>:

We have $T^{(1)}_F(\delta_{x_0} - F) = g'\left(\int x\, dF(x)\right)\left(x_0 - \int x\, dF(x)\right)$

$$\begin{aligned}T^{(1)}_{\hat{F}_n}(\delta_{x_0} - \hat{F}_n) &= g'\left(\int x\, d\hat{F}_n(x)\right)\left(x_0 - \int x\, d\hat{F}_n(x)\right)\\ &= g'(\bar{X}_n)(x_0 - \bar{X}_n),\end{aligned}$$

so that

$$\begin{aligned}\hat{\vartheta} &= \int \left[T^{(1)}_{\hat{F}_n}(\delta_x - \hat{F}_n)\right]^2 d\hat{F}_n(x)\\ &= \int \left[g'(\bar{X}_n)(x-\bar{X}_n)\right]^2 d\hat{F}_n(x)\\ &= \left[g'(\bar{X}_n)\right]^2 \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2\\ &= \left[g'(\bar{X}_n)\right]^2 \hat{\sigma}_n^2.\end{aligned}$$

<u>**Theorem**</u> (Bootstrap works for Hadamard differentiable functionals):

If T is Hadamard differentiable and $\vartheta = \int [T_F^{(1)}(\delta_x - F)]^2 dF(x) < \infty$, then

$$\sup_{x \in \mathbb{R}} \left| P_*\left(\sqrt{n}\left(T(\hat{F}_n^*) - T(\hat{F}_n)\right) \le x\right) - P\left(\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) \le x\right) \right| \to 0$$

in probability as $n \to \infty$.

In the above, we construct the bootstrap version $\hat{F}_n^*$ of $\hat{F}_n$ as

$$\hat{F}_n^* = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^*}, \quad \text{where} \quad X_1^*, \dots, X_n^* \mid X_1, \dots, X_n \overset{\text{ind}}{\sim} \hat{F}_n .$$

We now give the definition of Hadamard differentiability:

<u>Hadamard differentiability</u>

Let $\mathbb{D}$ be the space of linear combinations of probability distributions.

A functional $T: \mathbb{D} \to \mathbb{R}$ is Hadamard differentiable at $F \in \mathbb{D}$ in the direction $G \in \mathbb{D}$ if there exists a linear function $T_F^{(1)}: \mathbb{D} \to \mathbb{R}$ such that

$$\lim_{n \to \infty} \left| \frac{T(F + \varepsilon_n(G_n - F)) - T(F)}{\varepsilon_n} - T_F^{(1)}(G - F) \right| = 0$$

for every sequence $\{G_n\}$ such that $\|G_n - G\|_\infty \to 0$ as $n \to \infty$
and every sequence $\varepsilon_n \downarrow 0$ as $n \to \infty$.