Stat210B: Theoretical Statistics **Lecture Date: April 24, 2007**

Functional Delta Method

Lecturer: Michael I. Jordan *Scribe: Ariel Kleiner*

As seen in the previous lecture, many statistics can be written as functionals $\phi(P)$. Thus, it is natural to consider plug-in estimators of the form $\phi(\mathbb{P}_n)$. The functional delta method provides a way of approaching the asymptotics of such estimators. Heuristically, we would like to obtain an expansion of the following form (note that we assume linearity of ϕ'_P):

$$\begin{aligned}\phi(\mathbb{P}_n) - \phi(P) &= \frac{1}{\sqrt{n}}\phi'_P(\mathbb{G}_n) + R_n \\ &= \frac{1}{n}\sum_i \phi'_P(\delta_{x_i} - P) + R_n\end{aligned}$$

This approach suggests using the *Gateaux derivative*:

$$\begin{aligned}\phi'_P(\delta_x - P) &= \left.\frac{d}{dt}\right|_{t=0} \phi((1-t)P + t\delta_x) \\ &= IF_{\phi,P}(x) \\ &\equiv \text{Influence Function}\end{aligned}$$

That is, we have defined a partial derivative along one direction. If we can exchange integration and differentiation, then (again, heuristically), we expect that

$$\int \phi'_P(\delta_x - P)dP = 0.$$

Furthermore, if $\sqrt{n}R_n \xrightarrow{P} 0$, then, defining

$$\gamma^2 = \int IF_{\phi,P}(x)^2 dP,$$

we have from the Central Limit Theorem that

$$\sqrt{n}(\phi(\mathbb{P}_n) - \phi(P)) \xrightarrow{d} N(0, \gamma^2).$$

Definition 1. $\lambda^* \equiv \sup_x |IF_{\phi,P}(x)|$ is the *gross error sensitivity.*

Ideally, $\lambda^* < \infty$.

Example 2 (nonparametric mean)**.** Consider

$$\phi(F) = E_F[X] = \mu$$

and observe that

$$\begin{aligned}IF_{\phi,F}(x) &= \left.\frac{d}{dt}\right|_{t=0} \phi((1-t)F + t\delta_x) \\ &= x - \mu\end{aligned}$$

It is clear that $\lambda^* = \infty$ (i.e., the sample mean is not a robust estimator). Now, note that

$$\int IF_{\phi,F}(x)dF(x) = 0$$

and, defining $\hat{F}_n$ to be the empirical distribution function,

$$\begin{aligned}\phi(\hat{F}_n) - \phi(F) &= \bar{X}_n - \mu \\ &= \frac{1}{n}\sum_{i=1}^{n} IF_{\phi,F}(x_i)\end{aligned}$$

As a result, $R_n = 0, \forall n$. Additionally,

$$\gamma^2(F) = E_F(X-\mu)^2 = \mathrm{Var}(X)$$

and so

$$\sqrt{n}(\phi(\hat{F}_n) - \phi(F)) \xrightarrow{d} N(0, \mathrm{Var}(X))$$

by the Central Limit Theorem.

Example 3 (Cramér – von Mises)**.** Given some distribution F_0, let

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

Additionally, let $\delta_x(t) = 1_{t\geq x}$ and consider

$$\int ((1-t)F(z) + t\delta_x(z) - F_0(z))^2 dF_0(z).$$

Noting that we need only consider terms that are linear in t, it follows that

$$IF_{\phi,F}(x) = 2\int (F(z)-F_0(z))(\delta_x(z) - F(z))dF_0(z)$$

and so $\lambda^* \leq 2$ ($\Rightarrow$ robustness). Now,

$$\begin{aligned}R_n &= \int (\hat{F}_n - F_0)^2 dF_0 - \int (F-F_0)^2 dF_0 \\ &\quad -\frac{2}{n}\int \sum_i (F-F_0)(\delta_{x_i} - F)dF_0 \\ &= \int (\hat{F}_n - F)^2 dF_0 \\ &\leq \sup_x (\hat{F}_n(x) - F(x))^2.\end{aligned}$$

By Donsker (recall that F is the true underlying distribution) and the fact that the Brownian bridge is tight,

$$\sqrt{n}\sup_x |\hat{F}_n(x) - F(x)| = O_P(1).$$

Therefore, squaring,

$$n\sup_x (\hat{F}_n(x) - F(x))^2 = O_P(1)$$

and so

$$\sqrt{n}R_n \xrightarrow{P} 0,$$

thus yielding asymptotic normality.

If $F = F_0$, then the influence function is zero. In that case, we must consider higher-order terms in the von Mises expansion; thus, we expect a chi-square asymptotic distribution.

Example 4 (Mann-Whitney)**.** Given $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} (F, G)$, consider

$$T_n = \frac{1}{n^2} \sum_i \sum_j 1_{X_i \le Y_j}.$$

Observe that

$$P(X \le Y) = \int F dG$$

and

$$T_n = \int \hat{F}_n d\hat{G}_n.$$

We hope to be able to write

$$\sqrt{n}(T_n - \int F dG) \xrightarrow{d} N(0, \gamma^2(F, G)).$$

Now,

$$\begin{aligned} IF_{\phi,F,G}(x,y) &= \left.\frac{d}{dt}\right|_{t=0} \int ((1-t)F + t\delta_x) d((1-t)G + t\delta_y) \\ &= -2\int F dG + F(y) + 1 - G_-(x) \end{aligned}$$

Further calculation yields $E[IF] = 0$ as well as the value of γ^2. Additionally, $\sqrt{n}R_n$ goes to 0 in probability by the U-statistic theorem proved in previous lectures.

Example 5 (Z-estimators)**.** By definition, the true parameter value is a zero of $P\psi_\theta$, and so we define

$$P\psi_{\phi(P)} = 0.$$

We now compute the influence function via implicit differentiation. First, plugging the perturbed measure into the defintion of $\phi(P)$ above, we obtain

$$0 = ((1-t)P + t\delta_x)\psi_{\phi((1-t)P + t\delta_x)}.$$

Now, differentiating with respect to t at $t = 0$, we have

$$0 = (1-t)\left.\left(\frac{\partial}{\partial\theta} P\psi_\theta\right)\right|_{\theta=\phi(P)} \left.\left(\frac{d}{dt}\phi((1-t)P + t\delta_x)\right)\right|_{t=0} - P\psi_{\phi((1-t)P+t\delta_x)}\big|_{t=0} + \psi_{\phi((1-t)P+t\delta_x)}(x)\big|_{t=0}.$$

Setting all remaining t's in the above expression equal to zero,

$$0 = \left.\left(\frac{\partial}{\partial\theta} P\psi_\theta\right)\right|_{\theta=\phi(P)} IF_{\phi,P}(x) + \psi_{\phi(P)}(x).$$

Finally, solving for the influence function,

$$IF_{\phi,P}(x) = \left(-\frac{\partial}{\partial\theta} P\psi_\theta\Big|_{\theta=\phi(P)}\right)^{-1} \psi_{\phi(P)}(x).$$

Note that because we have freedom in choosing ψ, we can select it to be bounded, thus yielding a bounded influence function (i.e., we can construct the influence function by choosing ψ).

To control the remainder terms in the von Mises expansion, we must return to the arguments used to show asymptotic normality of Z-estimators in previous lectures (i.e., the above derivation has simply been a reorganization of our thinking).

We now introduce two other types of functional derivatives.

Definition 6 (Hadamard derivative)**.** Given normed linear spaces $\mathbb{D}_\phi, \mathbb{E}$, a map $\phi : \mathbb{D}_\phi \to \mathbb{E}$ is *Hadamard differentiable* if

$$\left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \phi'_\theta(h) \right\|_{\mathbb{E}} \to 0, \qquad t \downarrow 0$$

for some linear continuous operator ϕ'_θ, for all sequences $h_t \to h$.

Definition 7 (Fréchet derivative)**.** A map $\phi : \mathbb{D} \to \mathbb{E}$ is *Fréchet differentiable* if

$$\|\phi(\theta + h) - \phi(\theta) - \phi'_\theta(h)\|_{\mathbb{E}} = o(\|h\|_{\mathbb{D}})$$

as $\|h\|_{\mathbb{D}} \to 0$.

References

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.