Intuituion for Influence Function and von Mises Expansion

Yi-Cheng, Tai

National Yang Ming Chiao Tung University

February 23, 2023

Introduction

We summarized the intuition and motivation provided by Fisher and Kennedy (2021)[1] to illustrate the concept of influence function and its application.

The definition of influence function is motivated by the chain rule for multivariable function. They also provide the analogy with von Mises expansion and ordinary Taylor expansion.

Estimand as Functional of distribution

The parameter of our interest (estimand) could be represented as a functional of distribution, denotes as $\Psi(P)$, where P is a probability measure.

e.g. Let (Y_1, Y_2) be a bivariate random vector with probability measure P.

$$\Psi_1(P) = E_P(Y_1)$$
$$\Psi_2(P) = Cov(Y_1, Y_2) = E_P(Y_1 Y_2) - E_P(Y_1)E_P(Y_2)$$
$$\Psi_3(P) = E_P(Y_1|Y_2 = y), \text{ give } y$$

Plug-In Estimator

In non-parametric setting, we may estimate P with the empirical measure $\hat{P}_n$. Then the plug-in estimator of $\Psi(P)$ would be $\Psi(\hat{P}_n)$.

Usually, as sample size increases, the empirical measure gets closer to the true distribution. How do we quantify the improvement on the estimation with respect to the change of probability measure?

In ordinary Calculus, the change of y w.r.t the change of x at x_0, could be captured by derivative $\frac{dy}{dx}|_{x=x_0}$. The influence function plays the role of the derivative of Ψ w.r.t P.

Influence Function

Given a probability measure $\tilde{P}$ with density $\tilde{p}$, we define a path from P to $\tilde{P}$ as the probability measures P_ε with density

$$p_\varepsilon = p + \varepsilon(\tilde{p} - p), \varepsilon \in [0, 1]$$

Then the estimands on this curve can be denoted as a function defined on $[0, 1]$

$$\varepsilon \to \Psi(P_\varepsilon), \varepsilon \in [0, 1]$$

If we want to investigate the derivative of Ψ w.r.t P at $\tilde{P}$ along the path defined, we may want to calculate

$$\left.\frac{d\Psi(P_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=1}$$

Influence Function (Cont. 1)

As a motivation, we may consider the discrete distributions on support $\{o_1, \ldots, o_K\}$. Then by chain rule,

$$
\begin{aligned}
\left.\frac{d\Psi(P_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=1} &= \sum_{j=1}^{K} \left.\frac{\partial\Psi(P_\varepsilon)}{\partial p_\varepsilon(o_j)}\right|_{\varepsilon=1} \left.\frac{\partial p_\varepsilon(o_j)}{\partial\varepsilon}\right|_{\varepsilon=1} \\
&= \sum_{j=1}^{K} \left.\frac{\partial\Psi(P_\varepsilon)}{\partial p_\varepsilon(o_j)}\right|_{\varepsilon=1} \lim_{\varepsilon\to 1} \frac{p_\varepsilon(o_j) - \tilde{p}(o_j)}{\varepsilon - 1} \\
&= \sum_{j=1}^{K} \left.\frac{\partial\Psi(P_\varepsilon)}{\partial p_\varepsilon(o_j)}\right|_{\varepsilon=1} [\tilde{p}(o_j) - p(o_j)]
\end{aligned}
$$

The interpretation of $\left.\frac{\partial\Psi(P_\varepsilon)}{\partial p_\varepsilon(o_j)}\right|_{\varepsilon=1}$ seems to capture the desired meaning of influence function.

Influence Function (Cont. 2)

However, since there is a constraint on the mass function,

$$\sum_{j=1}^{K} p(o_j) = 1$$

The mass function p_ε could not be treated as variation-independent and $\frac{\partial \Psi(P_\varepsilon)}{\partial p_\varepsilon(o_j)}\Big|_{\varepsilon=1}$ is ill-defined.

Here is a formal definition of influence function and preserve the intuition obtained above:

The influence function $\phi(y, P)$ satisfies

$$\frac{\partial \Psi(P_\varepsilon)}{\partial \varepsilon}\Bigg|_{\varepsilon=1} = \int \phi(y, \tilde{P}) d(\tilde{P} - P)(y)$$

and

$$E_P(\phi(Y, P)) = 0$$

von Mises Expansion

By ordinary Taylor expansion,

$$\begin{aligned}\Psi(P) &= \Psi(\tilde{P}) + \left.\frac{\partial \Psi(P_\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=1}(0-1) + R_2(\tilde{P}, P) \\ &= \Psi(\tilde{P}) - \int \phi(y, \tilde{P}) d(\tilde{P} - P)(y) + R_2(\tilde{P}, P)\end{aligned}$$

It is referred to as distributional Taylor expansion or von Mises expansion of Ψ.

1-step Estimator

Since the influence function ϕ satisfies

$$E_P(\phi(Y, P)) = 0$$

The von Mises expansion in this case can be written as

$$\begin{aligned}
\Psi(P) &= \Psi(\tilde{P}) + \int \phi(y, \tilde{P}) dP(y) + R_2(\tilde{P}, P) \\
&= \Psi(\tilde{P}) + E_P(\phi(Y, \tilde{P})) + R_2(\tilde{P}, P) \\
&\approx \Psi(\tilde{P}) + n^{-1} \sum_{i=1}^{n} \phi(Y_i, \tilde{P}) + R_2(\tilde{P}, P)
\end{aligned}$$

$\hat{\Psi} = \Psi(\tilde{P}) + n^{-1} \sum_{i=1}^{n} \phi(Y_i, \tilde{P})$ is the so called 1-step estimator of $\Psi(P)$. The convergence of 1-step estimator $\hat{\Psi}$ depends on the average of i.i.d sample and remainder term.

Convergence of Remainder Term

The convergence of average of i.i.d sample could be proved by C.L.T. The convergence of R_2 is usually done on a case-by-case basis. It depends on two aspects:

- the distance between $\tilde{P}$ and P (convergence of $\tilde{P} \to P$)
- the smoothness of the functional Ψ

We focus on how to quantify the smoothness of Ψ. Note that when $\tilde{P}$ is far from P, the resulting 1-step estimator is usually closer to $\Psi(P)$ and the concept of smoothness would depend on the initial choice of $\tilde{P}$. To eliminate the dependency on $\tilde{P}$, we need to re-scale the step size in all possible paths.

$$P_\Delta = P + \frac{\Delta}{\|\tilde{P} - P\|_2}(\tilde{P} - P), \Delta \in [0, \|\tilde{P} - P\|_2]$$

Smoothness of Functional

Consider the von Mises expansion along the re-scaled path:

$$
\begin{aligned}
\Psi(P) &= \Psi(\tilde{P}) + \left.\frac{\partial \Psi(P_\Delta)}{\partial \Delta}\right|_{\Delta=\|\tilde{P}-P\|_2} (0 - \|\tilde{P} - P\|_2) \\
&\quad + \left.\frac{\partial^2 \Psi(P_\Delta)}{\partial \Delta^2}\right|_{\Delta=\Delta^*} (0 - \|\tilde{P} - P\|_2)^2 \\
&= \Psi(\tilde{P}) - \int \phi(y, \tilde{P}) d(\tilde{P} - P)(y) + \left.\frac{\partial^2 \Psi(P_\Delta)}{\partial \Delta^2}\right|_{\Delta=\Delta^*} \|\tilde{P} - P\|_2^2 \\
&= \Psi(\tilde{P}) + \int \phi(y, \tilde{P}) dP(y) + \left.\frac{\partial^2 \Psi(P_\Delta)}{\partial \Delta^2}\right|_{\Delta=\Delta^*} \|\tilde{P} - P\|_2^2
\end{aligned}
$$

, where $\Delta^* \in [0, \|\tilde{P} - P\|_2]$

Smoothness of Functional (Cont.)

Smoothness Condition:

$$\left.\frac{\partial^2 \Psi(P_\Delta)}{\partial \Delta^2}\right|_{\Delta=\Delta^*} = O(1) \text{ as } \Delta^* \to 0$$

Then,

$$\Psi(P) = \Psi(\tilde{P}) + \int \phi(y, \tilde{P}) dP(y) + O(\|\tilde{P} - P\|_2^2)$$

Thus,

$$\hat{\Psi} - \Psi(P) = O(\|\tilde{P} - P\|_2^2)$$

Hence, we may use the derivatives of Ψ w.r.t Δ to control the convergence of remainder term.

$$\left.\frac{\partial^j \Psi(P_\Delta)}{\partial \Delta^j}\right|_{\Delta=\Delta^*} = O(1) \text{ as } \Delta^* \to 0$$

Appendix: Score-Based Definition of the Influence Function

Consider the score function of P_ε at $\tilde{P}$.

$$\begin{aligned}
\tilde{s} &= \frac{\partial}{\partial \varepsilon} \log\{dP + \varepsilon(d\tilde{P} - dP)\}\Big|_{\varepsilon=1} \\
&= \frac{d\tilde{P} - dP}{dP + \varepsilon(d\tilde{P} - dP)}\bigg|_{\varepsilon=1} \\
&= \frac{d\tilde{P} - dP}{d\tilde{P}}
\end{aligned}$$

Appendix: Score-Based Definition of the Influence Function (Cont.)

Then, the influence function could be defined as a mean-zero random function ϕ satisfying the following equation:

$$\begin{aligned}
\left.\frac{\partial \Psi(P)}{\partial \varepsilon}\right|_{\varepsilon=1} &= \int \phi(y, \tilde{P}) d(\tilde{P} - dP)(y) \\
&= \int \phi(y, \tilde{P}) \frac{d\tilde{P} - dP}{d\tilde{P}}(y) d\tilde{P}(y) \\
&= \int \phi(y, \tilde{P}) \tilde{s}(y) d\tilde{P}(y) \\
&= E_{\tilde{P}} \left[\phi(Y, \tilde{P}) \tilde{s}(Y) \right]
\end{aligned}$$

This representation provides the geometric intuition for influence function.

References

A. Fisher and E. H. Kennedy.
Visually communicating and teaching intuition for influence functions.
The American Statistician, 75(2):162–172, 2021.