June 8, 2023

Gâteaux derivative in direction of the Dirac measure

Anders Munch

1 Distribution function

Recall the previous problem which we solved with a kernel density estimator, i.e., estimation of the parameter

$$\Psi(P) = P(X \leq x), \quad \text{for some fixed} \quad x \in \mathbb{R}.$$

Note that we can write

$$\Psi(P) = \mathbb{E}_P \left[\mathbb{1}_{(-\infty,x]}(X)\right] = \int \mathbb{1}_{(-\infty,x]}(z) \,\mathrm{d}P(z).$$

For some fixed $x_i \in \mathbb{R}$, calculate

$$\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \Psi(P + \varepsilon\delta_{x_i}).$$

2 Average treatment effect

We consider data on the form $O = (Y, A, X) \sim P$ where $A \in \{0, 1\}$ is a binary treatment indicator. The parameter of interest is

$$\Psi(P) = \int f_P(1, X)\mu_P(\mathrm{d}x),$$

where

$$f_P(a, x) = \mathbb{E}_P [Y \mid A = a, X = x], \quad \text{and} \quad \mu_P(\mathrm{d}x) = P(X \in \mathrm{d}x),$$

We want to calculate the directional derivative in the direction of the Dirac measure at $o_i = (y_i, a_i, w_i)$, i.e.,

$$\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \Psi(P_\varepsilon), \quad \text{where} \quad P_\varepsilon = P + \varepsilon\delta_{o_i}. \tag{1}$$

For simplicity we assume that Y and X are discrete. This means that we can write the measure P as the probability mass function

$$p(y, a, x) = P(Y = y, A = a, X = x),$$

and the Dirac measure at $o_i = (y_i, a_i, x_i)$ as the indicator function

$$\mathbb{1}_{o_i}(o) = \begin{cases} 1 & \text{if } o_i = o, \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$p_\varepsilon(o) = p(o) + \varepsilon\mathbb{1}_{o_i}(o), \quad \text{and} \quad \Psi(P) = \sum_x f_P(1, x)p(x).$$

We calculate (1) in a sequence of steps. The following two properties are useful:

$$\mathbb{1}_{o_i}(o) = \mathbb{1}_{y_i}(y)\mathbb{1}_{a_i}(a)\mathbb{1}_{x_i}(x), \tag{2}$$

$$\sum_o h(o)\mathbb{1}_{o_i}(o) = h(o_i), \quad \text{for any function} \quad h. \tag{3}$$

2.1

Use the relations

$$p(x) = \sum_y \sum_a p(y, a, x),$$

and

$$p(a, x) = \sum_y p(y, a, x),$$

and equation (2) to argue that

$$p_\varepsilon(x) = p(x) + \varepsilon \mathbb{1}_{x_i}(x).$$

$$p_\varepsilon(x, a) = p(x, a) + \varepsilon \mathbb{1}_{x_i}(x)\mathbb{1}_{a_i}(a).$$

2.2

Recall the product rule

$$\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \{h(\varepsilon)g(\varepsilon)\} = h'(0)g(0) + h(0)g'(0),$$

and use (3) to argue that

$$\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \Psi(P_\varepsilon) = \sum_x \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} f_{P_\varepsilon}(1, x)p(x) + f_P(1, x_i).$$

2.3

Recall the quotient rule

$$\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \left\{\frac{h(\varepsilon)}{g(\varepsilon)}\right\} = \frac{h'(0)g(0) - h(0)g'(0)}{g(0)^2}$$

and use the relation

$$P(Y = y \mid A = a, X = x) = \frac{p(y, a, x)}{p(a, x)},$$

together with the results from exercise 2.1 and equation (2) to argue that

$$\begin{aligned}
&\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} P_\varepsilon(Y = y \mid A = a, X = x) \\
&= \frac{\mathbb{1}_{y_i,a_i,x_i}(y,a,x)p(a,x) - p(y,a,x)\mathbb{1}_{a_i,x_i}(a,x)}{p(a,x)^2} \\
&= \frac{\mathbb{1}_{y_i,a_i,x_i}(y,a,x) - P(Y = y \mid A = a, X = x)\mathbb{1}_{a_i,x_i}(a,x)}{p(a,x)} \\
&= \{\mathbb{1}_{y_i}(y) - P(Y = y \mid A = a, X = x)\} \frac{\mathbb{1}_{a_i}(a)\mathbb{1}_{x_i}(x)}{p(a,x)}
\end{aligned}$$

2.4

Use the relation

$$\mathbb{E}_P\left[Y \mid A = a, X = x\right] = \sum_y yP(Y = y \mid A = a, X = x),$$

the previous exercise, and equation (3) to argue that

$$\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} f_{P_\varepsilon}(1,x) = \{y_i - f_P(1,x)\} \frac{a_i\mathbb{1}_{x_i}(x)}{p(1,x)}.$$

2.5

Define

$$\pi(x) = P(A = 1 \mid X = x)$$

and use the relation

$$P(A = a \mid X = x) = \frac{P(A = a, X = x)}{P(X = x)} = \frac{p(a,x)}{p(x)},$$

and the results from the previous exercise and exercise 2.2 to argue that

$$\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \Psi(P_\varepsilon) = \{y_i - f_P(1,x_i)\} \frac{a_i}{\pi(x_i)} + f_P(1,x_i).$$