Theoretical Statistics. Lecture 19.

Peter Bartlett

1. Functional delta method. [vdV20]
2. Differentiability in normed spaces:
   Hadamard derivatives. [vdV20]
3. Quantile estimates. [vdV21]

Recall: Delta method

Theorem: If $\phi : \mathbb{R}^k \to \mathbb{R}^m$ is differentiable at θ, and $\sqrt{n}(T_n - \theta) \rightsquigarrow T$, then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(T)$$

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) - \phi'_\theta(\sqrt{n}(T_n - \theta)) \xrightarrow{P} 0.$$

Here, ϕ'_θ is the derivative (linear map) satisfying

$$\phi(\theta + h) - \phi(\theta) = \phi'_\theta(h) + o(\|h\|)$$

for $h \to 0$.

Functional delta method

- What about more complex cases? For instance, what if we are interested in a property of the probability distribution $\phi(P)$, and we use an estimator $\phi(P_n)$?
- As a first example, we'll get some intuition by considering a "Taylor series" expansion about the value $\phi(P)$. But what does it mean for the Taylor series expansion of ϕ to exist?
- To make this rigorous, we need to consider ϕ as a map between normed linear spaces, and we need it to be appropriately differentiable.
- The right notion is *Hadamard differentiability*.

Functional delta method

Given $X_1, \ldots, X_n$ from a distribution P, and a parameter of interest $\phi(P)$, suppose that we estimate $\phi(P)$ by $\phi(P_n)$, where P_n is the empirical distribution. What is the asymptotic behavior of $\phi(P_n)$?

Define the derivative $\phi_P^{(k)}(H)$ of the map $t \mapsto \phi(P + tH)$ at $t = 0$, where H is a perturbation direction. Then if the derivatives exist we have a Taylor series expansion:

$$\phi(P+tH)-\phi(P) = t\phi'_P(H)+\frac{1}{2}t^2\phi_P^{(2)}(H)+\cdots+\frac{1}{m!}t^m\phi_P^{(m)}(H)+o(t^m).$$

Functional delta method

Substituting $t = 1/\sqrt{n}$ and $H = \mathbb{G}_n (= \sqrt{n}(P_n - P)X)$,

$$tH = \frac{1}{\sqrt{n}}\mathbb{G}_n = P_n - P,$$

gives the *von Mises expansion*,

$$\phi(P_n) - \phi(P) = \frac{1}{\sqrt{n}}\phi'_P(\mathbb{G}_n) + \frac{1}{2n}\phi_P^{(2)}(\mathbb{G}_n) + \cdots + \frac{1}{m!n^{m/2}}\phi_P^{(m)}(\mathbb{G}_n) + o(n^{-m/2}).$$

(This is not rigorous: it requires a stronger notion of differentiability, because the perturbation direction is now random: $H = \mathbb{G}_n$.)

Functional delta method

If ϕ'_P is a linear map, we have

$$\begin{aligned}\phi(P_n) - \phi(P) &= \frac{1}{\sqrt{n}}\phi'_P(\mathbb{G}_n) + o(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\phi'_P(\delta_{X_i} - P) + o(n^{-1/2}),\end{aligned}$$

where δ_{X_i} is the discrete distribution concentrated on X_i.

So if $\phi'_P(\delta_X - P)$ is mean zero and finite variance, we have that $\sqrt{n}(\phi(P_n) - \phi(P))$ is asymptotically normal.

Aside: influence functions

The function $x \mapsto \phi'_P(\delta_x - P)$ is the *influence function* of ϕ:

$$\begin{aligned} \phi'_P(\delta_x - P) &= \frac{d}{dt}\phi(P + t(\delta_x - P))\Big|_{t=0} \\ &= \frac{d}{dt}\phi((1-t)P + t\delta_x)\Big|_{t=0}. \end{aligned}$$

This measures the impact of changing P by mixing in a tiny amount of δ_{X_i}. It is important in robust statistics.

Example: Quantiles

Suppose we wish to estimate the pth quantile of P, where $p \in (0, 1)$. We'll write it as $\phi(F)$ for the cumulative distribution function F of P. If F is continuous at the appropriate point, we can define

$$\phi(F) = F^{-1}(p).$$

We need to calculate

$$\phi'_F(s_x - F) = \left. \frac{d}{dt}\phi((1-t)F + ts_x) \right|_{t=0},$$

where s_x is the cdf of δ_x, i.e., the step function $a \mapsto 1[a \geq x]$.

Example: Quantiles

Write $F_t = (1-t)F + ts_x$, then differentiate both sides of the equation $p = F_t(\phi(F_t))$ at $t = 0$ (ignoring the non-differentiability at $\phi(F_t) = x$):

$$\begin{aligned}
0 = & \left.\frac{d}{d_t} F_t(\phi(F_t))\right|_{t=0} \\
= & \left.\frac{d}{d_t}(1-t)F(\phi(F_t)) + ts_x(\phi(F_t))\right|_{t=0} \\
= & -F(\phi(F_t)) + (1-t)f(\phi(F_t))\phi'_{F_t}(s_x - F) + s_x(\phi(F_t))\Big|_{t=0} \\
= & -F(\phi(F)) + f(\phi(F))\phi'_F(s_x - F) + s_x(\phi(F)),
\end{aligned}$$

where f is the density of P.

Example: Quantiles

Rearranging, we have

$$\begin{aligned}\phi'_F(s_x - F) &= \frac{F(\phi(F)) - s_x(\phi(F))}{f(\phi(F))} \\ &= \frac{p - s_x(F^{-1}(p))}{f(F^{-1}(p))} \\ &= \begin{cases} \frac{p-1}{f(F^{-1}(p))} & \text{if } x \leq F^{-1}(p), \\ \frac{p}{f(F^{-1}(p))} & \text{if } x > F^{-1}(p). \end{cases}\end{aligned}$$

Example: Quantiles

So

$$\mathbf{E}\phi'_F(s_X - F) = \frac{p(p-1) + (1-p)p}{f(F^{-1}(p))} = 0$$

$$\begin{aligned} \operatorname{var} \phi'_F(s_X - F) &= \frac{p(1-p)^2 + (1-p)p^2}{f(F^{-1}(p))^2} \\ &= \frac{p(1-p)}{f(F^{-1}(p))^2}. \end{aligned}$$

And we have

$$\sqrt{n}(\phi(F_n) - \phi(F)) \rightsquigarrow N\left(0, \frac{p(1-p)}{(f(F^{-1}(p)))^2}\right).$$

Differentiability of functions in normed spaces

How do we make this rigorous? *Functional delta method.*

Theorem: Suppose $\phi : D \to E$, where D and E are normed linear spaces. Suppose the statistic $T_n : \Omega_n \to D$ satisfies $\sqrt{n}(T_n - \theta) \rightsquigarrow T$ for a random element T in $D_0 \subset D$.
If ϕ is *Hadamard differentiable at* θ *tangentially to* D_0 then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(T).$$

If we can extend $\phi' : D_0 \to E$ to a continuous map $\phi' : D \to E$, then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) = \phi'_\theta(\sqrt{n}(T_n - \theta)) + o_P(1).$$

Differentiability of functions in normed spaces

For the von Mises expansion, we considered

$$\phi'_P(H) = \frac{d}{dt}\phi(P + tH)\Big|_{t=0}$$

for some perturbation direction H. This is the Gateaux derivative.

Definition: $\phi : D \to E$ is *Gateaux differentiable* at $\theta \in D$ if

$$\forall h \in D,\ \exists \phi'_\theta(h) \in E,\ \text{s.t. as } t \to 0,\ \left\| \frac{\phi(\theta + th) - \phi(\theta)}{t} - \phi'_\theta(h) \right\| \to 0.$$

ϕ'_θ might not be (i) linear, or (ii) continuous (even if it's linear, it might not be continuous if D, E are infinite dimensional).

Differentiability of functions in normed spaces

Definition: $\phi : D \to E$ is *Hadamard differentiable* at $\theta \in D$ if

$\exists \phi'_\theta : D \to E$ (linear, continuous), $\forall h \in D$, if $t \to 0$, $\|h_t - h\| \to 0$, then

$$\left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \phi'_\theta(h) \right\| \to 0.$$

Gateaux requires the difference quotients to converge to some $\phi'_\theta(h)$ for each direction h; Hadamard requires a single ϕ'_θ that works for every direction h. It is equivalent to the convergence in the definition of Gateaux differentiability being uniform over h in a compact subset of D.

Differentiability of functions in normed spaces

Definition: $\phi : D \to E$ is *Fréchet differentiable* at $\theta \in D$ if

$$\exists \phi'_\theta : D \to E \text{ (linear, continuous)}, \forall h \in D, \text{ if } \|h\| \to 0, \text{ then}$$

$$\left\| \frac{\phi(\theta + h) - \phi(\theta) - \phi'_\theta(h)}{\|h\|} \right\| \to 0.$$

Hadamard requires the difference quotients to converge to zero for each direction, possibly with different rates for different directions; Fréchet requires the same rate for each direction. They are equivalent for $D = \mathbb{R}^d$.

Differentiability of functions in normed spaces

We'll consider Hadamard differentiability, but allow the weaker notion of tangential differentiability: the limiting directions h can be constrained.

Definition: $\phi : D \to E$ is *Hadamard differentiable* at $\theta \in D$ *tangentially to* $D_0 \subseteq D$ if

$$\exists \phi'_\theta : D_0 \to E \text{ (linear, continuous)}, \forall h \in D_0,$$

$$\text{if } t \to 0, \|h_t - h\| \to 0, \text{ then}$$

$$\left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \phi'_\theta(h) \right\| \to 0.$$

Differentiability of functions in normed spaces

Theorem: Suppose $\phi : D \to E$, where D and E are normed linear spaces. Suppose the statistic $T_n : \Omega_n \to D$ satisfies $\sqrt{n}(T_n - \theta) \rightsquigarrow T$ for a random element T in $D_0 \subset D$.

If ϕ is *Hadamard differentiable at* θ *tangentially to* D_0 then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(T).$$

If we can extend $\phi' : D_0 \to E$ to a continuous map $\phi' : D \to E$, then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) = \phi'_\theta(\sqrt{n}(T_n - \theta)) + o_P(1).$$

Differentiability of functions in normed spaces

Proof:

Based on continuous mapping theorem. Consider the maps

$$f_n(h) = \sqrt{n}\left(\phi\left(\theta + \frac{1}{\sqrt{n}}h\right) - \phi(\theta)\right).$$

Hadamard differentiability implies that for any sequence $h_n \to h \in D_0$, we have $f_n(h_n) \to \phi'_\theta(h)$. So $f_n(\sqrt{n}(T_n - \theta)) \rightsquigarrow \phi'_\theta(T)$.

(Second statement follows from continuous mapping theorem for the function $h \mapsto (\phi(h), \phi'_\theta(h))$.)

Differentiability of functions in normed spaces

The chain rule lets us determine Hadamard derivatives of a composition of maps.

Theorem: Suppose $\phi : D \to E$, $\psi : E \to F$, where D, E and F are normed linear spaces. If

1. ϕ is Hadamard differentiable at θ tangentially to D_0, and
2. ψ is Hadamard differentiable at $\phi(\theta)$ tangentially to $\phi'_\theta(D_0)$,

then $\psi \circ \phi : D \to F$ is Hadamard differentiable at θ tangentially to D_0, with derivative $\psi'_{\phi(\theta)} \circ \phi'_\theta$.

Example: Quantiles

Definition: The *quantile function* of F is $F^{-1} : (0,1) \to \mathbb{R}$,

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}.$$

[PICTURE]

For $p \in (0,1)$ and $x \in \mathbb{R}$,

$$F^{-1}(p) \leq x \;\Leftrightarrow p \leq F(x),$$

which implies the *quantile transformation*: for U uniform on $(0,1)$,

$$F^{-1}(U) \sim F.$$

Example: Quantiles

Also, $F(F^{-1}(p)) \geq p$, with equality unless F has a discontinuity at $F^{-1}(p)$. This implies the *probability integral transformation*: for $X \sim F$, $F(X)$ is uniform on [0,1] iff F is continuous on $\mathbb{R}$, because $F(X) = F(F^{-1}(U)) = U$ in that case.

Finally,

$$F^{-1}(F(x)) \leq x,$$

with equality unless F is flat to the left of x. Thus, F^{-1} is an inverse (i.e., $F^{-1}(F(x)) = x$ and $F(F^{-1}(p)) = p$ for all x and p) iff F is continuous and strictly increasing.

Empirical quantile function

For a sample with distribution function F, define the *empirical quantile function* as the quantile function F_n^{-1} of the empirical distribution function F_n.

$$F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\} = X_{n(i)},$$

where i is chosen such that

$$\frac{i-1}{n} < p \leq \frac{i}{n},$$

and $X_{n(1)}, \ldots, X_{n(n)}$ are the order statistics of the sample, that is, $X_{n(1)} \leq \cdots \leq X_{n(n)}$ and

$$\big(X_{n(1)}, \ldots, X_{n(n)}\big)$$

is a permutation of the sample $(X_1, \ldots, X_n)$.