The functional delta method

Patrick Breheny

August 30

Recap

- Last lecture, we introduced the influence function and demonstrated its use in assessing the robustness of an estimator to contaminating point masses
- This lecture, we will see how the influence function also allows us to perform inference and obtain central limit theorem-type results for statistical functionals

Overview

- In parametric statistics, we estimate θ and can then use the delta method to obtain distributional results for $T(\theta)$
- In nonparametric statistics, we estimate F and can then use the *functional delta method* to obtain distributional results for $T(F)$
- This lecture will be devoted to proving the functional delta method and illustrating its use

Lemma 1

- We begin by proving a simpler version of the functional delta method, assuming that $T(F)$ is a linear functional
- For the lemmas that follow, $T(F)$ is assumed to be a linear functional – the general case will follow
- **Lemma 1:** For any G,

$$\int L_F(x)dG(x) = T(G) - T(F)$$

- This result is similar to the fundamental theorem of calculus, only for functional calculus
- **Corollary:**

$$\int L_F(x)dF(x) = 0$$

Lemma 2

Lemma 2: Let $\tau^2 = \int L^2(x)dF(x)$. If $\tau^2 < \infty$,

$$\sqrt{n}\left\{T(\hat{F}) - T(F)\right\} \xrightarrow{\mathsf{d}} N(0, \tau^2)$$

Lemma 3

Lemma 3: Let $\hat{\tau}^2 = n^{-1}\sum_i \hat{L}^2(x_i)$. Then

$$\hat{\tau}^2 \xrightarrow{\text{P}} \tau^2$$
$$\frac{\widehat{SE}}{SE} \xrightarrow{\text{P}} 1,$$

where $\widehat{SE} = \hat{\tau}/\sqrt{n}$ and $SE = \sqrt{\mathbb{V}(T(\hat{F}))}$

Lemma 4

Lemma 4:

$$\frac{\sqrt{n}\left\{T(\hat{F}) - T(F)\right\}}{\hat{\tau}} \xrightarrow{\mathsf{d}} N(0,1)$$

General case

- We have arrived at a very useful result – for linear functionals
- Does this work for nonlinear functionals?
- The usual strategy for a proof like this is to take a Taylor series expansion to reduce the nonlinear problem to a linear problem

General case (cont'd)

- In the linear case, our results depended on the expression

$$T(\hat{F}) = T(F) + \frac{1}{n}\sum L_F(X_i)$$

- We can prove the general case by the same mechanism if we can write

$$T(\hat{F}) = T(F) + \frac{1}{n}\sum L_F(X_i) + o_P(1)$$

- The question, of course, is whether or not there exists a functional Taylor's theorem

General case (cont'd)

- The answer is that yes, there does (it is called the *von Mises expansion*), and to apply it, T needs to be Hadamard differentiable at F
- **Theorem:** If T is Hadamard differentiable at F, then

$$\frac{\sqrt{n}\left\{T(\hat{F}) - T(F)\right\}}{\hat{\tau}} \overset{\text{d}}{\longrightarrow} N(0,1)$$

The functional delta method

Thus, under appropriate regularity conditions, a $1-\alpha$ confidence interval for $\theta = T(F)$ is

$$\hat{\theta} \pm z_{\alpha/2}\widehat{SE}$$

where $\hat{\theta}$ is the plug-in estimate, $\widehat{SE} = n^{-1/2}\hat{\tau}$, and $\hat{\tau}^2 = n^{-1}\sum_i \hat{L}^2(x_i)$

The mean

Using the functional delta method to derive an asymptotic confidence interval for the mean, we have

- $\hat{\theta} = T(\hat{F}) = \bar{x}$
- $\hat{\tau}^2 = n^{-1} \sum_i \hat{L}^2(x_i) = n^{-1} \sum_i (x_i - \bar{x})^2$
- $\widehat{SE} = n^{-1/2} \hat{\tau}$
- And an asymptotic 95% confidence interval for θ is $\bar{x} \pm 1.96\widehat{SE}$ – nearly identical to the normal parametric interval

The variance

For the variance, we have

- $\hat{\theta} = T(\hat{F}) = n^{-1} \sum (x_i - \bar{x})^2$
- $\hat{\tau}^2 = n^{-1} \sum_i \hat{L}^2(x_i) = n^{-1} \sum_i \{(x_i - \bar{x})^2 - \hat{\sigma}^2\}^2$
- $\widehat{SE} = n^{-1/2} \hat{\tau}$

Homework

Homework: Compare the nonparametric confidence interval for the variance obtained from using the functional delta method to the normal-theory interval:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}\right],$$

where s^2 is the (unbiased) sample variance.

Conduct a simulation study to determine the coverage probability and average interval width of these two intervals.

(a) Carry out the above simulation with data generated from the standard normal distribution.
(b) Repeat using data generated from an exponential distribution with rate 1.
(c) Briefly, comment on the strengths and weaknesses of these two methods.

Homework

Homework: The `R` data set `quakes` contains (among other information) the magnitude of 1,000 earthquakes that have occurred near the island Fiji.

(a) Estimate the CDF for the magnitude of earthquakes in this region, along with a 95% confidence interval. Plot your results.

(b) Estimate and provide a 95% confidence interval for $F(4.9) - F(4.3)$.

(c) Estimate the variance of the magnitude, and provide a nonparametric 95% confidence interval for its value.

Numerical approximation

- The most difficult aspect of applying the functional delta method is the derivation of the influence function
- In situations where the functional of interest (and its influence function) may be complicated, we can still apply the delta method approximately using a numerical approach

Influence components

- We have seen that the confidence interval provided by the delta method depends only on $\{\hat{L}_i\}$, the so-called *influence components*, where $\hat{L}_i = \hat{L}(x_i)$, which are found by examining

$$\lim_{\epsilon \to 0} \frac{T(\hat{F}_i(\epsilon)) - T(\hat{F})}{\epsilon}$$

  where $\hat{F}_i(\epsilon) = (1 - \epsilon)\hat{F} + \epsilon\delta_i$

- We can obtain a numerical approximation to this limit by evaluating the above expression for a very small value of ϵ

Epsilon weights

- Note that $\hat{F}_i(\epsilon)$ places a point mass at every observed x value of

$$\left(\frac{1}{n}(1-\epsilon), \cdots, \frac{1}{n}(1-\epsilon)+\epsilon, \cdots, \frac{1}{n}(1-\epsilon)\right)$$
$$\left(\frac{1}{n}-\frac{\epsilon}{n}, \cdots, \frac{1}{n}+\frac{(n-1)\epsilon}{n}, \cdots, \frac{1}{n}-\frac{\epsilon}{n}\right)$$

- Denote these weights $\{w_{ij}\}$

Accuracy

- Now, for example, the i^{th} influence component for the variance can be calculated as

$$\hat{L}_i = \frac{\sum_j w_{ij}(x_j - \bar{x}_i)^2 - \sum_j \frac{1}{n}(x_j - \bar{x})^2}{\epsilon}$$

  where $\bar{x}_i = \sum_j w_{ij} x_j$

- The approximation is quite accurate:

```
> x <- rnorm(100)
> L <- calcL(x)
> L.approx <- approxL(x, eps=1e-6)
> mean(abs(L-L.approx))
[1] 9.295309e-07
```