Statistical functionals and influence functions

Patrick Breheny

August 28
Recap

In our first lecture, we introduced the empirical distribution function as a nonparametric way of estimating $F$, and showed that it had a number of very attractive properties:

- $\hat{F} \xrightarrow{\text{a.s.}} F$, regardless of $F$
- The nonparametric maximum likelihood estimator
- Able to derive confidence intervals and confidence bands that work for any $F$, any $n$, and all $x$
Why so much emphasis on estimating $F$?

Because $\hat{F}$ will play the same role in nonparametric estimation that $\hat{\theta}$ played in parametric estimation.

In parametric statistics, we find the most likely distribution of the data based on $\hat{F} = F(\hat{\theta})$.

This then provides estimates of our parameters of interest – usually $\theta$ itself or some function of $\theta$, such as $e^\theta$, but conceivably also something like the 75th percentile.
Clearly, nonparametric statistics does not provide estimates of parameters (or functions of parameters) . . . so what are we estimating in nonparametric statistics?

We estimate what are called *statistical functionals*: a functional $\theta = T(F)$ is any function of $F$

Many common descriptive statistics can be expressed as statistical functionals
Examples

- Mean: $T(F) = \int x \, dF(x)$
- Variance: $T(F) = \int (x - \mu)^2 \, dF(x)$
- Quantiles: $T(F) = F^{-1}(p)$
Estimation of statistical functionals

- Estimation of these quantities is straightforward: calculate the quantity you are interested in based on the nonparametric maximum likelihood estimator of $F$, the empirical distribution function $\hat{F}$:

$$\hat{\theta} = T(\hat{F})$$

- This idea is often called the *plug-in principle*, and its resulting estimate the *plug-in estimate*

  - Examples:
    - Mean: $\bar{x}$
    - Variance: $n^{-1} \sum (x_i - \bar{x})^2$
    - Quantile: Sample quantile

- In a sense, the plug-in principle is a nonparametric analog of the likelihood principle
Is the plug-in estimator a good estimator?

- A natural question: Is the plug-in estimator a good estimator?
- The Glivenko-Cantelli Theorem says that $\hat{F} \xrightarrow{a.s.} F$; does this mean that $T(\hat{F}) \xrightarrow{a.s.} T(F)$?
- The answer turns out to be a complicated “sometimes”; often “yes”, but not always
As an example of a case where the plug-in estimator is not consistent, consider density estimation, where $T(F) = F'(x)$:
So when is the plug-in estimator consistent?

It requires certain conditions on the smoothness (differentiability) of $T(F)$

But what does it mean to take the derivative of a function with respect to a function?

To answer this question, we will have to expand our notion of the derivative
The Gâteaux derivative

- The extension of the derivative that we will need is called the *Gâteaux derivative*
- The Gâteaux derivative of $T$ at $F$ in the direction $G$ is defined by

$$L_F(T; G) = \lim_{\epsilon \to 0} \left( \frac{T\{(1 - \epsilon)F + \epsilon G\} - T(F)}{\epsilon} \right)$$

- An equivalent way of stating the definition is to define $D = G - F$, and the above becomes

$$L_F(T; D) = \lim_{\epsilon \to 0} \left( \frac{T\{F + \epsilon D\} - T(F)}{\epsilon} \right)$$

- Either way, the definition boils down to

$$L_F(T) = \lim_{\epsilon \to 0} \left( \frac{T(F\epsilon) - T(F)}{\epsilon} \right)$$
Interpretation

- From a mathematical perspective, the Gâteaux derivative a generalization of the concept of a directional derivative to functional analysis.
- From a statistical perspective, it represents the rate of change in a statistical functional upon a small amount of contamination by another distribution $G$. 
As an example of how the Gâteaux derivative works, suppose \( F \) is a continuous CDF, and \( G \) is the distribution that places all of its mass at the point \( x_0 \).

What happens to the Gâteaux derivative of \( T(F) = f(x_0) \)?

\[
L_F(T; G) = \lim_{\epsilon \to 0} \left[ \frac{d}{dx} \{(1 - \epsilon)F(x) + \epsilon G(x)\}_{x=x_0} - \frac{d}{dx} F(x)_{|x=x_0} \right] \\
= \lim_{\epsilon \to 0} \left[ \frac{(1 - \epsilon)f(x_0) + \epsilon g(x_0) - f(x)}{\epsilon} \right] \\
= \infty
\]
So, even though \( F \) and \( F_\epsilon \) differ from each other only infinitesimally \( T(F) \) and \( T(F_\epsilon) \) differ from each other by an infinite amount.

Thus, the Glivenko-Cantelli theorem does not help us here: \( \sup_x \left| \hat{F}(x) - F(x) \right| \) may go to zero without \( T(\hat{F}) \to T(F) \).
It turns out that even Gâteaux differentiability is too weak to ensure that $T(\hat{F}) \rightarrow T(F)$

Even if the Gâteaux derivative exists, it may not exist in an entirely unique way, and this is the subtle idea introduced by Hadamard differentiability

A functional $T$ is Hadamard differentiable if, for any sequence $\epsilon_n \rightarrow 0$ and $D_n$ satisfying $\sup_x |D_n(x) - D(x)| \rightarrow 0$, we have

$$\frac{T(F + \epsilon_n D_n) - T(F)}{\epsilon_n} \rightarrow L_F(T; D)$$

If $T$ is Hadamard differentiable, then $T(\hat{F}) \xrightarrow{P} T(F)$
Another useful condition that — certainly, one that is easier to check than Hadamard differentiability — is that if the functional is bounded, then the plug-in estimate will converge to the true value.

**Homework:** Suppose that there exists a constant $C$ such that the following relation holds for all $G$:

$$|T(F) - T(G)| \leq C \sup_x |F(x) - G(x)|.$$

Show that $T(\hat{F}) \xrightarrow{a.s.} T(F)$. 
Contamination by a point mass

- The idea of contaminating a distribution with a small amount of additional data has a long history in statistics and the investigation of robust estimators.

- Statisticians usually do not work with the general Gâteaux derivative, but a special case of it called the influence function, in which $G$ places a point mass of 1 at $x$:

$$
\delta_x(u) = \begin{cases} 
0 & \text{if } u < x \\
1 & \text{if } u \geq x 
\end{cases}
$$
The influence function and empirical influence function

- The influence function is usually written as a function of $x$, and defined as

$$L(x) = \lim_{\epsilon \to 0} \left[ \frac{T\{(1 - \epsilon)F + \epsilon \delta_x\} - T(F)}{\epsilon} \right]$$

- A closely related concept is that of the *empirical influence function*:

$$\hat{L}(x) = \lim_{\epsilon \to 0} \left[ \frac{T\{(1 - \epsilon)\hat{F} + \epsilon \delta_x\} - T(\hat{F})}{\epsilon} \right]$$
Example: The mean

\[ L(x) = x - \mu \]

\[ \hat{L}(x) = x - \bar{x} \]
Linear functionals

- The mean is an example of a \textit{linear functional}: one in which

\[ T(F) = \int a(x) dF(x) \]

- Linear functionals are particularly easy to work with, as

\[ L(x) = a(x) - T(F) \]

\[ \hat{L}(x) = a(x) - T(\hat{F}) \]
Working directly from the definition each time is time-consuming.

Fortunately, the Gâteaux derivative has many of the same properties as ordinary derivatives – in particular, the chain rule.

Suppose our functional can be written in the form

\[ T(F) = a\{T_1(F), T_2(F), \ldots \}; \]

then

\[ L(x) = \sum_j \frac{\partial a}{\partial T_j} \bigg|_F L_j(x), \]

where \( L_j(x) \) is the influence function of \( T_j(F) \).
Example: Variance

\[ L(x) = (x - \mu)^2 - \sigma^2 \]
\[ \hat{L}(x) = (x - \bar{x})^2 - \hat{\sigma}^2 \]
Homework: Consider a random variable $X$ that is always positive. We are interested in the statistical functionals $\theta = \int \log(x) dF(x)$ and $\lambda = \log(\mu)$.

(b) What are the influence and empirical influence functions for $\theta$?

(d) What are the influence and empirical influence functions for $\lambda$?

(f) Do $\hat{\lambda}$ and $\hat{\theta}$ converge to the same number?

(g) Plot the empirical influence functions from parts (b) and (d). Label the point $x$ on the horizontal axis where $L(x) = 0$.

(h) Briefly, comment on the relative robustness of $\hat{\theta}$ and $\hat{\lambda}$ to outliers.
Example: Quantiles

\[ L(x) = \begin{cases} \frac{p-1}{f(\theta)} & \text{if } x \leq \theta \\ \frac{p}{f(\theta)} & \text{if } x > \theta \end{cases} \]

\[ \hat{L}(x) = ? \]
**Homework:** Show that, for $x > \theta$,

$$L(x) = \frac{p}{f(\theta)}$$
Note that the influence function for the median is unbounded, while the influence function for the mean is not.

**Homework:** Let $b(\epsilon) = \sup_x |T(F) - T(F_\epsilon)|$, where $F_\epsilon = (1 - \epsilon)F + \epsilon \delta_x$. The *breakdown point* of an estimator, $\epsilon^*$, is defined as $\epsilon^* = \inf\{\epsilon : b(\epsilon) = \infty\}$

a) Find the breakdown point of the mean
b) Find the breakdown point of the median