Empirical Distribution Function, Statistical Functionals, and Influence Function

Tianchen Qian

Advanced Theory Class, Dec 9, 2014

1. Empirical Distribution and Beyond

2. Influence Function and Functional Delta Method
   - Plug-in estimator
   - Functional Differentiation
   - Influence Function and Functional Delta Method

Empirical Distribution Functions

- Assume $X_1, \ldots, X_n$ are i.i.d. random variables with cdf F. Their empirical distribution is defined as

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x), \quad x \in \mathbb{R}.$$

- By WLLN and CLT, we have (for fixed x)

$$\hat{F}_n(x) \xrightarrow{P} F(x),$$
$$\sqrt{n}\left\{\hat{F}_n(x) - F(x)\right\} \xrightarrow{D} N(0, F(x)(1 - F(x))).$$

(Pointwise) Confidence Interval

- Based on

$$\sqrt{n}\left\{\hat{F}_n(x) - F(x)\right\} \xrightarrow{D} N(0, F(x)(1 - F(x))),$$

we can construct pointwise confidence interval $C(x)$ for $F(x)$, for fixed x. I.e., we have

$$P\{F(x) \in C(x)\} \geq 1 - \alpha.$$

Confidence Band

- To find a confidence band for $F(\cdot)$, we want to find $C(\cdot)$ such that
$$P\{F(x) \in C(x) \quad \forall x \in \mathbb{R}\} \geq 1 - \alpha.$$
- This is asking more than a pointwise confidence interval.
- We need some control of F on the entire $\mathbb{R}$.

Confidence Band (continued)

- **Glivenko-Cantelli Theorem** tells us that

$$\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F(x)\right| \overset{a.s.}{\to} 0, \quad \text{as } n \to \infty.$$

- A stronger result called the **Dvoretzky-Kiefer-Wolfowitz inequality**, or DKW inequality:

$$P\left\{\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F(x)\right| > \epsilon\right\} \leq 2e^{-2n\epsilon^2}.$$

- Note that this is a finite-sample result (holds for finite n).

Confidence Band (continued)

$$P\left\{\sup_{x\in\mathbb{R}}\left|\hat{F}_n(x)-F(x)\right|>\epsilon\right\}\le 2e^{-2n\epsilon^2}.$$

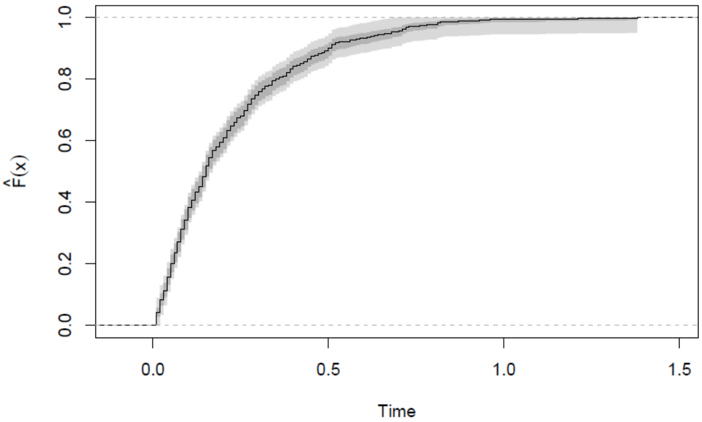
- Setting the right hand side of the DKW inequality equal to α, and we have

$$\epsilon=\sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}.$$

- So the confidence band for F is

$$\left(\hat{F}_n(x)-\sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)},\hat{F}_n(x)+\sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right),\quad x\in\mathbb{R}$$

Confidence Interval vs Confidence Band



http://web.as.uky.edu/statistics/users/pbreheny/621/F12/notes/8-23.pdf

Other Confidence Band

- There are other types of confidence bands, for example,

$$P\left\{\sup_{x\in\mathbb{R}}\left|\hat{F}_n(x)-F(x)\right|>\epsilon\right\}\leq 8(n+1)\,e^{-n\epsilon^2/32}.$$

  This is derived from Vapnik-Chervonenkis Theorem.
- See Theorem 2.43 and Example 2.45 of Wasserman, L. (2006). All of nonparametric statistics. Springer.

1 Empirical Distribution and Beyond

2 Influence Function and Functional Delta Method
- Plug-in estimator
- Functional Differentiation
- Influence Function and Functional Delta Method

Parameter as functionals

- We can treat the parameter of interest θ as a functional of the distribution function F:

$$\theta = T\left(F\right).$$

- Examples:
    - Mean: $T\left(F\right) = \int x dF\left(x\right)$
    - Variance: $T\left(F\right) = \int \left(x-\mu\right)^2 dF\left(x\right)$
    - Quantiles: $T\left(F\right) = F^{-1}\left(p\right)$
    - Mann-Whitney functional:
      $T\left(F,G\right) = P\left(X \leq Y \mid X \sim F, Y \sim G\right) = \int F\left(x\right) dG\left(x\right)$

Plug-in Principle

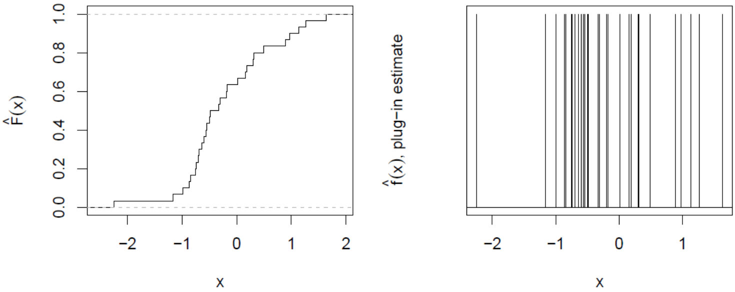
- Calculate the quantity based on empirical distribution function $\hat{F}$:
$$\hat{\theta} = T\left(\hat{F}\right).$$
- Examples:
  - Mean: $\bar{X}_n$
  - Variance: $n^{-1}\sum\left(X_i - \bar{X}_n\right)^2$
  - Quantile: Sample quantile
  - Mann-Whitney statistic: $T\left(\hat{F}_m, \hat{G}_n\right) = \int \hat{F}_m(x)\, d\hat{G}_n(x) = \frac{1}{mn}\sum_{i,j} I\left(X_i \leq Y_j\right)$

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 $\hat{F} \stackrel{a.s.}{\to} F$; does this mean that $T\left(\hat{F}\right) \stackrel{a.s.}{\to} T\left(F\right)$?
- Answer: sometimes.

Plug-in density estimation

- As an example where the plug-in estimator is not consistent, consider density estimation, where $T(F) = F'(x_0)$, for some x_0.



http://web.as.uky.edu/statistics/users/pbreheny/621/F12/notes/8-23.pdf

When is the plug-in estimator consistent?

- When is the plug-in estimator consistent?
  - Mean: $T(F) = \int x dF(x)$
  - Density at x_0: $T(F) = F'(x_0)$
  - Change F to $(1-\epsilon)F + \epsilon\delta_x$, how much does $T(F)$ change?
- It requires certain conditions on the smoothness (differentiability) of $T(F)$.
- To address this more clearly, we need to expand our notion of the derivative.

1. Empirical Distribution and Beyond
2. Influence Function and Functional Delta Method
    - Plug-in estimator
    - Functional Differentiation
    - Influence Function and Functional Delta Method

The Gateaux derivative

$$\text{“}\frac{dT(F)}{dF}\text{''} = \text{''} \lim_{\epsilon \to 0} \left[\frac{T\{F_\epsilon\} - T(F)}{\epsilon} \right]$$

- The **Gateaux derivative** of T at F in the direction of 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 definition: let $D = G - F$, then

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

Interpretation

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

- From a mathematical perspective, the Gateaux derivative is a generalization of the concept of 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.

Gateaux derivative of $T(F) = f(x_0)$

- As an example of how the Gateaux derivative works, suppose F is CDF of a continuous variable, and G is the distribution that places all of its mass at point x_0.

$$
\begin{aligned}
L_F(T;G) &= \lim_{\epsilon \to 0} \left[\frac{T\{(1-\epsilon)F + \epsilon G\} - T(F)}{\epsilon} \right] \\
&= \lim_{\epsilon \to 0} \left[\frac{\frac{d}{dx}\{(1-\epsilon)F(x) + \epsilon G(x)\}\,|_{x=x_0} - \frac{d}{dx}F(x)\,|_{x=x_0}}{\epsilon} \right] \\
&= \lim_{\epsilon \to 0} \left[\frac{(1-\epsilon)f(x_0) + \epsilon g(x_0) - f(x_0)}{\epsilon} \right] \\
&= \infty
\end{aligned}
$$

Glivenko-Cantelli does not imply convergence of estimators

- So, even though F and F_ϵ 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: $\sup_x \left|\hat{F}(x) - F(x)\right| \overset{a.s.}{\to} 0$ doesn't imply $T\left(\hat{F}\right) \to T(F)$.

Hadamard differentiability

- It turns out that even Gateaux differentiability is not enough for $T\left(\hat{F}\right) \to T(F)$.
- We need a stronger definition of differentiability.
- A functional T is **Hadamard differentiable** if for a $D(= G - F)$, there exists some $L_F(T; D)$, such that for any sequence $\epsilon_n \to 0$ and D_n satisfying $\sup_x |D_n(x) - D(x)| \to 0$, we have

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

Compare Gateaux and Hadamard differentiability

- Gateaux:

$$\frac{T\{F+\epsilon D\}-T(F)}{\epsilon} \to L_F(T;D)$$

- Hadamard:

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

as long as $\epsilon_n \to 0$ and $\sup_x |D_n(x)-D(x)| \to 0$

Sufficient conditions for $T\left(\hat{F}\right) \stackrel{a.s.}{\to} T(F)$

- If
  - T is Hadamard differentiable;
  - Or, T is a bounded functional (at F): $\exists C > 0$, such that

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

  Then $T\left(\hat{F}\right) \stackrel{a.s.}{\to} T(F)$.
- Now we have consistency. What about asymptotic normality?
  - Functional Delta Method.

1. Empirical Distribution and Beyond

2. Influence Function and Functional Delta Method
   - Plug-in estimator
   - Functional Differentiation
   - Influence Function and Functional Delta Method

Influence function and empirical influence function

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

- The **influence function** of the functional T at point x is defined as

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

  where δ_x is a point mass at x.
- A closely related concept is the empirical influence function:

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

Functional Delta Method (linear case)

- First consider a simple case: $T(F)$ is a linear functional: $T(\alpha F + \beta G) = \alpha T(F) + \beta T(G)$. E.g. $T(F) = \int h(x)\, dF(x)$ for some function h.
- In this case,

$$\begin{aligned}
L_F(x) &= \lim_{\epsilon \to 0} \left[\frac{T\{(1-\epsilon)F + \epsilon\delta_x\} - T(F)}{\epsilon} \right] \\
&= \lim_{\epsilon \to 0} \frac{(1-\epsilon)T(F) + \epsilon T(\delta_x) - T(F)}{\epsilon} \\
&= T(\delta_x) - T(F)
\end{aligned}$$

so for any G,

$$\begin{aligned}
\int L_F(x)\, dG(x) &= \int \{T(\delta_x) - T(F)\}\, dG(x) \\
&= \int T(\delta_x)\, dG(x) - \int T(F)\, dG(x) \\
&= T(G) - T(F)
\end{aligned}$$

Functional Delta Method (linear case, continued)

- **Lemma 1**: Assume T is a linear functional. For any G,

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

- This is similar to the fundamental theorem of calculus.
- **Corollary**:

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

Functional Delta Method (linear case, continued)

- **Lemma 2**: Assume T is a linear functional. Let $\tau^2 = \int L_F^2(x)\, dF(x)$. If $\tau^2 < \infty$, then

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

- Proof: By Lemma 1,

$$\begin{aligned}
\sqrt{n}\left\{T\left(\hat{F}\right) - T(F)\right\} &= \sqrt{n}\int L_F(x)\, d\hat{F}(x) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_F(X_i) \\
&\xrightarrow{D} N\left(0, \tau^2\right)
\end{aligned}$$

Functional Delta Method (linear case, continued)

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

$$\hat{\tau}^2 \xrightarrow{P} \tau^2$$

- **Lemma 4**: Assume T is a linear functional.

$$\frac{\sqrt{n}\left\{T\left(\hat{F}\right) - T(F)\right\}}{\hat{\tau}^2} \xrightarrow{D} N(0,1)$$

Functional Delta Method (general case)

- In the linear case, our result depends on the expression

$$\sqrt{n}\left\{T\left(\hat{F}\right)-T\left(F\right)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_F\left(X_i\right)$$

- We can prove the Delta Method for general T if we have

$$\sqrt{n}\left\{T\left(\hat{F}\right)-T\left(F\right)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_F\left(X_i\right)+o_P\left(1\right)$$

- This is a special case of the **von Mises expansion**, and the $o_P\left(1\right)$ requires some restrictions on T.

An intuitive derivation

- For distributions F and G, consider the map $\phi : t \mapsto T(F + tG)$. Assume the ordinary derivatives of $\phi(t)$ at $t = 0$ exist for $k = 1, 2, \ldots, m$. Denoting them by $\phi^{(k)}(0)$. By Taylor's theorem, we have

$$\phi(t) - \phi(0) = t\phi'(0) + \ldots + \frac{1}{m!} t^m \phi^{(m)}(0) + o(t^m)$$

- Substituting $t = 1/\sqrt{n}$ and $G = \sqrt{n}\left(\hat{F} - F\right)$, we obstain

$$T\left(\hat{F}\right) - T(F) = \frac{1}{\sqrt{n}} \phi'(0) + \ldots$$

- To have control on the remainder term, some bounds need to be established. See von Mises (1947).

Functional Delta Method (Theorem)

- Theorem: If T is Hadamard differentiable at F, then

$$\frac{\sqrt{n}\left\{T\left(\hat{F}\right)-T\left(F\right)\right\}}{\hat{\tau}^2}\xrightarrow{D}N\left(0,1\right),$$

where $\hat{\tau}^2=\frac{1}{n}\sum_i \hat{L}^2\left(X_i\right)$.

Take home message

- For a functional T, its Gateaux derivative w.r.t a point mass δ_x is the influence function for the plug-in estimator $T\left(\hat{F}\right)$:

$$\sqrt{n}\left\{T\left(\hat{F}\right) - T\left(F\right)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_F\left(X_i\right) + o_P\left(1\right)$$

- The Delta Method applies for plug-in estimators viewed as functionals: $T\left(\hat{F}\right)$, and this requires some smoothness on the functional T.

References for Functional Delta Method

- Van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge university press.
    - Chapter 20 for numerous examples
    - Section 20.1 on von Mises calculus (intuitive)
    - Section 20.2 on Hadamard derivatives
    - Theorem 20.8 on an even more general Delta Method (about normed vector spaces)
- Mises, R. V. (1947). On the asymptotic distribution of differentiable statistical functions. The annals of mathematical statistics, 309-348.

Bibliography

- Van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge university press. Chapter 20
- Lecture notes by Jon Wellner: http://www.stat.washington.edu/jaw/COURSES/580s/581/LECTNOTES/ch7.pdf
- Wasserman, L. (2006). All of nonparametric statistics. Springer. Chapter 2, 3.1
- Lecture notes by Patrick Breheny: http://web.as.uky.edu/statistics/users/pbreheny/621/F12/notes.html
- Gill, R. D., Wellner, J. A., & Præstgaard, J. (1989 & 1993). Non-and semi-parametric maximum likelihood estimators and the von mises method. Scandinavian Journal of Statistics, 97-128.