STAT/Q SCI 403: Introduction to Resampling Methods **Spring 2017**

Lecture 9: Introduction to the Bootstrap Theory

Instructor: Yen-Chi Chen

9.1 Statistical Functionals

To study how the bootstrap works, we first introduce the concepts of *statistical functionals*.

What is a functional? A functional is just a function of a function. Namely, it is a 'function' such that the input is another function and the output is a number. Formally speaking, a functional is a mapping $T : \mathcal{F} \mapsto \mathbb{R}$, where $\mathcal{F}$ is a collection of functions. A statistical functional is a mapping T such that you input a distribution (CDF) and it returns a number.

This sounds very complicated but actually, we have encountered numerous statistical functionals. Here are some examples.

- **Mean of a distribution.** The mean of a distribution is a statistical functional
$$\mu = T_{\mathsf{mean}}(F) = \int x dF(x).$$
When F has a PDF $p(x)$, $dF(x) = p(x)dx$ so the mean functional reduces to the form that we are familiar with:
$$\mu = T_{\mathsf{mean}}(F) = \int x dF(x) = \int x p(x) dx.$$
When F is a distribution of discrete random variables, we define
$$\int x dF(x) = \sum_x x P(x) \Longrightarrow \mu = T_{\mathsf{mean}}(F) = \sum_x x P(x),$$
where $P(x)$ is the PMF of the distribution F.
You may have noticed that if a random variable X has a CDF F, then
$$\mathbb{E}(X) = \int x dF(x) = T_{\mathsf{mean}}(F).$$
Therefore, for any function g,
$$\mathbb{E}(g(X)) = \int \omega(x) dF(x).$$
Using the function g, we introduce another functional T_ω such that
$$T_\omega(F) = \int \omega(x) dF(x).$$
Such a functional, T_ω, is called a *linear functional*.
- **Variance of a distribution.** The variance of a distribution is also a statistical functional. Let X be a random variable with CDF F. Then
$$\sigma^2 = T_{\mathsf{var}}(F) = \mathsf{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \int x^2 dF(x) - \left(\int x dF(x)\right)^2.$$

- **Median of a distribution.** Using the concept of a statistical functional, median and any quantile can be easily defined. The median of a distribution F is a point θ_{med} such that $F(\theta_{\mathsf{med}}) = 0.5$. Thus,

$$T_{\mathsf{med}}(F) = F^{-1}(0.5).$$

Note that when F is a CDF of a discrete random variable, F^{-1} may have multiple values. In this case, we define

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

Any quantile of a distribution can be represented in a similar way. For instance, the q-quantile ($0 < q < 1$) will be

$$T_{\mathsf{q}}(F) = F^{-1}(q).$$

As a result, the interquartile range (IQR) is

$$T_{\mathsf{IQR}}(F) = F^{-1}(0.75) - F^{-1}(0.25).$$

Why do we want to use the form of statistical functionals? One answer is: it elegantly describes a population quantity that we may be interested in. Recall that the statistical model about how the data is generated is that we observe a random sample $X_1, \cdots, X_n$ IID from an unknown distribution F. Thus, the distribution F is our model for the population. Because the statistical functionals map F into some real numbers, they can be viewed as quantities describing the features of the population. The mean, variance, median, quantiles of F are numbers characterizing the population. Thus, using statistical functionals, we have a more rigorous way to define the concepts of population parameters.

In addition to the above advantage, there is a very powerful features of statistical functionals–they provide a simple estimator to these population quantities. Recall that the EDF $\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leq x)$ is a good estimator of F. Thus, if we want to estimate a population quantity $\theta = T_{\mathsf{target}}(F)$, we can use $T_{\mathsf{target}}(\widehat{F}_n) = \widehat{\theta}_n$ as our estimator. Actually, many estimators do follow this form. For instance, in the case of estimating the mean $\mu = T_{\mathsf{mean}}(F)$, we often use the sample mean $\bar{X}_n$ as our estimator. However, if you plug-in $\widehat{F}_n$ into the statistical functional:

$$T_{\mathsf{mean}}(\widehat{F}_n) = \int x d\widehat{F}_n(x) = \sum_{i=1}^n X_i \frac{1}{n} = \sum_{i=1}^n \frac{X_i}{n} = \bar{X}_n.$$

This implies that the estimator from the statistical functional is the same as sample mean! Note that we in the above calculation, we use the fact that $\widehat{F}_n(x)$ is a distribution with whose PMF puts equal probability $(1/n)$ at $X_1, \cdots, X_n$. The estimator formed via replacing F by $\widehat{F}_n$ is called a *plug-in* estimator.

Similarly, we may estimate the variance $\sigma^2 = T_{\mathsf{var}}(F)$ via

$$T_{\mathsf{var}}(\widehat{F}_n) = \int x^2 d\widehat{F}_n(x) - \left(\int x d\widehat{F}_n(x)\right)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n-1}{n} S_n^2.$$

This estimator is very similar to the sample variance S_n^2 (they are asymptotically the same).

Using how we define the inverse of a CDF of a discrete random variable, we can define the estimator of median

$$T_{\mathsf{med}}(\widehat{F}_n) = \widehat{F}_n^{-1}(0.5)$$

and other quantiles of a distribution. And it turns out that this estimator is the sample median (and the corresponding sample quantiles)!

Therefore, the statistical functional provides an elegant way to define a population quantities as well as an estimator. And the plug-in estimator will be a good estimator if the statistical functional $T(\cdot)$ is 'smooth' with respect to the input function because we know that $\widehat{F}_n \to F$ in various ways so that the smoothness of T with respect the input will implies $T(\widehat{F}_n) \to T(F)$[1].

9.2 Bootstrap and Statistical Functionals

So far, we have not yet talked about the bootstrap. However, we have learned that the (empirical) bootstrap sample is a new random sample from the EDF $\widehat{F}_n$. The bootstrap sample forms another EDF called the bootstrap EDF, denoted as $\widehat{F}^*_n$. Namely, let $X^*_1, \cdots, X^*_n$ be a bootstrap sample. Then the bootstrap EDF is

$$\widehat{F}^*_n(x) = \frac{1}{n}\sum_{i=1}^n I(X^*_i \leq x).$$

Here is how the statistical functionals and the bootstrap is connected. In estimating the parameter $\theta = T_{\mathsf{target}}(F)$, we often use a plug-in estimate from the EDF $\widehat{\theta}_n = T_{\mathsf{target}}(\widehat{F}_n)$ (just think of how we estimate the sample mean). In this case, the bootstrap estimator, the estimator using the bootstrap sample, will be

$$\widehat{\theta}^*_n = T_{\mathsf{target}}(\widehat{F}^*_n),$$

another plug-in estimator but now we are plugging in the bootstrap EDF $\widehat{F}^*_n$.

Consistency of bootstrap variance estimator. How do we use the bootstrap to estimate the variance and construct a confidence interval? We keep generating bootstrap samples from the EDF $\widehat{F}_n$ and obtain several realizations of $\widehat{\theta}^*_n$'s. Namely, we generate

$$\widehat{\theta}^{*(1)}_n, \cdots, \widehat{\theta}^{*(B)}_n$$

and use their sample variance, $\widehat{\mathsf{Var}}_B(\widehat{\theta}^*_n)$, as an estimator of $\mathsf{Var}(\widehat{\theta}_n)$. Note that $\widehat{\mathsf{Var}}_B(\widehat{\theta}^*_n)$ is

$$\widehat{\mathsf{Var}}_B(\widehat{\theta}^*_n) = \frac{1}{B-1}\sum_{\ell=1}^N \left(\widehat{\theta}^{*(\ell)}_n - \bar{\widehat{\theta}}^*_{n,B}\right), \quad \bar{\widehat{\theta}}^*_{n,B} = \frac{1}{B}\sum_{\ell=1}^B \widehat{\theta}^{*(\ell)}_n.$$

When B is large, the sample variance of the bootstrap estimators

$$\widehat{\mathsf{Var}}_B(\widehat{\theta}^*_n) \approx \mathsf{Var}(\widehat{\theta}^*_n|\widehat{F}_n). \tag{9.1}$$

Note that $\cdot|\widehat{F}_n$ means *conditioned* on $\widehat{F}_n$ being fixed. The reason why here it converges to this conditioned variance is because when we generate bootstrap samples, the original EDF $\widehat{F}_n$ is fixed (and we are generating from it). Thus, the variance is conditioned on $\widehat{F}_n$ being fixed.

To argue that the bootstrap variance $\widehat{\mathsf{Var}}_B(\widehat{\theta}^*_n)$ is a good estimate of the original variance, we need to argue

$$\widehat{\mathsf{Var}}_B(\widehat{\theta}^*_n) \approx \mathsf{Var}(\widehat{\theta}^*_n|\widehat{F}_n) \approx \mathsf{Var}(\widehat{\theta}_n).$$

However, because of equation (9.1) and we can select B as large as we wish, so what really matters is

$$\mathsf{Var}(\widehat{\theta}^*_n|\widehat{F}_n) \approx \mathsf{Var}(\widehat{\theta}_n).$$

[1] Note that here we ignore lots of technical details. The smoothness of a 'functional' is an advanced topic in mathematics called *functional analysis*: https://en.wikipedia.org/wiki/Functional_analysis. There are formal ways of defining continuity of functionals and even 'differentiation' of functionals; see, e.g., https://en.wikipedia.org/wiki/G%C3%A2teaux_derivative.

Or more formally,

$$\frac{\mathsf{Var}(\widehat{\theta}^*_n|\widehat{F}_n)}{\mathsf{Var}(\widehat{\theta}_n)} \approx 1 \tag{9.2}$$

(people generally use the ratio expression because both quantities often converge to 0 when the sample size $n \to \infty$).

Therefore, we conclude that

> *as long as we can show that equation (9.2) holds, the bootstrap variance is a good estimate of the variance of the estimator $\widehat{\theta}_n$.*

Because $\widehat{\theta}_n = T_{\mathsf{target}}(\widehat{F}_n)$ is a statistic (a function of our random sample $X_1, \cdots, X_n$), its distribution is completely determined by the distribution $X_1, \cdots, X_n$ are sampling from, which is F, and the sample size n. This implies that the variance of $\widehat{\theta}_n$ is determined by F and n as well. Therefore, we can write

$$\mathsf{Var}(\widehat{\theta}_n) = \mathsf{Var}(T_{\mathsf{target}}(\widehat{F}_n)) = \mathbb{V}_{n,\mathsf{target}}(F).$$

And it turns out that we often have

$$\mathbb{V}_{n,\mathsf{target}}(F) \approx \frac{1}{n}\mathbb{V}_{1,\mathsf{target}}(F) \equiv \frac{1}{n}\mathbb{V}_{\mathsf{target}}(F).$$

Note that here $\mathbb{V}_{n,\mathsf{target}}(\cdot), \mathbb{V}_{\mathsf{targe}}(\cdot)$ are both again statistical functionals!

Because the bootstrap estimator $\widehat{\theta}^*_n = T_{\mathsf{target}}(\widehat{F}^*_n)$, its conditional variance will be

$$\mathsf{Var}(\widehat{\theta}^*_n|\widehat{F}_n) = \mathsf{Var}(T_{\mathsf{target}}(\widehat{F}^*_n)|\widehat{F}_n) = \mathbb{V}_{n,\mathsf{target}}(\widehat{F}_n) \approx \frac{1}{n}\mathbb{V}_{\mathsf{target}}(\widehat{F}_n).$$

Thus, as long as

$$\mathbb{V}_{\mathsf{target}}(\widehat{F}_n) \approx \mathbb{V}_{\mathsf{target}}(F), \tag{9.3}$$

equation (9.2) holds. Namely, the bootstrap variance estimate will be a good estimator of the variance of the true estimator[2].

Validity of bootstrap confidence interval. How about the validity of the bootstrap confidence interval? Here is a derivation showing that the consistency of bootstrap variance estimator implies the validity of bootstrap confidence interval.

For the bootstrap confidence interval, a simple way is first show that

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \sqrt{n}\left(T_{\mathsf{target}}(\widehat{F}_n) - T_{\mathsf{target}}(F)\right) \approx N(0, \mathbb{V}_{\mathsf{target}}(F)) \tag{9.4}$$

which implies

$$\sqrt{n}(\widehat{\theta}^*_n - \widehat{\theta}_n) = \sqrt{n}\left(T_{\mathsf{target}}(\widehat{F^*}_n) - T_{\mathsf{target}}(\widehat{F}_n)\right) \approx N(0, \mathbb{V}_{\mathsf{target}}(\widehat{F}_n)).$$

Thus, as long as the bootstrap variance converges, we also have the convergence of the entire distribution, implying the validity of a bootstrap confidence interval [3].

[2] A more formal way is to show that it converges in probability.

[3] To formally prove this, we need to show the convergence in terms of CDF of the difference. In more details, let $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta)$ and $Z^*_n = \sqrt{n}(\widehat{\theta}^*_n - \widehat{\theta}_n)$. We need to prove

$$\sup_t \left|P(Z^*_n \le t|\widehat{F}_n) - P(Z_n \le t)\right| \xrightarrow{P} 0.$$

Example: mean. We now consider a simple example: the mean of a distribution $T_{\mathsf{target}} = T_{\mathsf{mean}}$. The mean of a distribution has the form

$$\mu = T_{\mathsf{mean}}(F) = \int x dF(x).$$

The plug-in estimator is

$$\widehat{\mu}_n = T_{\mathsf{mean}}(\widehat{F}_n) = \int x d\widehat{F}_n(x) = \bar{X}_n$$

and the bootstrap estimator is

$$\widehat{\mu}_n^* = T_{\mathsf{mean}}(\widehat{F}_n^*) = \int x d\widehat{F}_n^*(x) = \bar{X}_n^*.$$

In it clearly from the Central Limit Theorem that

$$\sqrt{n}(\widehat{\mu}_n - \mu) \approx N(0, \mathsf{Var}(T_{\mathsf{mean}}(\widehat{F}_n)))$$

so equation (9.4) holds and

$$\sqrt{n}(\widehat{\mu}_n^* - \widehat{\mu}_n) \approx N(0, \mathsf{Var}(T_{\mathsf{mean}}(\widehat{F}_n^*)|\widehat{F}_n)).$$

In this case, we know that

$$\mathsf{Var}(T_{\mathsf{mean}}(\widehat{F}_n)) = \mathsf{Var}(\bar{X}_n) = \frac{1}{n}\mathsf{Var}(X_i) \Longrightarrow \mathbb{V}_{\mathsf{mean}}(F) = \mathsf{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}^2(X_i) = \int x^2 dF(x) - \left(\int x dF(x)\right)^2.$$

Therefore, the bootstrap variance is

$$\mathsf{Var}(T_{\mathsf{mean}}(\widehat{F}_n^*)|\widehat{F}_n) = \frac{1}{n}\mathbb{V}_{\mathsf{mean}}(\widehat{F}_n) = \int x^2 d\widehat{F}_n(x) - \left(\int x d\widehat{F}_n(x)\right)^2.$$

Because of the Law of Large Number,

$$\int x^2 d\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} \mathbb{E}(X_i^2) = \int x^2 dP(x)$$
$$\int x d\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mathbb{E}(X_i) = \int x dP(x).$$

Thus, [4]

$$\mathbb{V}_{\mathsf{mean}}(\widehat{F}_n) \xrightarrow{P} \mathbb{V}_{\mathsf{mean}}(F),$$

which shows that equation (9.3) holds and so is equation (9.2). Thus, the bootstrap variance estimator converges to the true variance estimator and we conclude that

$$\frac{\mathsf{Var}(T_{\mathsf{mean}}(\widehat{F}_n^*)|\widehat{F}_n)}{\mathsf{Var}(T_{\mathsf{mean}}(\widehat{F}_n))} \xrightarrow{P} 1.$$

As a result, the bootstrap variance estimator is consistent and the bootstrap confidence interval is also valid.

---

[4]Note that here we use the continuous mapping theorem: if f is a continuous function and random variable $A_n \xrightarrow{P} a_0$, then $f(A_n) \xrightarrow{P} f(a_0)$. Setting $f(x) = x^2$, we obtain the convergence of the second quantity.

9.3 Delta Method

In this section, we will talk about a very useful technique in handling the convergence–the *delta method*. We start with an example of proving consistency theorem of some bootstrap estimates.

Example: inverse of mean. Assume we are interested in the inverse of the population mean. Namely, the statistical functional we will be using is

$$T_{\mathsf{inv}}(F) = \frac{1}{\int x dF(x)} = \lambda.$$

This statistical functional was implicitly used when we the MLE of the rate parameter of an exponential distribution. The plug-in estimator (as well as the MLE of estimating an exponential model) is

$$\widehat{\lambda}_n = T_{\mathsf{inv}}(\widehat{F}_n) = \frac{1}{\int x d\widehat{F}_n(x)} = \frac{1}{\bar{X}_n}.$$

The bootstrap estimator is

$$\widehat{\lambda}^*_n = T_{\mathsf{inv}}(\widehat{F}^*_n) = \frac{1}{\int x d\widehat{F}^*_n(x)} = \frac{1}{\bar{X}^*_n}.$$

In the lab session, we have shown that this estimator follows asymptotically a normal distribution. But how do we show this? and how do we compute the variance of the estimator $\widehat{\lambda}_n$? Here is how the delta method will help us.

The Delta Method

Assume that we have a sequence of random variables $Y_1, \cdots, Y_n \cdots$ such that

$$\sqrt{n}(Y_n - y_0) \xrightarrow{D} N(0, \sigma^2_Y) \tag{9.5}$$

for some constants y_0 and σ^2_Y. Note that this implies that $\mathsf{Var}(Y_n) = \sigma^2_Y$. If a function f is differentiable at y_0, then using the Taylor expansion,

$$\sqrt{n}\left(f(Y_n) - f(y_0)\right) \approx \sqrt{n} f'(y_0) \cdot (Y_n - y_0) = f'(y_0)\sqrt{n}\left(Y_n - y_0\right).$$

Notice that $f'(y_0)$ is just a constant. Thus, this implies

$$\sqrt{n}\left(f(Y_n) - f(y_0)\right) \approx N(0, |f'(y_0)|^2\sigma^2_Y), \quad \mathsf{Var}(f(Y_n)) \approx \frac{1}{n}|f'(y_0)|^2\sigma^2_Y. \tag{9.6}$$

Now using equation (9.6) and identifying Y_n as $\bar{X}_n$ and $f(x)$ as $\frac{1}{x}$, we obtain

$$\sqrt{n}(\widehat{\lambda}_n - \lambda) = \sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mathbb{E}(X_i)}\right) \approx -\frac{1}{\mathbb{E}^2(X_i)}\sqrt{n}\left(\bar{X}_n - \frac{1}{\mathbb{E}^2(X_i)}\right) \approx N\left(0, \underbrace{\frac{1}{\mathbb{E}^4(X_i)}\mathsf{Var}(X_i)}_{=\mathbb{V}_{\mathsf{inv}}(F)}\right).$$

Using the fact that $\mathbb{E}(X_i) = \int x dF(x)$ and $\mathsf{Var}(X_i) = \int x^2 dF(x) - \left(\int x dF(x)\right)^2$, we obtain

$$\sqrt{n}(\widehat{\lambda}_n - \lambda) \approx N(0, \mathbb{V}_{\mathsf{inv}}(F)),$$

where
$$\mathbb{V}_{\mathsf{inv}}(F) = \frac{\int x^2 dF(x) - \left(\int x dF(x)\right)^2}{\left(\int x dF(x)\right)^4}.$$
So equation (9.4) holds and
$$\sqrt{n}(\widehat{\lambda}^*_n - \widehat{\lambda}_n) \approx N(0, \mathbb{V}_{\mathsf{inv}}(\widehat{F}_n)),$$
where
$$\mathbb{V}_{\mathsf{inv}}(\widehat{F}_n) = \frac{\int x^2 d\widehat{F}_n(x) - \left(\int x d\widehat{F}_n(x)\right)^2}{\left(\int x d\widehat{F}_n(x)\right)^4}$$
is the corresponding bootstrap variance component.

As long as $\int x dF(x) \neq 0$, each component in $\mathbb{V}_{\mathsf{inv}}(\widehat{F}_n)$ is a natural estimator of the corresponding component in $\mathbb{V}_{\mathsf{inv}}(F)$. Therefore, we conclude
$$\mathbb{V}_{\mathsf{inv}}(\widehat{F}_n) \overset{P}{\to} \mathbb{V}_{\mathsf{inv}}(F),$$
which shows that equation (9.3) holds, implying that the bootstrap variance estimator is consistent:
$$\frac{\mathsf{Var}(T_{\mathsf{inv}}(\widehat{F}^*_n)|\widehat{F}_n)}{\mathsf{Var}(T_{\mathsf{inv}}(\widehat{F}_n))} \overset{P}{\to} 1$$
and moreover, the bootstrap confidence interval is also valid.

9.4 Influence Function

9.4.1 Linear Functional

In the above derivations, we see many examples of statistical functionals that are of the form
$$T_\omega(F) = \int \omega(x) dF(x),$$
where g is a function. As we have mentioned, this type of statistical functionals are called *linear* functionals.

Linear functionals has a feature that the estimators
$$T_\omega(\widehat{F}_n) = \int \omega(x) d\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n g(X_i),$$
$$T_\omega(\widehat{F}^*_n) = \int \omega(x) d\widehat{F}^*_n(x) = \frac{1}{n}\sum_{i=1}^n g(X^*_i).$$

Moreover, a powerful feature of the linear functional is that for another CDF G, we always have
$$\begin{aligned} T_\omega(G) - T_\omega(F) &= \int \omega(x) dG(x) - T_\omega(F) \\ &= \int \omega(x) dG(x) - \int T_\omega(F) dF(x) \\ &= \int L_F(x) dG(x), \end{aligned}$$

where

$$L_F(x) = \omega(x) - T_\omega(F) \tag{9.7}$$

is called the *influence function* of the functional T_ω.

The influence is a power tool because when we replace G in the above derivation by $\widehat{F}_n$, we obtain

$$T_\omega(\widehat{F}_n) - T_\omega(F) = \int L_F(x) d\widehat{F}_n(x) = \frac{1}{n} \sum_{L_F(X_i)}.$$

Moreover,

$$\mathbb{E}(L_F(X_i)) = \int L_F(x) dF(x) = \int (\omega(x) - T_\omega(F))\, dF(x) = T_\omega(F) - T_\omega(F) = 0.$$

Thus, by central limit theorem,

$$\sqrt{n}\left(T_\omega(\widehat{F}_n) - T_\omega(F)\right) \approx N\left(0, \mathbb{V}_\omega(F) = \int L_F^2(x) dF(x)\right)$$

(you can check that the variance of $\sqrt{n}\left(T_\omega(\widehat{F}_n) - T_\omega(F)\right)$ is indeed $\int L_F^2(x)dF(x)$). Namely, for a linear functional T_ω, equation (9.4) always holds with

$$\mathbb{V}_\omega(F) = \int L_F^2(x) dF(x) = \int \left(\omega^2(x) - 2\omega(x) T_\omega(F) - T_\omega^2(F)\right) dF(x) = \int \omega^2(x) dF(x) = T_{\omega^2}(F). \tag{9.8}$$

Moreover,

$$\begin{aligned} \mathbb{V}_\omega(\widehat{F}_n) &= \int L_{\widehat{F}_n}^2(x) d\widehat{F}_n(x) = \int \left(\omega^2(x) - 2\omega(x) T_\omega(\widehat{F}_n) + T_\omega^2(\widehat{F}_n)\right) d\widehat{F}_n(x) \\ &= \int \omega^2(x) d\widehat{F}_n(x) - T_\omega^2(\widehat{F}_n). \end{aligned} \tag{9.9}$$

By Law of Large Number (and continuous mapping theorem),

$$T_\omega^2(\widehat{F}_n) \xrightarrow{P} T_\omega^2(F) = 0$$

if $\mathbb{E}(|\omega(X_i)|) = T_{|\omega|} < \infty$. And

$$\int \omega^2(x) d\widehat{F}_n(x) = T_{\omega^2}(\widehat{F}_n) \xrightarrow{P} T_{\omega^2}(F) = \mathbb{V}_\omega(F)$$

if $\mathbb{E}(\omega(X_i)^2) = T_{\omega^2}(F) < \infty$. Therefore, we conclude that when $T_{\omega^2}(F) < \infty$,

$$\mathbb{V}_\omega(\widehat{F}_n) = \int \omega^2(x) d\widehat{F}_n(x) - T_\omega^2(\widehat{F}_n) \xrightarrow{P} \mathbb{V}_\omega(F) + 0^2 = \mathbb{V}_\omega(F),$$

implying that the equation (9.3) holds. As a result, the bootstrap always works for the linear functional whenever $T_{\omega^2}(F) < \infty$.

9.4.2 Non-linear Functional

Although the linear functional has so many beautiful properties, many statistical functionals are not linear. For instance, the median

$$T_{\mathsf{med}}(F) = F^{-1}(0.5)$$

is not a linear functional. Therefore, our results of linear functional cannot be directly applied to analyze the median.

Then how can we analyze the properties of non-linear statistical functionals? One way to proceed is to generalize the notion of influence function. And here is the formal definition of the influence function.

Let δ_x be a point mass at location x. The *influence function* of a (general) statistical function $T_{\sf target}$ is

$$L_F(x) = \lim_{\epsilon\to 0} \frac{T_{\sf target}((1-\epsilon)F + \epsilon\delta_x) - T_{\sf target}(F)}{\epsilon}. \tag{9.10}$$

Some of you may find equation (9.10) very familiar; it seems to be taking a derivative. And yes – it is a derivative of a functional with respect to a function. This type of derivative is called *Gâteaux derivative*[5], a type of derivative of functionals. You can check that applying equation(9.10) to a linear functional leads to an influence function as we defined previously.

A powerful feature of this generalized version of influence function is that when the statistical functional $T_{\sf target}$ is 'smooth[6]', equation (9.8) and (9.9) hold in the sense that

$$\mathbb{V}_{\sf target}(F) = \int L_F^2(x)dF(x), \quad \mathbb{V}_{\sf target}(\widehat{F}_n) = \int L^2_{\widehat{F}_n}(x)d\widehat{F}_n(x) \tag{9.11}$$

and, moreover, equation (9.4) holds. Note that $L_{\widehat{F}_n}(x)$ is defined via replacing F by $\widehat{F}_n$ in equation (9.10). That is, when the statistical functional $T_{\sf target}$ is smooth, we only need to verify

$$\int L^2_{\widehat{F}_n}(x)d\widehat{F}_n(x) \approx \int L_F^2(x)dF(x)$$

to argue the validity of bootstrap consistency.

Example: median. Why median follows a normal distribution? Here we will show this using the influence function. The influence function of the functional $T_{\sf med}$ is

$$L_F(x) = \frac{1}{2p(F^{-1}(0.5))},$$

where p is the PDF of F (you can verify it). Thus, equation (9.4) implies

$$\sqrt{n}\left(\underbrace{T_{\sf med}(\widehat{F}_n)}_{\text{sample median}} - \underbrace{T_{\sf med}(F)}_{\text{population median}}\right) \approx N\left(0, \frac{1}{4p^2(F^{-1}(0.5))}\right).$$

Note that $F^{-1}(0.5) = T_{\sf med}(F)$ is the median of F. So this shows not only the asymptotic normality of sample median but also its limiting variance, which is inversely related to the PDF at the median.

The influence function is also related to the robustness of an estimator[7] and plays a key role in the semi-parametric statistics[8]. You would encounter it several times if you want to pursue a career in statistics.

[5] https://en.wikipedia.org/wiki/G%C3%A2teaux_derivative.

[6] More precisely, we need it to be Hadamard differentiable with respect to the L_∞ metric $d(F,G) = \sup_x |F(x) - G(x)|$; see https://en.wikipedia.org/wiki/Hadamard_derivative

[7] https://en.wikipedia.org/wiki/Robust_statistics#Influence_function_and_sensitivity_curve

[8] https://en.wikipedia.org/wiki/Semiparametric_model