Estimation of statistical functionals

(notes by Pavel Chigansky)

Let $X_1, \dots, X_n$ be an i.i.d. sample from an unknown distribution F on $\mathbb{R}^d$. Rather than estimating F itself, it is often required to estimate some *functional* of F, that is, a map from the space of distribution functions to $\mathbb{R}^k$. Common examples of such statistical functionals are

$$
\begin{aligned}
\mu_F &= \int_{\mathbb{R}} x dF(x), && \text{(mean)} \\
\sigma_F^2 &= \int_{\mathbb{R}} (x-\mu_F)^2 dF(x), && \text{(variance)} \\
q_F(p) &= \inf\Big\{x \in \mathbb{R} : F(x) \ge p\Big\}, && \text{(p-quantile[1]).}
\end{aligned}
$$

If a functional can be applied to purely discrete distributions, a reasonable estimator can be obtained by replacing F with the empirical distribution $\widehat{F}_n$. It is easy to see that for the examples above, the empirical mean, variance and quantiles are obtained (Problem 1). This chapter explores some of the properties of such *plug-in* or *substitution* estimators. For simplicity the case $d = k = 1$ will be considered, but the same theory applies in higher dimension.

1. Nonparametric Delta-method

Delta-method is a common technique of constructing consistent estimators and confidence sets for smooth statistical functionals.

1.1. A refresh on parametric Delta-method. Let us first recall the standard parametric setup, when F belongs to a parametric family $\mathcal{F}_\Theta = (F_\theta)_{\theta\in\Theta}$ with $\Theta \subseteq \mathbb{R}^m$. In this case, any functional of F can be viewed as a function of θ,

$$q(\theta) := T(F_\theta), \quad \theta \in \Theta.$$

Typically, under mild regularity conditions, such as identifiability and smoothness in θ, standard estimation methods produce a sequence of estimators $\widehat{\theta}_n$, which is

lecture notes for "Advanced Statistical Models B" course.

[1] $F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \ge p\}$ is a generalized inverse: when F is strictly increasing, this reduces to the usual inverse.

$\sqrt{n}$-consistent and asymptotically normal with the limit covariance matrix $\Sigma(\theta)$,

$$\sqrt{n}\big(\widehat{\theta}_n - \theta\big) \xrightarrow[n\to\infty]{d(\mathbb{P}_F)} N\big(0, \Sigma(\theta)\big). \tag{1.1}$$

Then, if $q(\theta)$ has a continuous gradient $\nabla q(\theta)$, the plug-in estimator $q(\widehat{\theta}_n)$ is also $\sqrt{n}$-consistent and asymptotically normal for $q(\theta)$,

$$\sqrt{n}\big(q(\widehat{\theta}_n) - q(\theta)\big) \xrightarrow[n\to\infty]{d(\mathbb{P}_F)} N\Big(0, \nabla q(\theta)\Sigma(\theta)\nabla q(\theta)^\top\Big), \quad \forall F \in \mathcal{F}_\Theta. \tag{1.2}$$

If the limit covariance matrix has a continuous inverse, this limit allows to construct asymptotic confidence sets, such as ellipsoids, boxes, etc. (Problem 3). In dimension one, $d = k = 1$, the interval with endpoints

$$q(\widehat{\theta}_n) \pm \frac{1}{\sqrt{n}}\big|q'(\widehat{\theta}_n)\big|\sqrt{\Sigma(\widehat{\theta}_n)}z_{1-\alpha/2}$$

is an asymptotic $1-\alpha$ confidence interval.

1.2. Linear functionals. It is instructive to consider first the elementary case of *linear* functionals of the form

$$T(F) = \int_{\mathbb{R}} \psi(x)dF(x), \tag{1.3}$$

where $\psi : \mathbb{R} \mapsto \mathbb{R}$ is a known function. The plug-in estimator in this case is the empirical mean of ψ,

$$T(\widehat{F}_n) = \int_{\mathbb{R}^d} \psi(x)d\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^n \psi(X_j).$$

By the strong LLN, this estimator is consistent, if $\mathbb{E}_F|\psi(X_1)| < \infty$. Furthermore, if $\mathbb{E}_F\psi(X_1)^2 < \infty$, it is asymptotically normal at rate $\sqrt{n}$,

$$\sqrt{n}\big(T(\widehat{F}_n) - T(F)\big) \xrightarrow[n\to\infty]{d(\mathbb{P}_F)} N\big(0, V_F\big),$$

where the asymptotic variance is

$$V_F = \mathrm{Var}_F(\psi(X_1)) = \int (\psi(x) - T(F))^2 dF(x). \tag{1.4}$$

The plug-in estimator $V_{\widehat{F}_n}$ for V_F is consistent (Problem 2) and hence, by Slutsky's lemma, the interval with endpoints at

$$T(\widehat{F}_n) \pm \frac{1}{\sqrt{n}}\sqrt{V_{\widehat{F}_n}} z_{1-\alpha/2} \tag{1.5}$$

is an asymptotic $1-\alpha$ confidence interval for $T(F)$.

1.3. The influence function. The integrand in (1.4) can be identified with the following generalization of the directional derivative.

DEFINITION 1.1. *Let F and G be distribution functions. A functional T is Gâteaux differentiable at F in the direction $\Delta := G - F$ if there exists a linear functional $\Delta \mapsto \dot{T}_F(\Delta)$, called the Gâteaux derivative, such that*

$$\frac{T(F_\varepsilon) - T(F)}{\varepsilon} \xrightarrow[\varepsilon \to 0]{} \dot{T}_F(\Delta),$$

where $F_\varepsilon = (1-\varepsilon)F + \varepsilon G = F + \varepsilon\Delta$.

For the linear functional in (1.3),

$$T(F_\varepsilon) - T(F) = \int \psi\big((1-\varepsilon)dF + \varepsilon dG\big) - \int \psi dF = \varepsilon \int \psi d\Delta,$$

and therefore $T(F)$ is Gâteaux differentiable with the derivative

$$\dot{T}_F(\Delta) = \int \psi d\Delta = T(G) - T(F). \tag{1.6}$$

DEFINITION 1.2. *The Gâteaux derivative of T at F in the direction* [2] *$\delta_x - F$ is called the influence function,*

$$L_F(x) := \dot{T}_F(\delta_x - F).$$

In view of (1.6), the influence function for the linear functional (1.3) is given by

$$L_F(x) = T(\delta_x) - T(F) = \psi(x) - T(F),$$

and the expression for the limit variance (1.4) takes the form

$$V_F = \int L_F(x)^2 dF(x). \tag{1.7}$$

Analogous formula can be obtained for linear functionals of multivariate distributions.

1.4. Nonlinear functionals. The results above were derived by elementary means. Do they remain valid beyond the simple linear functionals of the form (1.3)? An affirmative answer to this question can be given in a greater generality, following the same pattern as in the parametric case. To this end, a stronger notion of derivative is required.

DEFINITION 1.3. *A functional T is Hadamard differentiable at F in the direction $\Delta := G - F$ if there exists a linear functional $\Delta \mapsto \dot{T}_F(\Delta)$, called the Hadamard derivative, such that for any path of distributions G_ε such that $\|G_\varepsilon - G\|_\infty \to 0$,*

$$\frac{T(F_\varepsilon) - T(F)}{\varepsilon} \xrightarrow[\varepsilon \to 0]{} \dot{T}_F(\Delta), \tag{1.8}$$

where $F_\varepsilon = (1-\varepsilon)F + \varepsilon G_\varepsilon$.

---

[2] δ_x is the point measure at x, i.e. $\delta_x(A) = \mathbf{1}_{\{x \in A\}}$ for any measurable subset $A \subseteq \mathbb{R}$. With a convenient abuse of notations, the same symbols will be used interchangeably for measures and their distribution functions.

If T is Hadamard differentiable, then, in particular, convergence in (1.8) holds for the constant path $G_\varepsilon := G$. Hence it is also Gâteaux differentiable with the same derivative. The converse, however, is not necessarily true.

THEOREM 1.4. *Suppose T is Hadamard differentiable at F in all directions, then*

$$\sqrt{n}\big(T(\widehat{F}_n) - T(F)\big) \xrightarrow[n\to\infty]{d(\mathbb{P}_F)} N(0, V_F),$$

where V_F is given by (1.7) *with the influence function*

$$L_F(x) = \dot{T}_F(\delta_x - F).$$

PROOF. (a rough sketch) Recall that Donsker's theorem asserts that

$$\sqrt{n}(\widehat{F}_n - F) \xrightarrow[n\to\infty]{w(\mathbb{P}_F)} \overline{B}\circ F, \tag{1.9}$$

where convergence is weak with respect to supremum norm [3] and $\overline{B}$ is the Brownian bridge, that is, the Gaussian process with zero mean and covariance function

$$\mathbb{E}\overline{B}(x)\overline{B}(y) = x\wedge y - xy, \quad x, y \in [0,1]. \tag{1.10}$$

It can be shown that if T is Hadamard differentiable at F in all directions, then the weak convergence (1.9) implies

$$\sqrt{n}\big(T(\widehat{F}_n) - T(F)\big) = \frac{T\big(F + \frac{1}{\sqrt{n}}\sqrt{n}(\widehat{F}_n - F)\big) - T(F)}{\frac{1}{\sqrt{n}}} \xrightarrow[n\to\infty]{d(\mathbb{P}_F)} \dot{T}_F(\overline{B}\circ F).$$

Since $\dot{T}_F$ is a linear functional, the limit has zero mean Gaussian distribution. It remains to derive the expression for its variance.

If the linear functional $\dot{T}_F$ is sufficiently regular, it has the form of integral operator

$$\dot{T}_F(\Delta) = \int g(x) d\Delta(x),$$

for some function g. If $\lim_{x\to\pm\infty}\Delta(x) = 0$, integrating by parts gives

$$\dot{T}_F(\Delta) = -\int \Delta(x) dg(x), \tag{1.11}$$

[3] Recall that a sequence of random processes X_n converges to a random process X weakly with respect to a norm d, denoted $X_n \xrightarrow{w} X$, if $\mathbb{E}T(X_n) \to \mathbb{E}T(X)$ for any bounded functional T, continuous in d.

if g is smooth enough. Since the Brownian bridge vanishes at 0 and 1, the process $\overline{B}\circ F$ vanishes at $\pm\infty$ and consequently

$$
\begin{aligned}
&\mathbb{E}_F\Big(\dot{T}_F(\overline{B}\circ F)\Big)^2 = \mathbb{E}_F\Big(\int(\overline{B}\circ F)(x)dg(x)\Big)^2 = \\
&\int\int \mathbb{E}_F(\overline{B}\circ F)(x)(\overline{B}\circ F)(y)dg(x)dg(y) = \\
&\int\int \Big(F(x\wedge y)-F(x)F(y)\Big)dg(x)dg(y) \stackrel{\dagger}{=} \\
-&\int\int g(x)d\Big(F(x\wedge y)-F(x)F(y)\Big)dg(y) = \\
-&\int\Big(\int_{-\infty}^{y} g(x)dF(x)-F(y)\int gdF\Big)dg(y) \stackrel{\dagger}{=} \\
&\int g(y)d\Big(\int_{-\infty}^{y} g(x)dF(x)-F(y)\int gdF\Big) = \\
&\int g^2dF-\Big(\int gdF\Big)^2 = \int\Big(g(x)-\int gdF\Big)^2 dF(x) = \\
&\int\Big(\int g(y)d(\mathbf{1}_{\{y\ge x\}}-F(y))\Big)^2 dF(x) = \int \dot{T}_F(\delta_x-F)^2 dF(x) = V_F,
\end{aligned}
$$

where † was obtained by integrating by parts. □

EXAMPLE 1.5. Assume that F has a continuous positive density function f. Fix a number $p\in(0,1)$ and consider the quantile functional

$$T(F)=\min\{x\in\mathbb{R}:F(x)\ge p\}=F^{-1}(p).$$

Note that this functional is nonlinear and cannot be put in the form (1.3). Let G_ε be a path of distribution functions, such that $\|G_\varepsilon-G\|_\infty\to 0$ for some distribution G. For brevity, let $q=T(F)$ and $q_\varepsilon=T\big((1-\varepsilon)F+\varepsilon G_\varepsilon\big)$. We aim at finding the limit $\lim_{\varepsilon\to 0}(q_\varepsilon-q)/\varepsilon$.

Since F is continuous, by the definition of quantile,

$$(1-\varepsilon)F(q_\varepsilon)+\varepsilon G_\varepsilon(q_\varepsilon)\ge p \quad\text{and}\quad (1-\varepsilon)F(q_\varepsilon)+\varepsilon G_\varepsilon(q_\varepsilon-)\le p. \tag{1.12}$$

Subtracting $F(q)=p$ from the first equation gives

$$F(q_\varepsilon)-F(q)\ge \varepsilon F(q_\varepsilon)-\varepsilon G_\varepsilon(q_\varepsilon),$$

or, equivalently,

$$\int_q^{q_\varepsilon} f(t)dt \ge \varepsilon\Big(F(q_\varepsilon)-G(q_\varepsilon)\Big)+\varepsilon\Big(G(q_\varepsilon)-G_\varepsilon(q_\varepsilon)\Big). \tag{1.13}$$

Similarly, the second equation in (1.12) implies the bound

$$\int_q^{q_\varepsilon} f(t)dt \le \varepsilon\Big(F(q_\varepsilon)-G(q_\varepsilon-)\Big)+\varepsilon\Big(G(q_\varepsilon-)-G_\varepsilon(q_\varepsilon-)\Big). \tag{1.14}$$

This implies $\int_q^{q_\varepsilon} f(t)dt \xrightarrow[\varepsilon\to 0]{} 0$ and, since f is positive, $q_\varepsilon\to q$.

Moreover, the lower bound (2.10) implies

$$\frac{q_\varepsilon - q}{\varepsilon}\frac{1}{q_\varepsilon - q}\int_q^{q_\varepsilon} f(t)dt \ge \Big(F(q_\varepsilon) - G(q_\varepsilon)\Big) + \|G - G_\varepsilon\|_\infty$$

and, since f is continuous,

$$\varliminf_{\varepsilon\to 0} \frac{q_\varepsilon - q}{\varepsilon} f(q) \ge F(q_\varepsilon) - \varlimsup_{\varepsilon\to 0} G(q_\varepsilon).$$

Similarly, the upper bound (1.14) gives

$$\varlimsup_{\varepsilon\to 0} \frac{q_\varepsilon - q}{\varepsilon} f(q) \le F(q_\varepsilon) - \varliminf_{\varepsilon\to 0} G(q_\varepsilon -).$$

These two estimates imply

$$T_F(F-G) = \lim_{\varepsilon\to 0}\frac{q_\varepsilon - q}{\varepsilon} = \frac{F(q)-G(q)}{f(q)},$$

whenever q is a continuity point of G. The influence function is obtained by evaluating this functional at the distribution $G(u) := \mathbf{1}_{\{u\ge x\}}$

$$L_F(x) = \frac{p - \mathbf{1}_{\{q\ge x\}}}{f(q)}, \quad x\in\mathbb{R}\setminus\{q\}.$$

The corresponding variance is

$$V_F = \int_{\mathbb{R}} L_F(x)^2 dF(x) = \frac{1}{f(q)^2}\int_{\mathbb{R}} (p-\mathbf{1}_{\{q\ge x\}})^2 dF(x) = \frac{p(1-p)}{f(q)^2} = \frac{p(1-p)}{f(F^{-1}(p))^2}.$$

The plug-in estimator for $T_p(F)$ is the empirical quantile

$$T_p(\widehat{F}_n) = \inf\big\{x\in\mathbb{R} : \widehat{F}_n(x)\ge p\big\} = \inf\Big\{x\in\mathbb{R} : \sum_{j=1}^n \mathbf{1}_{\{X_{(j)}\le x\}} \ge pn\Big\} = X_{(\lceil np\rceil)}.$$

To be able to construct an asymptotic confidence interval, we need a density estimator $\widehat{f}_n$, consistent with respect to the supremum norm. In the following chapters, we will see how such estimators can be constructed for smooth densities. Then

$$X_{(\lceil np\rceil)} \pm \frac{z_{1-\alpha/2}}{\sqrt{n}}\frac{\sqrt{p(1-p)}}{\widehat{f}_n(X_{(\lceil np\rceil)})}$$

is an asymptotic confidence interval with coverage probability $1-\alpha$. ■

Computer experiment. Implement the asymptotic confidence interval from Problem 7. Run $M = 10{,}000$ MC trials to approximate the actual coverage probability and the mean interval length for the sample of size $n = 100$ from some bivariate distribution. Compare to the target coverage probability. Repeat for the sample size of $n = 1000$. Explain the obtained results.

Exercises.

PROBLEM 1. Find the plug-in estimators for the mean, variance and quantile.

PROBLEM 2. Prove that the plug-in estimator for the functional V_F defined in (1.4) is strongly consistent,

$$V_{\widehat{F}_n} \xrightarrow[n\to\infty]{\mathbb{P}_F-a.s.} V_F,$$

for any ψ and F such that $\int \psi^2 dF < \infty$.

PROBLEM 3. Let $X_1,\dots,X_n \sim N(\mu,\sigma^2)$ where $\theta = (\mu,\sigma^2) \in \mathbb{R}\times\mathbb{R}_+$ is the unknown parameter. Apply parametric Delta method to the m.l.e. $\widehat{\theta}_n$ for θ in order to construct asymptotic $1-\alpha$ confidence intervals for the following functionals.

(1) The fourth moment $\mathbb{E}_\theta X_1^4$.

(2) The p-th quantile.

(3) The probability $\mathbb{P}_\theta(X_1 \in [a,b])$ where $a<b$ are fixed real numbers.

PROBLEM 4. Find the plug-in estimator for the functional

$$T(F) = \mathbb{P}_F(X_1 \in [a,b]) = F(b) - F(a),$$

where $a<b$ are fixed numbers. Find the influence function and the limit variance. Construct $1-\alpha$ asymptotic confidence interval for $T(F)$.

PROBLEM 5.

(1) Show the linear functional (1.3) is Hadamard differentiable at F in the direction $\delta_x - F$ for all $x\in\mathbb{R}$ if ψ is continuous and bounded. Find the derivative $L_F(x)$ in this case.

(2) Give an example of ψ with $\int|\psi|dF<\infty$, so that (1.3) is not Hadamard differentiable at F in direction $\delta_x - F$ for some x. Is it Gâteaux differentiable ? If so, find the derivative.

PROBLEM 6. Let $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} F$ and consider estimation problem for the functional

$$T(F) = h\big(T_1(F),\dots,T_k(F)\big),$$

where $h:\mathbb{R}^k \mapsto \mathbb{R}$ is continuously differentiable and $T_j(F)$ are linear functionals of the form (1.3) with weight functions ψ_j, such that T_j's are Hadamard differentiable. Construct asymptotic confidence interval for $T(F)$, using the nonparametric Delta-method.

PROBLEM 7. Consider the problem of estimating the correlation coefficient

$$\rho_F = \frac{\mathrm{Cov}_F(X_1,Y_1)}{\sqrt{\mathrm{Var}_F(X_1)\,\mathrm{Var}_F(Y_1)}}$$

from the random sample $(X_1,Y_1),\dots,(X_n,Y_n)\sim F$, where F is a distribution on $\mathbb{R}^2$.

(1) Specify the plug-in estimator.

(2) Extending the result from the previous problem to distributions on the plane $\mathbb{R}^2$, prove that the influence function is

$$L_F(x,y) = \widetilde{x}\widetilde{y} - \frac{1}{2}\rho_F(\widetilde{x}^2 + \widetilde{y}^2),$$

where

$$\widetilde{x} = \frac{x - \int s dF(s,t)}{\sqrt{\int s^2 dF(s,t) - (\int s dF(s,t))^2}}, \quad \widetilde{y} = \frac{y - \int t dF(s,t)}{\sqrt{\int t^2 dF(s,t) - (\int t dF(s,t))^2}}.$$

Hint: the calculations are somewhat cumbersome

(3) Detail the plug-in estimator for the limit variance.

PROBLEM 8. Using nonparametric Delta-method, construct $1-\alpha$ confidence interval for the functional

$$T(F) = \int_0^1 (F(x) - x)^2 dx,$$

which measures the $L^2([0,1])$ proximity of a distribution F on the interval $[0,1]$ to the uniform distribution.