Avd. Matematisk statistik

KTH Teknikvetenskap

Sf 2955: Computer intensive methods :
STATISTICAL FUNCTIONALS, PLUG-IN ESTIMATES AND INFLUENCE FUNCTIONS

Timo Koski

1 Functionals of Distribution Functions

Let X be a random variable and let a distribution function F on the real line be defined as

$$F(x) = \mathbf{P}\left(X \leq x\right), \quad -\infty < x < \infty$$

and we assume that the true distribution function is a member of a class of distribution functions $\mathcal{M}$. This is basically a nonparametric statistical model, i.e., we do not assume the existence of a finite set of parameters that uniquely define the members of $\mathcal{M}$.

Often the quantity θ we are interested in estimating can be viewed as

$$\theta = T(F)$$

and we say that θ is a **statistical functional** (on $\mathcal{M}$). By this we mean that T maps the function F to a real number θ, or, as one writes in mathematics,

$$\mathcal{M} \ni F \stackrel{T}{\mapsto} \theta \in R.$$

This is readily understood by means of examples.

Example 1.1 Probability of an interval Let F be a distribution function and set for real numbers $a < b$

$$\theta = T(F) = F(b) - F(a).$$

We can also write this as

$$T(F)=\int_a^b dF(x)=\begin{cases}\int_a^b f(x)dx & X \text{ is a continuous r.v.}\\ \sum_{k:a<x_k\leq b} p\left(x_k\right) & X \text{ is a discrete r.v.}\end{cases} \tag{1.1}$$

■

Example 1.2 Mean $\theta=E[X]$, and thus

$$T(F)=\int_{-\infty}^{\infty} x dF(x)=\begin{cases}\int_{-\infty}^{\infty} x f(x)dx & X \text{ is a continuous r.v.}\\ \sum_k x_k p\left(x_k\right) & X \text{ is a discrete r.v.}\end{cases} \tag{1.2}$$

■

Example 1.3 r-th moment θ is the r-th moment and then

$$T(F)=\int_{-\infty}^{\infty} x^r dF(x),$$

where the right hand side can be written analogously to the right hand side in (1.2).

■

Example 1.4 Variance The variance $\theta=V(X)$ and

$$\theta=T(F)=\int_{-\infty}^{\infty}\left(x-\int_{-\infty}^{\infty} x dF(x)\right)^2 dF(x). \tag{1.3}$$

■

Example 1.5 Median θ is the median, if

$$\theta=T(F)=F^{-1}\left(\frac{1}{2}\right)\Leftrightarrow F(\theta)=\frac{1}{2}.$$

■

Example 1.6 Skewness Let $\mu = E[X]$ and $\sigma^2 = V[X]$. Then the **skewness** of X is measured by

$$\kappa = \frac{E(X-\mu)^3}{\sigma^3} = \theta. \tag{1.4}$$

Thus

$$T(F) = \frac{\int_{-\infty}^{\infty} (x-\mu)^3 \, dF(x)}{\left(\int_{-\infty}^{\infty} (x-\mu)^2 \, dF(x)\right)^{3/2}}. \tag{1.5}$$

This is a measure of the lack of symmetry of the distribution F.

■

Example 1.7 Covariance The covariance $\theta = \mathrm{Cov}(X,Y)$, now

$$F(x,y) = \mathbf{P}(X \le x, Y \le y)$$

$$F_X(x) = \mathbf{P}(X \le x, Y \le \infty), F_Y(y) = \mathbf{P}(X \le \infty, Y \le y)$$

and

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(x - \int_{-\infty}^{\infty} x dF_X(x)\right)\left(y - \int_{-\infty}^{\infty} y dF_Y(y)\right) dF(x,y) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy dF(x,y) - \int_{-\infty}^{\infty} x dF_X(x) \cdot \int_{-\infty}^{\infty} y dF_Y(y) \\
&= t_3 - t_1 t_2 = a(t_1, t_2, t_3).
\end{aligned}$$

We set

$$T_1(F) = \int_{-\infty}^{\infty} x dF_X(x), \quad T_2(F) = \int_{-\infty}^{\infty} y dF_Y(y)$$

$$T_3(F) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy dF(x,y).$$

Then

$$\theta = a(T_1(F), T_2(F), T_3(F)). \tag{1.6}$$

■

Example 1.8 Quantiles Let $F(x)$ be strictly increasing with a density. Let $0 \leq p \leq 1$ and then

$$T(F) = F^{-1}(p)$$

is the p:th quantile.

$\blacksquare$

Example 1.9 Mann-Whitney Functional Let $F(x)$ and $G(x)$ be two distribution functions and let $X \sim F$, $Y \sim G$ be independent random variables. Then

$$\theta = T(F,G) = P\left(X \leq Y\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} dF(x)dG(y).$$

is the Mann-Whitney functional.

As a check of the formula

$$P\left(X \leq Y\right) = \int_{-\infty}^{\infty} P\left(X \leq y \mid Y = y\right) dG(y)$$

and independence gives

$$= \int_{-\infty}^{\infty} P\left(X \leq y\right)) dG(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} dF(x)dG(y).$$

$\blacksquare$

Example 1.10 The chi-squared functional Let $A_l \subseteq R$ be a partition of the real line into k cells , i.e.,

$$A_l \cap A_r = \emptyset, l \neq r, \quad , \cup_{l=1}^{k} A_l = R.$$

Let $p_l, l = 1, 2, \ldots, k$ be a given probability distribution and F be a distribution function. Next

$$\mathbf{P}\left(A_l\right) = \int_{A_l} dF(x).$$

Then we define the chi-squared functional as

$$T(F) = \sum_{l=1}^{k} p_l^{-1}\left(\int_{A_l} dF(x) - p_l\right)^2.$$

$\blacksquare$

2 Empirical Distribution Function

We start with the definition. Let $X_1, \dots, X_n$ have the distribution $F \in \mathcal{M}$. The empirical distribution function is simply

$$\widehat{F}_n(x) = \frac{1}{n} \times (\text{ the number of } X_i \leq x).$$

More formalistically we can introduce the following.

Definition 2.1 The **empirical distribution function** $\widehat{F}_n(x)$ for $x \in R$ is defined by

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} U(x - X_i), \tag{2.1}$$

where $U(x)$ is the Heaviside function

$$U(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases} \tag{2.2}$$

■

Hence we see that the empirical distribution puts the probability mass $\frac{1}{n}$ on every data point $X_i = x_i$. Some first properties of $\widehat{F}_n(x)$ are given in the next theorem.

Proposition 2.1 Let $X_1, \dots, X_n$ be I.I.D. and have the distribution F. Then it holds that

1. For any fixed x

$$E_F\left(\widehat{F}_n(x)\right) = F(x), \tag{2.3}$$

   and

$$V_F\left(\widehat{F}_n(x)\right) = \frac{F(x)(1 - F(x)}{n}. \tag{2.4}$$

2. **Glivenko-Cantelli Theorem**

$$\sup_x \mid \widehat{F}_n(x) - F(x) \mid \overset{a.s.}{\rightarrow} 0, \tag{2.5}$$

   as $n \rightarrow 0$.

3. **Dvoretzky-Kiefer-Wolfowitz (DKW) Inequality** For any $\varepsilon > 0$

$$\mathbf{P}\left(\sup_x \mid \widehat{F}_n(x) - F(x) \mid > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}, \tag{2.6}$$

as $n \to 0$.

■

By Chebyshev′s inequality we have for any $\varepsilon > 0$ in view of (2.3) that

$$\mathbf{P}\left(\mid \widehat{F}_n(x) - F(x) \mid > \varepsilon\right) \leq \frac{1}{\varepsilon^2} V_F\left(\widehat{F}_n(x)\right)$$

and thus (2.4) gives

$$\mathbf{P}\left(\mid \widehat{F}_n(x) - F(x) \mid > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\frac{F(x)(1-F(x)}{n},$$

and this shows that we have convergence in probability

$$\widehat{F}_n(x) \xrightarrow{p} F(x),$$

as $n \to 0$.

3 Plug-in Estimates of Statistical Functionals

The class of estimators of $\theta = T(F)$ mostly analysed in this course is the **plug-in estimator** to be defined next.

Let $x_1, \ldots, x_n$ be a sample of I.I.D. random variables $X_1, \ldots, X_n$, respectively. Then

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} U\left(x - x_i\right). \tag{3.1}$$

Definition 3.1 $x_1, \ldots, x_n$ is a sample of I.I.D. random variables. The **plug-in estimator** $\widehat{\theta}_n$ of $\theta = T(F)$ on basis of $x_1, \ldots, x_n$ is defined by

$$\widehat{\theta}_n = T\left(\widehat{F}_n\right). \tag{3.2}$$

■

Example 3.1 The Mean Let $\theta = E[X]$. Then we get by example 1.2

$$T\left(\widehat{F}_n\right) = \int_{-\infty}^{\infty} x d\widehat{F}_n.$$

But then we observe that $\widehat{F}_n$ is a discrete probability distribution with the mass $\frac{1}{n}$ at every sample point, and hence we can apply the corresponding case in the right hand side of equation (1.2) to obtain

$$T\left(\widehat{F}_n\right) = \sum_{i=1}^{n} x_i \frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

For those familiar with the Heaviside function $U(x)$ we write

$$T\left(\widehat{F}_n\right) = \int_{-\infty}^{\infty} x d\widehat{F}_n = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} x dU\left(x - x_i\right).$$

The 'derivative' of the Heaviside function $U(x)$ is the Dirac functional, $\delta(x)$, in the sense that for any (sufficiently regular) function $\varphi(x)$

$$\int_{-\infty}^{\infty} \varphi(x) dU\left(x\right) = \int_{-\infty}^{\infty} \varphi(x)\delta(x) dx = \varphi(0).$$

Therefore we get for any i

$$\int_{-\infty}^{\infty} x dU\left(x - x_i\right) = \int_{-\infty}^{\infty} x\delta(x - x_i) dx = x_i.$$

Hence

$$\widehat{\theta}_n = T\left(\widehat{F}_n\right) = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}, \tag{3.3}$$

which gives the plug-in estimate of the mean as the arithmetic mean of the samples.

■

Example 3.2 The Variance Let $\theta = V[X]$. Then we have in example 1.4

$$\theta = T(F) = \int_{-\infty}^{\infty} \left(x - \int_{-\infty}^{\infty} x dF(x) \right)^2 dF(x)$$

$$= \int_{-\infty}^{\infty} x^2 dF(x) - \left(\int_{-\infty}^{\infty} x dF(x) \right)^2 .$$

Thus

$$T\left(\widehat{F}_n\right) = \int_{-\infty}^{\infty} x^2 d\widehat{F}_n - \left(\int_{-\infty}^{\infty} x d\widehat{F}_n \right)^2 .$$

By the plug-in example 3.1 above we get

$$= \int_{-\infty}^{\infty} x^2 d\widehat{F}_n - \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2 .$$

When we apply the same reasoning about the Heaviside function $U(x)$ as in the preceding example, we get

$$\int_{-\infty}^{\infty} x^2 d\widehat{F}_n = \frac{1}{n} \sum_{i=1}^{n} x_i^2 .$$

Thus from (3.3)

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2 .$$

As an algebraic identity we get

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \overline{x} \right)^2 . \tag{3.4}$$

■

Example 3.3 Skewness From equation (1.5) we get using (3.3) and (3.4)

$$\widehat{\kappa} = T\left(\widehat{F}_n\right) = \frac{\int_{-\infty}^{\infty} (x-\mu)^3 \, d\widehat{F}_n(x)}{\left(\int_{-\infty}^{\infty} (x-\mu)^2 \, d\widehat{F}_n(x) \right)^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^{n} \left(x_i - \overline{x} \right)^3}{\widehat{\sigma}^3} . \tag{3.5}$$

■

Example 3.4 Covariance In view of example 1.7 and the expression in (3.6) we have

$$\begin{aligned}\widehat{\theta} &= a\left(T_1\left(\widehat{F}_n\right), T_2\left(\widehat{F}_n\right), T_3\left(\widehat{F}_n\right)\right) \\ &= T_3\left(\widehat{F}_n\right) - T_1\left(\widehat{F}_n\right) T_2\left(\widehat{F}_n\right),\end{aligned} \tag{3.6}$$

and

$$T_1\left(\widehat{F}_n\right) = \frac{1}{n}\sum_{i=1}^{n} x_i, T_2\left(\widehat{F}_n\right) = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

$$T_3\left(\widehat{F}_n\right) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

Then we get, as is well known, by an algebraic manipulation that

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right).$$

■

Example 3.5 Estimation of Quantiles From example 1.8 we get the p:th quantile

$$T(F) = F^{-1}(p).$$

Then the plug-in estimate of the quantile is

$$T\left(\widehat{F}_n\right) = \widehat{F}_n^{-1}(p).$$

However, the empirical distribution function is not invertible, and we use

$$\widehat{F}_n^{-1}(p) = \inf_x \{x \mid \widehat{F}_n(x) \geq p\}.$$

■

4 Linear Functionals and Linear Estimators

Any statistical functional $T(F)$ of the form

$$T(F) = \int_{-\infty}^{\infty} a(x)\, dF(x)$$

is a **linear functional**. The functional T is called linear, as it satisfies

$$T(aF + bG) = aT(F) + bT(G) \tag{4.1}$$

for any two distribution functions F and G and any real numbers a and b. We have that

$$T(F) = \begin{cases} \int_{-\infty}^{\infty} a(x) f(x) dx & X \text{ is a continuous r.v.} \\ \sum_k a(x_k)\, p(x_k) & X \text{ is a discrete r.v.} \end{cases}$$

As in the preceding we find that

$$T\left(\widehat{F}_n\right) = \int_{-\infty}^{\infty} a(x)\, d\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} a(x_i)\,. \tag{4.2}$$

This is the plug-in representation of a linear statistical functional and is called a **linear estimator**. We shall now check a few examples of linear functionals and linear estimators.

Example 4.1 Probability of an interval In example 1.1 we introduced for real numbers $a < b$

$$\theta = T(F) = F(b) - F(a).$$

This is a linear functional, since we can write it as

$$T(F) = \int_a^b dF(x) = \int_{-\infty}^{\infty} I_{]a,b]}(x) dF(x)$$

where $a(x) = I_{]a,b]}(x)$ is the indicator function

$$I_{]a,b]}(x) = \begin{cases} 1 & \text{if } a < x \leq b \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Therefore the corresponding linear estimator (of $F(b) - F(a)$) is by (4.2)

$$F(\widehat{b) - F}(a) = \frac{1}{n} \sum_{i=1}^{n} a(x_i) = \frac{1}{n} \sum_{i=1}^{n} I_{]a,b]}(x_i) = \frac{\text{number of } x_i \text{ in }]a,b]}{n}$$

i.e., the relative frequency of the samples x_i hitting $]a, b]$.

$\blacksquare$

Example 4.2 The Variance with Known Mean Let in example 1.4 the mean $\mu = \int_{-\infty}^{\infty} x dF(x)$ be known. Then the variance is

$$T(F) = \int_{-\infty}^{\infty} (x - \mu)^2 \, dF(x),$$

which is a linear functional with

$$a(x) = (x - \mu)^2 .$$

The plug-in estimate of variance to be obtained by means of (4.2) is now

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

$\blacksquare$

Further linear estimators are the plug-in estimators of mean and r-th mean.

Proposition 4.3 Let $\theta = T(F) = \int_{-\infty}^{\infty} a(x) \, dF(x)$ be a linear functional and $\widehat{\theta} = T\left(\widehat{F}_n\right)$. Then we have the following:

1.
$$E_F\left[\widehat{\theta}\right] = \theta, \tag{4.4}$$
i.e., $\widehat{\theta}$ is unbiased.

2.
$$\widehat{\text{bias}}_{\text{boot}} = 0$$

Proof: We prove the latter assertion. By definition of $\widehat{\text{bias}}_{\text{boot}}$ applied to the linear plug-in estimator we get

$$\widehat{\text{bias}}_{\text{boot}} = E_{\widehat{F}_n}\left(\frac{1}{n} \sum_{i=1}^{n} a\left(X_i^*\right)\right) - \frac{1}{n} \sum_{i=1}^{n} a\left(x_i\right)$$

and as the bootstrap variables are identically distributed,

$$= E_{\widehat{F}_n} a\left(X_1^*\right) - \frac{1}{n} \sum_{i=1}^{n} a\left(x_i\right)$$

but $a(X_1^*)$ is a discrete random variable that assumes the values $\{a(x_i)\}_{i=1}^n$ with the respective probabilities $1/n$, we get

$$= \frac{1}{n}\sum_{i=1}^{n} a(x_i) - \frac{1}{n}\sum_{i=1}^{n} a(x_i) = 0.$$

■

5 Other Non-linear Estimators, Quadratic Estimators

Example 5.1 Estimation of the chi-squared functional The chi-squared functional in example 1.10 is not linear, but inside it

$$T_l(F) = \int_{A_l} dF(x) = \int_{-\infty}^{\infty} I_{A_l}(x) dF(x)$$

is a linear functional, with $I_{A_l}(x)$ being the indicator function of the cell A_l. Therefore we get from (4.2) that

$$T_l\left(\widehat{F}_n\right) = \frac{1}{n}\sum_{i=1}^{n} I_{A_l}(x_i) = \frac{\text{number of } x_i \text{ in } A_l}{n},$$

which is a generalization of the estimate in example 4.1 above. We write

$$\frac{n_l}{n} = \frac{\text{number of } x_i \text{ in } A_l}{n}.$$

Then we have for the the chi-squared functional

$$T\left(\widehat{F}_n\right) = \sum_{l=1}^{k} p_l^{-1}\left(\int_{A_l} d\widehat{F}_n(x) - p_l\right)^2$$

$$= \sum_{l=1}^{k} p_l^{-1}\left(\frac{n_l}{n} - p_l\right)^2 = \sum_{l=1}^{k}\left(n^2 p_l\right)^{-1}\left(n_l - np_l\right)^2.$$

Now we observe that

$$nT\left(\widehat{F}_n\right) = \sum_{l=1}^{k} \frac{\left(n_l - np_l\right)^2}{np_l}$$

is the familiar **chi-square statistic** [1] for testing whether the hypothesis $P(A_i) = p_i, i = 1, \ldots, k$ holds w.r.t. the observed sample.

■

Example 5.2 Estimation of Mann-Whitney Functional In example 1.9 above we introduced the Mann-Whitney functional

$$\theta = T(F, G) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} dF(x) dG(y).$$

This is not a linear functional but we may use the lessons learned there. If $\widehat{F}_n$ and $\widehat{G}_n$ are the the empirical distributions based on a sample of I.I.D. random variables $X_1, \ldots, X_n$ and a sample of I.I.D. random variables $Y_1, \ldots, Y_m$, respectively, then the plug-in estimate of θ is

$$\widehat{\theta} = T(\widehat{F}_n, \widehat{G}_m) = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n} \sum_{i=1}^{n} I_{]-\infty, y_j]}(x_i),$$

where $I_{]-\infty, y_i]}(x)$ is the indicator function of $]-\infty, y_j]$, i.e.,

$$I_{]-\infty, y_j]}(x) = \begin{cases} 1 & \text{if } -\infty < x \leq y_j \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

■

An estimator $\widehat{\theta}\,(X_1, \ldots, X_n)$ that can be written in the form

$$\widehat{\theta}\,(X_1, \ldots, X_n) = \mu + \frac{1}{n} \sum_{i=1}^{n} \alpha\,(X_i) + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j<i} \beta\,(X_i, X_j) \tag{5.2}$$

is called a **quadratic estimator**. We can obtain quadratic estimators by suitable expansions of non-linear plug-in estimators and by approximating these by dropping out other than the linear and quadratic terms.

[1] c.f., the collection of formulas, p. 7. in
`http://www.math.kth.se/matstat/gru/FS/fs_5B1501_v05.pdf`

Example 5.3 The Plug-in Estimate of Variance as a Quadratic Estimator The plug-in estimate of variance in (3.4) is

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 . \tag{5.3}$$

Then we get that with $m = E\left(X\right)$,

$$\widehat{\sigma}^2 = \mu + \frac{1}{n}\sum_{i=1}^{n}\alpha\left(x_i\right) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j<i}\beta\left(x_i, x_j\right),$$

where

$$\mu = \frac{n-1}{n}\sigma^2, \alpha\left(x\right) = \frac{n-1}{n}\left(\left(x-m\right)^2 - \sigma^2\right)$$

and

$$\beta\left(x, x^{'}\right) = -2(x-m)(x^{'}-m).$$

The check of this is left for the reader.

■

6 Influence Functions

The influence function is an analytic tool used to approximate the standard error of a plug-in estimator. We give a formal definition, which requires a preliminary definition. Let δ_t be the distribution function that puts all probability mass in the point t. Or,

$$\delta_t(x) = \begin{cases} 0 & x < t \\ 1 & t \le x. \end{cases} \tag{6.1}$$

This is the Heaviside function $U(x-t)$, but the notation $\delta_t(x)$ is more common in statistics.

Definition 6.1 The **influence function** $L_T(t; F)$ of T is defined as

$$L_T(t; F) = \lim_{\epsilon \to 0}\frac{T\left((1-\epsilon)F + \epsilon\delta_t\right) - T\left(F\right)}{\epsilon}. \tag{6.2}$$

■

Intuitively, this is the derivative of $T(F)$ in the direction of δ_t. That is, if we write

$$F_\epsilon = (1-\epsilon)F + \epsilon\delta_t,$$

then, assuming existence of the derivative,

$$L_T(t;F) = \frac{d}{d\epsilon} T\left(F_\epsilon\right)|_{\epsilon=0} \,. \tag{6.3}$$

Definition 6.2 The **empirical influence function** $\widehat{L}_T(t)$ of T is defined as $\widehat{L}_T(t) = L_T(t;\widehat{F}_n)$, i.e.,

$$\widehat{L}_T(t) = \lim_{\epsilon \to 0} \frac{T\left((1-\epsilon)\widehat{F}_n + \epsilon\delta_t\right) - T\left(\widehat{F}_n\right)}{\epsilon}. \tag{6.4}$$

■

We shall now show how to use the influence functions to find more about the linear statistical functionals.

Proposition 6.1 Let $T(F) = \int_{-\infty}^{\infty} a\left(x\right) dF(x)$ be a linear functional. Then we have the following:

1. $$L_T(t;F) = a(t) - T(F), \quad \widehat{L}_T(t) = a(t) - T\left(\widehat{F}_n\right). \tag{6.5}$$

2. For any distribution function G

$$T(G) = T(F) + \int_{-\infty}^{\infty} L_T(t;F)dG(t). \tag{6.6}$$

3. $$\int_{-\infty}^{\infty} L_T(t;F)dF(t) = 0.$$

4. Let

$$\tau^2 = \int_{-\infty}^{\infty} L_T(t;F)^2 dF(t).$$

Then

$$\tau^2 = \int_{-\infty}^{\infty} \left(a(t) - T(F)\right)^2 dF(t)$$

and if $\tau^2 < \infty$, then

$$\sqrt{n}\left(T\left(\widehat{F}_n\right) - T(F)\right) \to N(0, \tau^2),$$

as $n \to \infty$.

5. Let

$$\widehat{\tau}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{L}_T\left(X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(a\left(X_i\right) - T\left(\widehat{F}_n\right)\right)^2.$$

Then

$$\widehat{\tau}_n^2 \xrightarrow{p} \tau^2,$$

and if $\widehat{\mathbf{se}} = \frac{\widehat{\tau}_n}{\sqrt{n}}$ and $\mathbf{se} = \sqrt{V\left(T\left(\widehat{F}_n\right)\right)}$, then

$$\frac{\widehat{\mathbf{se}}}{\mathbf{se}} \xrightarrow{p} 1,$$

as $n \to \infty$.

We note that $\frac{\widehat{\tau}_n^2}{\sqrt{n}}$ is the estimated standard error of T.
Proof:

1. In (6.2) we look at the ratio

$$\frac{T\left((1-\epsilon)F + \epsilon\delta_t\right) - T\left(F\right)}{\epsilon}.$$

By linearity (4.1)

$$T\left((1-\epsilon)F + \epsilon\delta_t\right) = (1-\epsilon)T\left(F\right) + \epsilon T\left(\delta_t\right).$$

We have

$$T(\delta_t) = \int_{-\infty}^{\infty} a\left(x\right) d\delta_t(x) = a(t).$$

(Compare for the Dirac functional). Thus we have

$$\frac{T\left((1-\epsilon)F + \epsilon\delta_t\right) - T\left(F\right)}{\epsilon} = \frac{(1-\epsilon)T\left(F\right) + \epsilon a(t) - T\left(F\right)}{\epsilon}$$

$$= \frac{-\epsilon T\left(F\right) + \epsilon a(t)}{\epsilon} = a(t) - T\left(F\right).$$

Hence $L_T(t; F) = a(t) - T(F)$. The proof of the expression for $\widehat{L}_T(t)$ is identical.

2. Let us observe that by the first case of this proof

$$\int_{-\infty}^{\infty} L_T(t;F)dG(t) = \int_{-\infty}^{\infty} \left(a(t) - T(F)\right) dG(t)$$

$$= \int_{-\infty}^{\infty} a(t)dG(t) - \int_{-\infty}^{\infty} T(F)dG(t)$$

$$= T(G) - T(F),$$

since $\int_{-\infty}^{\infty} T(F)dG(t) = T(F)\int_{-\infty}^{\infty} dG(t) = T(F)$, as G is a distribution function.

3. This is now obvious, since by the above

$$\int_{-\infty}^{\infty} L_T(t;F)dF(t) = \int_{-\infty}^{\infty} \left(a(t) - T(F)\right) dF(t)$$

$$= \int_{-\infty}^{\infty} a(t)dF(t) - \int_{-\infty}^{\infty} T(F)dF(t)$$

$$= \int_{-\infty}^{\infty} a(t)dF(t) - T(F)\int_{-\infty}^{\infty} dF(t) = T(F) - T(F) = 0.$$

4. We apply (6.6) to get

$$\begin{aligned} T(\widehat{F}_n) &= T(F) + \int_{-\infty}^{\infty} L_T(t;F)d\widehat{F}_n(t) \\ &= T(F) + \frac{1}{n}\sum_{i=1}^{n} L_T(X_i;F). \end{aligned} \tag{6.7}$$

Thus

$$\sqrt{n}\left(T\left(\widehat{F}_n\right) - T(F)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_T(X_i;F).$$

By the preceding steps of this proof we know that for any i

$$E\left[L_T(X_i;F)\right] = \int_{-\infty}^{\infty} L_T(t;F)dF(t) = 0.$$

since X_i are I.I.D.. Then

$$V\left[L_T(X_i;F)\right] = E\left[L_T(X_i;F)^2\right] = \int_{-\infty}^{\infty} L_T(t;F)^2 dF(t) = \tau^2.$$

If $\tau^2 < \infty$, then the central limit theorem gives that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_T(X_i;F) \xrightarrow{d} N(0,\tau^2),$$

as $n \to \infty$, as was to be proven.

5. We are to consider

$$\widehat{\tau}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{L}_T\left(X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(a\left(X_i\right) - T\left(\widehat{F}_n\right)\right)^2$$

and rewrite this as

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\left(a\left(X_i\right) - T\left(F\right)\right) - \left(T(F) - T\left(\widehat{F}_n\right)\right)\right)^2$$

$$\begin{aligned} = \frac{1}{n}\sum_{i=1}^{n}\left(a\left(X_i\right) - T\left(F\right)\right)^2 - 2\left(T(F) - T\left(\widehat{F}_n\right)\right)\frac{1}{n}\sum_{i=1}^{n}\left(a\left(X_i\right) - T\left(F\right)\right) \\ + \left(T(F) - T\left(\widehat{F}_n\right)\right)^2, \end{aligned} \tag{6.8}$$

since $\frac{1}{n}\sum_{i=1}^{n}\left(T(F) - T\left(\widehat{F}_n\right)\right)^2 = \left(T(F) - T\left(\widehat{F}_n\right)\right)^2$. We consider first the mixed term in (6.8), i.e.,

$$\left(T(F) - T\left(\widehat{F}_n\right)\right)\frac{1}{n}\sum_{i=1}^{n}\left(a\left(X_i\right) - T\left(F\right)\right)$$

We have above in (6.7) shown the following

$$T(\widehat{F}_n) - T(F) = \frac{1}{n}\sum_{i=1}^{n} L_T(X_i;F)$$

which thus equals

$$= \frac{1}{n} \sum_{i=1}^{n} \left(a\left(X_i\right) - T\left(F\right) \right).$$

By the law of large numbers, as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} L_T(X_i; F) \overset{a.s.}{\to} E\left[L_T(X_1; F)\right] = \int_{-\infty}^{\infty} L_T(t; F) dF(t) = 0.$$

Thus the mixed term converges almost surely to zero, as $n \to \infty$ and in fact the limit of the third term in (6.8) has been treated by this, too, i.e.,

$$\left(T(F) - T\left(\widehat{F}_n\right)\right)^2 \to 0$$

as $n \to \infty$. It remains to note that the first term in (6.8) is

$$\frac{1}{n} \sum_{i=1}^{n} \left(a\left(X_i\right) - T\left(F\right) \right)^2 = \frac{1}{n} \sum_{i=1}^{n} L_T\left(X_i, F\right)^2$$

The law of large numbers gives, as $a\left(X_i\right) - T\left(F\right)$ are I.I.D. random variables that

$$\frac{1}{n} \sum_{i=1}^{n} L_T\left(X_i, F\right)^2 \overset{p}{\to} E\left[L_T(X_1; F)^2\right] = \tau^2,$$

and by the preceding

$$\tau^2 = \int_{-\infty}^{\infty} \left(a(x) - T(F)\right)^2 dF(x).$$

The final statement to be proved is left for the reader.

■

We point first out a few examples of influence functions and the properties found above.

Example 6.2 The Influence Function for the Probability of an Interval In examples 1.1 and 4.1 we studied for real numbers $a < b$

$$\theta = T(F) = F(b) - F(a) = \int_{-\infty}^{\infty} I_{]a,b]}(x) dF(x),$$

where $a(x) = I_{]a,b]}(x)$ is the indicator function of the half-open interval $]a, b]$. From (6.5) we obtain the influence functions

$$\begin{aligned} L_{F(b)-F(a)}(x; F) &= I_{]a,b]}(x) - F(b) - F(a), \\ \widehat{L}_{F(b)-F(a)}(x) &= I_{]a,b]}(x) - \frac{1}{n} \sum_{i=1}^{n} I_{]a,b]}\left(x_i\right). \end{aligned} \tag{6.9}$$

Thus

$$\widehat{\tau}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(I_{]a,b]}\left(X_i\right) - \frac{1}{n} \sum_{i=1}^{n} I_{]a,b]}\left(X_i\right) \right)^2$$

and

$$\begin{aligned} \widehat{\tau}_n^2 &\xrightarrow{p} \int_{-\infty}^{\infty} \left(I_{]a,b]}(x) - F(b) - F(a)\right)^2 dF(x) \\ &= \left(F(b) - F(a)\right)\left(1 - \left(F(b) - F(a)\right)\right), \end{aligned}$$

as is readily seen. Thus

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} L_{F(b)-F(a)}(X_i; F) \xrightarrow{d} N(0, \left(F(b) - F(a)\right)\left(1 - \left(F(b) - F(a)\right)\right)),$$

as $n \to \infty$.

■

Example 6.3 Variance with Known Mean In example 1.4 we assume that $\mu = \int_{-\infty}^{\infty} x dF(x)$ is known and then

$$T(F) = \int_{-\infty}^{\infty} (x - \mu)^2 dF(x),$$

Then by (6.5) we obtain the influence functions

$$L_V(x; F) = (x - \mu)^2 - \sigma^2, \quad \widehat{L}_V(x) = (x - \mu)^2 - \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{6.10}$$

Then it follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} L_V(X_i; F) \xrightarrow{d} N\left(0, \tau^2\right),$$

where

$$\tau^2 = \int_{-\infty}^{\infty} \left((x-\mu)^2 - \sigma^2\right)^2 dF(x),$$

which is consistently estimated by

$$\widehat{\tau}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{L}_V(X_i)^2 = \frac{1}{n}\sum_{i=1}^{n} \left((X_i-\mu)^2 - \sigma^2 - \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right)^2.$$

■

We make one further simple finding in the proof above explicit.

Corollary 6.4 If $T(F)$ is a linear functional, then

$$\lim_{n\to\infty} T(\widehat{F}_n) \overset{a.s.}{=} T(F).$$

■

We needed in this the representation (6.6), which holds exactly for linear functionals but can hold approximately for other functionals, as will be discussed next.

7 A Series Expansion and The Nonparametric Delta Method

It can be shown under reasonable regularity conditions that

$$T(\widehat{F}_n) = T(F) + \frac{1}{n}\sum_{i=1}^{n} L_T(X_i;F) + O_p\left(\frac{1}{n}\right), \tag{7.1}$$

where

$$L_T(x;F) = \psi(x) - \int_{-\infty}^{\infty} \psi(x)dF(x)$$

for some integrable function $\psi(x)$. The formula (7.1) can be seen as kind of Taylor expansion of the statistical functional T.

Then we approximate by the linear estimator

$$T\left(\widehat{F}_n\right) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} L_T(X_i;F),$$

and we get

$$V_F\left(T(\widehat{F}_n)\right) \approx V_F\left(T(F) + \frac{1}{n}\sum_{i=1}^{n} L_T(X_i;F)\right)$$

$$= \frac{1}{n}V_F\left(L_T(X;F)\right) = \frac{1}{n}E_F\left(L_T(X;F)^2\right),$$

since $E_F\left(L_T(X;F)\right) = 0$.

In view of theorem 6.1 and the Slutzky theorem we get with $\widehat{\mathbf{se}} = \frac{\widehat{\tau}_n}{\sqrt{n}}$ that

$$\frac{\left(T(F) - T\left(\widehat{F}_n\right)\right)}{\widehat{\mathbf{se}}} \to N(0,1),$$

as $n \to \infty$. We call the approximation

$$\frac{\left(T(F) - T\left(\widehat{F}_n\right)\right)}{\widehat{\mathbf{se}}} \approx N(0,1)$$

the **non-parametric delta method** . By force of this approximation we get a large sample confidence interval for $T(F)$. The procedure should work well for nonlinear functionals that admit the expansion in (7.1).

An asymptotic confidence interval with the degree of confidence $1-\alpha$ is given by

$$T\left(\widehat{F}_n\right) \pm \lambda_{\alpha/2}\widehat{\mathbf{se}},$$

where $\lambda_{\alpha/2}$ is the $\alpha/2$ -quantile of $N(0,1)$.

Example 7.1 An asymptotic confidence interval for the mean $T(F) = \int_{-\infty}^{\infty} x dF(x)$, $T(\widehat{F}_n) = \overline{X}$. This is a linear functional, and from theorem 6.1 we get

$$L_T(x;F) = a(x) - T(F) = x - \theta, \widehat{L}_T(x) = a(x) - T\left(\widehat{F}_n\right) = x - \overline{X}.$$

Therefore

$$\widehat{\tau}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{L}_T\left(X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \widehat{\sigma}^2,$$

which is the familiar (example 3.2) plug-in estimate of variance Thus

$$\overline{X} \pm 1.96\frac{\widehat{\sigma}}{\sqrt{n}}$$

is a pointwise asymptotic 95% confidence interval for θ^2 .

■

8 Exercises on Influence Functions

Statistical functionals can be of the form, c.f. (3.6),

$$T(F) = a\left(T_1(F), \ldots, T_m(F)\right)$$

with a real valued function $a\left(t_1, \ldots, t_m\right)$. If the chain rule of multivariable calculus holds for $a\left(t_1, \ldots, t_m\right)$ we get

$$L_T(x; F) = \sum_{i=1}^{m} \frac{\partial a}{\partial t_i} L_{T_i}(x; F).$$

Example 8.1 We continue the example 1.7. We have

$$T_1(F) = \int_{-\infty}^{\infty} x dF_X(x), \quad T_2(F) = \int_{-\infty}^{\infty} y dF_Y(y)$$

$$T_3(F) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy dF(x, y).$$

and augment these by

$$T_4(F) = \int_{-\infty}^{\infty} x^2 dF_X(x), \quad T_5(F) = \int_{-\infty}^{\infty} y^2 dF_Y(y).$$

We define

$$a\left(t_1, t_2, t_3, t_4, t_5\right) = \frac{t_3 - t_1 t_2}{\sqrt{t_4 - t_1^2}\sqrt{t_5 - t_2^2}}.$$

[2]This as the nonparametric version of the approximative method in section 12.3 in the collection of formulas, p. 5. in
`http://www.math.kth.se/matstat/gru/FS/fs_5B1501_v05.pdf`

Then the coefficient of correlation $\rho_{X,Y}$ is

$$\rho_{X,Y} = T(F) = a\left(T_1(F), T_2(F), T_3(F), T_4(F), T_5(F)\right).$$

An Exercise Show that the influence function of $\rho_{X,Y}$ is

$$L_T((x,y);F) = \tilde{x}\tilde{y} - \frac{1}{2}T(F)\left(\tilde{x} - \tilde{y}\right),$$

where

$$\tilde{x} = \frac{x - \int x dF_X(x)}{\sqrt{\int x^2 dF_X(x) - \left(\int x dF_X(x)\right)^2}}$$

and

$$\tilde{y} = \frac{y - \int y dF_Y(y)}{\sqrt{\int y^2 dF_Y(y) - \left(\int y dF_Y(y)\right)^2}}.$$

■

Example 8.2 Quantiles Let $F(x)$ be strictly increasing with the density $f(x)$. Let $0 \leq p \leq 1$ and then

$$\theta = T(F) = F^{-1}(p)$$

is the p:th quantile. The influence function is

$$L_T(x;F) = \begin{cases} \frac{p-1}{f(\theta)} & x \leq \theta \\ \frac{p}{f(\theta)} & x > \theta. \end{cases}$$

Hint: if $F_\epsilon = (1-\epsilon)F + \epsilon\delta_t$, then by (6.3) we need to compute

$$\frac{d}{d\epsilon}T\left(F_\epsilon\right)\mid_{\epsilon=0} = \frac{d}{d\epsilon}F_\epsilon^{-1}(p).$$

Since

$$p = F_\epsilon\left(T\left(F_\epsilon\right)\right).$$

we get

$$0 = \frac{d}{d\epsilon}F_\epsilon\left(T\left(F_\epsilon\right)\right).$$

and continue.

Thus the asymptotic variance of $T\left(\widehat{F}_n\right)$ is

$$V_F\left(T(\widehat{F}_n)\right) = \frac{1}{n}E_F\left(L_T(X;F)^2\right) = \frac{1}{n}\int_{-\infty}^{\infty} L_T(x;F)^2 dF(x) = \frac{p(1-p)}{nf^2(\theta)}.$$

$\blacksquare$

9 Sources

- A.C. Davison & D. Hinkley: *Bootstrap Methods and their Applications*, Cambridge University Press, 9th Printing, 2007.
- R. Serfling: *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, 1980.
- L. Wasserman: *All of Nonparametric Statistics.* , Springer 2006.
- L. Wasserman: *All of Statistics. A Concise Course in Statistical Inference*, Springer 2010.