

Chapter 23

Statistical Estimates Obtained by the Minimal Distances Method

The goals of this chapter are to:

- Consider the problem of parameter estimation by the method of minimal distances,
- Study the properties of the estimators.

Notation introduced in this chapter:

Notation	Description
w°	Brownian bridge
$F_\theta(x) = F(x, \theta)$	Distribution function with parameter θ
$p_\theta(x) = p(x, \theta)$	Density of $F_\theta(x)$

23.1 Introduction

In this chapter, we consider minimal distance estimators resulting from using the \mathfrak{N} -metrics and compare them with classical M -estimators. This chapter, like Chap. 22, is not directly related to quantitative convergence criteria, although it does demonstrate the importance of \mathfrak{N} -metrics.

23.2 Estimating a Location Parameter: First Approach

Let us begin by considering a simple case of estimating a one-dimensional location parameter. Assume that

$$\mathcal{L}(x, y) = \mathcal{L}(x - y)$$

is a strongly negative definite kernel and

$$N(F, G) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(x, y) dR(x) dR(y), \quad R = F - G,$$

is the corresponding kernel defined on the class of distribution functions (DFs). As we noted in Chap. 22, $\mathfrak{N}(F, G) = \mathcal{N}^{1/2}(F, G)$ is a distance on the class $\mathbf{B}(\mathcal{L})$ of DFs under the condition

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(x, y) dF(x) dF(y) < \infty.$$

Suppose that x_1, \dots, x_n is a random sample from a population with DF $F_\theta(x) = F(x - \theta)$, where $\theta \in \Theta \subset \mathbb{R}^1$ is an unknown parameter (Θ is some interval, which may be infinite). Assume that there exists a density $p(x)$ of $F(x)$ (with respect to the Lebesgue measure). Let $F_n^*(x)$ be the empirical distribution based on the random sample, and let θ^* be a minimum distance estimator of θ , so that

$$N(F_n^*, F_{\theta^*}) = \min_{\theta \in \Theta} N(F_n, F_\theta) \quad (23.2.1)$$

or

$$\theta^* = \operatorname{argmin}_{\theta \in \Theta} N(F_n^*, F_\theta). \quad (23.2.2)$$

We have

$$\begin{aligned} N(F_n^*, F_\theta) &= \frac{2}{n} \sum_{j=1}^n \int_{-\infty}^{\infty} \mathcal{L}(x_j - \theta - y) p(y) dy \\ &\quad - \frac{1}{n^2} \sum_{ij} \mathcal{L}(x_i - x_j) \\ &\quad - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(x - y) p(x) p(y) dx dy. \end{aligned}$$

Suppose that $\mathcal{L}(u)$ is differentiable and \mathcal{L} and p are such that

$$\begin{aligned} \int_{-\infty}^{\infty} \mathcal{L}(x) p'(x + \theta) dx &= \frac{d}{d\theta} \int_{-\infty}^{\infty} \mathcal{L}(x - \theta) p(x) dx \\ &= - \int_{-\infty}^{\infty} \mathcal{L}'(x - \theta) p(x) dx. \end{aligned} \quad (23.2.3)$$

Then, (23.2.2) implies that θ^* is the root of

$$\frac{d}{d\theta} N(F_n^*, F_\theta)|_{\theta=\theta^*} = 0$$

or

$$\sum_{j=1}^n \int_{-\infty}^{\infty} \mathcal{L}'(x_j - \theta^* - v) p(v) dv = 0. \tag{23.2.4}$$

Since the estimator θ^* satisfies the equation

$$\sum_{j=1}^n g_1(x_j - \theta) = 0, \tag{23.2.5}$$

where

$$g_1(x) = \int_{-\infty}^{\infty} \mathcal{L}'(x - v) p(v) dv,$$

it is an M -estimator.¹ It is well known [see, e.g., Huber (1981)] that (23.2.4) [or (23.2.5)] determines a consistent estimator only if

$$\int_{-\infty}^{\infty} g_1(x) p(x) dx = 0,$$

that is,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}'(u - v) p(u) p(v) dudv = 0. \tag{23.2.6}$$

We show that if (23.2.3) holds, then (23.2.6) does as well. The integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u - v) p(u + \theta) p(v + \theta) dudv = \int_{-\infty}^{\infty} \mathcal{L}(u - v) p(u) p(v) dudv$$

does not depend on θ . Therefore,

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u - v) p(u + \theta) p(v + \theta) dudv = 0. \tag{23.2.7}$$

¹See, for example, Huber (1981) for the definition and properties of M -estimators.

On the other hand,

$$\begin{aligned}
 & \frac{d}{d\theta} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u-v) p(u+\theta) p(v+\theta) dudv \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u-v) p'(u+\theta) p(v+\theta) dudv \\
 & \quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u-v) p(u+\theta) p'(v+\theta) dudv \\
 &= 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u-v) p'(u+\theta) p(v+\theta) dudv.
 \end{aligned}$$

Here, we used the equality $\mathcal{L}(u-v) = \mathcal{L}(v-u)$. Comparing this with (23.2.7), we find that for $\theta = 0$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u-v) p'(u) p(v) dudv = 0. \tag{23.2.8}$$

However,

$$\begin{aligned}
 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u-v) p'(u) p(v) dudv &= \int_{-\infty}^{\infty} \left(\frac{d}{du} \int_{-\infty}^{\infty} \mathcal{L}(u-v) p(v) dv \right) p(u) du \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}'(u-v) p(u) p(v) dudv.
 \end{aligned}$$

Consequently [see (23.2.8)],

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u-v) p(u) p(v) dudv = 0,$$

which proves (23.2.6).

We see that the minimum \mathfrak{N} -distance estimator is an M -estimator, and the necessary condition for its consistency is automatically fulfilled.

The standard theory of M -estimators shows that the asymptotic variance of θ^* [i.e., the variance of the limiting random variable of $\sqrt{n}(\theta^* - \theta)$ as $n \rightarrow \infty$] is

$$\sigma_{\theta^*}^2 = \frac{\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \mathcal{L}'(u-v)p(v)dv \right]^2 p(u)du}{\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}''(u-v)p(u)p(v)dudv \right]^2},$$

where we assumed the existence of \mathcal{L}'' and that the differentiation can be carried out under the integral. Note that when the parameter space Θ is compact, it is clear from geometric considerations that $\theta^* = \operatorname{argmin}_{\theta \in \Theta} N(F_n^*, F_\theta)$ is unique for sufficiently large n .

23.3 Estimating a Location Parameter: Second Approach

We now consider another method for estimating a location parameter θ . Let

$$\theta' = \operatorname{argmin}_{\theta \in \Theta} N(F_n^*, \delta_\theta), \tag{23.3.1}$$

where δ_θ is a distribution concentrated at the point θ and F_n^* is an empirical DF. Proceeding as in Sect. 23.2, it is easy to verify that θ' is a root of

$$\sum_{j=1}^n \mathcal{L}'(x_j - \theta) = 0, \tag{23.3.2}$$

and so it is a classic M -estimator. A consistent solution of (23.3.2) exists only if

$$\int_{-\infty}^{\infty} \mathcal{L}'(u)p(u)du = 0. \tag{23.3.3}$$

What is a geometric interpretation of (23.3.3)? More precisely, how is the measure parameter δ_θ related to the family parameter, that is, to the DF F_θ ? This must be the same parameter, that is, for all θ_1 we must have

$$N(F_\theta, \delta_\theta) \leq N(F_\theta, \delta_{\theta_1}).$$

Otherwise,

$$\frac{d}{d\theta_1} N(F_\theta, \delta_{\theta_1})|_{\theta_1=\theta} = 0.$$

It is easy to verify that the last condition is equivalent to (23.3.3). Thus, (23.3.3) has to do with the accuracy of parameterization and has the following geometric interpretation. The space of measures with metric \mathfrak{N} is isometric to some simplex in a Hilbert space. In this case, δ -measures correspond to the extreme points (vertices)

of the simplex. Consequently, (23.3.3) signifies that the vertex closest to the measure with DF F_θ corresponds to the same value of the parameter θ (and not to some other value θ_1).

23.4 Estimating a General Parameter

We now consider the case of an arbitrary one-dimensional parameter, which is approximately the same as the case of a location parameter. We just carry out formal computations assuming that all necessary regularity conditions are satisfied.

Let x_1, \dots, x_n be a random sample from a population with DF $F(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}^1$. Assume that $p(x, \theta) = p_\theta(x)$ is the density of $F(x, \theta)$. The estimator

$$\theta^* = \operatorname{argmin}_{\theta \in \Theta} N(F_n^*, F_\theta)$$

is an M -estimator defined by the equation

$$\frac{1}{n} \sum_{j=1}^n g(x_j, \theta) = 0, \quad (23.4.1)$$

where

$$g(x, \theta) = \int_{-\infty}^{\infty} \mathcal{L}(x, v) p'_\theta(v) dv - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u, v) p_\theta(u) p'_\theta(v) du dv.$$

Here, $\mathcal{L}(u, v)$ is a negative definite kernel, which does not necessarily depend on the difference of arguments, and the prime $'$ denotes the derivative with respect to θ . As in Sect. 23.2, the necessary condition for consistency,

$$E_\theta g(x, \theta) = 0,$$

is automatically fulfilled. The asymptotic variance of θ^* is given by

$$\sigma_{\theta^*}^2 = \frac{\operatorname{Var} \left(\int_{-\infty}^{\infty} \mathcal{L}(x, v) p'_\theta(v) dv \right)}{\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(u, v) p'_\theta(u) p'_\theta(v) du dv \right)^2}.$$

We can proceed similarly to Sect. 23.3 to obtain the corresponding results in this case. Since the calculations are quite similar, we do not state these results explicitly. Note that to obtain the existence and uniqueness of θ^* for sufficiently large n , we do not need standard regularity conditions such as the existence of variance, differentiability of the density with respect to θ , and so on. These are used only to obtain the estimating equation and to express the asymptotic variance of the estimator.

In general, from the construction of θ^* we have

$$N(F_n^*, F_{\theta^*}) \leq N(F_n^*, F_\theta) \text{ a.s.,}$$

and hence

$$\begin{aligned} E_\theta N(F_n^*, F_{\theta^*}) &\leq E_\theta N(F_n^*, F_\theta) \\ &= \frac{1}{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(x, y) dF(x, \theta) dF(y, \theta) \xrightarrow{n \rightarrow \infty} 0. \end{aligned} \tag{23.4.2}$$

In the case of a bounded kernel \mathcal{L} , the convergence is uniform with respect to θ . In this case it is easy to verify that $nN(F_n^*, F_\theta)$ converges to

$$- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(x, y) dw^\circ(F(x, \theta)) dw^\circ(F(y, \theta))$$

as $n \rightarrow \infty$, where w° is the Brownian bridge.

23.5 Estimating a Location Parameter: Third Approach

Let us return to the case of estimating a location parameter. We will present an example of an estimator obtained by minimizing the \mathfrak{N} -distance, which has good robust properties. Let

$$\mathcal{L}_r(x) = \begin{cases} |x| & \text{for } |x| < r \\ r & \text{for } |x| \geq r, \end{cases}$$

where $r > 0$ is a fixed number. The famous Pólya criterion² implies that the function $f(t) = 1 - \frac{1}{r}\mathcal{L}_r(t)$ is the characteristic function of some probability distribution. Consequently, $\mathcal{L}_r(t)$ is a negative definite function. This implies that for a sufficiently large sample size n there exists an estimator θ^* of minimal \mathfrak{N}^r distance, where \mathcal{N}^r is the kernel constructed from $\mathcal{L}_r(x - y)$. If the distribution function $F(x - \theta)$ has a symmetric unimodal density $p(x - \theta)$ that is absolutely continuous and has a finite Fisher information

$$I = \int_{-\infty}^{\infty} \left(\frac{p'(x)}{p(x)} \right)^2 p(x) dx,$$

then we conclude by (23.4.2) that θ^* is consistent and is asymptotically normal. The estimator θ^* satisfies (23.2.5), where

²See, for example, Lukacs (1969).

$$g_1(x) = \int_{-\infty}^{\infty} \mathcal{L}'(x - v)p(v)dv$$

and

$$\mathcal{L}'(u) = \begin{cases} 0 & \text{for } |u| \geq r, \\ 1 & \text{for } 0 < u < r, \\ 0 & \text{for } u = 0, \\ -1 & \text{for } -r < u < 0. \end{cases}$$

This implies that θ^* has a bounded influence function and, hence, is B -robust.³

Consider now the estimator θ' obtained by the method discussed in Sect. 23.3. It is easy to verify that this estimator is consistent under the same assumptions. However, θ' satisfies the equation

$$\sum_{j=1}^n \mathcal{L}'(x_j - \theta) = 0,$$

so that it is a trimmed median. It is well known that a trimmed median is the most B -robust estimator in the corresponding class of M -estimators.⁴

23.6 Semiparametric Estimation

Let us now briefly discuss semiparametric estimation. This problem is similar to that considered in Sect. 23.4, except that here we do not assume that the sample comes from a parametric family. Let x_1, \dots, x_n , be a random sample from a population given by DF $F(x)$, which belongs to some distribution class \mathcal{P} . Suppose that the metric \mathfrak{N} is generated by the negative definite kernel $\mathcal{L}(x, y)$ and that $\mathbf{P} \subset \mathcal{B}(\mathcal{L})$. $\mathcal{B}(\mathcal{L})$ is isometric to some subset of the Hilbert space \mathfrak{H} . Moreover, Aronszajn's theorem implies that \mathfrak{H} can be chosen to be minimal in some sense. In this case, the definition of \mathfrak{N} is extended to the entire \mathfrak{H} .

We assume that the distributions under consideration lie on some “nonparametric curve.” In other words, there exists a nonlinear functional φ on \mathfrak{H} such that the distributions F satisfy the condition

$$\varphi(F) = c = \text{const.}$$

The functional φ is assumed to be smooth. For any $H \in \mathfrak{H}$

³See Hampel et al. (1986).

⁴See Hampel et al. (1986).

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{N(F + tH, G) - N(F, G)}{t} &= 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}(x, y) d(G(x) - F(x)) dH(y) \\ &= \langle \text{grad } N(F, G), H \rangle, \end{aligned}$$

where G is fixed.

Under the parametric formulation of Sect. 23.4, the equation for θ has the form

$$\frac{d}{d\theta} N(F_\theta, F_n^*) = 0,$$

that is,

$$\left\langle \text{grad } N(F, F_n^*)|_{F=F_\theta}, \frac{d}{d\theta} F_\theta \right\rangle = 0.$$

Here, the equation explicitly depends on the gradient of the functional $N(F, F_n^*)$. However, under the nonparametric formulation, we work with the conditional minimum of the functional $N(F, F_n^*)$, assuming that F lies on the surface $\varphi(F) = C$. Here, our estimator is

$$\tilde{F}^* = \underset{F \in \{F: \varphi(F) = c\}}{\text{argmin}} N(F, F_n^*).$$

According to general rules for finding conditional critical points, we have

$$\text{grad } N(\tilde{F}^*, \tilde{F}_n^*) = \lambda \text{grad } \phi(\tilde{F}^*), \quad (23.6.1)$$

where λ is a number. Thus, in the general case, (23.6.1) is an eigenvalue problem. This is a general framework of semiparametric estimation.

References

- Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust statistics: the approach based on influence functions. Wiley, New York
 Huber P (1981) Robust statistics. Wiley, New York
 Lukacs E (1969) Characteristic functions. Griffin, London