A Short Course on Set Estimation

Antonio Cuevas
Departamento de Matemáticas
Universidad Autónoma de Madrid, Spain
antonio.cuevas@uam.es

Abstract

The basic aim of set estimation is to reconstruct (i.e, to estimate, in statistical terms) a compact set $S \subset \mathbb{R}^d$ from a random sample of points $X_1, \ldots, X_n$ whose distribution is related to S. The most typical situation corresponds to the case in which S is the (compact) support of a probability distribution P_X and the data X_i are sample points drawn from P_X. This is the *support estimation problem*. In other cases S is a level set of type $S = \{f \geq c\}$, where f is a density function, $c > 0$ is a given constant, and the data X_i are independent random observations with density f. This is the *level set estimation problem*.

Some variants of these two basic problems can be also considered, including the estimation of a set S under shape restrictions (convexity, r-convexity, cone-convexity, star-shape, etc.) as well as the estimation of some relevant functionals (volume, surface area,...) or subsets (boundary, medial axis,...) of S.

In this short course, whose estimated duration is six hours, we aim at proving a general overview of the state of the art in these subjects, with especial emphasis on the mathematical methods and the main techniques used in this field. Most proofs will be omitted or drastically sketched. Some practical applications will be outlined.

The course schedule will be roughly as follows:

1. Set estimation: **statement of the problem, some basic definitions**. Description of the general aims and techniques in set estimation. Connections with other statistical methodologies.

2. **All-purposes support estimators:** the sample, the Devroye-Wise estimator, plug-in estimators. Asymptotic results and applications in quality control, clustering and dimension reduction methods.

3. Support estimation in the **convex case**: the convex hull. Excess mass estimators.

4. Some **convexity-type restrictions**: r-convexity, rolling-ball properties, positive reach, cone-convexity. The volume function and the polynomial volume restriction.

5. Support estimation under **convexity-type restrictions**: a summary of theoretical results and applications.

6. Estimation of the **boundary measure** (perimeter, surface area...): a short overview of methods, results and applications.

7. **The level set estimation problem**: plug-in approach, granulometric smoothing and excess mass approach. Some theoretical results and applications.

8. Some current developments and **suggestions for further research**.