Parallel sets in density estimation: theoretical results with a glance at possible applications and further developments

Elena Villa

The parallel set at distance \( r \) of any closed subset \( A \) in \( \mathbb{R}^d \) (also named \( r \)-neighbourhood or Minkowski enlargement of \( A \)) is the set so defined:

\[
A_{\oplus r} = \{ x \in \mathbb{R}^d : \text{dist}(x, A) \leq r \}.
\]

Many real phenomena may be modelled as random closed sets in \( \mathbb{R}^d \), of different Hausdorff dimensions. Of particular interest are cases in which their Hausdorff dimension, say \( n \), is strictly less than \( d \), such as fiber processes, boundaries of germ-grain models, and \( n \)-facets of random tessellations. The mean density, say \( \lambda_n \), of a random closed set \( \Theta_n \) in \( \mathbb{R}^d \) with Hausdorff dimension \( n \) is defined to be the density of the measure \( \mathbb{E}[\mathcal{H}^n(\Theta_n \cap \cdot)] \) on \( \mathbb{R}^d \), whenever it is absolutely continuous with respect to the \( d \)-dimensional Hausdorff measure \( \mathcal{H}^d \). A crucial problem is the pointwise estimation of \( \lambda_n(x) \).

In this talk we consider a generalization of the so-called naïve estimator of the pdf of a random variable; namely, under suitable regularity assumptions on \( \Theta_n \), generally fulfilled in applications, it can be shown that \( \lambda_n(x) = \lim_{r \to 0} \frac{P(x \in \Theta_{n,r})}{b_d n^{d-n} r^d} \), for a.e. \( x \in \mathbb{R}^d \), which suggests the following estimator of \( \lambda_n(x) \), given an i.i.d. random sample \( \Theta_1^n, \Theta_2^n, \ldots, \Theta_N^n \) of \( \Theta_n \):

\[
\hat{\lambda}_n(x) := \frac{\# \{ i : x \in \Theta_{n,r}^i \}}{N b_d n^{d-n} r^d}.
\]

By the very definition of \( \hat{\lambda}_n(x) \), it turns out that it is of easy computational evaluation, and so of potential interest in applications. A series of recent theoretical results on \( \hat{\lambda}_n(x) \) (optimal bandwidth, consistency, asymptotic normality, large deviation principles), might be taken as a starting point for further developments. Finally, we shall discuss the estimation based on a single observation in the case of \( \Theta_n \) Boolean model.