Model selection for density estimation with $\mathbb{L}_2$-loss

Lucien Birgé

Université Paris VI
Laboratoire de Probabilités et Modèles Aléatoires
U.M.R. C.N.R.S. 7599

Vienna — July 2008

There exists an enormous number of papers dealing with density estimation on $[0,1]$ with $\mathbb{L}_2$-loss from n i.i.d. observations with unknown density s but most of them are restricted (directly or indirectly) to bounded densities. For instance an Hölderian density on $[0,1]$ is necessarily bounded and this is also the case for densities in Besov spaces $B^{\alpha}_{p,\infty}([0,1])$ when $\alpha > 1/p$. This is no more true when $\alpha < 1/p$ which accounts for the fact that the problem has rarely been considered in this case. Similar boundedness restrictions hold for papers on model selection and aggregation of estimators.

The purpose of this talk is to elucidate the problem of estimating arbitrary densities s belonging to $\mathbb{L}_2([0,1))$. We first show that, when $s \in \mathbb{L}_\infty([0,1))$ and the estimator $\hat{s}$ is bound to belong to some D-dimensional model $\overline{S}$, the best universal (valid for arbitrary s and $\overline{S}$) risk bound we can expect is of the form

$$\mathbb{E}\left[\|\hat{s}-s\|^2\right] \le C\left[\inf_{t\in\overline{S}}\|t-s\|^2 + n^{-1}D\|s\|_\infty\right],$$

which clearly indicates that problems may occur when $s \notin \mathbb{L}_\infty([0,1))$. For instance, the minimax risk over a D-dimesional model may be infifnite. The situation is even more complicated with many models.

We shall introduce a new method for model selection when the underlying density to be estimated is arbitrary in $\mathbb{L}_2([0,1))$, therefore possibly unbounded, with applications to estimation selection, partition selection for histograms and adaptive estimation in Besov spaces for $\alpha < 1/p$. This method is fairly general since it can cope with arbitrary families of (possibly non-linear) finite dimensional models. As a counterpart, it is quite abstract since the resulting estimators are purely theoretical ones, therefore providing only theoretical (rather than practical) risk bounds. It nevertheless leads to results that are presently not reachable by more concrete methods.

A reference paper for non-specialists:

BIRGÉ, L. (2006). Statistical estimation with model selection. *Indagationes Mathematicae* **17**, 497-537 (2006).

For specialists we suggest:

BIRGÉ, L. and MASSART, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds.), 55-87. Springer-Verlag, New York.

BIRGÉ, L. (2006). Model selection via testing : an alternative to (penalized) maximum likelihood estimators. *Ann. Inst. Henri Poincaré, Probab. et Statist.* **42**, 273-325.

MASSART, P. (2007). Concentration Inequalities and Model Selection. In *Lecture on Probability Theory and Statistics, Ecole d'Eté de Probabilités de Saint-Flour XXXIII - 2003* (J. Picard, ed.). Lecture Note in Mathematics, Springer-Verlag, Berlin.

RIGOLLET, T. and TSYBAKOV, A.B. (2007). Linear and convex aggregation of density estimators. *Math. Methods of Statis.* **16**, 260-280.