Asymptotic Statistics

Lecture 9

Minimax lower bounds

Rates of convergence

We say that a sequence of estimators T_n for a parameter θ has rate of convergence $r_n \to \infty$ with respect to the (pseudo)metric d if

$$r_n d(T_n, \theta) = O_{P_\theta}(1).$$

For instance implied by

$$E_\theta d^p(T_n, \theta) \leq C r_n^{-p}$$

for some $p, C > 0$.

Examples

- In smooth, parametric models, have seen various estimators for which $\sqrt{n}(T_n - \theta) \rightsquigarrow T$ for some random vector T. Then T_n has rate of convergence $\sqrt{n}$ w.r.t. the Euclidean distance.
- If $f \in C^\beta$ and $\int |f^{(\beta)}(x)|^2\,dx < \infty$, then for a properly tuned kernel estimator $\hat{f}_n$, we have

$$\int E_f(\hat{f}_n(x) - f(x))^2\,dx \leq Cn^{-2\beta/(1+2\beta)}.$$

Then $\hat{f}_n$ has a rate of convergence $n^{-\beta/(1+2\beta)}$ w.r.t. the L^2-distance

$$d(f_1, f_2) = \Big(\int (f_1(x) - f_2(x))^2\,dx\Big)^{1/2}.$$

Converse question

Question: what is the best rate that any estimator can have?

Role of uniformity: suppose $\Theta = \{\theta_0, \theta_1\}$, consider $T_n \equiv \theta_0$. Then T_n has rate ∞ if $\theta = \theta_0$, but rate 0 if $\theta = \theta_1$.

Conclusion: more meaningful to ask what the best possible rate is uniformly over a range of parameters.

Examples of uniform rate results

- For many estimators in smooth, parametric models, it actually holds that
$$\sup_{\theta \in \Theta} E_\theta \| T_n - \theta \|^2 \leq C/n.$$
(See location model example lecture notes.)
- Let $\mathcal{F}_{m,M}$ be the collection of all $f \in C^\beta$ such that $\int |f^{(\beta)}(x)|^2 \, dx \leq M$, then for a properly tuned kernel estimator $\hat{f}_n$, we have
$$\sup_{f \in \mathcal{F}_{m,M}} \int E_f(\hat{f}_n(x) - f(x))^2 \, dx \leq C n^{-2\beta/(1+2\beta)}.$$

Minimax lower bounds

Goal, in a given statistical model, prove lower bound results of the form

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} E_\theta d^p(\hat{\theta}, \theta) \geq C r_n^{-p},$$

where the infimum is over all estimators $\hat{\theta}$.

This is a minimax lower bound for the risk $E_\theta d^p(\hat{\theta}, \theta)$.

It implies that no estimator can have a rate convergence faster than r_n, uniformly over Θ.