26.1 Introduction

Recall from last time,

$$\begin{aligned} f_0 &= 0 \\ f_1 &= Lh_n^{\beta} K(x - x_0/h_n),\ x \in [0,1] \\ h_n &= c_o n^{\frac{1}{2\beta+1}} \end{aligned}$$

We needed to find a parameter that adds enough smoothness but still remains in the function class. So we chose

$$d(f_0, f_1) \geq 2s = cn^{\frac{\beta}{2\beta+1}} \text{with } \phi_n = n^{\frac{\beta}{2\beta+1}} \text{ and } K(P_0, P_1) \leq \alpha < \infty.$$

We saw previously that $\alpha = \mathcal{O}(nh_n^{2\beta+1}) \Rightarrow h_n = \mathcal{O}(n^{\frac{1}{2\beta+1}})$.

What if, instead of considering a particular point, we want to use the L_2 distance between these functions?

Then

$$\begin{aligned} d(f_0, f_1) &= \left(\int f_1^2(x) dx \right)^{1/2} \\ &= Lh_n^{\beta} \left(\int K^2 (\frac{x - x_0}{h_n} dx \right)^{1/2} \\ &= Lh_n^{\beta+1/2} \left(\int K^2(u) du \right)^{1/2} \\ &= \mathcal{O}(h_n^{\beta+1/2}) = \mathcal{O}(n^{-1/2}) \\ &\Rightarrow \phi_n = n^{-1/2} \end{aligned}$$

goes to zero more quickly than the upper bound. No known estimator achieves this lower bound, so either a better estimator exists, or this technique is not tight enough.

In particular, it is not sufficient to use 2 hypotheses f_0, f_1 and $\max\{f_0, f_1\}$ as a proxy for $\sup \Theta$. This motivates the following technique.

26.2 Main result (KL version)

If $m \geq 2, \theta_0, ..., \theta_m \in \Theta$, then

1. $d(\theta_i, \theta_k) \geq 2s > 0 \forall j, k$
2. $P_j << P_0$ (always true if, e.g., every msr has a density wrt Lebesgue measure)

3. $\frac{1}{m}\sum_{j=1}^{m} K(P_j, P_0) \leq \alpha \log m$ and
4. $0 < \alpha < 1/8$,

using parameter 0 as the base case. Then

$$\inf_{\widehat{\theta}} \sup_{\theta \in \Theta} P_\theta(d(\widehat{\theta}, \theta) \geq s) \geq C_1(m) C_2(\alpha) > 0 \; \forall m$$

The following theorem concerns minimax nonparametric KDE, but allows the dimension to increase.

Theorem 26.1 (McDonald). *Let $X_1, ..., X_n \sim$ iid f and define*

1. $f : \mathbb{R}^d \to \mathbb{R}^+$
2. $f \in \mathcal{N}_p(\beta, c)$, *the Nikolsky class:*
   - (a) $D^s f$ *exists if* $|s| \leq \lfloor \beta \rfloor$
   - (b) $\int (D^s f(x+t) - D^s f(x))^p dx \leq C^p ||t||_1^{p(\beta - |s|)}$
   - (c) $\int f = 1$

Then

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{N}} \mathbb{E}\left[\left(\frac{n^\beta}{d^d} \right)^{\frac{1}{2\beta+d}} ||f - \widehat{f}||_p \right] \geq c$$

Unlike the previous result, this holds when both $n, d \to \infty$.

Ingredients

1. Need to find $m+1$ densities in $\mathcal{N}_p(\beta, c)$
2. Then show $||f_i - f_j||_p \geq 2c\phi_{nd} \; \forall i, j$
3. $\phi_{nd} = (\frac{d^d}{n^\beta})^{\frac{1}{2\beta+d}}$
4. show $\frac{1}{m}\sum_{j=1}^{m} K(P_j, P_0) \leq \alpha \log m$

Proof Sketch

What are these densities? Have to be a bit more careful than previous case (in which f_1 was just a slight perturbation of the null distribution) to ensure the perturbed functions are still densities *and* smooth enough to be in the Nikolsky class.

$$\Gamma_0 \in W^{\beta, 1/2} \cap C^\infty(\mathbb{R})$$

This is the key difference:

$$\Gamma(u) = dC \prod_{i=1}^{d} \Gamma_0(u_i)$$

We need this d factor to get the right rate.

For any integer $m \geq 2$, let $\gamma_{m,j}(x) = m^{-\beta}\Gamma(mx - j)$, a d-dim vector $j \in J = \{1, ..., m\}^d$.

The distributions are:

$$f_0 = \mathcal{N}_d(0, \sigma I)$$
$$f_\omega(x) = f_0(x) + \sum_{j \in J} \omega(j)\gamma_{m,j}(x),$$

where ω is a binary vector of size $m^d, \omega(j) \in \{0, 1\}$

For details, see McDonald (2017).

Example 26.2. *$m = 4, d = 2, J = \{(1, 1), (1, 2), (2, 1), ..., (4, 4)\}$.*

$\Rightarrow f_\omega$ has 0 to 16 $\Rightarrow 2^{16}$ densities

Note In general, for m hypotheses we will always choose elements according to the vertices of binary hypercube, then apply Lemma 26.3:

Lemma 26.3 (Varshamov-Gilbert)**.** *Let $m \geq 8$. Then $\exists D \subseteq \{\omega\}$ s. t. $\forall \omega, \omega' \in D, H(\omega, \omega') \geq \frac{m^d}{8}$ and $|D| \geq \exp(m^d/8)$*

1. Can show f_ω is smooth enough if $m > \left[dC(C)^d\right]^{1/\beta}$ (if m too small, the edges of the hypercube are too small to get useful perturbations)
2. to show appropriate separation, need to look at
$$\begin{aligned} f_\omega - f_{\omega'}||_p &= ||\sum_{j \in J}(\omega(j) - \omega'(j))\gamma_{m,j}||_p \\ &= m^{-\beta - d/p}H^{1/2}(\omega, \omega')||\Gamma||_p, \end{aligned}$$
where H is the Hamming dist. (Hamming dist separation is a generic part of all m-hypothesis proofs.)
We will shrink class to D, throwing away alternatives that are "too close".
Then
$$m^{-\beta - d/p}H^{1/2}(\omega, \omega')||\Gamma||_p \geq m^{-\beta - d/p}(\frac{m^d}{8})^{1/p}dC||\Gamma_0||_p^d = 8^{-1/p}dm^{-\beta}C||\Gamma_0||_p^d$$
3. $K(P_\omega, P_0) \leq \cdots \leq C^d nd^2 m^{-2\beta}$. Need to choose m to annihilate n if we want the $\alpha \log m$ bound (and satisfy the many other conditions):
So we choose m s. t.
$$nC^d d^2 m^{-2\beta} \leq \alpha \log |D| \Rightarrow m \leq \left[C_1 d^2 n C_2^d\right]^{\frac{1}{2\beta + d}}$$
4. Finally, combine these terms:
set
$$m = ||\Gamma_0||_p^{\frac{d+1}{\beta}} \kappa(d^2 n)^{\frac{1}{2\beta + d}}.$$
Plug in m to get
$$\geq 2C\phi_{nd}, \phi_{nd} = (d^d n^{-\beta})^{\frac{1}{2\beta + d}}.$$

References

McDonald, D. (2017), "Minimax Density Estimation for Growing Dimension," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, **54**, 194–203.