SDS8102: Advanced Mathematical Statistics

Week 12 – Lecture 2
November 19, 2026
Lecturer: Alexandre Tsybakov

---

5 Unbiased risk estimation. Cross-validation (contd.)

Stone (1984)[1] was the first to investigate the issue of optimality in connection with cross-validation. He proved that the integrated squared error of the estimator $\tilde{p}_n^{CV}$ is asymptotically equivalent to that of some oracle which is, however, different from the kernel oracle with bandwidth h^*.The integrated squared error is defined as

$$ISE(h) = \int (\tilde{p}_{n,h} - p)^2,$$

where $\tilde{p}_{n,h}$ is the kernel estimator with bandwidth h. Stone (1984) proved that

$$\frac{ISE(\hat{h}_{CV})}{\min_{h>0} ISE(h)} \xrightarrow{a.s.} 1, \qquad \text{as } n \to \infty, \tag{1}$$

under some assumptions (the density p is bounded, the kernel is compactly supported, essentially nonnegative, and satisfies the Hölder condition). However, the definition of ISE does not involve taking expectations. The minimum $\min_{h>0} ISE(h)$ in (1) is not attained at the oracle bandwidth h^* but rather at some random $\tilde{h}$. Recall that the risk we wish to minimize is

$$MISE(h) = \mathbb{E}(ISE(h)).$$

The result of Stone does not imply any optimality property in terms of $MISE$.
To prove that the estimator is adaptive, we need a *uniform* bound on the *expected value* of ISE:

$$\sup_{p \in \mathcal{P}_S(\beta,L)} \mathbb{E}(ISE(\hat{h})) \le C n^{-\frac{2\beta}{2\beta+1}},$$

where the construction of $\hat{h}$ is entirely data-driven (does not depend on β, L). Stone's result does not allow us to achieve this goal. However, a more recent work of Dalelane (2005)[2] provides a finite sample bound, which is sufficient for our purposes.

Theorem 4 (Adapted from Dalelane (2005))**.** *Let $p(x) \le p_{max} < \infty$, h. Assume that kernel K is in $L_2(\mathbb{R})$, and its Fourier transform $\hat{K}$ is symmetric, unimodal, $\hat{K} \ge 0$, and $\mathrm{supp}(\hat{K}) \subseteq [-a, a]$ for some $a > 0$. Let*

$$p_n^* = \tilde{p}_{n,\tilde{h}_{CV}},$$

[1] Stone, C.J. (1984) An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates. *The Annals of Statistics*, 12, 1285-1297.

[2] Dalelane, C. (2005) Exact oracle inequality for a sharp adaptive kernel density estimator. `arXiv:math/0504382`

where $\tilde{h}_{CV} \in \operatorname{arginf}_{h>1/n} CV(h)$. *Then, for any* $0 < \delta < 1$, $n \geq 1$,

$$\mathbb{E}\big[\|p_n^* - p\|_2^2\big] \leq \left(1 + \frac{C}{n^\delta}\right) \inf_{h>1/n} \underbrace{\mathbb{E}\big[\|\tilde{p}_{n,h} - p\|_2^2\big]}_{MISE(h)} + \frac{C(\log n)^{5/2}}{n^{1-\delta}},$$

where $C = C(p_{max}, K, \delta)$.

Remarks:

- We can turn $\min_{h>1/n}$ to $\min_{h>0}$ since, due to Theorem 3, the variance term is equal to $\frac{1}{nh}\int K^2(u)du$. So, except for pathological cases, the minimum is not achieved at $h \leq 1/n$.
- Theorem 4 does not cover kernel estimators in great generality. The spline kernel, the Gaussian kernel, the Epanechnikov kernel, and the rectangular kernel do not satisfy the assumptions of the theorem, but the sinc kernel and the Pinsker kernel do.
- Theorem 4 provides an example of what is called an *oracle inequality*. It tells us that the risk of an estimator is less than or equal to the risk of the oracle plus some small residual term. If we let δ approach 0, the residual can be arbitrarily close to $\frac{(\log n)^{5/2}}{n}$. This is of smaller order than the typical nonparametric rate $n^{-\frac{2\beta}{2\beta+1}}$.

Corollary 1 (Adaptivity of p_n^* on the scale of Sobolev classes)**.** *Let* K *be the sinc kernel. Then*

$$\sup_{p \in \mathcal{P}_S(\beta, L)} \mathbb{E}\big[\|p_n^* - p\|_2^2\big] \leq C n^{-\frac{2\beta}{2\beta+1}}, \qquad \forall \beta > 1/2, \quad L > 0.$$

where C *is a constant that only depends on* β *and* L.

Recall that p_n^* does not depend on β or L. It depends only on the data.

Proof: We check the assumptions of Theorem 4. The sinc kernel satisfies all the assumptions on K in Theorem 4. It remains only to show that there exists $p_{max} > 0$ depending only on β, L such that $p(x) \leq p_{max}$ for all x. Indeed,

$$\begin{aligned}\int |\phi(w)|dw &= \int_{|w|\leq 1} |\phi(w)|dw + \int_{|w|>1} |\phi(w)|dw \\ &\leq 2 + \int_{|w|>1} \frac{|w|^\beta |\phi(w)|}{|w|^\beta} dw.\end{aligned}$$

Use the Cauchy-Schwarz inequality to further bound the last integral by

$$\left(\int_{|w|>1} |w|^{-2\beta} dw\right)^{1/2} \left(\int |w|^{2\beta} |\phi(w)|^2 dw\right)^{1/2}.$$

The first integral in this expression is bounded by a constant if $\beta > 1/2$, and the second one is less than or equal to $2\pi L^2$ since $p \in \mathcal{P}_S(\beta, L)$. Thus, $\int |\phi(w)|dw$ is bounded by a constant $C_0 = C_0(\beta, L) < \infty$, so that we can take $p_{max} = C_0(\beta, L)$ whenever $\beta > 1/2$.

We are now in a position to apply Theorem 4. Taking the maxima over $p \in \mathcal{P}_S(\beta, L)$ of both sides of the oracle inequality in that theorem we obtain

$$\sup_{p\in\mathcal{P}_S(\beta,L)} \mathbb{E}\big[\|p_n^* - p\|_2^2\big] \le \left(1 + \frac{C'}{n^\delta}\right) \sup_{p\in\mathcal{P}_S(\beta,L)} \inf_{h>1/n} \mathbb{E}\big[\|\tilde{p}_{n,h} - p\|_2^2\big] + \frac{C'(\log n)^{5/2}}{n^{1-\delta}},$$

where $C' > 0$ is a constant that only depends on β, L and δ. Remark that the rate optimal bandwidth $\bar{h} = n^{-\frac{1}{2\beta+1}}$ is larger than $\frac{1}{n}$. Hence,

$$\inf_{h>1/n} \mathbb{E}\big[\|\tilde{p}_{n,h} - p\|_2^2\big] \le \mathbb{E}\big[\|\tilde{p}_{n,\bar{h}} - p\|_2^2\big].$$

On the other hand, Theorem 3 implies

$$\sup_{p\in\mathcal{P}_S(\beta,L)} \mathbb{E}\big[\|\tilde{p}_{n,\bar{h}} - p\|_2^2\big] \le C_1 n^{-\frac{2\beta}{2\beta+1}}, \qquad \forall \beta > 0, L > 0,$$

since the sinc kernel is a super-kernel. Here, the positive constant C_1 depends only on β and L. Thus, we have

$$\sup_{p\in\mathcal{P}_S(\beta,L)} \mathbb{E}\big[\|p_n^* - p\|_2^2\big] \le C_1(1 + C')n^{-\frac{2\beta}{2\beta+1}} + C'\frac{(\log n)^{5/2}}{n^{1-\delta}}.$$

Finally, choose a small enough δ such that $1 - \delta > \frac{2\beta}{2\beta+1}$ to control the remainder term. The corollary follows. ∎

6 Minimax lower bound on the risk

In Theorem 1 it was proved that, for all $x \in \mathbb{R}$, $n \ge 1$, under appropriate conditions on K and h, there exists $C > 0$ such that

$$\sup_{p\in\mathcal{P}(\beta,L)} \mathbb{E}[(\hat{p}_n(x) - p(x))^2] \le Cn^{-\frac{2\beta}{2\beta+1}}.$$

Our goal now is to prove that this rate is optimal in minimax sense, that is, there exists $c > 0$ such that

$$\inf_{T_n} \sup_{p\in\mathcal{P}(\beta,L)} \mathbb{E}[(T_n(x) - p(x))^2] \ge cn^{-\frac{2\beta}{2\beta+1}}. \tag{2}$$

Here, T_n ranges over all density estimators.

Without loss of generality, take $x = 0$, and write for brevity $t_n = T_n(0)$. Write the expectation sign as $\mathbb{E}_p$, with index p, to emphasize that the expected value is taken over the distribution of a sample from density p.

For any $p_0, p_1 \in \mathcal{P}(\beta, L)$, we may write

$$\begin{aligned}\sup_{p\in\mathcal{P}(\beta,L)} \mathbb{E}_p[(t_n - p(0))^2] &\ge \max\{\mathbb{E}_{p_0}[(t_n - p_0(0))^2], \mathbb{E}_{p_1}[(t_n - p_1(0))^2]\} \\ &\ge \frac{1}{2}[\mathbb{E}_{p_0}[(t_n - p_0(0))^2] + \mathbb{E}_{p_1}[(t_n - p_1(0))^2]]. \end{aligned} \tag{3}$$

Write the expected values as integrals:

$$\mathbb{E}_p(t_n - p(0))^2 = \int \dots \int (t_n(x_1, x_2, \dots, x_n) - p(0))^2 \prod_{i=1}^{n} p(x_i) dx_i.$$

Let $\mathcal{X} = (x_1, \dots, x_n)$, and $\pi_n(x) = \prod_{i=1}^n p(x_i)$. Similarly, let $\pi_{0,n}$, $\pi_{1,n}$ be the joint densities corresponding to the chosen densities p_0 and p_1. The expression in (3) is then equal to

$$\begin{aligned}
&\frac{1}{2}\left[\int (t_n - p_0(0))^2 \pi_{0,n}(\mathcal{X}) d\mathcal{X} + \int (t_n - p_1(0))^2 \pi_{1,n}(\mathcal{X}) d\mathcal{X}\right] \\
\geq &\frac{1}{2}\int \left[(t_n - p_0(0))^2 + (t_n - p_1(0))^2\right] \min\{\pi_{0,n}(\mathcal{X}), \pi_{1,n}(\mathcal{X})\} d\mathcal{X} \\
\geq &\frac{1}{4}(p_1(0) - p_0(0))^2 \int \min\{\pi_{0,n}(\mathcal{X}), \pi_{1,n}(\mathcal{X})\} d\mathcal{X}.
\end{aligned}$$

The last step comes from the elementary inequality

$$u^2 + v^2 \geq \frac{(u-v)^2}{2}.$$

In view of these remarks, to prove (2) it suffices to find densities p_0, p_1, such that

(i) $p_0, p_1 \in \mathcal{P}(\beta, L) = \{p \text{ is a density, } p \in \Sigma(\beta, L)\}$,

(ii) $|p_0(0) - p_1(0)| \geq c n^{-\frac{\beta}{2\beta+1}}$,

(iii) $\int \min\{\pi_{0,n}(\mathcal{X}), \pi_{1,n}(\mathcal{X})\} d\mathcal{X} \geq c'$.

Here, c, c' are positive constants.

Let us sketch the idea of constructing such p_0, p_1 (the details are available in the supplementary material, see Section 8 below). Take p_0 as a density, which is positive on the whole the real line and such that $p_0 \in \Sigma(\beta, L/2)$. For example, these properties are satisfied for the density of gaussian distribution $\mathcal{N}(0, \sigma^2)$ with σ^2 chosen big enough (check this as an exercise). Construct p_1 by adding a small perturbation to p_0:

$$p_1(x) = p_0(x) + h^\beta K\left(\frac{x}{h}\right),$$

where h is of the order $n^{-\frac{1}{2\beta+1}}$, the support of K is $[-\frac{1}{2}, \frac{3}{2}]$, K is infinitely differentiable and belongs to $\Sigma(\beta, L/2)$, $K(0) > 0$, and $\int K(u) du = 0$. These properties of K imply that p_1 is a probability density for all h smaller than a constant $h_0 > 0$, and $p_0, p_1 \in \mathcal{P}(\beta, L)$ for $h < h_0$.

Set now $h = \alpha n^{-\frac{1}{2\beta+1}}$, with $\alpha < h_0$. Then p_0, p_1 satisfy the above three conditions (i) – (iii):

(i) Proof available in the supplementary material, see Section 8 below.

(ii) This item easily follows by plugging in the value of h:

$$(p_0(0) - p_1(0))^2 = h^{2\beta} K^2(0) = K^2(0) \alpha^\beta n^{-\frac{2\beta}{2\beta+1}}.$$

(iii) The proof of this item is more involved. We will need the following two lemmas.

Lemma 2. *Let π_0 and π_1 be two probability densities on $\mathbb{R}^n$. Then*

$$\int \min(\pi_0, \pi_1) = 1 - \frac{1}{2}\int |\pi_0 - \pi_1| \tag{4}$$

$$\geq 1 - \frac{1}{2}\sqrt{\chi^2(\pi_0, \pi_1)}, \tag{5}$$

where

$$\chi^2(\pi_0, \pi_1) \stackrel{\text{def}}{=} \int \frac{(\pi_0 - \pi_1)^2}{\pi_0}$$

is the χ^2**-divergence** *between π_0 and π_1.*

Proof: Since π_0 and π_1 are probability densities,

$$2 = \int \pi_0 + \int \pi_1 = 2\int \min(\pi_0, \pi_1) + \int |\pi_0 - \pi_1|.$$

which implies (4). To show (5), apply the Cauchy-Schwarz inequality to obtain

$$\int |\pi_0 - \pi_1| = \int \frac{1}{\sqrt{\pi_0}}|\pi_0 - \pi_1|\sqrt{\pi_0} \leq \sqrt{\int \frac{1}{\pi_0}(\pi_0 - \pi_1)^2}.$$

Plug it back into (4) to complete the proof. ∎

Lemma 3. *Let $\pi_{j,n}(\mathcal{X}) = \prod_{i=1}^n p_j(x_i)$, $j = 0, 1$. Then*

$$\chi^2(\pi_{0,n}, \pi_{1,n}) = (\chi^2(p_0, p_1) + 1)^n - 1.$$

Proof: Left as an exercise.

With the two lemmas we are ready to prove (iii). In view of Lemma 2, it suffices to bound $\frac{1}{2}\sqrt{\chi^2(\pi_{0,n}, \pi_{1,n})}$ from above by a constant strictly less than one. First write $\chi^2(p_0, p_1)$ as

$$\begin{aligned}
\int \frac{(p_0 - p_1)^2}{p_0} &= \int_{-h/2}^{3h/2} \left[\frac{h^{2\beta}K^2(\frac{x}{h})}{p_0(x)}\right] dx \\
&\leq \frac{h^{2\beta+1}}{\min_{x\in[-h/2, 3h/2]} p_0(x)} \int K^2(u)du \\
&\leq \frac{h^{2\beta+1}}{\min_{x\in[-h_0/2, 3h_0/2]} p_0(x)} \int K^2(u)du \\
&\leq C_* h^{2\beta+1},
\end{aligned}$$

where the constant $C_* > 0$ depends only on p_0 and K. Plugging in the choice of h we obtain

$$\chi^2(p_0, p_1) \leq C_* \alpha^{2\beta+1} n^{-1}$$

Therefore, appling Lemma 3 we find

$$\chi^2(\pi_{0,n}, \pi_{1,n}) \leq (C_* \alpha^{2\beta+1} n^{-1} + 1)^n - 1 \leq \exp(C_* \alpha^{2\beta+1}) - 1.$$

The last step uses the inequality $1 + x < e^x$. Now, we choose α small enough to guarantee that $\exp(C_*\alpha^{2\beta+1}) - 1 < 1$. Then, applying Lemma 2 we obtain

$$\int \min(\pi_{0,n}, \pi_{1,n}) \geq 1 - \frac{1}{2} = \frac{1}{2},$$

and condition (iii) follows. ∎

Remark: We only proved the lower bound at $x = 0$. But by shifting the construction of p_0 and p_1 we can show that the result holds for all $x \in \mathbb{R}$. Thus, we have proved the following theorem.

Theorem 5. *Let $\beta > 0$ $L > 0$. There exists a constant $c_* > 0$ that only depends on β and L such that, for all $x \in \mathbb{R}$, $n \geq 1$,*

$$\inf_{T_n} \sup_{p \in \mathcal{P}(\beta,L)} \mathbb{E}_p[(T_n(x) - p(x))^2] \geq c_* n^{-\frac{2\beta}{2\beta+1}}.$$

Remark: Since the choice of x is arbitrary and c_* does not depend on x, we can equivalently put $\inf_{x\in\mathbb{R}}$ before the minimax risk.

7 Supplementary material 1: Other density estimators

7.1 Modified kernel density estimators

The main modification is to choose a bandwidth $\hat{h} = \hat{h}(x, X_1, \ldots, X_n)$ depending on x and on the sample, i.e., to consider the estimator

$$\hat{p}_n^*(x) = \frac{1}{n\hat{h}} \sum_{i=1}^{n} K\left(\frac{X_i - x}{\hat{h}}\right). \tag{6}$$

The dependence on x might be of interest, for example, if a density is highly oscillating on some interval, and very smooth in other places. We would need larger h for smooth parts and smaller h for oscillating parts.

Examples.

(a) *Cross-validation*: Here, $\hat{h} = \hat{h}_{CV}$ depends on the data, but not on x.

(b) *Nearest neighbor estimator* is defined by (6) where $\hat{h}$ is the distance from x to its k_nth nearest neighbor among $X_1, \ldots, X_n$. Here, k_n plays the role of smoothing parameter and needs to satisfy $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

Another modification is given by the *transformed density estimator*. Let $T : \mathbb{R} \to \mathbb{R}$ be a function in $C^\infty(\mathbb{R})$. Define $Y_i = T(X_i)$, and let $\hat{p}_n^Y(\cdot)$ be the kernel density estimator constructed with the transformed variables Y_i instead of X_i. The transformed density estimator is then defined by

$$p_{n,T}(x) = \hat{p}_n^Y(T(x))T'(x).$$

One example of transformation is $T(x) = \arctan(x)$. It maps the data from the real line to a bounded interval. This might help to reduce the effects of too erratic behavior of kernel estimator at the tails of the distribution where the observations are rare. It is essentially close to choosing larger bandwidth at the tails.

7.2 Orthogonal series estimator

Assume $\{\varphi_j\}$ is an orthonormal basis on $\mathbb{R}$ (or orthonormal basis on a bounded interval of $\mathbb{R}$), and that the density p is in the corresponding L_2 space. Let c_j be the Fourier coefficients of p with respect to the orthonormal basis $\{\varphi_j\}$:

$$c_j = \int p(x)\varphi_j(x)dx.$$

To estimate c_j, replace the density using the empirical distribution function:

$$\hat{c}_j = \int \varphi_j(x)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n}\varphi_j(X_i).$$

Clearly, $\hat{c}_j$ is an unbiased estimator of c_j and it converges to c_j with rate $\frac{1}{\sqrt{n}}$. Assume that $p(x)$ can be represented as

$$p(x) = \sum_{j=1}^{\infty} c_j\varphi_j(x)$$

where the series converges pointwisely. Then, the orthogonal series estimator (or projection estimator) of $p(x)$ is defined by truncating the series and replacing c_j's by their estimators $\hat{c}_j$:

$$p_n(x) = \sum_{j=1}^{N}\hat{c}_j\varphi_j(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\varphi_j(X_i)\varphi_j(x).$$

This estimator was proposed by Čencov (1962)[3]. If we define $K_N(X_i, x)$ to be $\sum_{j=1}^{N}\varphi_j(X_i)\varphi_j(x)$, we can see its resemblance with the kernel density estimator. The integer $N = N_n$ plays the role of smoothing parameter and needs to satisfy $N_n \to \infty$, $N_n/n \to 0$.

7.3 Nonparametric MLE type estimators

Consider the likelihood function $p \mapsto \prod_{i=1}^{n} p(X_i)$. It does not make sense to maximize it over all densities p, because we can construct densities concentrating on the data points, so the maximum of the likelihood function would be at infinity. Thus, we should restrict the density p to some class of functions.

Suppose we know that p belongs to a nonparametric class of functions $\mathcal{P}$, for instance, $\mathcal{P} = \mathcal{P}(\beta, L)$. Maximizer of the likelihood function over this class is hard to construct. There exist only few remarkable examples of nonparametric classes where the maximum likelohood estimator has a simple form. One of them is

$$\mathcal{P} = \{\text{monotone decreasing densities on } [0, \infty)\}.$$

The problem of maximizing the likelihood on this class has an explicit solution. Consider the empirical distribution function F_n. Take its smallest concave majorant $\bar{F}_n$, and then take the derivative of $\bar{F}_n$ to obtain the MLE (called the Grenander estimator). Note that like the histogram

[3]Cencov, N.N. (1962) Estimation of an Unknown Distribution Density from Observations. *Soviet Mathematics*, 3, 1559-1562.

estimator, the Grenander estimator is not smooth. It is piecewise constant since $\bar{F}_n$ is piecewise linear.

One can also consider penalized versions of MLE type estimators. We will use the log-likelihood from now on, because maximizing $\prod_{i=1}^n p(X_i)$ is equivalent to minimizing $-\frac{1}{n}\sum_{i=1}^n \log p(X_i)$. The estimator by Geman and Hwang (1982)[4] is defined as a solution of the problem

$$\min_{p\geq 0}\left[-\frac{1}{n}\sum_{i=1}^n \log p(X_i) + \lambda \|p^{(\ell)}\|_2^2\right]. \tag{7}$$

Here $\lambda > 0$ is the smoothing parameter. For example, one may choose λ by cross-validation. In practice, integer ℓ should typically be less than or equal to 2. Good and Gaskins (1971)[5] proposed yet another penalized estimator that solves

$$\min_{p\geq 0}\left[-\frac{1}{n}\sum_{i=1}^n \log p(X_i) + \lambda_1 \int p + \lambda_2 \mathcal{I}(p)\right],$$

where $\mathcal{I}(p)$ is the Fisher information $\int \frac{(p')^2}{p}$, and $\lambda_1, \lambda_2 > 0$ are smoothing parameters. If we introduce the notation $u = \sqrt{p}$, then the problem becomes

$$\min_{u\geq 0}\left[-\frac{2}{n}\sum_{i=1}^n \log u(X_i) + \lambda_1 \|u\|_2^2 + 4\lambda_2 \|u'\|_2^2\right],$$

which can be numerically solved by similar means as (7) with $\ell = 1$.

8 Supplementary material 2: Construction of p_0 and p_1

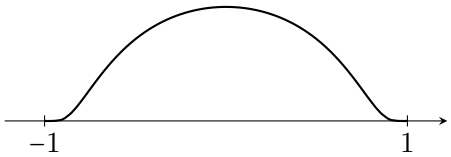
Let p_0 be a probability density, which is positive on $\mathbb{R}$ and belongs to $\mathcal{P}(\beta, L/2)$. Let

$$p_1(x) = p_0(x) + h^\beta K\left(\frac{x}{h}\right), \quad \forall x \in \mathbb{R}.$$

Here, $K \in C^\infty(\mathbb{R})$ is defined as follows. We first introduce an auxiliary function

$$K_0(u) = \begin{cases} \exp\left(-\dfrac{1}{1-u^2}\right), & |u| \leq 1, \\ 0, & |u| > 1. \end{cases}$$

Note that $K_0 \in C^\infty(\mathbb{R})$, and $\mathrm{supp}(K_0) = [-1, 1]$. The graph of function K_0 has the following form:



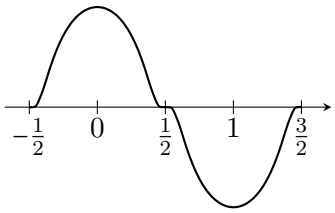
[4] Geman, S. and Hwang, C-R. (1982) Nonparametric Maximum Likelihood Estimation by the Method of Sieves, *The Annals of Statistics*, 10, 401-414.

[5] Good, I. J., and Gaskins, R. A. (1971) Nonparametric roughness penalties for probability densities. *Biometrika*, 2, 255-277

Set

$$g(u) = K_0(2u) - K_0\big(2(u-1)\big), \quad \forall u \in \mathbb{R}.$$

The graph of function g is as follows:



This function g has the following properties.

(a) $g(0) > 0$;

(b) $\int_{\mathbb{R}} g(u)\,du = 0$;

(c) $g \in C^\infty(\mathbb{R})$, hence $g \in \Sigma(\beta, L')$ for some $L' > 0$ (check this as an exercise).

We define the function $K : \mathbb{R} \to \mathbb{R}$ as follows:

$$\boxed{K \stackrel{\text{def}}{=} ag,}$$

where $a > 0$ is a constant. For sufficiently small a we have $K \in \Sigma(\beta, L/2)$. This holds thanks to the property (c) of function g. In what follows, we assume that a is chosen small enough to grant $K \in \Sigma(\beta, L/2)$. We use this function K in the definition of p_1.

We now check that p_0 and p_1 belong to $\mathcal{P}(\beta, L)$.

- The fact that p_0 belongs to $\mathcal{P}(\beta, L)$ is obvious since, by definition, it is a probability density, and $p_0 \in \Sigma(\beta, L/2)$, which implies that $p_0 \in \Sigma(\beta, L)$.
- Let us show that p_1 is a probability density for all $h > 0$ small enough.

  (1) First, the fact that
  $$\int_{\mathbb{R}} p_1(x)\,dx = 1$$
  follows immediately from property (b) of function g.

  (2) Second, p_1 is posititive everywhere on $\mathbb{R}$ for all $h > 0$ small enough. Indeed, we have that, for all $x \in [-h/2, 3h/2]$,
  $$\begin{aligned} p_1(x) &\geq \min_{t \in [-h/2, 3h/2]} p_0(t) - h^\beta \sup_{t \in [-h/2, 3h/2]} |K(t/h)| \\ &= \min_{t \in [-h/2, 3h/2]} p_0(t) - h^\beta \sup_{t \in \mathbb{R}} |K(t)|. \end{aligned}$$
  Since p_0 is continuous and $p_0(0) > 0$, there exists $h_0 > 0$ such that $p_1(x) > 0$ for every $x \in [-h/2, 3h/2]$ if h is smaller than h_0. For all $x \notin [-h/2, 3h/2]$, $p_1(x) = p_0(x) > 0$.

We conclude that p_1 is a probability density for $h < h_0$.

- Let us show that $p_1 \in \Sigma(\beta, L)$. For $\ell = \lfloor \beta \rfloor$ we have

$$p_1^{(\ell)}(x) = p_0^{(\ell)}(x) + h^{\beta-\ell} K^{(\ell)}\left(\frac{x}{h}\right).$$

Since $p_0 \in \Sigma(\beta, L/2)$ and $K \in \Sigma(\beta, L/2)$, we obtain that, for any $x, x' \in \mathbb{R}$,

$$\begin{aligned}
\left|p_1^{(\ell)}(x) - p_1^{(\ell)}(x')\right| &\le \left|p_0^{(\ell)}(x) - p_0^{(\ell)}(x')\right| \\
&\quad + h^{\beta-\ell}\left|K^{(\ell)}\left(\frac{x}{h}\right) - K^{(\ell)}\left(\frac{x'}{h}\right)\right| \\
&\le \frac{L}{2}|x - x'|^{\beta-\ell} + \frac{L}{2} h^{\beta-\ell}\left|\frac{x - x'}{h}\right|^{\beta-\ell} \\
&= L|x - x'|^{\beta-\ell}.
\end{aligned}$$

Conclusion: There exists $h_0 > 0$ such that $p_1 \in \mathcal{P}(\beta, L)$ for all $h < h_0$.