Mathematical Foundations of Infinite-Dimensional Statistical Models:

4.1 Definitions and Basic Approximation Theory

이종진

Seoul National University

ga0408@snu.ac.kr

July 06, 2018

Notation and Preliminaries

- $L_p(A), L_\infty(A)$
- C(A) denotes the space of continuous functions on A
- $C_u(A)$ denotes the space of uniformly continuous functions on A.
- $C_{per}(A)$ denotes the space of continuous periodic functions on A,

  f(a)=f(b), when A is a half-open bounded interval with endpoints a,b
- The L_p - Sobolev space of order $m \in \mathbb{N}$

  $$H_p^m(A) = \{f \in L^p : D^j f \ \in L^p \forall j = 1, \ldots, m : \|f\|_{H_p^m(A)} \equiv \|f\|_p + \|D^m f\|_p < \infty\}$$

Notation and Preliminaries : Weak derivative

- For $A \subseteq \mathbb{R}$ an interval, a function $f \in L^p(A)$ is said to be weakly differentiable if there exists a locally integrable function $Df-$ the weak derivative of $f-$ such that

$$\int_A f(u)\phi'(u)du = -\int_A Df(u)\phi(u)du$$

  for every infinitely differentiable function ϕ of compact support in the interior of A.

- The L_p - Sobolev space of order $m \in \mathbb{N}$

$$H_p^m(A) = \{f \in L^p : D^j f \ \in L^p \forall j = 1, \dots, m : \|f\|_{H_p^m(A)} \equiv \|f\|_p + \|D^m f\|_p < \infty\}$$

Notation and Preliminaries

- $C^m(A)$,

  $$C^m(A) = \{f^{(j)} \in C_u(A) : \forall j = 1, \ldots, m : \|f\|_{C^m(A)} \equiv \|f\|_\infty + \|f^{(m)}\|_p < \infty\}$$

- $C^\infty(A)$ denotes the space of infinitely differentiable functions defined on A
- $C_0^\infty(A)$ is the subspace which consists of all $\phi \in C^\infty(A)$ that have compact support in the interior of A.
- Schwartz space $S(\mathbb{R})$ consists of all functions $f \in C^\infty(\mathbb{R})$ such that all derivatives $f^{(\alpha)}, \alpha \geq 0$, exist and decay at $\pm\infty$ faster than any inverse polynomial.

Notation and Preliminaries : Convolution

- For measurable functions f,g, defined on $\mathbb{R}$, their convolution is

$$f * g(x) \equiv \int_{\mathbb{R}} f(x-y)g(y)dy, x \in \mathbb{R}$$

- If $f \in L^P(\mathbb{R}), g \in L^q(\mathbb{R}), 1 \leq p, q \leq \infty$ such that $1/p + 1/q = 1$

  then, $f * g \in C(\mathbb{R})$ and $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$ (Hölder's inequality)

- If $f \in L^P(\mathbb{R}), g \in L^1(\mathbb{R})$

  then, $f * g$ is well defined a.e., and $\|f * g\|_p \leq \|f\|_p \|g\|_1$ (Minkowski's inequality)

- If $f \in C(\mathbb{R}), g \in L^1(\mathbb{R})$

  then, $f * g$ is defined everywhere, and $f * g \in C(\mathbb{R})$

Notation and Preliminaries : Convolution

- We can also define

$$f * \mu(x) = \int_{\mathbb{R}} f(x-y)d\mu(y), \quad x \in \mathbb{R}$$

for $\mu \in M(\mathbb{R})$, where $M(\mathbb{R})$ denotes the spaces of finite signed measures on $\mathbb{R}$, and one has likewise $\|f * \mu\|_p \leq \|f\|_p |\mu|(\mathbb{R})$, where $|\mu|$ is the total variation measure of μ.

Notation and Preliminaries : Fourier transform

- For a function $f \in L^1(\mathbb{R})$, Fourier transform

$$\mathcal{F}[f](u) \equiv \hat{f}(u) = \int_{\mathbb{R}} f(x)e^{itx}dx, u \in \mathbb{R}$$

- If $f \in L^1(\mathbb{R})$ is such that $\hat{f} \in L^1(\mathbb{R},)$ the Fourier inversion theorem

$$\mathcal{F}^{-1}(\hat{f}) \equiv \frac{1}{2\pi}\int_{\mathbb{R}} e^{iu\cdot}\hat{f}(u)du = f \quad a.e.$$

- One immediately has

$$\|\hat{f}\|_\infty \leq \|f\|_1, \quad \|f\|_\infty \leq \frac{1}{2\pi}\|\hat{f}\|_1$$

the (inverse) Fourier transformation is injective from L^1 to L^∞

Notation and Preliminaries : Fourier transform

- $f \in L^1 \cap L^2$, Plancherel's theorem states that

$$\|f\|_2 = \frac{1}{\sqrt{2\pi}}\|\hat{f}\|_2, \quad \langle f, g\rangle = \frac{1}{2\pi}\langle \hat{f}, \hat{g}\rangle$$

  and $\sqrt{2\pi}\mathcal{F}$ extends continuously to an isometry from L^2 to L^2

- Some basic properties
  1. $\mathcal{F}[f(\cdot - k)](u) = e^{-iku}\hat{f}(u)$
  2. $\mathcal{F}[f(a\cdot)](u) = a^{-1}\hat{f}(u/a), \quad a > 0$
  3. $\mathcal{F}[f * g](u) = \hat{f}(u)\hat{g}(u), \quad \mathcal{F}[f * \overline{f(-\cdot)}](u) = |\hat{f}(u)|^2$
  4. $\frac{d^N}{(du)^N}\hat{f}(u) = \int_{\mathbb{R}} f(x)(-ix)^N e^{-ixu}dx, (iu)^N\hat{f}(u) = \mathcal{F}\left[D^N f\right](u)$

- Fourier transform $\mathcal{F}$ maps the Schwartz space $\mathcal{S}(\mathbb{R})$ into itself.

Notation and Preliminaries : Fourier transform

- Instead of $\mathbb{R}$ the group is $(0, 2\pi]$ with addtion modulo 2π, we have similar results.
- In particular, any 2π-periodic $f \in L^2((0, 2\pi])$, decomposes into its Fourier series

$$f = \sum_{k \in \mathbb{Z}} c_k e^{ik\cdot}, \text{in } L^2((0, 2\pi]), \quad c_k \equiv c_k(f) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx, \{c_k\} \in \ell_2$$

- in fact, $f \mapsto \{c_l\}$ gives a Hilbert space isometry between $L^2((0, 2\pi])$ and ℓ_2 If, further, $\{c_k\} \in \ell_1$, then the Fourier series of f converges a.e. on $(0, 2\pi]$ (pointwise if f is continuous)

- Fourier inversion and Fourier series can be linked to each other by the Poisson summation formula: if $f \in L^1(\mathbb{R})$, then the periodised sum

$$S(x) = \sum_{l \in \mathbb{Z}} f(x + 2\pi l), \quad x \in (0, 2\pi]$$

converges a.e., belongs to $L^1((0, 2\pi])$ and the Fourier coefficients of S are given by

$$c_k(S) = \frac{1}{2\pi} \hat{f}(k) = \mathcal{F}^{-1}[f](-k)$$

Notation and Preliminaries : The Schwartz space

- Define a countable familiy of seminorms on $\mathcal{S}(\mathbb{R})$

$$\|f\|_{m,r} = \max_{\alpha \leq r} \left\| \left(1 + |\cdot|^2\right)^m f^{(\alpha)} \right\|_\infty, \quad m, r \in \mathbb{N} \cup \{0\}$$

  these seminorms provide a metrisable locally convex topology on $\mathcal{S}(\mathbb{R})$

- $\mathcal{S}(\mathbb{R})$ is complete, and the set $C_0^\infty(\mathbb{R})$ is dense in $\mathcal{S}(\mathbb{R})$

- $\mathcal{S}^* \equiv \mathcal{S}(\mathbb{R})^*$ is the topological dual space of $\mathcal{S}(\mathbb{R})$, tempered distributions, or Schwartz distributions, equipped with the weak topology: $T_n \to T$ in $\mathcal{S}^*$ if $T_n(\phi) \to T(\phi)$ for every $\phi \in \mathcal{S}$
- Weak differentiation is continuous operation from $\mathcal{S}^*$ to $\mathcal{S}^*$
- μ is a finite measure or any signed measure of at most polynomial growth at $\pm\infty(|\mu|(x; |x| < R) \lesssim (1 + |R|^2)^l$ for all $R > 0$, some $l \in \mathbb{N})$ then, the action of μ on $\mathcal{S}(\mathbb{R})$ defines an element of $\mathcal{S}^*(\mathbb{R})$.

$$f \mapsto \int_{\mathbb{R}} f d\mu$$

- any $f \in L^p$ acting on $\mathcal{S}(\mathbb{R})$ by $\phi \mapsto \int f\phi$ defines an element of $\mathcal{S}^*(\mathbb{R})$

Notation and Preliminaries : The Schwartz space

- Fourier transform of $T \in \mathcal{S}^*$ as the element $\mathcal{F}T$ of $\mathcal{S}^*$ whose action on $\mathcal{S}$ is given by

$$\phi \mapsto \mathcal{F}T(\phi) = T(\hat{\phi})$$

- For $T = f \in L^1$,

$$\int_{\mathbb{R}} \hat{f}(u)\phi(u)du = \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-iux} f(x)\phi(u)dudx = \int_{\mathbb{R}} f(u)\hat{\phi}(u)du$$

  In particular, the Fourier transform maps $\mathcal{S}^*$ continuously onto itself, and $\mathcal{F}^{-1}[\mathcal{F}T] = T$ in $\mathcal{S}^*$.

- We can by the same principles define periodic Schwartz distributions. For A any interval (0,a]

Approximate Identities

- Convolution with Kernels
- For $f : \mathbb{R} \to \mathbb{R}$, we can define the convolution

$$K_h * f(x) = \int_{\mathbb{R}} K_h(x-y)f(y)dy = \int_{\mathbb{R}} f(x-y)K_h(y)dy = f * K_h(x)$$

  of f with a suitably 'localised' kernel function

$$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right), \quad h > 0, x \in \mathbb{R}$$

  where K is typically chosen to be bounded and integrable and in particular satisfies $\int_{\mathbb{R}} K(x)dx = 1$
- so as $h \to 0$ the function K_h looks more and more like a point mass δ_0 at 0

$$f * K_h \sim f * \delta_0 = \int_{\mathbb{R}} f(x-y)d\delta_0(y) = f(x)$$

Approximate Identities

Proposition (4.1.1)

Let $f : \mathbb{R} \to \mathbb{R}$, *be a measurable function, and let* $K \in L^1$ *satisfy* $\int_{\mathbb{R}} K(x)dx = 1$.

1. *If* f *is bounded on* $\mathbb{R}$ *and continuous at* $x \in \mathbb{R}$, *then* $K_h * f(x)$ *converges to* $f(x)$ *as* $h \to 0$.
2. *If* f *is bounded and uniformly continuous on* $\mathbb{R}$, *then* $\|K_h * f - f\|_\infty \to 0$ *as* $h \to 0$
3. *If* $f \in L^p$ *for some* $1 \le p < \infty$, *then* $\|K_h * f - f\|_p \to 0$ *as* $h \to 0$

Ortho-normal basis

- Let $\mathcal{L} \subset \mathbb{Z}$ be an index set. A family of functions $\{e_l : l \in \mathcal{L}\} \subset L^2(A)$ is called an ortho-normal basis.
  1. $\langle e_k, e_l \rangle = 0$ whenever $k \neq l$ and $\langle e_l, e_l \rangle = \|e_l\|_2^2 = 1$ otherwise.
  2. if the linear span
  $$\left\{ \sum_{l \in \mathcal{L}} c_l e_l : c_l \in \mathbb{R} \right\}$$
  is norm-dense in $L^2(A)$

Projection kernel

- V the closed subspace of $L^2(A)$ genertated by the linear span of $\{e_l : l \in \mathcal{L}'\}$ for some subset $\mathcal{L}' \subset \mathcal{L}$,
- $\pi_V(f)$ is the best L^2 -approximation of f from the subspace V .

$$\pi_V(f)(x) = \sum_{l\in\mathcal{L}'} \langle f, e_l \rangle\, e_l(x) = \int_A \sum_{l\in\mathcal{L}'} e_l(x) e_l(y) f(y) dy$$

- we define the projection kernel

$$\pi_V(f)(x) = \int_A K_V(x,y) f(y) dy, \quad K_V(x,y) = \sum_{l\in\mathcal{L}'} e_l(x) e_l(y)$$

- We shall now discuss some classical examples of ortho-normal bases of L^2 including some basic historical examples of wavelet bases, which will be introduced in full generality later.

The Trigonometric Basis

- If $A = (0, 1]$, then the trigonometric basis of $L^2((0,1])$ consists of the complex trigonometric polynomials

$$\left\{ e_l = e^{2\pi i l} = \cos(2\pi l \cdot) + i \sin(2\pi l \cdot) : l \in \mathbb{Z} \right\}$$

- The partial sums can be represented as

$$S_N(f)(x) = \sum \langle f, e_l \rangle \, e_l(x) = \int_0^1 D_N(x-y) f(y) dy = D_N * f(x)$$

where

$$D_N(x) = \sum_{|l| \leq N} e^{2\pi i l x} = \frac{\sin((2N+1)\pi x)}{\sin(\pi x)}$$

- Dirichlet kernel, Proposition 4.1.1. does not hold, D_N is not bounded uniformly in $L^1(A)$
- Convergence of $S_N(f) \to f$ in $L^p(A), p \neq 2$, or in $C_u(A)$ does not hold in general
- One way around this problem is based Fejer kernel

$$F_m = \frac{1}{m+1} \sum_{k=0}^{m} D_k$$

The Haar Basis

- $\left\{\phi_{jk} \equiv 2^{j/2}\phi\left(2^j(\cdot) - k\right), k \in \mathbb{Z}\right\}, \quad j \in \mathbb{N} \cup \{0\}$
- Partition $\mathbb{R}$ into dyadic intervals $(k/2^j, (k+1)/2^j]$
- $K_j(x, y) = 2^j K\left(2^j x, 2^j y\right) = \sum_{k\in\mathbb{Z}} 2^j \phi\left(2^j x - k\right)\phi\left(2^j y - k\right) = \sum_{k\in\mathbb{Z}} \phi_{jk}(x)\phi_{jk}(y)$
- It has some comparable approximation properties

Proposition (4.1.2)

Let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function, and let K be the Haar projection kernel.

1. *If f is bounded on $\mathbb{R}$ and continuous at $x \in \mathbb{R}$, then $K_j(f)(x)$ converges to $f(x)$ as $j \to \infty$.*
2. *If f is bounded and uniformly continuous on $\mathbb{R}$, then $\|K_j(f) - f\|_\infty \to 0$ as $j \to \infty$*
3. *If $f \in L^p$ for some $1 \leq p < \infty$, then $\|K_j(f) - f\|_p \to 0$ as $j \to \infty$*

The Haar Basis

- $K_j(f)$

$$K_j(f) = K_0(f) + \sum_{l=0}^{j-1} \left(K_{l+1}(f) - K_l(f)\right)$$

an elementary computation shows that

$$K_{l+1}(f) - K_l(f) = \sum_{k\in\mathbb{Z}} \langle \psi_{lk}, f \rangle \, \psi_{lk}$$

where $\psi = 1_{[0,1/2]} - 1_{(1/2,1]}, \psi_{lk}(x) = 2^{l/2}\psi\left(2^l x - k\right)$

$$f = \sum_{k\in\mathbb{Z}} \langle \phi_k, f \rangle \, \phi_k + \sum_{l=0}^{\infty} \sum_{k\in\mathbb{Z}} \langle \psi_{lk}, f \rangle \, \psi_{lk}$$

- $\{\phi_k, \psi_{lk} : k \in \mathbb{Z}, l \in \mathbb{N} \cup \{0\}\}$ forms an ortho-normal basis of L^2 known as the Haar basis.

The Shannon Basis

- Consider a function $f \in \mathcal{V}_\pi$, where $\mathcal{V}_\pi$ is the space of continuous functions $f \in L^2$ which have (distributional) Fourier transform $\hat{f}$ supported in $[-\pi, \pi]$
- express $\hat{f}$ with trigonometric basis.

$$\hat{f} = \sum_{k \in \mathbb{Z}} c_k e^{ik(\cdot)}, \qquad \text{in } L^2([-\pi, \pi])$$

with Fourier coefficients given by

$$c_k = c_k(\hat{f}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-iku} \hat{f}(u) du = f(-k)$$

the last identity following from (4.5) if $f \in L^1$

$$f(x) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} c_k \int_{-\pi}^{\pi} e^{iu(k+x)} du$$
$$= \sum_{k \in \mathbb{Z}} f(k) \frac{\sin \pi(x-k)}{\pi(x-k)}$$

The Shannon Basis

- $\phi(X) = sin(\pi x)/(\pi x)$
- $\hat{\phi} = 1_{[-\pi,\pi]}$, and its interger translates of the function ϕ are ortho-normal in L^2. (by Plancherel's theorem.)
- $\{\phi_k = \phi(\cdot - k) : k \in \mathbb{Z}\}$ is an ortho-normal in $\mathcal{V}_\pi$
- $\left\{\phi_{jk} = 2^{j/2}\phi\left(2^j(\cdot) - k\right) : k \in \mathbb{Z}\right\}$ span $\mathcal{V}_{2^j\pi}$
- The projection of $f \in L^2(\mathbb{R})$ onto $\mathcal{V}_{2^j\pi}$ is

$$\Pi_{\mathcal{V}_{2^j\pi}}(f) = \sum_k \langle \phi_{jk}, f \rangle \phi_{jk}$$

- Like, haar basis, we can telescope these projections, by set ψ

$$\psi = \mathcal{F}^{-1}\left[1_{[-2\pi,-\pi]} + 1_{[\pi,2\pi]}\right]$$

- the functions $\left\{\psi_{lk} = 2^{l/2}\psi\left(2^l \cdot -k\right) : k \in \mathbb{Z}\right\}$ form an ortho-normal basis for $W_l = \mathcal{V}_{2^l\pi} \ominus \mathcal{V}_{2^{l-1}\pi}$
- f

$$f = \sum_k \langle \phi_k, f\rangle \phi_k + \sum_{l=0}^{\infty}\sum_k \langle \psi_{lk}, f\rangle \psi_{lk}$$

- the ortho-normal 'Shannon' basis $\{\phi_k, \psi_{lk} : k \in \mathbb{Z}, l \in \mathbb{N} \cup \{0\}\}$
- We would like to construct ortho-normal bases of L^2 that are in a sense interpolating' between the Haar and Shannon bases, and this is what leads to wavelet theory, as we shall see later.

Approximation in Sobolev Spaces by General Integral Operators

- Consider the general framework of integral operators.
- $f \mapsto K_h(f) = \int_{\mathbb{R}} K_h(\cdot, y) f(y) dy = \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{\cdot}{h}, \frac{y}{h}\right) f(y) dy, \quad h > 0$
- Calderon-Zygmund operators, with the obvious notational conversion $h = 2^{-j}$

Proposition (4.1.3)

Let $f : \mathbb{R} \to \mathbb{R}$ *be a measurable function, let* K_h *be as above and suppose that* $\int_{\mathbb{R}} \sup_{v \in \mathbb{R}} |K(v, v-u)| du < \infty$, $\int_{\mathbb{R}} K(x, y) dy = 1$ *for every* $x \in \mathbb{R}$. *Then we have*

1. *If* f *is bounded on* $\mathbb{R}$ *and continuous at* $x \in \mathbb{R}$, *then* $K_h(f)(x)$ *converges to* $f(x)$ *as* $h \to 0$.
2. *If* f *is bounded and uniformly continuous on* $\mathbb{R}$, *then* $\|K_h(f) - f\|_{\infty} \to 0$ *as* $h \to 0$
3. *If* $f \in L^p$ *for some* $1 \leq p| < \infty$, *then* $\|K_h(f) - f\|_p \to 0$ *as* $h \to 0$

- To investigate further approximation properties we shall impose following conditions.
  - $(\boldsymbol{M})$: $c_N(K) \equiv \int_{\mathbb{R}} \sup_{v\in\mathbb{R}} |K(v, v-u)||u|^N du < \infty$
  - $(\boldsymbol{P})$: For every $v \in \mathbb{R}$ and $k = 1, \ldots, N-1$

$$\int_{\mathbb{R}} K(v, v+u)du = 1 \quad \text{and} \quad \int_{\mathbb{R}} K(v, v+u)u^k du = 0$$

Proposition (4.1.5)

Let K be a kernel that satisfies Condition $(\boldsymbol{M}), (\boldsymbol{P})$ for some $N \in \mathbb{N}$ and let

$$c(m, K) = c_m(K) \int_0^1 \frac{(1-t)^{m-1}}{(m-1)!} dt$$

for any integer $m \leq N$.

1. *If $f \in H_p^m(\mathbb{R}), 1 \leq p < \infty$, then*

$$\|K_h(f) - f\|_p \leq c(m, K) \|D^m f\|_p h^m$$

2. *If $f \in C^m(\mathbb{R})$, then*

$$\|K_h(f) - f\|_\infty \leq c(m, K) \left\| f^{(m)} \right\|_\infty h^m$$

Littlewood-paley Decomposition

- The main idea behind the Haar and Shannon bases of L^2 was a partition of unity either in the time or the frequency domain.
- However, the functions used in the partition are not smooth or indicatios of intervals.
- Use smooth functions, relaxing the requirement of orthgonality of the functions involved.

Littlewood-paley Decomposition

- Take $\phi \in \mathcal{S}(\mathbb{R})$ to be symmetric function such that

$$\hat{\phi} \in C_0^\infty(\mathbb{R}), \quad \text{supp}(\hat{\phi}) \in [-1, 1], \quad \hat{\phi} = 1 \text{ on } \left[-\frac{3}{4}, \frac{3}{4}\right]$$

Define, moreover,

$$\hat{\psi} = \hat{\phi}\left(\frac{\cdot}{2}\right) - \hat{\phi} \quad \textit{equivalentto} \quad \psi = 2\phi(2\cdot) - \phi$$

so that $\hat{\psi}$ is supported in $\left\{2^{-1} \leq |u| \leq 2\right\}$. If we set $\psi_{2^{-j}} = 2^j \psi\left(2^j .\right)$, then $\overline{\psi_{2^{-j}}} = \hat{\psi}\left(\cdot/2^j\right)$, and by a telescoping sum, for every $u \in \mathbb{R}$

$$\hat{\phi}(u) + \sum_{j=0}^{\infty} \hat{\psi}\left(u/2^j\right) = \lim_{J\to\infty}\left(\hat{\phi}(u) + \sum_{j=0}^{J-1} \hat{\psi}\left(u/2^j\right)\right) = \lim_{J} \hat{\phi}\left(u/2^J\right) = 1$$

Littlewood-paley Decomposition

- For f with Fourier transform $\hat{f}(u)$ and every $u \in \mathbb{R}$,

$$\hat{f}(u) = \hat{f}(u)\hat{\phi}(u) + \sum_{i=0}^{\infty} \hat{\psi}\left(u/2^{j}\right)\hat{f}(u)$$

- Then, f is

$$f = f * \phi + \sum_{j=0}^{\infty} f * \psi_{2-j} = \lim_{J \to \infty} f * \phi_{2-j}$$

where $\phi_{2-} - J = 2^{J}\phi\left(2^{J}\cdot\right)$

- since $\hat{\phi}(0) = 1$, we see that $\int \phi = 1$, and since $\phi \in \mathcal{S}(\mathbb{R})$, we conclude from Proposition 4.1 .1 that the last limit holds in L^{p} whenever $f \in L^{p}$
- Moreover, $\int_{\mathbb{R}} x^{k}\phi(x)dx$ equals zero for every $k \in \mathbb{N}$ because $D^{k}\mathcal{F}[\phi](0)$ does, so Proposition 4.1 .5 applies for every N with $h = 2^{-j}$.