FDA Homework 3

Seokjun Choi

November 8th, 2019

1 Chapter 12

1.1 Problem 2

Show that for every eigenvalue λ of a bounded operator L, we have $|\lambda| \leq ||L||_{\mathcal{L}}$.

In problem's statement, L is being assumed to be spectral decomposable (i.e. self-adjoint, compact(or more strongly, Hilbert-Schmidt) operator), so I will work with these assumptions.

Firstly I claim that $||L||$ or $-||L||$ is eigenvalue of L. Without loss of generality, assume first case and denote $\lambda_1 = ||L|| = \sup\{< Lf, f >: ||f|| = 1\}$. Let $\{f_n\} \in \mathcal{H}$ such that $||f_n|| = 1$, $< Tf_n, f_n > \to \lambda_1$ and $Tf_n \to g$ for some $g \in \mathcal{H}$. (such $\{f_n\}, g$ exist because L is compact and $\mathcal{H}$ is complete.) Then

$$||Lf_n - \lambda_1 f_n||^2 = ||Lf_n||^2 - 2\lambda_1 < Lf_n, f_n > +\lambda_1^2 ||f_n||^2$$

$$\leq ||L||^2 ||f_n||^2 - 2\lambda_1 < Lf_n, f_n > +\lambda_1^2 ||f_n||^2$$

$$\leq \lambda_1^2 - 2\lambda_1 < Lf_n, f_n > +\lambda_1^2$$

$$\leq 2\lambda_1^2 - 2\lambda_1^2 \to 0$$

So $Lf_n \to g$, $\lambda_1 f_n \to g$, and under continuity of L from the assumptions, we get $\lambda_1 g = Lg$. And we also verify $g \neq 0$ because if 0, it becomes $\lambda_1 = ||L|| = 0$, contradiction. thus λ_1 is eigenvalue of L. (The proof for $||L|| = -\lambda_1$ case is similar.)

Next I claim one more thing that above $\lambda_1 = \max |\lambda|$ over all λs which are eigenvalues of L. Without loss of generality, consider only the case all eigenvalues of L is nonnegative. (if not, change sign of it and its pair eigenfunction together.) Assume the claim is false, then there are λ^* and the pair eigenfuction v^* whose norm is 1. Then, $Lv^* = \lambda^* > \lambda_1 = ||L|| = \sup_{||v||=1} ||Lv||$, we have contradiction (to sup and definition of operator norm). So, all eigenvalues of L are smaller then λ_1, and it is what we want, $|\lambda| \leq ||L|| = \lambda_1$.

1.2 Problem 5

Assume that $X_1, ..., X_N$ are iid element of $L^2[0,1]$ with $E||X_n||^4 < \infty$ and whose first p eigenvalue are distinct. Prove that

$$|N < \hat{v}_j - v_j, v_j > | = O_P(1) \text{ for } j = 1, ..., p$$

Why is this a seemingly unusual convergence rate? (Hint: $| < \hat{v}_j - v_j, v_j > | = \frac{1}{2}||\hat{v}_j - v_j||$)

With hint of the problem, I'll show that $N||\hat{v}_j - v_j||^2$ is bounded in probability sense, or $O_P(1)$. (constant 1/2 does not matter in this context.)

Then by theorem 12.2.1 on our book, under our assumptions, we know that

$$\limsup_{N \to \infty} NE[||\hat{v}_j - v_j||^2] < C$$

for some C and $j = 1, ..., p$. On other hand, by Chebyshev's inequality,

$$Pr(N||(\hat{v}_j - v_j)||^2 > \alpha) < \frac{NE[||\hat{v}_j - v_j||^2]}{\alpha}$$

The result of above theorem says that right-hand side is bounded in probability sense for all $\alpha > 0$. Then using definition of boundedness in probability to left-hand side, we get $N||\hat{v}_j - v_j||^2$ is bounded in probability, which we want.

Since this proof (and the base theorem) depend on the strong condition that whole points of function are completely observable, the inner product term(view as 'weak distance' heuristically) is bounded of rate $O(N^{-1})$, unusual(-ly good) rate comparing to ordinary parametric rate $O(N^{-1/2})$.

1.3 Problem 6

Prove Theorem 12.1.3: Let $x, y \in \mathcal{H}$. Then $||x \otimes y||_{\mathcal{H} \otimes \mathcal{H}} = || < y, . > x||_{\mathcal{S}}$

When we view tensor as operator, $x \otimes y(.) = < y, . > x$ holds. Using this fact, when $\{e_i\}$ are orthonormal basis of $\mathcal{H}$, by the definition of Hilbert-Schmidt norm equipped on $\mathcal{S}$,

$$|| < y, . > x||_{\mathcal{S}}^2 = \sum_{i=1}^{\infty} ||(< y, z > x)e_i||^2 \text{ for } \forall z \in \mathcal{H}$$

then by Parseval's identity,

$$= || < y, z > x||_{\mathcal{H}}^2 \text{ for } \forall z \in \mathcal{H}$$

then by above relation between inner product and tensor product and since z is arbitrary,

$$= ||x \otimes y||_{\mathcal{H} \otimes \mathcal{H}}^2$$

1.4 Problem 7

Suppose that the data $X_n(t) : t \in [0, 1], 1 \leq n \leq N$ are expressed using an orthonormal basis $e_1, ..., e_J$:

$$X_n(t) = \sum_{j=1}^{J} x_{n_j} e_j(t)$$

In this case, the EFPC's, $\hat{v}_i(t)$ can also be expressed as

$$\hat{v}_i(t) = \sum_{j=1}^{J} \hat{v}_{ij} e_j(t)$$

Explain how to obtain the coefficient $\hat{v}_{ij}$ from the x_{nj}. Justify your answer.

I cannot ensure the purpose of this problem. So I will give some outline to get $\hat{v}$.

For notational convenience, I'll omit hat-expression (except some variable) despite almost all variables are sample value.

1. Fit to N function-objects $X_1, ..., X_N$ using basis $\{e_j\}$ in (normally) L^2 with data points x_{nj}, and get $X_n(t) = \sum_m c_{nm} B_m(t)$ form where B_m is m-th basis and t becomes data points of each curve.
   In practice, do some transformation corresponding to chosen basis (for example, if choose Fourier basis, then get Fourier series expression algorithm-matically.) and keep the coefficients as form of matrix.

2. Get sample mean function $\hat{\mu}(t) = N^{-1} \sum_{n=1}^{N} X_n(t)$. we get it easily by just calculating means over each coefficients of each basis respectively and taking sum, so $\hat{\mu}(t) = \sum_m (N^{-1} \sum_n c_{nm}) B_m(t)$.
   In practice, we only need to keep $\bar{c}_m = N^{-1} \sum_n c_{nm}$.

3. Get sample covariance operator (if need, adjust mean to 0.) $\hat{c}(t, s) = N^{-1} \sum_{n=1}^{N} (X_n(t) - \hat{\mu}(t))(X_n(s) - \hat{\mu}(s))$.
   In practice, calculate $\tilde{c}_{nm} = c_{nm} - \bar{c}_m$ and get $\tilde{c}_{m_1}^T \tilde{c}_{m_2}$, then the $(m_1 m_2)$-th element becomes the coefficient of $B_{m_1}(t) B_{m_2}(t)$.

4. Find eigenpairs, $\lambda_j v_j(t) = \int_0^1 \hat{c}(t, s) v_j(s) ds$.
   In practice, find λ_js and v_js by solving $\lambda_j \int v_j(t) B_{m_1}(t) dt = \int \int \hat{c}(t, s) v_j(s) B_{m_1}(t) ds dt$. In detail, we replace integral with sum, $\lambda_j v_j(t) = \sum_s \hat{c}(t, s) v_s$, so solve $\lambda_j v_{jm_1} = \sum_{m_3} \tilde{c}_{m_1}^T \tilde{c}_{m_3} v_{jm_3}$.

At last step, (replace notation j to i and now use j for index of basis) we get $\hat{v}_{ij}$ values which we want.

1.5 Problem 12

Under the same assumptions as in Problem 12.8.5, shows that, for $j \neq k$ and $1 \leq j \leq p$,

$$< \hat{v}_j - v_j, v_k > = \frac{< \hat{C} - C, \hat{v}_j \otimes v_k >}{\hat{\lambda}_j - \lambda_k}$$

What can you conclude about the asymptotic distribution of $N^{-1/2} < \hat{v}_j - v_j, v_k >$?
By theorem 12.3.2 on our book, under our assumptions we know that

$$N^{1/2}(\hat{v}_j - v_j) = \sum_{i \neq j} \frac{1}{\lambda_j - \lambda_i} < \sqrt{N}(\hat{C} - C), v_i \otimes v_j > v_i + o_P(1)$$

Take both side to $< ., v_k >$ and neglect ignorable term, then

$$N^{1/2} < \hat{v}_j - v_j, v_k > = \sum_{i \neq j} \frac{1}{\lambda_j - \lambda_i} < \sqrt{N}(\hat{C} - C), v_i \otimes v_j >< v_i, v_k >$$

then, the last inner product term becomes 1 only $i = k$, and otherwise 0. So we can rewrite it without summation symbol as

$$N^{1/2} < \hat{v}_j - v_j, v_k > = \frac{\sqrt{N} < \hat{C} - C, v_k \otimes v_j >}{\lambda_j - \lambda_k}$$

multiply $\sqrt{N}$ to both sides. And, because $\hat{\lambda}_j \to \lambda_j$ and $\hat{v}_j \to v_j$ as $N \to \infty$ in probability sense. (For showing eigenfunction part convergence in detail, using problem 5's result considering the form dividing problem's expression by N. And for eigenvalue part convergence, do like procedure of solving problem 5 using the result of eigenvalue part of theorem 12.2.1 and Chebyshev's inequality.)
So we can replace λ_j, v_j with $\hat{\lambda}_j, \hat{v}_j$ with adding only ignorable terms in right-hand side, and if we disregard them, we get

$$< \hat{v}_j - v_j, v_k > = \frac{< \hat{C} - C, \hat{v}_j \otimes v_k >}{\hat{\lambda}_j - \lambda_k}$$

which we want.
(comment: Can I commute tensor product term in inner product? Intuitively, in general, may be No. But in this case, thinking only above equation and view $C, \hat{C}, v_k \otimes \hat{v}_j$ as operator of $\mathcal{H}_1 \times \mathcal{H}_2 \to \mathcal{H}$ where $\mathcal{H} = \mathcal{H}_1 = \mathcal{H}_2$, let's consider the domain-permuted(?) operator of $C, \hat{C}, \hat{v}_j \otimes v_k$ such that $\mathcal{H}_2 \times \mathcal{H}_1 \to \mathcal{H}$. then, $C, \hat{C}$ is invariant because of their definition, so are LHS's $v_j, \hat{v}_j, v_k$ and RHS's $\hat{\lambda}_j, \lambda_k$. And only last thing changes to correspond to $\hat{v}_j \otimes v_k$. So I may commute with tensor product in our special context.)
For finding distribution, using this result and apply CLT to $\sqrt{N}(\hat{C} - C)$ (like theorem 12.3.1), or taking the result of corollary12.3.1 directly, we get the asymptotic distribution of $N^{-1/2} < \hat{v}_j - v_j, v_k >$ as

$$N(0, N^{-1/2} < C_j, v_j \otimes v_k >)$$

where C_j is of the corollary's, i.e.

$$C_j = \sum_{i \neq j} \sum_{l \neq j} \frac{< \Gamma, v_i \otimes v_j \otimes v_l \otimes v_j >}{(\lambda_j - \lambda_i)(\lambda_j - \lambda_l)} (v_i \otimes v_l)$$

$$\Gamma = E[[(X - \mu) \otimes (X - \mu) - C] \otimes [(X - \mu) \otimes (X - \mu) - C]]$$

and μ is mean function of X.