FDA Homework 4

Seokjun Choi

November 29th, 2019

1 Chapter 4

1.1 Problem 1

Consider the design matrix X in (4.5). Show that if X has rank p, then X^TX is non-singular.

Firstly note that X^TX is symmetric for any case of X. So I'll show that X^TX is positive definite, which is equivalent statement of non-singularity. (For verifying this equivalence, use spectral decomposition to symmetric positive definite matrix and observe all eigenvalues should be non zero.)

Assume $n > p$, an ordinary situation. But it is direct from below observation. For $v \in \mathcal{R}^p$ and $v \neq 0$,

$$v^TX^TXv = \langle Xv, Xv \rangle_{\mathcal{R}^n} > 0$$

Last inequality follows from the fact that because X is rank p linear transformation from $\mathcal{R}^p$ to $\mathcal{R}^n, n > p$, only $v = 0$ can makes $Xv = 0$, but by assumption, $v \neq 0$ thus $Xv \neq 0$. then combining the definition of inner-product, $\langle a, a \rangle \geq 0$ for all $a \in \mathcal{H}$ and $\langle a, a \rangle = 0$ iff $a = 0$.

1.2 Problem 2

Consider the linear model (4.6) and the least squares estimator (4.7). Suppose x is a deterministic matrix of rank p and the errors ϵ_i are uncorrelated with variance σ_ϵ^2. Show that $E[\hat{\beta}] = \beta$ and $Var[\hat{\beta}] = \sigma_\epsilon^2(X^TX)^{-1}$.

Under the context and notation of book's and this problem, $\epsilon \sim [0, diag(\sigma_\epsilon^2)]$ and X^TX is invertible since X is rank p and by result of problem 1. Then using (4.7),

$$\hat{\beta} = (X^TX)^{-1}X^TY = (X^TX)^{-1}X^T(X\beta + \epsilon) = \beta + (X^TX)^{-1}X^T\epsilon)$$

then since $E(\epsilon) = 0$,

$$E(\hat{\beta}) = E(\beta + (X^TX)^{-1}X^T\epsilon)) = \beta$$

And

$$Var(\hat{\beta}) = Var(\beta + (X^TX)^{-1}X^T\epsilon) = Var((X^TX)^{-1}X^T\epsilon)$$
$$= (X^TX)^{-1}X^TVar(\epsilon)X(X^TX)^{-1} = (X^TX)^{-1}X^T\sigma_\epsilon^2IX(X^TX)^{-1} = \sigma_\epsilon^2(X^TX)^{-1}$$

2 Chapter 5

2.1 Problem 1

Show that for any functions $\varphi_1, \varphi_2, ..., \varphi_k$, the $K \times K$ matrix I_φ with the entries $\varphi_{kl} = \int \varphi_k(t)\varphi_l(t)dt, 1 \leq k, l \leq K$, is nonnegative definite, i.e. for any real numbers $x_1, x_2, ..., x_K$,

$$\sum_{k,l=1}^{K} \varphi_{kl}x_kx_l \geq 0$$

For becoming this problem to be proper, there should be a assumption: "each φ_i is in $\mathcal{L}^2$", rather than "any function φ". Because if not, the value $\varphi_{kk} = \int \varphi_k \varphi_k = \int \varphi_k^2$ may be not well defined. (φ_{kk} may become ∞.)

Then, with inner product and norm of $\mathcal{L}^2$, observe that for any $x_i \in \mathcal{R}$,

$$||\sum_i^K x_i \varphi_i||_{\mathcal{L}^2}^2 = \langle \sum_k^K x_k \varphi_k, \sum_l^K x_l \varphi_l \rangle_{\mathcal{L}^2} = \sum_k^K \sum_l^K \langle x_k \varphi_k, x_l \varphi_l \rangle_{\mathcal{L}^2}$$

$$= \sum_k^K \sum_l^K \int x_k x_l \varphi_k(t) \varphi_l(t) dt = \sum_k^K \sum_l^K x_k x_l \int \varphi_k(t) \varphi_l(t) dt = \sum_k^K \sum_l^K x_k x_l \varphi_{kl}$$

And above norm value $||.|| \geq 0$ by definition of norm.

And incidentally, we get what we want, $\sum_k^K \sum_l^K x_k x_l \varphi_{kl} \geq 0$.

2.2 Problem 2

Show that if $\{u_j, j \geq 1\}$ and $\{v_i, i \geq 1\}$ are base in $\mathcal{L}^2([0,1])$. (not necessarily orthonormal), then

$$\{v_i(s) u_j(t), 0 \leq s, t \leq 1, i, j \geq 1\}$$

is a basis in $\mathcal{L}^2([0,1] \times [0,1])$.
Show that if $\{u_j, j \geq 1\}$ and $\{v_i, i \geq 1\}$ are both orthonormal systems, then above equation is an orthonormal system as well.

I start with some comments. First, it seems that there are many methods to solve this problem, and considering tensor-product space is one of them. But I don't choose the way because I think that it seems to use this one thing's result for proving original one thing. Instead, I try to direct proof for this problem.

Second, I use Fubini's theorem on Lebesgue measurable function, famous and elementary one in Lebesgue integration theory, but not having been dealt with in our course. So I may have to prove it to use, but because proof of the theorem is too long to bring this report, I only write down the statements of the theorem.

Theorem (Fubini's). *Suppose $f(x,y)$ is in $\mathcal{L}^1$ on $\mathcal{R}^{d_1} \times \mathcal{R}^{d_2}$. Then for almost every $y \in \mathcal{R}^{d_2}$:*

- *for fixed y, the slice f^y is in $\mathcal{L}^1(\mathcal{R}^{d_1})$, such that $f^y(x) = f(x,y)$.*
- *The function defined by $\int_{\mathcal{R}^{d_1}} f^y(x) dx$ is in $\mathcal{L}^2(\mathcal{R}^{d_2})$.*
- $\int_{\mathcal{R}^{d_2}} (\int_{\mathcal{R}^{d_1}} f(x,y) dx) dy = \int_{\mathcal{R}^{d_1} \times \mathcal{R}^{d_2}} f$

Note that if replace $\mathcal{R}$ with $[0,1] \subset \mathcal{R}$, above theorem holds.

Then let's start to solve this problem.

Since $\mathcal{L}^2$ is separable Hilbert space, I need to show only that for any $f \in \mathcal{L}^2([0,1] \times [0,1])$, f has an expression of linear combination of $\{v_i(s) u_j(t)\}$.

Although the direction of this problem says that each basis are not necessarily orthonormal respectively, we already know well that there exists 1-1 correspond linear transformation from $\mathcal{H}$ to $\mathcal{H}$ between given non-orthonormal basis and newly given orthonormal basis. (For detail, if need to make orthonormal basis newly, using gram-schmidt process, we can get (may be infinite but theoretically have no problem) linear-equation system between non-orthonormal and also get orthonormal basis and linear transformation which maps them. if there is other given orthonormal system, just project original non-orthonormal basis to them. Projection operator's uniqueness are guaranteed.) So without any loss of generality, I assume that $\{u_j, j \geq 1\}$ and $\{v_i, i \geq 1\}$ are orthonormal basis in $\mathcal{L}^2([0,1])$ for simplicity of the proof.

Let $f \in \mathcal{L}^2([0,1] \times [0,1])$ and write f as $f(t,s)$ using variable of $t, s \in [0,1]$ respectively. Note that, since the domain of f is finite measure space, if $f \in \mathcal{L}^2$, automatically $f \in \mathcal{L}^1$. Then, by fix t such that the slice f^t of f is in $L^1([0,1]) \cap \mathcal{L}^2([0,1])$, get $f^t(s)$, using Fubini's theorem (the first dot statement guarantees that almost every t, f^t satisfying $\mathcal{L}^1$ condition) and the fact that originally $f \in \mathcal{L}^2([0,1] \times [0,1])$. Then, since f^t in

$L^2([0,1])$, using parseval's identity with second coordinate's orthonormal basis $\{v_i, i \geq 1\}$, we get expression like

$$f^t(s) = \sum_{i \geq 1} a_i(t) v_i(s)$$

where $a_i(t) = \langle f^t, v_i \rangle = \int_{[0,1]} f^t(s) v_i(s) ds$. Note that the coefficients are depend on t value.

And observe that

$$||a_i(t)||^2_{\mathcal{L}^2} = \int_{[0,1]} |\int_{[0,1]} f^t(s) v_i(s) ds|^2 dt \leq \int_{[0,1]} (\int_{[0,1]} |f^t(s) v_i(s)| ds)^2 dt$$

$$\leq \int_{[0,1]} (||f^t(s)||^{1/2}_{\mathcal{L}^2([0,1])} ||v_i(s)||^{1/2}_{\mathcal{L}^2([0,1])})^2 dt = ||f^t(s)||_{\mathcal{L}^2([0,1])} ||v_i(s)||_{\mathcal{L}^2([0,1])} < \infty$$

so $a_i(t)$ in $\mathcal{L}^2([0,1])$. Then apply parseval's identity with first coordinate's orthonormal basis $\{u_j, j \geq 1\}$ to $a_i(t)$, get $a_i(t) = \sum_{j \geq 1} b_j u_j$ where $b_j = \langle a_i, u_j \rangle$. then we get following expression

$$f^t(s) = \sum_{i \geq 1} a_i(t) v_i(s) = \sum_{i \geq 1} \sum_{j \geq 1} b_j u_j(t) v_i(s)$$

that we want.

And since $||u_j(t) v_i(s)||_{\mathcal{L}^2([0,1] \times [0,1])} = ||u_j(t)||_{\mathcal{L}^2([0,1])} ||v_i(s)||_{\mathcal{L}^2([0,1])} = 1$, $\{u_j(t) v_i(s)\}$ is in $\mathcal{L}^2([0,1] \times [0,1])$ and by above result, becomes basis of $\mathcal{L}^2([0,1] \times [0,1])$.

Let's verify $\{u_j(t) v_i(s)\}$ are orthonormal. At just above, we see $||u_j(t) v_i(s)||_{\mathcal{L}^2([0,1] \times [0,1])} = 1$. And observe that using the Fubini theorem (at second equality),

$$\langle u_j v_i, u_k v_l \rangle_{\mathcal{L}^2([0,1] \times [0,1])} = \int_{[0,1] \times [0,1]} u_j v_i u_k v_l = \int_{[0,1]} \int_{[0,1]} u_j(t) v_i(s) u_k(t) v_l(s) dt ds$$

$$= \int_{[0,1]} v_i(s) v_l(s) (\int_{[0,1]} u_j(t) u_k(t) dt) ds = (\int_{[0,1]} v_i(s) v_l(s) ds)(\int_{[0,1]} u_j(t) u_k(t) dt)$$

$$= \langle u_j, u_k \rangle_{\mathcal{L}^2([0,1])} \langle v_i, v_l \rangle_{\mathcal{L}^2([0,1])}$$

Since $\{u_j\}, \{v_i\}$ are orthonormal basis,

$$\langle u_j v_i, u_k v_l \rangle_{\mathcal{L}^2([0,1] \times [0,1])} = \begin{cases} 1, & \text{if } j = k, i = l \\ 0, & \text{otherwise} \end{cases}$$

So $\{u_j(t) v_i(s)\}$ are orthonormal basis.

3 Chapter 6

3.1 Problem 5

Assume Y_n are independent Bernoulli random variables with mean $E[Y_n] = p_n = logit^{-1}(X_n^T \beta)$ and variance $Var(Y_n) = p_n(1 - p_n)$, as in Example 6.1.2. Find the estimating equation (6.6), i.e. replace μ etc with their corresponding values.

I will use the notation of chapter 6.1 of book, especially of example 6.1.2's and following material's.

Since $E[Y_n] = p_n = logit^{-1}(X_n^T \beta) = \frac{e^{X_n^T \beta}}{1 + e^{X_n^T \beta}}$, put $\theta_n = logit(p_n) = logit(logit^{-1}(X_n^T \beta)) = X_n^T \beta$ to get canonical form of distribution. And because $Y_n \sim Ber(p_n) = Bin(1, p_n)$, if we continue to follow the book's expression (6.2) for exponential family, i.e. $f(y|\theta, \phi) = exp\{\frac{\theta y - b(\theta)}{a(\phi)} + c(y, \phi)\}$, we get $a(\phi) = 1$ and $b(\theta_n) = log(1 + e^{\theta_n})$. (For more detail, see Example 6.1.2 considering $n = 1$ case.) Then $\mu = b'(\theta_n) = \frac{e^{\theta_n}}{1 + e^{\theta_n}}$ and $g^{-1} = b'$.

Then, from the log-likelihood function $l(\theta(\beta))$ of distributions in exponential family, the estimation equation of this model becomes

$$\frac{\partial l(\theta(\beta))}{\partial \beta} = \sum_{n=1}^{N} \frac{\partial \theta_n}{\partial \beta} \frac{Y_n - b'(\theta_n)}{a(\phi)} = 0$$

and by plugging above things,

$$\sum_{n=1}^{N} (\frac{\partial}{\partial \beta}(X_n^T \beta))(Y_n - \frac{e^{\theta_n}}{1+e^{\theta_n}}) = 0$$

$$\sum_{n=1}^{N} X_n (Y_n - \frac{e^{X_n^T \beta}}{1+e^{X_n^T \beta}}) = 0$$

The last equation is what we want. (In practice, find β satisfying this equation numerically.)

4 Chapter 6

4.1 Problem 6

Consider a Gaussian process $Z(t)$ in $\mathcal{L}^2([0,1])$ with mean 0 and covariance C. Suppose we also have a second process $X(t) := \mu(t) + Z(t)$. Let $v_j(t)$ be the eigenfunctions of C and λ_j the eigenvalues.

a. Write down the joint density of $\{\langle Z, v_1\rangle, ..., \langle Z, v_m\rangle\}$ for some fixed $m \in \mathcal{N}$. Write down the joint density of $\{\langle X, v_1\rangle, ..., \langle X, v_m\rangle\}$.

For notational convenience, I omit .(t) for expression of functions.

Since $Z \sim N(0, C)$, by definition of functional distribution in weak sense, $\langle Z, x\rangle \sim N(\langle 0, x\rangle, \langle C(x), x\rangle)$ for all $x \in \mathcal{H} = \mathcal{L}^2([0,1])$. So, with eigenfunctions $\{v_i\}$, we get multivariate normal distribution for

$$\{\langle Z, v_i\rangle\}_{i=1,2,...,m} \sim Normal_m(\begin{bmatrix} \langle 0, v_1\rangle \\ \langle 0, v_2\rangle \\ ... \\ \langle 0, v_m\rangle \end{bmatrix}, \begin{bmatrix} \langle C(v_1), v_1\rangle & \langle C(v_1), v_2\rangle & ... & \langle C(v_1), v_m\rangle \\ \langle C(v_2), v_1\rangle & \langle C(v_2), v_2\rangle & ... & \langle C(v_2), v_m\rangle \\ ... & & & \\ \langle C(v_m), v_1\rangle & \langle C(v_m), v_2\rangle & ... & \langle C(v_m), v_m\rangle \end{bmatrix})$$

For simplicity, denote the covariance matrix as Σ_m. Then we simply write above as

$$\{\langle Z, v_i\rangle\}_{i=1,2,...,m} \sim Normal_m(0, \Sigma_m)$$

And the pdf of multivariate normal is well known. If denote $\{\langle Z, v_i\rangle\}_{i=1,2,...,m}$ as vector z_m, then

$$f_m(z_m) = \frac{1}{(\sqrt{2\pi})^m det(\Sigma_m)} exp(-\frac{1}{2}(z_m - 0)^T \Sigma_m^{-1}(z_m - 0))$$

Likewise, since $X := \mu + Z$,

$$\{\langle X, v_i\rangle\}_{i=1,2,...,m} \sim Normal_m(\begin{bmatrix} \langle \mu, v_1\rangle \\ \langle \mu, v_2\rangle \\ ... \\ \langle \mu, v_m\rangle \end{bmatrix}, \begin{bmatrix} \langle C(v_1), v_1\rangle & \langle C(v_1), v_2\rangle & ... & \langle C(v_1), v_m\rangle \\ \langle C(v_2), v_1\rangle & \langle C(v_2), v_2\rangle & ... & \langle C(v_2), v_m\rangle \\ ... & & & \\ \langle C(v_m), v_1\rangle & \langle C(v_m), v_2\rangle & ... & \langle C(v_m), v_m\rangle \end{bmatrix})$$

with denoting the mean vector as μ_m covariance matrix as Σ_m. Then we simply write above as

$$\{\langle X, v_i\rangle\}_{i=1,2,...,m} \sim Normal_m(\mu_m, \Sigma_m)$$

and pdf

$$f_m(x_m) = \frac{1}{(\sqrt{2\pi})^m det(\Sigma_m)} exp(-\frac{1}{2}(x_m - \mu_m)^T \Sigma_m^{-1}(x_m - \mu_m))$$

where $\{\langle X, v_i \rangle\}_{i=1,2,\ldots,m}$(as vector) $= x_m$

b. You can obtain the density of $\{\langle X, v_i \rangle\}$ with respect to $\{\langle Z, v_i \rangle\}$, by taking their ratio. Write down this ratio.

By direct calculation for ratio of two pdfs of (a) with input common Y_m, we get

$$exp(-\frac{1}{2}((Y_m - \mu_m)^T \Sigma_m^{-1} (Y_m - \mu_m) - Y_m^T \Sigma_m^{-1} Y_m))$$

$$= exp(-\frac{1}{2}(-\mu_m^T \Sigma_m^{-1} Y_m - Y_m^T \Sigma_m^{-1} \mu_m + \mu_m^T \Sigma^{-1} \mu_m))$$

c. Suppose you tried to take the limit $m \to \infty$ of the ratio you obtained in (b). What requirement on μ do you need to ensure the limit exist and is finite?

Simply we need $Y_m^T \Sigma_m^{-1} \mu_m < \infty$ for all $Y \in \mathcal{R}^m$ as $m \to \infty$. For detail, since we are assuming that C is proper covariance operator, the eigenvalue λ_m of C vanishes as m goes to ∞. So if $Y_m^T \mu_m$ is bounded as $m \to \infty$, we can ensure the existence of the limit.

d. Based on the above, form a hypothesis about when the distribution of X is orthogonal/equivalent to the distribution of Z.

Since $X(t) := \mu(t) + Z(t)$ in this problem's setting, $E(XZ) = Cov(X, Z) + E(X)E(Z) = Cov(X, Z)$ in weak sense (for simplicity, I omit overt $\langle .x \rangle, x \in \mathcal{L}^2([0, 1])$ notation.) could not be 0 except the degenerate case. So We cannot form a hypothesis about orthogonal conditions.

For equivalence condition, $\langle \Sigma_m Y_m, \mu_m \rangle = 0$ for all Y_m and for all $m \in \mathcal{N}$ Or, at limit as $m \to \infty$, $\langle Cy, \mu \rangle = 0$ for all $y \in \mathcal{L}^2([0, 1])$. Or more simply, $\mu = 0$.