Chapter 5: Two sample inference for the mean & covariance functions

5.1 Equality of mean functions

Goal: Given two samples (The samples can have different structures):

$X_1, \dots, X_N$ & $X_1^*, \dots, X_M^*$

$X_i(t) = \mu(t) + \varepsilon_i(t) \qquad 1 \le i \le N$

$X_i^*(t) = \mu^*(t) + \varepsilon_i^*(t) \qquad 1 \le i \le M$

• error functions can be different

We want to test:

$H_0: \mu = \mu^*$ in L^2

$H_A: \mu \neq \mu^*$

Assumptions:

(i) $X_1, \dots, X_N$ & $X_1^*, \dots, X_M^*$ are independent

(ii) $\varepsilon_1, \dots, \varepsilon_N \overset{iid}{\sim}$ $E[\varepsilon_1(t)] = 0$ & $E\|\varepsilon_1\|^4 < \infty$

(iii) $\varepsilon_1^*, \dots, \varepsilon_M^* \overset{iid}{\sim}$ $E[\varepsilon_i^*(t)] = 0$ & $E\|\varepsilon_i^*\|^4 < \infty$

* Don't need to assume ε_i & ε_i^* follow the same distribution

Method I:

• $\bar{X}_N(t) = \frac{1}{N}\sum_{i=1}^N X_i(t)$ & $\bar{X}_M^*(t) = \frac{1}{M}\sum_{i=1}^M X_i^*(t)$ are unbiased estimators of $\mu(t)$ & $\mu^*(t)$, respectively

So a natural statistic would be:

$$U_{N,M} = \frac{NM}{N+M} \int_0^1 \{\bar{X}_N(t) - \bar{X}_M^*(t)\}^2 \, dt$$

where we reject H_0 if $U_{N,M}$ is large

Finally by continuity thm

$$\bar{X}_N(t) - \bar{X}^*_M(t) = \sqrt{\frac{M}{N+M}} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i(t) - \mu(t)) - \sqrt{\frac{N}{N+M}} \frac{1}{\sqrt{M}} \sum_{i=1}^{M} (X_i^*(t) - \mu^*(t))$$

$$\xrightarrow{D} \Gamma(t) = \sqrt{1-\theta}\, \Gamma_1(t) - \sqrt{\theta}\, \Gamma_2(t)$$

where Γ is a Gaussian process w/ zero mean &

Covariance $(1-\theta)\, c(t,s) + \theta\, c^*(t,s)$

So $U_{N,M} \xrightarrow{D} \int_0^1 \Gamma^2(t)\, dt$

Thm 5.1

If assumptions (i - iii) hold, $\frac{N}{N+M} \to \theta$ $\quad 0 \le \theta \le 1$ & H_0 holds, then

$$U_{N,M} \xrightarrow{D} \int_0^1 T^2(t)\,dt$$

where $\{T(t) : 0 \le t \le 1\}$ is a Gaussian process such that

- $E[T(t)] = 0$
- $E[T(t)T(s)] = (1-\theta)\,c(t,s) + \theta\, c^*(t,s)$

$\qquad c(t,s) = Cov[X_1(t), X_1(s)] \qquad c^*(t,s) = Cov[X_i^*(t), X_i^*(s)]$

Pf:

By assumption (i), $X_1, \dots, X_N$ & $X_1^*, \dots, X_M^*$ are independent samples

$\Rightarrow \frac{1}{\sqrt{N}} \sum_{i=1}^N (X_i - \mu)$ & $\frac{1}{\sqrt{M}} \sum_{i=1}^M (X_i^* - \mu^*)$ are also independent

$$U_{N,M} = \frac{NM}{N+M} \int_0^1 \{\bar{X}_N(t) - \bar{X}_M(t)\}^2 dt \overset{(*)}{=} \frac{NM}{N+M} \int_0^1 \{(\bar{X}_N(t) - \mu(t)) - (\bar{X}_M^*(t) - \mu^*(t))\}^2 dt$$

(*) if H_0 holds $\mu(t) = \mu^*(t)$

$$= \int_0^1 \left\{ \sqrt{\tfrac{M}{N+M}} \tfrac{1}{\sqrt{N}} \sum_{i=1}^N (X_i(t) - \mu(t)) - \sqrt{\tfrac{N}{N+M}} \tfrac{1}{\sqrt{M}} \sum_{i=1}^M (X_i^*(t) - \mu^*(t)) \right\}^2 dt$$

By Thm 2.1 $\quad \frac{1}{\sqrt{N}} \sum_{i=1}^N (X_i(t) - \mu(t)) \xrightarrow{D} T_1(t)$

$\qquad \frac{1}{\sqrt{M}} \sum_{i=1}^M (X_i^*(t) - \mu^*(t)) \xrightarrow{D} T_2(t)$

where T_1 & T_2 are independent Gaussian processes with 0 means & covariances C & C^*

By assumption $\sqrt{\frac{M}{N+M}} \to \sqrt{(1-\theta)}$ & $\sqrt{\frac{N}{N+M}} \to \sqrt{\theta}$

Therefore by Slutsky's Thm

$$\begin{pmatrix} \sqrt{\frac{M}{N+M}} \frac{1}{\sqrt{N}} \sum_{i=1}^N (X_i(t) - \mu(t)) \\ \sqrt{\frac{N}{N+M}} \frac{1}{\sqrt{M}} \sum_{i=1}^M (X_i^*(t) - \mu^*(t)) \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \sqrt{1-\theta}\, T_1(t) \\ \sqrt{\theta}\, T_2(t) \end{pmatrix}$$

Thm 5.2 If assumptions (i - iii) hold, $\frac{N}{N+M} \to \theta$ $0 \le \theta \le 1$

and $\int_0^1 \{\mu(t) - \mu^*(t)\}^2 dt > 0$

Then $U_{N,M} \xrightarrow{P} \infty$

Pf:

$$U_{N,M} = \frac{NM}{N+M}\int_0^1 \{\bar{X}_N(t) - \bar{X}_M(t)\}^2 dt = \int_0^1 \Big\{\Big(\sqrt{\tfrac{M}{N+M}}\tfrac{1}{\sqrt{N}}\sum_{i=1}^N (X_i(t) - \mu(t)) + \sqrt{\tfrac{MN}{N+M}}\mu(t)\Big) - \Big(\sqrt{\tfrac{N}{M+N}}\tfrac{1}{\sqrt{M}}\sum_{i=1}^M (X_i^*(t) - \mu^*(t)) + \sqrt{\tfrac{MN}{N+M}}\mu^*(t)\Big)\Big\}^2 dt$$

$$= \int_0^1 \Big\{\Big(\sqrt{\tfrac{M}{N+M}}\tfrac{1}{\sqrt{N}}\sum_{i=1}^N (X_i(t) - \mu(t)) - \sqrt{\tfrac{N}{M+N}}\tfrac{1}{\sqrt{M}}\sum_{i=1}^M (X_i^*(t) - \mu^*(t))\Big) + \Big(\sqrt{\tfrac{MN}{N+M}}(\mu(t) - \mu^*(t))\Big)\Big\}^2 dt$$

$$= \int_0^1 \Big[\Big\{\sqrt{\tfrac{M}{N+M}}\tfrac{1}{\sqrt{N}}\sum_{i=1}^N (X_i(t) - \mu(t)) - \sqrt{\tfrac{N}{M+N}}\tfrac{1}{\sqrt{M}}\sum_{i=1}^M X_i^*(t) - \mu^*(t)\Big\}^2 + \Big\{\sqrt{\tfrac{M}{N+M}}\tfrac{1}{\sqrt{N}}\sum_{i=1}^N (X_i(t) - \mu(t)) - \sqrt{\tfrac{N}{N+M}}\tfrac{1}{\sqrt{M}}\sum_{i=1}^M (X_i^*(t) - \mu^*(t))\Big\}\Big(\sqrt{\tfrac{MN}{N+M}}(\mu(t) - \mu^*(t))\Big)\Big] dt$$

$$+ \int_0^1 \frac{MN}{M+N}(\mu(t) - \mu^*(t))^2 dt$$

$$= \underbrace{\frac{MN}{M+N}}_{\to\infty} \underbrace{\int_0^1 \{\mu(t) - \mu^*(t)\}^2 dt}_{>0} + O_p(1)$$

Thus $U_{N,M} \longrightarrow \infty$

How do we approximate $\int_0^1 T^2(t)\,dt$?

By Karhunen-Loève expansion

$$T(t) = \sum_{k=1}^{\infty} \tau_k^{1/2} N_k \phi_k(t)$$

where
- $N_k \overset{iid}{\sim} N(0,1)$
- $\left.\begin{matrix} \tau_1 \geq \tau_2 \geq \cdots \\ \phi_1, \phi_2, \ldots \end{matrix}\right\}$ are the eigenvalues & eigen functions of the operator determined by $c(1-\theta) + c^* \theta$

$$\Rightarrow \int_0^1 T^2(t)\,dt = \int_0^1 \left\{ \sum_{k=1}^{\infty} \tau_k^{1/2} N_k \phi_k(t) \right\}^2 dt$$

$$= \int_0^1 \left[\tau_1 N_1^2 \phi_1^2(t) + \tau_1 \tau_2 N_1 N_2 \phi_1(t) \phi_2(t) + \cdots \right] dt$$

$$= \tau_1 N_1^2 \underbrace{\int_0^1 \phi_1^2(t)\,dt}_{=1} + \tau_1 \tau_2 N_1 N_2 \underbrace{\int_0^1 \phi_1(t) \phi_2(t)\,dt}_{=0} + \cdots$$

$$= \sum_{k=1}^{\infty} \tau_k N_k^2$$

* So to approximate $\int_0^1 T^2(t)\,dt$, we only need to estimate τ_k's

We can use the estimated eigenvalues, $\hat{\tau}_k$'s, of the empirical covariance function

$$\hat{Z}_{N,M}(t,s) = \frac{M}{M+N} \frac{1}{N} \sum_{i=1}^{N} \{X_i(t) - \bar{X}_N(t)\}\{X_i(s) - \bar{X}_N(s)\} + \frac{N}{M+N} \frac{1}{M} \sum_{i=1}^{M} \{X_i^*(t) - \bar{X}_M^*(t)\}\{X_i^*(s) - \bar{X}_M^*(s)\}$$

Then $\int_0^1 T^2(t)\,dt \approx \sum_{k=1}^{d} \hat{\tau}_k N_k^2$ for d "large enough"

Thus combining this w/ Thm's S.1 & S.2, method I is asymptotically consistent

5.4 Equality of Covariance Operators

Let $X_1, \dots, X_N$ & $X_1^*, \dots, X_M^*$ be two independent samples of functions that are iid mean 0 elements of L^2

Let $C(x) = E[\langle X, x\rangle X]$ & $C^*(x) = E[\langle X^*, x\rangle X^*]$ be the covariance operators

Goal: Test $H_0: C = C^*$ vs. $H_A: C \neq C^*$

If we assume X & X^* are Gaussian then this test also tests ~~the~~ equality in distribution

Let $\hat{R}$ be the empirical covariance operator of the pooled data

$$\hat{R}(x) = \frac{1}{N+M}\left[\sum_{i=1}^{N} \langle X_i, x\rangle X_i + \sum_{j=1}^{M} \langle X_j^*, x\rangle X_j^*\right]$$

$$= \hat{\theta}\,\hat{C}(x) + (1-\hat{\theta})\,\hat{C}^*(x) \qquad \hat{\theta} = \frac{N}{N+M}$$

Then $\hat{R}$ has $N+M$ eigenfunctions, $\tilde{\phi}_k$

Let $\hat{\lambda}_k = \frac{1}{N}\sum_{n=1}^{N} \langle X_n, \hat{\phi}_k\rangle^2 \qquad \hat{\lambda}_k^* = \frac{1}{M}\sum_{m=1}^{M} \langle X_m^*, \hat{\phi}_k\rangle^2$

be the sample variances of the coefficients of X & X^* w/ respect to the orthonormal system $\{\hat{\phi}_k : 1 \le k \le N+M\}$

$$\hat{T} = \frac{N+M}{2}\,\hat{\theta}(1-\hat{\theta}) \sum_{i,j=1}^{p} \frac{\langle (\hat{C}-\hat{C}^*)\hat{\phi}_i, \hat{\phi}_j\rangle^2}{(\hat{\theta}\hat{\lambda}_i + (1-\hat{\theta})\hat{\lambda}_i^*)(\hat{\theta}\hat{\lambda}_j + (1-\hat{\theta})\hat{\lambda}_j^*)}$$

Thm 5.5 Suppose X & X^* are Gaussian elements of L^2 such that $E\|X\|^4 < \infty$ & $E\|X^*\|^4 < \infty$. Also suppose $\hat{\theta} \to \theta \in (0,1)$ as $N \to \infty$

Then:

$$\hat{T} \xrightarrow{D} \chi^2_{\frac{p(p+1)}{2}} \qquad N, M \to \infty$$

* So we can compare $\chi^2_{p(p+1)/2}$ critical values to $\hat{T}$ to test $H_0: C = C^*$

Method II

- We use projections onto the space determined by the leading eigenfunctions of the operator $Z = (1-\theta)C + \theta C^*$.

- We assume that eigenvalues of Z,

$$\tau_1 > \tau_2 > \cdots > \tau_d > \tau_{d+1}$$

and

corresponding eigenfunctions, $\varphi_1 \cdots \varphi_{d+1}$.

- We want to project the obs. onto the space spanned by $\varphi_1 \cdots \varphi_d$.

Since functions are unknown, we project $\bar{X}_N - \bar{X}_M^*$ into the linear space spanned by $\hat{\varphi}_1 \cdots \hat{\varphi}_d$.

Let $\hat{a}_i = \langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_i \rangle \quad 1 \le i \le d$

$$\hat{a} = (\hat{a}_1 \cdots \hat{a}_d)^T.$$

- Goal: under the conditions of Theorem 5.1,

$\sqrt{\dfrac{N \cdot M}{N+M}} \cdot \hat{a}$ is approximately d-variate normal up to some random signs.

Method II.

The asymptotic variance of $\sqrt{\frac{N \cdot M}{N+M}} \cdot \hat{a}$ is

$$Q(i,j) = (1-\theta)E\langle X_1-\mu, \varphi_i\rangle\langle X_1-\mu, \varphi_j\rangle + \theta E\langle X_i^*-\mu^*, \varphi_i\rangle\langle X_i^*-\mu^*, \varphi_j\rangle$$

$$= \int_0^1\int_0^1 (1-\theta)E\left[(X_1(t)-\mu(t))\varphi_i(t)(X_1(s)-\mu(s))\varphi_j(s)\right]dt\,ds$$

$$+ \int_0^1\int_0^1 \theta \cdot E\left[(X_i^*(t)-\mu^*(t))\varphi_i(t)(X_i^*(s)-\mu^*(s))\varphi_j(s)\right]dt\,ds$$

$$= \int_0^1\int_0^1 \left((1-\theta)C(s,t) + \theta C^*(s,t)\right)\varphi_i(t)\varphi_j(s)\,dt\,ds$$

$$= \int_0^1\int_0^1 Z(t,s)\varphi_i(t)\varphi_j(s)\,dt\,ds$$

$$= \int_0^1\int_0^1 \left(\sum_{K=1}^{\infty} \tau_K \cdot \varphi_K(t)\varphi_K(s)\right)\varphi_i(t)\varphi_j(s)\,dt\,ds$$

$$= \int_0^1\int_0^1 \sum_{K=1}^{\infty} \tau_K \varphi_K(t)\varphi_i(t)\varphi_K(s)\varphi_j(s)\,dt\,ds$$

$$= \sum_{K=1}^{\infty} \tau_K \int_0^1 \varphi_K(t)\varphi_i(t)\,dt \int_0^1 \varphi_K(s)\varphi_j(s)\,ds.$$

$$= \begin{cases} \tau_i & \text{if } i=j \\ 0 & \text{if } i\neq j \end{cases}$$

$$\Rightarrow \quad T_{N,M}^{(1)} = \frac{N \cdot M}{N+M}\sum_{K=1}^{d} \hat{a}_K^2 / \hat{\tau}_K$$

$$T_{N,M}^{(2)} = \frac{N \cdot M}{N+M}\sum_{K=1}^{d} \hat{a}_K^2$$

$\boxed{\text{Theorem 5.3}}$

If H_0 and (5.3) – (5.5), (5.6) and (5.9) holds,

then $T^{(1)}_{N,M} \xrightarrow{d} \chi^2(d)$.

$T^{(2)}_{N,M} \xrightarrow{d} \sum_{K=1}^{d} \tau_K \cdot N_K^2$, N_i's indep. stnd. normal random variables.

proof)

For proof, we need several steps.

By CLT for sums of iid random vectors in $\mathbb{R}^d$,

$$\sqrt{\frac{N \cdot M}{N+M}}\, a \triangleq \sqrt{\frac{N \cdot M}{N+M}} \left[\langle \bar{X}_N - \bar{X}_M^*, \varphi_1 \rangle, \cdots \langle \bar{X}_N - \bar{X}_M^*, \varphi_d \rangle \right]^T \quad \cdots\cdots \text{①}$$

$$\xrightarrow{d} N_d(0, Q).$$

By Lemma 2.2 and Lemma 2.3,

Lemma 2.2 $\Rightarrow$ $|\hat{\tau}_i - \tau_i| \leq \|\hat{Z} - Z\|_{\mathcal{L}}$ $\forall i$, (2.3)

$\Rightarrow \max_i |\hat{\tau}_i - \tau_i| \leq \|\hat{Z} - Z\|_{\mathcal{L}} \leq \|\hat{Z} - Z\|_S$

Recall (2.3): $\|\psi\|_{\mathcal{L}} \leq \|\psi\|_S$

Lemma 2.3 $\Rightarrow$ $\|\hat{\varphi}_i - \hat{c}_i \varphi_i\| \leq \|\hat{Z} - Z\|_{\mathcal{L}} \leq \|\hat{Z} - Z\|_S$.

Since $\|\hat{Z} - Z\|_S = \iint (\hat{Z}_{N,M}(t,s) - Z(t,s))^2 \, dt\, ds = O_p(1)$,

$\therefore \max_i |\hat{\tau}_i - \tau| = o_p(1) \quad \cdots\cdots \text{②}$

$\max |\hat{\varphi}_i - \hat{c}_i \varphi_i\| = o_p(1) \quad \cdots\cdots \text{③}$

Since

$N \int_0^1 (\bar{X}_N(t) - \mu(t))^2 dt = O_p(1)$, $M \int_0^1 (\bar{X}_M(t) - \mu^*(t))^2 dt = O_p(1)$. $\cdots$ ④

$\max_i \left(\frac{NM}{N+M}\right)^{1/2} |\langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_i - \hat{c}_i \varphi_i \rangle| = o_p(1)$. by ③ & ④

So,

$$\sqrt{\frac{N \cdot M}{N+M}}\, \hat{a} \xrightarrow{P} \sqrt{\frac{N \cdot M}{N+M}}\, a \xrightarrow{d} N_d(0, Q) \Rightarrow \left[\begin{array}{l} T^{(2)}_{N,M} \xrightarrow{d} \sum_{K=1}^{d} \tau_K \cdot N_K^2 \\ \\ T^{(1)}_{N,M} \xrightarrow{d} \chi^2_d \text{ by ② } \hat{\tau}_i \xrightarrow{P} \tau_i \end{array} \right.$$

§5.4 Equality of covariance operators

- $\{X_1 \cdots X_N\}$ & $\{X_1^* \cdots X_M^*\}$: iid mean zero elements of $\mathcal{L}^2$ and indep.

  $C(x) = E[\langle X, x\rangle X]$, $C^*(x) = E[\langle X^*, x\rangle X^*]$

  $H_0 : C = C^*$ versus $H_A : C \neq C^*$

- Assume that X and X^* are Gaussian elements of $\mathcal{L}^2$.

  $\Rightarrow$ $C = C^*$ implies the equality in distribution

  $\Rightarrow$ under the additional assumption of NORMALITY,

  H_0 states that X_i have the same distribution as the X_j^*.

- $\hat{R}(x)$: empirical cov. operator of the pooled data.

  $$\hat{R}(x) = \frac{1}{N+M}\left\{\sum_{i=1}^{N}\langle X_i, x\rangle X_i + \sum_{j=1}^{M}\langle X_j^*, x\rangle X_j^*\right\}$$

  $$= \hat{\theta}\hat{C}(x) + (1-\hat{\theta})\hat{C}^*(x), \quad x \in \mathcal{L}^2$$

  where $\hat{\theta} = \dfrac{N}{N+M}$

- $\hat{R} \Rightarrow \{\hat{\varphi}_k, \ 1 \leq k \leq N+M\}$ eigenfunctions

  $$\hat{\lambda}_k = \frac{1}{N}\sum^{N}\langle X_n, \hat{\varphi}_k\rangle^2$$

  $$\hat{\lambda}_k^* = \frac{1}{M}\sum_{m=1}^{M}\langle X_m^*, \hat{\varphi}_k\rangle^2$$

Remark:

$\hat{\lambda}_k$ and $\hat{\lambda}_k^*$ are NOT the eigenvalues of $\hat{C}$ and $\hat{C}^*$.

Theorem 5.5

Suppose X and X^* are Gaussian elements of $\mathcal{L}^2$ s.t

$E\|X\|^4 < \infty$ and $E\|X^*\|^4 < \infty$. Suppose also that $\hat{\theta} \to \theta \in (0,1)$, as $N \to \infty$

Then, $\hat{T} \xrightarrow{d} \chi^2_{p(p+1)/2}$, $N, M \to \infty$.

proof) Random operators,

$C_i(x) = \langle X_i, x \rangle X_i$, $C_j^*(x) = \langle X_j^*, x \rangle X_j^*$, $x \in \mathcal{L}^2$.

Under H_0, C_i and C_j^* have the same mean C and same covariance operator $\mathcal{B}(\psi)$.

$$\begin{aligned}\mathcal{B}(\psi) &= E[\langle C_i - C, \psi \rangle_S (C_i - C)] \\ &= E[\langle C_i - C, \psi \rangle_S C_i] - E[\langle C_i - C, \psi \rangle_S C] \\ &= E[\langle C_i, \psi \rangle_S C_i - \langle C, \psi \rangle_S C_i - \langle C_i, \psi \rangle_S C + \langle C, \psi \rangle_S C] \\ &= \underbrace{E[\langle C_i, \psi \rangle_S C_i]}_{\circledast_1} - \underbrace{\langle C, \psi \rangle_S C}_{\text{under } H_0\text{, same for both samples}}, \quad \psi \in \mathcal{S}\end{aligned}$$

$\circledast_1 \Rightarrow$ $E[\langle C_i, \psi \rangle_S C_i]$

$$\begin{aligned}&= E[\sum^{\infty} \langle C_i(e_n), \psi(e_n) \rangle C_i] \\ &= \sum E[\langle \langle X_i, e_n \rangle X_i, \psi(e_n) \rangle \langle X_i, e_n \rangle X_i] \\ &= \sum E[\langle X_i, e_n \rangle^2 \langle X_i, \psi(e_n) \rangle X_i]\end{aligned}$$

$\circledast$ Recall : (2.2)

$\langle \psi_1, \psi_2 \rangle_S = \sum_{i=1}^{\infty} \langle \psi_1(e_i), \psi_2(e_i) \rangle$

Since X_i and X_j^* have the same distribution under $H_0: C = C^*$,

$$E[\langle X_i, e_n \rangle^2 \langle X_i, \psi(e_n) \rangle X_i] = E[\langle X_j^*, e_n \rangle^2 \langle X_j^*, \psi(e_n) \rangle X_j^*].$$

For using Theorem 2.1 (25p), we need $E\|C_i\|_S^2 < \infty$.

Since $E\|X\|^4 < \infty$ and $E\|X^*\|^4 < \infty$, $E\|C_i\|_S^2 < \infty$ and $E\|C_i^*\|_S^2 < \infty$.

$$\left[\because\ E\|C_i\|_S^2 = E\sum_{n=1}^{\infty}\|C_i\|^2 = E\sum\|\langle X_i, e_n\rangle X_i\|^2 = E\Big[\|X_i\|^2\sum_{n=1}^{\infty}|\langle X_i, e_n\rangle|^2\Big] \underset{\text{parseval's}}{=} E\|X_i\|^4 < \infty\right]$$

By theorem 2.1,

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(C_n - C) = \frac{N}{\sqrt{N}}\cdot\frac{1}{N}\sum_{n=1}^{N}(C_n - C) = \sqrt{n}(\hat{C} - C) \xrightarrow{d} Z_1$$

Likewise, $\sqrt{n}(\hat{C}^* - C^*) \xrightarrow{d} Z_2$.

where Z_1 and Z_2 are Gaussian random element with the same covariance C_1 and C_2 respectively. Under H_0, $C_1 = C_2 = \mathcal{C}$

For every $1 \le i,j \le p$, introduce the random variables,

$$W_{N,M}(i,j) = \langle [(N+M)\hat{\theta}(1-\hat{\theta})]^{1/2}(\hat{C} - \hat{C}^*)\hat{c}_i\hat{\phi}_i,\ \hat{c}_j\hat{\phi}_j\rangle.$$

so that

$$\hat{T} = \frac{N+M}{2}\hat{\theta}(1-\hat{\theta})\sum_{i,j}^{p}\frac{\langle(\hat{C} - \hat{C}^*)\hat{\phi}_i, \hat{\phi}_j\rangle^2}{(\hat{\theta}\hat{\lambda}_i + (1-\hat{\theta})\hat{\lambda}_i^*)(\hat{\theta}\hat{\lambda}_j + (1-\hat{\theta})\hat{\lambda}_j^*)}.$$

$$= \sum_{i,j}^{p}\frac{W_{N,M}^2(i,j)}{2(\hat{\theta}\hat{\lambda}_i + (1-\hat{\theta})\hat{\lambda}_i^*)(\hat{\theta}\hat{\lambda}_j + (1-\hat{\theta})\hat{\lambda}_j^*)}$$

Under H_0,

under H_0, $C = C^*$

$$W_{N,M}(i,j) = \langle[(N+M)\hat{\theta}(1-\hat{\theta})]^{1/2}(\hat{C} - C - \hat{C}^* + C^*)\hat{c}_i\hat{\phi}_i,\ \hat{c}_j\hat{\phi}_j\rangle$$

$$= \langle\{\underbrace{[(N+M)\hat{\theta}(1-\hat{\theta})]}_{=(1-\hat{\theta})N}{}^{\frac{1}{2}}(\hat{C} - C) - [\underbrace{\ \ ''\ \ }_{=\hat{\theta}M}]^{\frac{1}{2}}(\hat{C}^* - C^*)\}\hat{c}_i\hat{\phi}_i,\ \hat{c}_j\hat{\phi}_j\rangle$$

$$\doteq \langle[\underbrace{(1-\hat{\theta})^{\frac{1}{2}}N^{\frac{1}{2}}(\hat{C} - C)}_{\xrightarrow{d} Z_1} - \underbrace{\hat{\theta}^{\frac{1}{2}}M^{\frac{1}{2}}(\hat{C}^* - C^*)}_{\to Z_2}]\underbrace{\hat{c}_i\hat{\phi}}_{\xrightarrow{P} V_i},\ \underbrace{\hat{c}_j\hat{\phi}_j}_{\xrightarrow{P} V_j}\rangle$$ by Theorem 2.7.

⊛ V_i's eigenfunction of C and C^*

$$\therefore W_{N,M}(i,j) \xrightarrow{d} \langle((1-\hat{\theta})^{\frac{1}{2}}Z_1 - \hat{\theta}^{\frac{1}{2}}Z_2)V_i, V_j\rangle = \langle ZV_i, V_j\rangle$$

Thus,

$$\hat{T} = \frac{\sum_{i,j}^{p} W_{N,M}^{2}(i,j)}{2(\hat{\theta}\hat{\lambda}_j + (1-\hat{\theta})\hat{\lambda}_i^{*})(\hat{\theta}\hat{\lambda}_j + (1-\hat{\theta})\hat{\lambda}_j^{*})}$$

$$\xrightarrow{d} \frac{\sum_{i,j}^{p} \langle Z(v_i), v_j \rangle^2}{2(\theta\lambda_i + (1-\theta)\lambda_i)(\theta\lambda_i + (1-\theta)\lambda_j)}$$

$$= \frac{\sum_{i,j}^{p} \langle Z(v_i), v_j \rangle^2}{2\lambda_i\lambda_j}$$

$$= \sum_{k=1}^{n} \frac{\langle Z(v_k), v_k \rangle^2}{2\lambda_k^2} + \sum_{k<n} \frac{\langle Z(v_k), v_n \rangle^2 + \langle Z(v_n), v_k \rangle^2}{2\lambda_k\lambda_n}$$

For conveniency, represent Z in terms of V_{ij} defined by

$$V_{ij}(x) = \langle v_i, x \rangle v_j$$

By Lemma 5.1,

$$Z = (1-\hat{\theta})^{\frac{1}{2}} Z_1 - \hat{\theta}^{\frac{1}{2}} Z_2 \overset{d}{=} \sqrt{2}\sum_{i=1}^{\infty} \lambda_i \zeta_{ii} V_{ii} + \sum_{i<j} \sqrt{\lambda_i\lambda_j}\, \zeta_{ij} (V_{ij} + V_{ji})$$

where ζ_{ij} are iid stnd. normal

※ Remark : Lemma 5.1,

Under the assumptions of Thm 5.5,

$$\mathcal{C} = \sum_{i=1}^{\infty} (\sqrt{2}\lambda_i)^2 \langle V_{ii}, \cdot \rangle_S V_{ii} + \sum_{i<j} \lambda_i\lambda_j \langle V_{ij} + V_{ji}, \cdot \rangle_S (V_{ij} + V_{ji}).$$

$$\Rightarrow Z(v_k) = \sqrt{2}\lambda_k \zeta_{kk} v_k + \sum_{k<j} \sqrt{\lambda_k\lambda_j}\, \zeta_{kj} v_j + \sum_{i<k} \sqrt{\lambda_i\lambda_k}\, \zeta_{ik} v_i$$

$$\Rightarrow \langle Z(v_k), v_n \rangle = \begin{cases} \sqrt{2}\lambda_k \zeta_{kk} & \text{if } k=n \\ \sqrt{\lambda_k\lambda_n}\, \zeta_{kn} & \text{if } k<n \\ \sqrt{\lambda_k\lambda_n}\, \zeta_{nk} & \text{if } k>n \end{cases}$$

Thus,

$$\hat{T} \xrightarrow{d} \sum_{k=1}^{p} \zeta_{kk}^2 + \sum_{k<p} \frac{\lambda_k\lambda_n\zeta_{kn}^2 + \lambda_k\lambda_n\zeta_{nk}^2}{2\lambda_k\lambda_n} = \sum_{k=1}^{p} \zeta_{kk}^2 + \sum_{k<p} \frac{\zeta_{kn}^2 + \zeta_{nk}^2}{2}$$

$$\overset{d}{=} \sum_{k=1}^{p} \zeta_{kk}^2 + \sum_{k<p} \zeta_{kn}^2 \overset{d}{=} \chi^2_{p(p+1)/2}$$

ζ_{ij} : iid stnd normal

Proof:

Let $C_i(x) = \langle X_i, x\rangle X_i$ & $C_j^* = \langle X_j^*, x\rangle X_j^*$ $x \in L^2$

So the C_i & C_j^* are sequences of iid elements in the Hilbert space S & because they are covariance operators, they are Hilbert-Schmidt operators (Chpt 2)

Under H_0, $E[C_i] = E[C_j^*] = C$ & they will have the same covariance operator:

$$G(\Psi) = E[\langle C_i - C, \Psi\rangle_S (C_i - C)] = E[\langle C_i, \Psi\rangle_S C_i] - \langle C, \Psi\rangle C \quad \Psi \in S$$

Because the space of H-S operators, S, is separable

$$E[\langle C_i, \Psi\rangle_S C_i] = E\left[\sum_{n=1}^{\infty} \langle C_i(e_n), \Psi(e_n)\rangle C_i\right]$$

$$= \sum_{n=1}^{\infty} E\left[\langle\langle X_i, e_n\rangle X_i, \Psi(e_n)\rangle \langle X_i, e_n\rangle X_i\right] \qquad C_i = \langle X_i, x\rangle X_i$$

$$= \sum_{n=1}^{\infty} E\left[\langle X_i, e_n\rangle^2 \langle X_i, \Psi(e_n)\rangle X_i\right] \qquad \langle ax, y\rangle = a\langle x, y\rangle$$

Under H_0

$$E[\langle X_i, e_n\rangle^2 \langle X_i, \Psi(e_n)\rangle X_i] = E[\langle X_j^*, e_n\rangle^2 \langle X_j^*, \Psi(e_n)\rangle X_j^*]$$

Next we want to apply the CLT to $\hat{C}$ & $\hat{C}^*$, but to do so we need to verify $E\|C_i\|_S^2 < \infty$

$$E\|C_i\|_S^2 = E\|\langle X_i, x\rangle X_i\|^2 = E\left[\sum_{n=1}^{\infty} \|\langle X_i, e_n\rangle X_i\|^2\right] = E\left[\|X_i\|^2 \underbrace{\sum_{n=1}^{\infty} |\langle X_i, e_n\rangle|^2}_{\|X_i\|^2 \text{ Parseval's' equality}}\right]$$

$$= E[\|X_i\|^2 \|X_i\|^2]$$

$$= E\|X_i\|^4 \overset{<\infty}{<\infty} \quad \text{(by assumption)}$$

Thus we can apply the CLT

So by CLT $\sqrt{N}(\hat{C} - C) \xrightarrow{D} Z_1$ $\quad \sqrt{M}(\hat{C}^* - C^*) \xrightarrow{D} Z_2$

where Z_1 & Z_2 are independent Gaussian elements of S with covariance operator G which we defined above

For all $1 \le i, j \le p$ define

$$W_{N,M}(i,j) = \langle [(N+M)\hat{\theta}(1-\hat{\theta})]^{1/2}(\hat{C} - \hat{C}^*)\hat{c}_i\hat{\phi}_i, \hat{c}_j\hat{\phi}_j \rangle$$

so
$$\hat{T} = \frac{\sum_{i,j=1}^{p} W_{N,M}^2(i,j)}{2(\hat{\theta}\hat{\lambda}_i + (1-\hat{\theta})\hat{\lambda}_i^*)(\hat{\theta}\hat{\lambda}_j + (1-\hat{\theta})\hat{\lambda}_j^*)}$$

Under H_0: $C = C^*$

$$\Rightarrow W_{N,M}(i,j) = \langle [\sqrt{1-\hat{\theta}}\sqrt{N}(\hat{C} - C) - \sqrt{\hat{\theta}}\sqrt{M}(\hat{C}^* - C^*)]\hat{c}_i\hat{\phi}_i, \hat{c}_j\hat{\phi}_j \rangle$$

By Thm 2.7 $\hat{c}_i\hat{\phi}_i \xrightarrow{P} v_i$ $\quad v_i$ = eigenfunctions of $C = C^*$

So using $\sqrt{N}(\hat{C} - C) \xrightarrow{D} Z_1$, $\sqrt{M}(\hat{C}^* - C^*) \to Z_2$

$\hat{\theta} \to \theta$

$$\Rightarrow W_{N,M}(i,j) \xrightarrow{D} \langle \sqrt{1-\theta}\, Z_1 - \sqrt{\theta}\, Z_2 v_i, v_j \rangle$$

Let $Z = \sqrt{1-\theta}\, Z_1 - \sqrt{\theta}\, Z_2$

$$\Rightarrow W_{N,M}(i,j) \xrightarrow{D} \langle Z v_i, v_j \rangle$$

Z_1 & Z_2 both have covariance operator G

$\Rightarrow Z$ also has covariance operator G

$$G_Z(\Psi) = E\left[\langle \sqrt{1-\theta}\, C_i - \sqrt{1-\theta}\, C + \sqrt{\theta}\, C_i - \sqrt{\theta}\, C, \Psi \rangle (\sqrt{1-\theta}\, C_i - \sqrt{1-\theta}\, C + \sqrt{\theta}\, C_i - \sqrt{\theta}\, C) \right]$$
$$= E[1-\theta \langle C_i - C, \Psi \rangle (C_i - C)] + E[\theta \langle C_i - C, \Psi \rangle (C_i - C)]$$
$$= (1-\theta)G(\Psi) + \theta G(\Psi)$$
$$= G(\Psi)$$

Thus
$$\hat{T} \xrightarrow{D} \frac{\sum_{i,j=1}^{p} \langle Z(v_i), v_j \rangle^2}{2(\theta\lambda_i + (1-\theta)\lambda_i)(\theta\lambda_j + (1-\theta)\lambda_j)} = \frac{\sum_{i,j=1}^{p} \langle Z(v_i), v_j \rangle^2}{2\lambda_i\lambda_j}$$

$$= \underbrace{\sum_{k=1}^{n} \frac{\langle Z(v_k), v_k \rangle^2}{2\lambda_k^2}}_{\text{Case } i=j} + \underbrace{\sum_{k<n} \frac{\langle Z(v_k), v_n \rangle^2}{2\lambda_k\lambda_n} + \sum_{k>n} \frac{\langle Z(v_k), v_n \rangle^2}{2\lambda_k\lambda_n}}_{\text{Cases } i \ne j}$$

$$= \sum_{k=1}^{n} \frac{\langle Z(v_k), v_k \rangle^2}{2\lambda_k^2} + \sum_{k<n} \frac{\langle Z(v_k), v_n \rangle^2 + \langle Z(v_n), v_k \rangle^2}{2\lambda_k\lambda_n}$$

Now define $V_{ij}(x) = \langle v_i, x \rangle v_j$

Lemma S.1 Under assumptions of Thm 5.5

$$G = \sum_{i=1}^{\infty} (\sqrt{2}\lambda_i)^2 \langle V_{ii}, \cdot \rangle_S V_{ii} + \sum_{i<j} \lambda_i \lambda_j \langle V_{ii} + V_{ji}, \cdot \rangle_S (V_{ij} + V_{ji})$$

By Lemma S.1

$$Z \overset{D}{=} \sqrt{2} \sum_{i=1}^{\infty} \lambda_i \zeta_{ii} V_{ii} + \sum_{i<j} \sqrt{\lambda_i \lambda_j}\, \zeta_{ij} (V_{ij} + V_{ji}) \qquad \zeta_{ij} \overset{iid}{\sim} N(0,1)$$

Now v_i form a basis $\Rightarrow$ when $i \neq j$ $V_{ij}(x) = 0$

$$\Rightarrow Z(v_k) = \sqrt{2}\lambda_k \zeta_{kk} v_k + \sum_{k<j} \sqrt{\lambda_k \lambda_j}\, \zeta_{kj} v_j + \sum_{i<k} \sqrt{\lambda_i \lambda_k}\, \zeta_{ik} v_i$$

$$\Rightarrow \langle Z(v_k), v_n \rangle = \begin{cases} \sqrt{2}\lambda_k \zeta_{kk} & k = n \\ \sqrt{\lambda_k \lambda_n}\, \zeta_{kn} & k < n \\ \sqrt{\lambda_k \lambda_n}\, \zeta & k > n \end{cases}$$

So finally

$$\hat{T} \xrightarrow{D} \sum_{k=1}^{p} \frac{(\sqrt{2}\lambda_k \zeta_{kk})^2}{2\lambda_k^2} + \sum_{k<p} \frac{(\sqrt{\lambda_k \lambda_n}\, \zeta_{kn})^2 + (\sqrt{\lambda_k \lambda_n}\, \zeta_{nk})^2}{2\lambda_k \lambda_n}$$

$$= \sum_{k=1}^{p} \zeta_{kk}^2 + \sum_{k<p} \frac{\zeta_{kn}^2 + \zeta_{nk}^2}{2}$$

$$\overset{D}{=} \sum_{k=1}^{p} \underbrace{\zeta_{kk}^2}_{\chi^2_1} + \sum_{k<p} \underbrace{\zeta_{kn}^2}_{\chi^2_1} \overset{D}{=} \chi^2_{\frac{p(p+1)}{2}}$$