再生核ヒルベルト空間における正規性の検定

島根大・総合理工学研究科　牧草 夏実
島根大・総合理工学研究科　数理科学領域　内藤 貫太

P を確率分布とするとき，データ $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ に基づく

$$\text{帰無仮説 } H_0 : P = N(m_0, \Sigma_0) \ \text{ v.s. } \ \text{対立仮説 } H_1 : P \neq N(m_0, \Sigma_0)$$

の検定を "正規性の検定" という．1 変量正規性の検定としては，シャピロ・ウィルク検定，コルモゴロフ・スミノフ検定，歪度と尖度によるオムニバス検定などがよく知られており，多変量でも同様に様々な検定方法が議論されている．本発表では，この正規性の検定の数理的な一般化を考え，ヒルベルト空間に値を取る確率変数に対する正規性の検定について，検定統計量を構築し，その漸近挙動について報告する．

ヒルベルト空間 $\mathcal{H}$ に値を取る確率変数 X が正規分布に従うとは，任意の $h \in \mathcal{H}$ に対して，$\mathcal{H}$ における内積 $\langle X, h \rangle_{\mathcal{H}}$ が 1 変量の正規分布の確率変数となることである．また，任意の $h, h' \in \mathcal{H}$ に対し，$\langle m, h \rangle_{\mathcal{H}} = \mathbb{E}_X \left[\langle X, h \rangle_{\mathcal{H}} \right]$，$\langle \Sigma h, h' \rangle_{\mathcal{H}} = \mathrm{cov} \left(\langle X, h \rangle_{\mathcal{H}}, \langle X, h' \rangle_{\mathcal{H}} \right)$ を満たす $m \in \mathcal{H}$ と作用素 $\Sigma : \mathcal{H} \to \mathcal{H}$ が一意に存在し，m，Σ をそれぞれ，X の期待値，共分散作用素といい，X の従う正規分布を $N(m, \Sigma)$ と表す．

$\mathcal{H}$ 上の 2 つの確率分布 P と $N(m_0, \Sigma_0)$ との違いは

$$\Delta(P, N(m_0, \Sigma_0)) = \sup_{f \in \mathcal{F}} \left| \mathbb{E}_{X \sim P} \left[f(X) \right] - \mathbb{E}_{Y \sim N(m_0, \Sigma_0)} \left[f(Y) \right] \right| \tag{1}$$

によって測られることが知られている．ここで，$\mathcal{F}$ は $\mathcal{H}$ 上の実数値関数のクラスである．characteristic kernel k の再生核ヒルベルト空間 $H(k)$ に対し，$\mathcal{B}_1(k) = \{ f \in H(k) \mid ||f||_{H(k)} \leq 1 \}$ を $H(k)$ の単位球とする．(1) の $\mathcal{F}$ として $\mathcal{B}_1(k)$ を用いることで，P と $N(m_0, \Sigma_0)$ の MMD（Maximum Mean Discrepancy）

$$\Delta(P, N(m_0, \Sigma_0)) = ||\mu(P) - \mu(N(m_0, \Sigma_0))||_{H(k)}$$

によってそれらの違いを見ることができる．ここで，$\mu(P) = \mathbb{E}_{X \sim P} \left[k(X, \cdot) \right]$，$\mu(N(m_0, \Sigma_0)) = \mathbb{E}_{X \sim N(m_0, \Sigma_0)} \left[k(X, \cdot) \right]$ は確率分布 P，$N(m_0, \Sigma_0)$ のカーネル k による埋め込みである．$\mathcal{H}$ におけるデータ $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ を用いて，$\mu(P)$ は $\widehat{\mu}(P) = (1/n) \sum_{i=1}^n k(X_i, \cdot)$ によって推定され，$(\widetilde{m}, \widetilde{\Sigma})$ を (m_0, Σ_0) の一致推定量すると，$\mu(N(m_0, \Sigma_0))$ は $\mu(N(\widetilde{m}, \widetilde{\Sigma}))$ によって推定される．したがって，$\Delta^2 = ||\mu(P) - \mu(N(m_0, \Sigma_0))||_{H(k)}^2$ は $\widehat{\Delta}^2 = \left|\left| \widehat{\mu}(P) - \mu(N(\widetilde{m}, \widetilde{\Sigma})) \right|\right|_{H(k)}^2$ によって推定される．さらに，$\mu(N(m, \Sigma))$ が (m, Σ) に関して連続であるとき，$\widehat{\Delta}^2$ は Δ^2 の一致推定量となっている．

本発表では特に，帰無仮説 $H_0 : P = N(m_0, \Sigma_0)$ のもとで，$n\widehat{\Delta}^2$ と漸近同等となる統計量を導き，退化 U 統計量の結果に帰着させることで $n\widehat{\Delta}^2$ の漸近分布が得られることについて報告する．