再生核ヒルベルト空間における正規性の検定の非正規性のもとでの挙動

島根大・総合理工学研究科　牧草 夏実
島根大・数理科学科　内藤 貫太

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，$\langle m, h\rangle_{\mathcal{H}} = \mathbb{E}[\langle X, h\rangle_{\mathcal{H}}]$，$\langle \Sigma h, h'\rangle_{\mathcal{H}} = \mathrm{cov}(\langle X, h\rangle_{\mathcal{H}}, \langle X, h'\rangle_{\mathcal{H}})$ 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．

P を $\mathcal{H}$ 上の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．ここで，$m_0 = \mathbb{E}[X_1]$，$\Sigma_0 = V[X_1]$ である．

$\mathcal{H}$ 上の 2 つの分布 P と $N(m_0, \Sigma_0)$ は正定値カーネル k を用いて，それぞれのカーネル平均埋め込みである $\mu(P) = \mathbb{E}_{X\sim P}[k(\cdot, X)]$ と $\mu(N(m_0, \Sigma_0)) = \mathbb{E}_{X\sim N(m_0,\Sigma_0)}[k(\cdot, X)]$ によって k に対応する再生核ヒルベルト空間 $H(k)$ 上に表現できる．Kellner and Celisse (2015) では，この P と $N(m_0, \Sigma_0)$ の違いを，Maximum Mean Discrepancy (MMD)

$$\Delta^2 = \|\mu(P) - \mu(N(m_0, \Sigma_0))\|^2_{H(k)}$$

によって測ることで検定統計量

$$\widehat{\Delta}^2 = \left\| \frac{1}{n}\sum_{i=1}^{n} k(\cdot, X_i) - \mu(N(\widehat{m}, \widehat{\Sigma})) \right\|^2_{H(k)}$$

を構築した．ここで，$\widehat{m}$，$\widehat{\Sigma}$ はそれぞれ m_0，Σ_0 の一致推定量である．2017 年度統計関連学会連合大会において，この検定の帰無仮説の下での漸近分布が，独立な自由度 1 の χ^2 分布の重み付き無限和になることを報告した．

本発表では，この $\widehat{\Delta}^2$ に基づく正規性の検定の非正規性の下での挙動について報告する．理論的結果として，非正規性の下での検定統計量の漸近分布が正規分布となることを示し，検定の一致性が導かれる．また，局所対立仮説の下での漸近分布が，独立な自由度 1 の非心 χ^2 分布の重み付き無限和となることについて報告する．さらに，そもそも MMD 自体が正規性からのずれに関してどの程度敏感に反応しているかについて，$\mathcal{H}$ をユークリッド空間として，多変量歪度・尖度との比較を行う．最後に，検定の挙動に関する数値的検証の結果を報告し，この検定の性能評価について知見を与える．