

| | |
|------|--|
| 題名 | New precise model of studentized principal components |
| 掲載雑誌 | Communications in Statistics - Theory and Methods (2022), DOI:10.1080/03610926.2022.2084110 |
| 著者 | 小林靖之 (小林靖之研究室) |
| 概要 | <p>最初に、次元数p,標本数nの基準データの標本共分散行列の標本固有値l_k・固有ベクトルf_kに基づき、標本マハラノビス距離(SMD)は次元数p個のスチューデント化主成分要素(SPC)の2乗和で表せる。このl_kとf_kは確率的に変動するため、SMDは異常診断の性能が劣化する性質がある。SMDが異常に大きい場合の原因因子の究明には、SMDを標本固有値ごとに分解したSPCの統計的に推定や検定が重要である。しかし、小標本数におけるSPCの近似確率モデルはなかった。</p> <p>そこで本論文では、母固有値・固有ベクトルを推定することなしに小標本数におけるSPCの高精度な確率モデルを提案した。本提案モデルは、標本数nとSPCの番号k(最大固有値に相当する番号を$k=1$とし固有値の大きい順に付番)を用いた簡便な式と、自由度$n-k$のt分布に従う確率変数の積として、番号kのSPCを表せる。既存モデルでは標本共分散行列から母固有値・固有ベクトルの推定等の高度な予備計算が必要だが、本提案モデルでは不要である。</p> <p>なお、本提案モデルは理論モデルとして、隣接しあう母固有値が大きく離れている条件を仮定し導出された。しかし、コンピュータによる数値計算の結果、本提案モデルは隣接しあう母固有値が近いような弱い条件であっても、様々な次元数pと標本数nで良い近似性能を示した。</p> <p>実用的な応用例として、SMDから無限大のnに相当する母マハラノビス距離(PMD)への補正に本提案モデルを適用したところ、既存の補正モデルよりもPMDへの補正性能が高いとわかった。</p> <p>本提案モデルの適用により、判別分析やクラスター分析、射影追跡(projection pursuit)におけるSPCの統計的検定が可能となり、さらにガウス混合モデルにおけるExpectation-Maximizationアルゴリズムの改善も期待できる。</p> <p>論文を公開ページはこちら：https://www.tandfonline.com/doi/full/10.1080/03610926.2022.2084110</p> |