パラメトリック · ノンパラメトリック · セミパラメトリック

確率変数 X_i が連続型の密度関数 $f(x-\mu)$ $(i=1,\cdots,n)$ をもち, 各 X_i は互いに独立である 1 標本の単純なモデルを例に採って説明する．ここでは $f(x)$ は 0 について対称，すなわち，$f(-x)=f(x)$ を仮定するものとする．このとき, $E(X_i)=\int_{-\infty}^{\infty} xf(x-\mu)dx=\mu$ となり, μ は平均である．興味ある仮説検定の帰無仮説として $H_0:\mu=\mu_0$ を考える．この仮説検定として t 検定がよく知られている．また, 位置パラメータ μ の点推定量は標本平均 $\tilde{\mu}=\bar{X}_n$ で与えられ, 基本的な統計解析法としてよく使われる．これらの t 検定, 標本平均は, $f(x)$ が正規分布の密度関数であるとき, すなわち標本データが正規分布に従うとき最良な手法となり, **パラメトリック論**による手法と呼ばれている．パラメトリック論による手法にはいくつかの欠点がある, 第一に, データの中に異常値 (外れ値) を含む場合には, 検定結果や推定結果が異常値を含まない場合に比べて大きく変わってくる．すなわち, データ解析の結果が異常値に振り回され易い．第二に $f(x)$ が裾の重い分布のときには推測法として良くない．第三に, 標本サイズが大きくないとき $f(x)$ が既知でなければ検定の有意水準が固定できないため, パラメトリック論による検定手法は使うことが難しい．

$f(x)$ が未知としても検定または推定が行える便利な手法を**ノンパラメトリック論**による手法と呼び, 5.5 節で紹介する順位に基づく方法が代表的である．順位統計量を T とするとき, この値が大きいならば H_0 を棄却する．H_0 の下で 定数 c_0 に対して $P(T\geq c_0)$ は $f(x)$ に依存しない, すなわち, この順位検定は標本データの従う分布が未知であっても検定が行えることとなる．その後, 漸近理論 ($n\to\infty$ とした理論) により順位検定は $f(x)$ が正規分布のときは t 検定より少し悪く, $f(x)$ が正規分布から離れた分布のときは t 検定より検出力が良いことが示されている．特に裾の重い分布に対しては順位検定は有効であることが解っている．さらに n が大きくなくても漸近理論と同じ結論であることが計算機シミュレーション (模擬実験) により検証できる．このように $f(x)$ が正規分布のときは, 正規分布のときの最良方法に少し劣るが $f(x)$ が正規分布から離れた分布であるときに正規分布のときの最良方法より良い手法を, **分布に関する頑健性** (ロバスト) をもつ手法と呼んでいる．すなわち順位検定法は分布に関する頑健性をもっている．また, 解析結果が異常値にあまり影響を受けない手法を, **異常値に関する頑健性**をもつ手法と呼んでいる．順位による位置パラメータ μ の推定はホッジス • レーマン (Hodges and Lehmann (1963)) によって提案され理論も

同時に述べられている．順位検定法の場合と同じように，この順位推定法は標本平均 $\tilde{\mu} = \bar{X}_n$ による推定との比較において $f(x)$ が正規分布のときは標本平均による推定より少し悪いが，$f(x)$ が正規分布以外の分布であるときには標本平均による推定より良いことが示された．またこの順位推定法が異常値に関する頑健性をもっていることが，フーバー (Huber (1981)) のテキストで理論的に示されている．

$f(x)$ が正規分布の ε 近傍 $U_\varepsilon \equiv \{f(x) = (1-\varepsilon)\varphi(x) + \varepsilon h(x) : \varphi(x)$ は標準正規分布の密度関数，$h(x)$ はすべての x に対して $h(-x) = h(x)$ を満たすある密度関数 $\}$ の中にあるときの近似的に最良な手法をパラメトリックとノンパラメトリックの中間という意味で**セミパラメトリック論**による手法と呼ばれている．フーバー (Huber (1964)) は 関数 $\psi(x) = x$ ($|x| < b$ のとき)$; = b$ ($x \geq b$ のとき)$; = -b$ ($x \leq -b$ のとき) を導入し方程式 $\sum_{i=1}^n \psi(X_i - \theta) = 0$ の解 $\hat{\theta}$ を μ の点推定量として提案し，これを M 推定と呼んだ．ε に依存して b をうまく採れば $f(x)$ が U_ε に入るとき M 推定が近似的に最良となることを示し，その後の研究成果を フーバーはテキスト (Huber (1981)) でレビューしている．一方，検定についても関数 $\psi(x)$ を使った M 検定を提案することができ，フーバーの M 推定のときと同じように標本データの分布が正規分布の ε 近傍に入るときに近似的に最良であることを示すことができる．すなわち，$f(x)$ が近傍 U_ε に入るとき平均に対する近似的に最良な検定である．さらに M 推定も M 検定も分布と異常値に関する頑健性をもっている．

生のデータが手元にあるときパラメトリック，ノンパラメトリック，セミパラメトリック論による 3 つの手法のうちいずれを適用すればよいかは，第 5 章で紹介する正規性の検定と分布の探索から解る．

更なる詳細へリンク