博士（理学）　橋　本　智　雄

学 位 論 文 題 名

Cramér-von Mises-Watson type statistics for goodness of fit with unknown parameter

（未知母数をもつCramér-von Mises-Watson型適合度検定統計量）

学位論文内容の要旨

数理統計学の推測理論のうちで、経験分布関数$F_n(x)$に基ずく統計推論は理論的にもまた応用面においても重要な分野である。本論文はこの経験分布関数に基ずく適合度検定$H_0: F(x)=F_0(x,\theta)$のうち、いわゆるCramér-von Mises-Watson　型検定統計量

$$W_n^2 = n\int_{-\infty}^{\infty} \{F_n(x)-F_0(x)\}^2 dF_0(x) \qquad (1)$$

$$U_n^2 = n\int_{-\infty}^{\infty} \{F_n(x)-F_0(x)-\int_{-\infty}^{\infty}[F_n(y)-F_0(y)]dF_0(y)\}^2 dF_0(x) \qquad (2)$$

の漸近分布を論じたものである。本論文は二部構成になっている。

第一部は、仮設した分布の母数θが未知、したがってデータからθを推定したあとで(推定量を$\hat\theta$とする）適合度検定$H: F(x)=F_0(x,\hat\theta)$にかける際の検定統計量$\hat U_n^2$ の漸近分布を詳しく研究したものである。Watson (1961, 1962)はランダムサンプル$x_1, x_2, \ldots, x_n$ が与えられた分布$F(x,\theta)$から抽出されたものかを調べる検定統計量として（２）を提案し、その分布、性質について詳細に研究した。本研究はDarling の研究にヒントを得て、この統計量U_n^2をθが未知の場合に拡張した統計量

$$\hat U_n^2 = n\int_{-\infty}^{\infty} \{F_n(x)-F_0(x,\hat\theta)-\int_{-\infty}^{\infty}[F_n(y)-F_0(y,\hat\theta)]dF_0(y,\hat\theta)\}^2 dF_0(x,\hat\theta) \qquad (3)$$

極限分布について詳しく調べたものである。このような場合の分布は一般にdistribution-free にならないため、推定量$\hat\theta$の選び方、すなわち推定量としてどのような性質をもつものを考えるべきか、また仮定した分布の型$F(x,\theta)$に依存するため、難しい問題となる。

§1　は統計量$\hat U_n^2$が順序統計量の和として表現されることを利用して、推定量$\hat\theta$がsuper-efficientのとき、$\hat U_n^2$の極限分布がU_n^2 の極限分布に帰着できることを示した。

この超有効性の条件は強すぎるので、§2、§3ではCramérの意味での正則性があるときの場合、および有効推定量の場合の漸近分布について考察した。この際、極限分布を議論するときの基本となる確率過程$X_n(t)\to X(u)$の法則収束を示し、共分散$\rho(u,v)$を計算した。

ここで$X(u)$は平均0、共分散$\rho(u, v)$をもつガウス過程である。主な結果は次の定理にまとめられる。

『定理：有効推定量が存在するとき、$\hat{U}_n^2$の極限分布は

$$\lim_{n\to\infty} P(\hat{U}_n^2 \le x) = P\{\int_0^1 X^2(u)du \le x\}$$
』

$\hat{\theta}$が最尤推定量の場合も条件を少しつけるだけで、ほぼ同じ結果が得られる。

実際の極限分布はKac-Siegertの表現により、Fredholm 型積分方程式の固有値を求め、さらにFredholm行列式D（λ）を求めなくてはならないが、本論文の§5ではこれを求める方法が論じられる。究極的にはこのD(2it)を反転して、密度関数p(x)を得ることになるが、これをexplicitに書き表すことは今のところ困難である。第一部の最後では、経験分布関数の特性関数を利用して、位置と尺度の変位にかんする適合度検定の検出力の議論が考察される。これは統計学的見地からはきわめて重要と考えるが、結果は次のようになる。

『定理：経験特性関数の実部に基ずいた正規性検定においては、検出力関数$L_n(t)$は漸近的に位置の変位と無関係に定まる。虚部についても、同様なことが成立する。』

本論文の第二部はデータが打ち切り標本（第1種センサリング）-この種のサンプルは寿命分布における解析モデルとして生物統計、工業統計の分野でしばしば用いられるものであるが-における正規性検定のための統計量としてCramér-von Mises-Watson型統計量W_{pn}^2, U_{np}^2 の極限分布を論じたものである。第一部と同様に母集団分布が未知母数$\hat{\theta}$を含む場合について$\hat{W}_{pn}^2, \hat{U}_{np}^2$ を考察し、これの漸近分布を求める方法について詳細に研究されている。

まず§7ではセンサードサンプルにおける正規分布の未知母数θが位置母数のとき、最尤推定量（MLE）の近似解$\hat{\theta}$を求め、これの漸近的挙動について理論的に重要な結果をえた。すなわち

『定理：（i）$\hat{\theta}$ は漸近的不偏性をもつ。(ii) $\hat{\theta}$ は一致推定量である。』

証明には、条件付期待値$E_c(\hat{\theta})$と分散$Var_c(\hat{\theta})$を計算することが用いられた。次いで、§8では経験過程$\hat{y}_n(t)=\sqrt{n}\,[\hat{F}_n(t)-t]$の弱収束に関する議論が展開され、先のMLE $\hat{\theta}$がDurbinのガウス過程に弱収束するための条件を満足することを示し、そして極限過程$\hat{y}(t)$の共分散$\rho(s, t)$を実際に計算した。さらに、$\hat{U}_{np}^2$を定義する経験過程$\hat{z}_n(t)$についても同様な考察をして、その極限過程$\hat{z}(t)$の共分散$\rho_1(s, t)$も exact に計算した。その際にPettitt(1976)が一般の分布について示した共分散の結果が間違いであることも指摘してある。§9は第一部の§5と同じように、$\hat{W}_p^2, \hat{U}_p^2$の分布を求める際の積分方程式のFredholm行列式、固有値、固有関数についての研究が述べられ、固有値がW_p^2 の固有値を用いて陽に書き表されることが示されている。最後に、この検定がparameter-freeですらないことが述べられるが、これはセンサードされたデータが如何に複雑であるかということを反映したものとみることができる。

応用統計の上では、結果を実際に利用するため数値的近似法の開発が残されている。しかしこれを解決するにはまったく別のアプローチを必要とするように思われる。

学位論文審査の要旨

主 査　教 授　上 見 錬太郎　　副 査　教 授　久保田 幸 次
副 査　教 授　津 田 一 郎
副 査　教 授　鈴 木 義一郎（文部省統計数理研究所）

学 位 論 文 題 名

Cramér-von Mises-Watson type statistics for goodness of fit with unknown parameter

（未知母数をもつCramér-von Mises-Watson型適合度検定統計量）

数理統計学の推測理論のうちで、経験分布関数$F_n(x)$に基ずく統計推論は理論的にもまた応用面においても重要な分野である。本論文はこの経験分布関数に基ずく適合度検定$H_0: F(x)=F_0(x, \theta)$のうち、いわゆるCramér-von Mises-Watson　型検定統計量W_n^2, U_n^2の漸近分布を論じたものである。本論文は二部構成になっている。

第一部は、仮設した分布の母数θが未知、したがってデータからθを推定したあとで(推定量を$\hat{\theta}$とする）適合度検定$H: F(x)=F_0(x, \hat{\theta})$にかける際の検定統計量$\hat{U}_n^2$ の漸近分布を詳しく研究したものである。Watson (1961, 1962) はランダムサンプル$x_1, x_2, \ldots, x_n$ が与えられた分布$F(x, \theta)$から抽出されたものかを調べる検定統計量としてU_n^2を提案し、その分布、性質について詳細に研究した。本研究はDarling の研究にヒントを得て、この統計量U_n^2をθが未知の場合に拡張した統計量

$$\hat{U}_n^2 = n\int_{-\infty}^{\infty} \{F_n(x)-F_0(x;\hat{\theta})-\int_{-\infty}^{\infty} [F_n(y)-F_0(y;\hat{\theta})]dF_0\}^2 dF_0(x;\hat{\theta})$$

の極限分布について詳しく調べたものである。このような場合の分布は一般にdistribution-free にならないため、すなわち仮定した分布の型$F_0(x;\hat{\theta})$に依存するため推定量$\hat{\theta}$の選び方からはじまって、その解析はかなり難しい問題となる。

§1　は統計量U_n^2が順序統計量の和として表現されることを利用して、推定量$\hat{\theta}$がsuper-efficientのとき、$\hat{U}_n^2$の極限分布がU_n^2 の極限分布に帰着できることを示した。

この超有効性の条件は強すぎるので、§2、§3ではCramérの意味での正則性があるときの場合、および有効推定量の場合の漸近分布について考察した。この際、極限分布を議論するときの基本となる確率過程$X_n(t) \to X(u)$の法則収束を示し、共分散$\rho(u, v)$を計算した。ここで$X(u)$は平均0、共分散$\rho(u, v)$をもつガウス過程である。主な結果は次の定理にまとめられる。

「定理：有効推定量が存在するとき、$\hat{U}_n^2$の極限分布は

$$\lim_{n\to\infty} P(\hat{U}_n^2 \le x) = P\{\int_0^1 X^2(u)du \le x\}$$

」

$\hat{\theta}$が最尤推定量の場合も条件を少しつけるだけで、ほぼ同じ結果が得られる。

実際の極限分布はKac-Siegertの表現により、Fredholm 型積分方程式の固有値を求め、さらにFredholm行列式$D(\lambda)$を求めなくてはならないが、本論文の§5ではこれを求める方法が論じられる。究極的にはこの$D(2it)$を反転して、密度関数$p(x)$を得ることになるが、これをexplicitに書き表すことは今のところ困難である。第一部の最後では、経験分布関数の特性関数を利用して、位置と尺度の変位にかんする適合度検定の検出力の議論が考察される。これは統計学的見地からはきわめて重要と考えるが、結果は次のようになる。

『定理：経験特性関数の実部に基ずいた正規性検定においては、検出力関数$L_n(t)$は漸近的に位置の変位と無関係に定まる。虚部についても、同様なことが成立する。』

本論文の第二部はデータが打ち切り標本（第1種センサリング）における正規性検定のための統計量としてCramér-von Mises-Watson型統計量の極限分布を論じたものである。第一部と同様に母集団分布が未知母数θを含む場合について$\hat{W}_{pn}^2, \hat{U}_{np}^2$ を考察し、これの漸近分布を求める方法について詳細に研究されている。

まず§7ではセンサードサンプルにおける正規分布の未知母数θが位置母数のとき、最尤推定量（MLE）の近似解$\hat{\theta}$を求め、これの漸近的挙動について理論的に重要な結果を得た。

『定理：（i）$\hat{\theta}$ は漸近的不偏性をもつ。(ii) $\hat{\theta}$ は一致推定量である。』

証明には、条件付期待値$E_c(\hat{\theta})$ と分散$Var_c(\hat{\theta})$ を計算することが用いられた。次いで、§8では経験過程$\hat{y}_n(t)=\sqrt{n}\{\hat{F}_n(t)-t\}$の弱収束に関する議論が展開され、先のMLE $\hat{\theta}$がDurbinのガウス過程に弱収束するための条件を満足することを示し、そして極限過程$\hat{y}(t)$の共分散$\rho(s,t)$を実際に計算した。さらに、$\hat{U}_{np}^2$を定義する経験過程$\hat{z}_n(t)$についても同様な考察をして、その極限過程$\hat{z}(t)$の共分散$\rho_1(s,t)$も exact に計算した。その際にPettitt (1976)が一般の分布について示した共分散の結果が間違いであることも指摘してある。§9は第一部の§5と同じように、$\hat{W}_p^2, \hat{U}_p^2$の分布を求める際の積分方程式のFredholm行列式、固有値、固有関数についての研究が述べられ、固有値がW_p^2 の固有値を用いて陽に書き表されることが示されている。最後に、この検定がparameter-freeですらないことが述べられるが、これはセンサードされたデータが如何に複雑であるかということを反映したものとみることができる。

以上本論文によって$\hat{W}_{pn}^2, \hat{U}_{pn}^2$の極限分布はガウス過程の汎関数の分布として表わされることが明確になった。

よって著者は、北海道大学博士（理学）の学位を授与される資格あるものと認める。