修士論文 要旨

修正経験尤度比統計量に関する研究

数理情報学専攻 48096230 向田裕人

指導教員 駒木文保 教授

1 はじめに

パラメトリックな統計的モデル $p(x;\psi,\lambda)$ (ψ : 1 次, λ : d 次) からのサンプル $\mathbf{x}=(x_1,\ldots,x_n)$ を用いて, 検定問題 $H_0:\psi=\psi_0$ v.s $H_1:\psi\neq\psi_0$ を考える際, 検定統計量として対数尤度比統計量 $L(\psi)=2\{l(\hat{\psi},\hat{\lambda})-l(\psi,\hat{\lambda}_\psi)\}$ を用いた方法が広く用いられる. ここで, $l(\psi,\lambda)$ は対数尤度関数, $\hat{\psi},\hat{\lambda}$ は最尤推定量, $\hat{\lambda}_\psi$ は ψ が既知の下の λ の最尤推定量である. この統計量は, H_0 が真ならば $L(\psi)\xrightarrow{d}\chi^2_{(1)}$ となるので受容域の近似的な設定がしやすいことや, H_0,H_1 がともに単純仮説の場合には最強力検定になるという利点がある. また, この統計量の収束の精度を上げるための, 統計量の修正方法についても様々な研究がなされている.

本論文では, [3] に基づいた符号付き根号対数尤度比統計量の修正方法について, observed geometry による修正項の持つ情報幾何学的な意味を与えた. さらに, 経験尤度に対する拡張方法と, exponential-tilt モデルの場合における検定の有用性を数値実験によって示した.

2 修正項の情報幾何学的考察

本節では, λ も 1 次元, つまり $d=1$ であるとする.

2.1 符号付き根号対数尤度比統計量の修正

ψ が真のパラメータ値のとき, 符号付き根号対数尤度比統計量 $R=\mathrm{sgn}(\hat{\psi}-\psi)\sqrt{L(\psi)}$ は次の性質を持つ.

定理 2.1. $\Pr(-r\leq R\leq r)=\Pr(-r\leq T\leq r)+O(n^{-1})$ $(T\sim\mathrm{N}(0,1))$ が成立する.

定理 2.2. $\mathbf{x}$ の従う分布 $f(\mathbf{x};\psi,\lambda)$ について, ある (近似的な) 補助統計量 a が存在し, $(\hat{\psi},\hat{\lambda},a)$ が最小十分統計量となる時, $R^*=R+\frac{1}{R}\log\left(\frac{u_I u_P}{R}\right)$ は, $\Pr(-r\leq R^*\leq r)=\Pr(-r\leq T\leq r)+O(n^{-3/2})$ が成立する [3]. ここで,

$$u_I:=\frac{l^\dagger_{\hat{\psi}}(\psi)-l_{\lambda\hat{\psi}}(\psi,\hat{\lambda}_\psi)^T l_{\lambda\hat{\lambda}}(\psi,\hat{\lambda}_\psi)^{-1}l^\dagger_{\hat{\lambda}}(\psi)}{|\hat{j}_{\psi\psi}-\hat{j}_{\psi\lambda}\hat{j}^{-1}_{\lambda\lambda}\hat{j}_{\lambda\psi}|^{1/2}},$$

$$u_P:=\frac{|l_{\lambda\hat{\lambda}}(\psi,\hat{\lambda}_\psi)|}{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}|\hat{j}_{\lambda\lambda}|^{1/2}},$$

$$l^\dagger(\psi):=l(\hat{\psi},\hat{\lambda})-l(\psi,\hat{\lambda}_\psi),$$

$$j(\psi,\lambda):=\begin{pmatrix}j_{\psi\psi}(\psi,\lambda) & j_{\psi\lambda}(\psi,\lambda)\\ j_{\lambda\psi}(\psi,\lambda) & j_{\lambda\lambda}(\psi,\lambda)\end{pmatrix}=-\begin{pmatrix}l_{\psi\psi}(\psi,\lambda) & l_{\psi\lambda}(\psi,\lambda)\\ l_{\lambda\psi}(\psi,\lambda) & l_{\lambda\lambda}(\psi,\lambda)\end{pmatrix}$$

であり, $l(\psi,\lambda)$ の下付き添え字は, その添え字で偏微分することを表す. また, $\hat{j}=j(\hat{\psi},\hat{\lambda})$ である.

よって, 統計量 R を用いた検定の精度は $O(n^{-1})$ で, R^* と修正すると精度が $O(n^{-3/2})$ となることが分かる.

2.2 Observed Geometry

情報幾何学とは, 統計的モデルを多様体とみなし, 幾何学的な構造を乗せて解析する方法であり, これを用いれば, 統計量や統計量の従う分布を展開して得られた値について, 各項の持つ意味や性質を直感的に理解できる. 情報幾何学の一つである observed geometry[2] では, 統計的モデル $p(x;\theta)$ について, $(\hat{\theta},a)$ ($\hat{\theta}$: θの最尤推定量) が最小十分統計量となる (近似的な) 補助統計量 a の存在を仮定し, Riemann 計量として observed information $j_{rs}(\theta)=-\frac{\partial^2}{\partial\theta^r\partial\theta^s}l(\theta)$ を導入する.

2.3 修正項の展開

修正項について次の性質が成り立つ.

補題 2.1. u_P,u_I は, パラメータについての非線形変数変換 $\psi'=g(\psi),\lambda'=h(\psi,\lambda)$ について不変である.

この性質から, ψ と λ が直交していると考えてよく, すると各修正項の展開式は,

$$u_P=1+\underbrace{\frac{1}{2\not{j}_{\lambda\lambda}}\overset{-1}{\not{H}}_{\lambda\lambda\psi}(\hat{\psi}-\psi)}_{\text{(i)}}+O_p(n^{-1}),$$

$$u_I=\underbrace{\not{j}^{1/2}_{\psi\psi}(\hat{\psi}-\psi)+\frac{1}{2}\not{j}^{-1/2}_{\psi\psi}\overset{-1}{\not{\Gamma}}_{\psi\psi\psi}(\hat{\psi}-\psi)^2}_{\text{(ii)}}$$

$$+\underbrace{\not{j}^{-1/2}_{\psi\psi}\overset{0}{\not{H}}_{\psi\psi\lambda}(\hat{\psi}-\psi)(\hat{\lambda}-\lambda)}_{\text{(iii)}}+O_p(n^{-1})$$

となる. ここで, $\not{g}$ という表記は, 関数 $g(\psi,\lambda;\hat{\psi},\hat{\lambda},a)$ について, 最尤推定量 $(\hat{\psi},\hat{\lambda})$ に真のパラメータ値 (ψ,λ) を代入したものを表し, $\overset{-\alpha}{\not{\Gamma}}$ は α-接続, $\not{H}$ は Euler–Schouten 曲率テンソルである. このことから, (i),(ii),(iii) の項について次のような解釈ができる.

(i) : -1-接続を導入したとき, $\lambda=(\mathrm{const.})$ という部分多様体が真のパラメータ (ψ,λ) 上でどのくらい曲がっているかを考慮する項.

(ii) : expected geometry[1] における α-divergence $D_\alpha(\theta,\theta')$ について, g を Fisher 計量としたとき,

$$
\begin{aligned}
&\mathrm{sgn}(\hat{\psi}-\psi)\sqrt{2D_3((\psi,\lambda),(\hat{\psi},\lambda))} \\
&= g_{\psi\psi}^{1/2}(\hat{\psi}-\psi)+\frac{1}{2}g_{\psi\psi}^{-1/2}{}^{-1}\Gamma_{\psi\psi\psi}(\hat{\psi}-\psi)^2+O_p(n^{-1})
\end{aligned}
$$

となることから，最尤推定量 $\hat{\psi}$ と真のパラメータ ψ の距離を考慮する項．

(iii) ：0-接続を導入したとき，$\psi=(\mathrm{const.})$ という部分多様体が，真のパラメータ (ψ,λ) 上でどのくらい曲がっているかを考慮する項．

3 経験尤度への応用

3.1 経験尤度

経験尤度とは，$x_1,\dots,x_n \sim F_0, x_i \in \mathbb{R}^d$ について，F_0 を確率 w_i で x_i を取る多項分布 F で離散近似して考える方法で，$L(F)=\prod_{i=1}^n w_i$ で定義される．

これを用いた検定を考えると，たとえば，$E[X]=\theta$ を検定する際は，プロファイル経験尤度比統計量は，

$$
\mathcal{R}(\theta)=\max\left\{\prod_{i=1}^n nw_i \,\middle|\, \sum_{i=1}^n w_iX_i=\theta, w_i\geq 0, \sum_{i=1}^n w_i=1\right\}
$$

となる．これは $E[X]=\theta$ が真ならば $-2\log\mathcal{R}(\theta)\xrightarrow{d}\chi^2_{(d)}$ が成立する．このように，パラメトリックモデルにおける尤度関数で成り立つ性質は，経験尤度でも成立することが多いことが知られている [4].

3.2 Exponential-tilt モデルにおける修正

確率変数 $X,Y\in\mathbb{R}$ について $X\sim g(x)$, $Y\sim h(x)=\exp(\alpha+\beta x)g(x)$ で表されるモデルを exponential-tilt モデルという．このモデルから $x_1,\dots,x_n\sim g(x)$, $y_1,\dots,y_m\sim h(x)$ とサンプルが得られた時，$g(x)$ を多項分布で離散近似して経験尤度を計算すると，

$$
l(x,y;\beta,\theta)=\beta\sum_{i=1}^m(y_i-z_{n+m})+\sum_{i=1}^{n+m-1}\theta_i+\Phi(\beta,\theta) \quad (1)
$$

となる．ここで，$\theta_i=\log\frac{p_i}{p_{n+m}}$, $z_i=x_i\ (i=1,\dots,n)$, $z_i=y_{i-n}\ (i=n+1,\dots,n+m)$, $\Phi(\beta,\theta)=-n\log(1+\sum_{i=1}^{n+m-1}\mathrm{e}^{\theta_i})-m\log(1+\sum_{i=1}^{n+m-1}\mathrm{e}^{\theta_i+\beta(z_i-z_{n+m})})$ である．これは $(n+m,n+m)$-指数型分布族の対数尤度関数の形であるので，容易に修正項の計算ができる．

3.3 数値実験

(1) 式を用いた R,R^* による検定の比較実験の，包含確率の結果を図 1 に，検出力の結果を図 2 に記す．この実験では，$g(x)=0.3\mathrm{N}(0,1)+0.7\mathrm{N}(1,1)$，$n=m$，$\beta$ の真の値を β_0 とし，R,R^* を用いて帰無仮説を有意水準 5% で検定した．帰無仮説は，包含確率の比較の場合は $\beta=\beta_0$，検出力の比較の場合は $\beta=0$ とした．

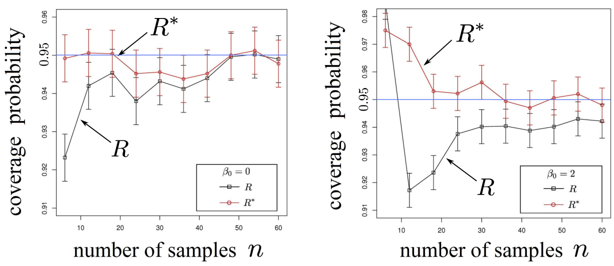

図 1. 包含確率の比較（左：$\beta_0=0$，右：$\beta_0=2$）.

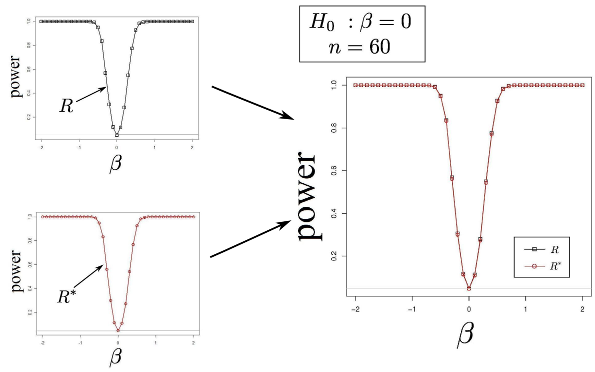

図 2. 検出力の比較（$n=60$）.

検定統計量として R^* を用いたほうが，包含確率の結果が有意水準の設定と整合のとれた結果となっている．また，検出力に差は見られなかった．

4 まとめと今後の課題

本論文では，符号付き根号対数尤度比統計量について，修正項の情報幾何学的意味を与えた．また，経験尤度への拡張について，exponential-tilt モデルの際の修正方法と，数値実験による有用性の考察を行った．

今後の課題としては，分布収束の精度の改善についての情報幾何学的考察や，exponential-tilt モデル以外のモデルへの応用と検証，収束の改善についての理論的な証明などが挙げられる．

参考文献

[1] S. Amari: *Differential-Geometry Methods in Statistics*, Springer Lecture Notes in Statistics, California, 1985.

[2] S. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen and C. R. Rao: *Differential Geometry in Statistical Inference*, Institute of Mathematical Statistics, California, 1987.

[3] O. E. Barndorff-Nielsen: Inference on full or partial parameters based on the standardized signed log likelihood ratio, *Biometrika*, Vol. 73 (1986), pp. 307–322.

[4] A. B. Owen: *Empirical Likelihood*, Chapman and Hall/CRC, New York, 2001.