尤度比統計量に代わる KL 統計量の可能性

中央大学　柳本 武美

0. 序

密度関数が $p(\boldsymbol{y};\theta), \boldsymbol{y}\in\mathcal{X}\subset R^p, \theta\in\Theta\subset R^q$ とする。実現値を $\boldsymbol{x}\in\mathcal{X}$ として、適当な推定値を $\hat\theta$、最尤推定値を $\hat\theta_M$ と書く。ここでは、θ の信頼領域（区間）あるいは検定問題を考える。今日もっとも汎用される統計量は尤度比統計量

$$\mathrm{LR}(\hat\theta_M,\theta) = \log\{p(\boldsymbol{x};\hat\theta_M)/p(\boldsymbol{x};\theta)\}$$

である。一方ここで議論したい統計量は

$$\mathrm{KL}(\hat\theta,\theta) = \mathrm{E}\left[\log\{p(\boldsymbol{y};\hat\theta)/p(\boldsymbol{y};\theta)\}\,|\,p(\boldsymbol{y};\hat\theta)\right]$$

であり、以下では KL 統計量と呼ぶ。改めて考察しても、尤度比統計量が KL 統計量よりも有用であるとの証拠は見あたらない。そこで、 KL 統計量の方が有用ではなかろうかとの仮説の下にに考察を進めている。

1. 予備的な議論

尤度比統計量が有用であるとの証拠を見つけるのは難しい。

a) もし密度関数が指数分布族であれば、$\mathrm{KL}(\hat\theta_M,\theta) = \mathrm{LR}(\hat\theta_M,\theta)$ が成り立つ。

b) もし標本の次元 p が高くて、最尤推定量が正規近似できるような場合には a) が適用される。

一方で、尤度比統計量には欠点がある。

c) 適当に推定量を選ぶと

$$\mathrm{E}\{\mathrm{KL}(\hat\theta,\theta) - \mathrm{LR}(\hat\theta,\theta)\,|\,p(\boldsymbol{x};\theta)\} = 0 \tag{1.1}$$

が成り立つことがある。この場合に、敢えて最尤推定量を使う根拠はない。

d) 良い推定量がある場合に KL 統計量ではそのまま適用できるが、尤度比統計量の修正は難しい。

2. 信頼区間 (検定) としての議論

前節の 2 つの統計量を用いて、信頼区間が適当な c_α に対して $T < c_\alpha$ で定義できるとする。もし信頼区間が点推定量と対応がついていれば、推測方式としてより自然であると思われる。KL 統計量を使うと、$100(1-\alpha)\%$ 信頼区間は c_α と $\hat\theta$ のみで表されるが、尤度比統計量を使うと一般には $\boldsymbol{x}$ に依存する。これを検定量として用いると、$\hat\theta(\boldsymbol{x}_1) > \hat\theta(\boldsymbol{x}_2) > \theta_0$ で $\boldsymbol{x}_1$ が棄却域に入らない一方で $\boldsymbol{x}_2$ が棄却域に入ることがあり得る。この結果、検出力が向上すれば良いが実際のところは分からない。

極限として 0% 信頼区間を構成すると通常点推定値が得られる。もし最尤推定量より良い推定量が得られる場合には、より良い 0% 信頼区間が得られることになる。この場合、一般に $100(1-\alpha)\%$ についても同様であろうと推認される。

[例] $\boldsymbol{x}$ が $N(\mu, \sigma^2)$ からの大きさ p の標本ベクトルとする。ここで、$H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2$ を考える。教科書の σ^2 の推定量は標本分散 s^2 であるが、最尤推定量は $s^2(p-1)/p$ となる。標本平均で平均を推定すると、最尤推定量は良くない。実際

$$\mathrm{KL}((\bar{\boldsymbol{x}}, s^2), (\bar{\boldsymbol{x}}, \sigma_0^2)) = p\{s^2/\sigma_0^2 - \log(s^2/\sigma_0^2) - 1\}$$

となり、$\mathrm{LR}((\bar{\boldsymbol{x}}, \hat{\sigma}_M^2), (\bar{\boldsymbol{x}}, \sigma_0^2))$ は、$\mathrm{KL}((\bar{\boldsymbol{x}}, s^2), (\bar{\boldsymbol{x}}, \sigma_0^2)) - \{s^2/\sigma_0^2 - p\log((p-1)/p)\}$ と表される。前者はガンマ分布の密度関数からも自然な式であるが、後者の第 2 項は両側検定量としては無理がある。

3. Bayes モデルにおける検定

本研究の動機は、Bayes モデルの発展的適用を計ることにある。多層ガンマ回帰モデルを考える。第 k 層 第 i 観測値 x_{ki} がガンマ分布 $\mathrm{Ga}(\mu_{ki}, \tau)$ に従うと仮定する。但し、回帰モデルとして対数線型モデル $\mu_{ki} = \exp(\alpha_k + \beta z_{ki})$ を仮定する。この密度関数は

$$p(\boldsymbol{x}; \boldsymbol{\alpha}, \beta, \tau) = \prod_k \left[\prod_i \left\{\frac{\tau^\tau}{\Gamma(\tau)} \frac{x_{ki}^{\tau-1}}{\exp\{\tau(\alpha_k + \beta z_{ki})\}} \exp\{-\tau x_{ki} \exp(-\alpha_k - \beta z_{ki})\}\right\}\right] \quad (3.1)$$

と表される。但し、$\boldsymbol{\alpha}$ は α_k の母数ベクトルである。

ここで我々の主たる関心は共通スロープ β の推定と検定にあり、勿論他の母数にも副次的な関心を持つ。このモデルは、一般化線型モデルに含まれるので、最尤法を基盤にした推測が標準である (例えば Lawless, 1982 Wiley)。しかし、層の数が多い (例えば 5 以上) であれば、Bayes 法が有力と見なされている。実際母数 α_k, に事前分布を仮定するために変換 $\eta_k = \exp(-\alpha_k)$ を行った上で $\eta_k \sim Ga(1/h, n_k\delta)$ と定義する。通常の経験 Bayes 法 (の拡張型) では、η を事後平均で推定する。また、他の母数は周辺密度 (尤度)

$pm(\boldsymbol{x}; \beta, \tau, h, \delta) =$

$$\prod_k \left[\frac{\Gamma(n_k(\tau+\delta))}{\Gamma(\tau)^{n_k}\Gamma(n_k\delta)} \frac{\tau^{n_k\tau}(n_k\delta)^{n_k\delta}\tilde{x}_k^{n_k(\tau-1)}}{\{n_k(\tau+\delta)\}^{n_k(\tau+\delta)}} \exp\{n_k\delta\log(h) + n_k(\tau+\delta)\log(\hat{\eta}_k)\}\right]$$

を用いて推定する。この推定量は予想されるように β についてもモデル全体の母数についても、最尤推定量に比べて大変良い性能を示す (箕田・鎌倉・柳本 2007 read at JSCS)。

問題は、β の検定 (信頼区間) の構成にある。この母数は、いわゆる超母数と言うよりモデルの母数である。実際、$\boldsymbol{\alpha}$ の要素に事前分布を仮定したのは、変量効果と見なしたためではなくて、良い推定量を導出するために操作的に導入している。超母数 δ についての検定では周辺尤度比検定も使い易い (例えば、Yanagimoto and Yanagimoto, 1987 Technometrics) が、β についてはその意味が取りにくい。実際、経験 Bayes 法自身が Bayes 推論の枠組みの中で議論がある。一方で、他の母数を固定して $\boldsymbol{\alpha}$ のみの推定を考えると、Bayes 推定量であるので、式 (1.1) が成立する。従って、形式的な尤度比統計量は使えない。実際に予備的な simulation でそのパワーが低いことを確かめている。

そうすると、KL 統計量を用いることが動機づけられることになる。周辺尤度最大化を改善する方法を試みて、β の推定量を改善することはできた。