非正規分布における標本平均絶対偏差の漸近相対効率
Asymptotic relative efficiency of sample mean absolute deviance under non-normality

数学専攻　熊倉　徹
KUMAKURA, Toru

1　はじめに

データの散らばり具合を表す指標として, 平均絶対偏差と標準偏差がある. 平均絶対偏差は解釈が容易であるが, 代数的操作が難解である. 一方, 標準偏差は代数的操作が比較的簡易であり, 現在の数理統計学において広く扱われているが, 外れ値に大きな影響を受けロバストでない. この 2 つの指標は主に推定の漸近相対効率という観点から, Eddington (1914), Fisher (1920), Tukey (1961) などにより比較がなされ, 現在では標準偏差を一般的に用いている. しかしながら, その漸近的評価はしばしば議論の対象となる.

本論文では, この漸近相対効率のよる比較とその批判についての理論的詳細を整理するとともに, 検定論における Pitman の漸近相対効率からの解釈を与え, さらにいくつかの母集団分布における検証を行う.

2　漸近相対効率

この節では, 推定と検定における漸近相対効率の定義を述べ, 3 節で述べる標本標準偏差の標本平均絶対偏差に対する漸近相対効率の導出に必要な理論を概説する.

2.1　推定の漸近相対効率

定義 2.1
$\{S_n\}$ を $\sqrt{n}[S_n - g(\theta)] \xrightarrow{L} N\left(0, \sigma_1^2\right)$ となる $g(\theta)$ の推定量の列とし, $\{T_n\}$ を $\sqrt{n}[T_n - g(\theta)] \xrightarrow{L} N\left(0, \sigma_2^2\right)$ となる $g(\theta)$ の推定量の列とすると, $\{S_n\}$ の $\{T_n\}$ に対する漸近相対効率 (Asymptotic Relative Efficiency, ARE) は

$$\mathrm{ARE}(S_n|T_n) = \frac{\sigma_2^2}{\sigma_1^2} \tag{2.1}$$

で定義される.

Fisher (1920), Tukey (1961) は, 漸近相対効率の観点から標本標準偏差と標本平均絶対偏差における優劣を議論している. しかしながら, そこで与えている漸近相対効率は上記の定義とは異なり, 批判の対象となることもある.

2.2　検定の漸近相対効率

2.1 節で与えた推定の漸近相対効率に対し, 仮説検定における漸近相対効率のうち Pitman の漸近相対効率について述べる. 詳細は柳川 (1982), 前園 (2001) などを参照されたい.

定義 2.2
$\{S_m\}$, $\{T_n\}$ を帰無仮説 $H_0 : \theta = 0$ を固定したときの対立仮説 $H_{1m} : \theta_m = \dfrac{\delta}{\sqrt{m}}\,(\delta > 0)$ を検定する検定統計量の系列とする．両検定が有意水準 $\alpha\,(0 < \alpha < 1)$ を漸近的にみたすこと，すなわち

$$\lim_{m\to\infty} P_0\,(S_m \geq s_{m_\alpha}) = \lim_{n\to\infty} P_0(T_n \geq t_{n_\alpha}) = \alpha$$

が成り立つと仮定する．$\{m\}, \{n\}$ を自然数の増加系列であるとし，$m \to \infty$ で $n \to \infty$ も同時に成り立つとする．以後は $\lim\limits_{m\to\infty}$ として記号を統一する．所与の定数 $\beta\,(0 < \beta < 1)$ に対して

$$\begin{aligned}\lim_{m\to\infty} \beta_m(\theta_m) &= \lim_{m\to\infty} P_{\theta_m}(S_m \geq s_{m_\alpha}) = \beta \\ &= \lim_{m\to\infty} P_{\theta_m}(T_n \geq t_{n_\alpha}) = \lim_{m\to\infty} \beta_n(\theta_m)\end{aligned}$$

が成立すると仮定する．これらの条件のもとで $\lim\limits_{m\to\infty} \dfrac{n}{m}$ が存在し，α や β が無関係な定数であるとき

$$\mathrm{ARE_P}(S|T) = \lim_{m\to\infty} \frac{n}{m} \tag{2.2}$$

を検定統計量 S の検定統計量 T に対する Pitman の漸近相対効率と呼ぶ．

これは同じ検出力を持つためのサンプルサイズの比の極限である．Pitman の漸近相対効率 $\mathrm{ARE}(S|T)$ が 1 より大きいとき検定統計量 S のほうが検定統計量 T より良いと判断でき，1 より小さいときはその逆がいえる．しかし (2.2) 式をそのまま扱うことは困難である．以下の定理を用いると，正則条件のもとで Pitman の漸近相対効率を比較的容易に求めることができる．

定理 2.1
$\{S\}$, $\{T\}$ が後述する 4 つの正則条件を満たす検定統計量とする． このとき，Pitman の漸近相対効率 $\mathrm{ARE_P}(S|T)$ は

$$\mathrm{ARE_P}(S|T) = \left[\frac{e(S)}{e(T)}\right]^2 \tag{2.3}$$

と表せる．

(条件 1) 以下の式

$$\lim_{m\to\infty} P_\theta\left(\frac{S_m - \mu_m(\theta)}{\sigma_m(\theta)} \leq x\right) = \Phi(x), \qquad \lim_{m\to\infty} P_\theta\left(\frac{T_n - \mu_n(\theta)}{\sigma_n(\theta)} \leq x\right) = \Phi(x)$$

が成立し，この収束は $H_0 : \theta = 0$ の近傍で一様収束である．$\mu_m(\theta)$, $\sigma_m^2(\theta)$ は P_θ のもとでの $\{S_m\}$ の期待値と分散であり，同様に $\mu_n(\theta)$, $\sigma_n^2(\theta)$ は P_θ のもとでの $\{T_n\}$ の期待値と分散である．なお，$\Phi(x)$ を標準正規分布の分布関数とする．

(条件 2) 導関数

$$\mu'_m(\theta) = \frac{d\mu_m(\theta)}{d\theta}, \qquad \mu'_n(\theta) = \frac{d\mu_n(\theta)}{d\theta}$$

が存在し，かつ，$H_0 : \theta = 0$ で連続である．

(条件 3) 以下の式

$$\lim_{m\to\infty} \frac{\mu'_m(0)}{\sqrt{n\sigma_m^2(0)}} = e(S), \qquad \lim_{m\to\infty} \frac{\mu'_n(0)}{\sqrt{n\sigma_n^2(0)}} = e(T)$$

となる正の定数 $e(S), e(T)$ が存在する．この $e(S)$, $e(T)$ を効率とよぶ．

(条件 4) 以下の式

$$\theta_m = \frac{\delta}{\sqrt{m}}\,(\delta > 0), \qquad \lim_{m\to\infty}\frac{\sigma_m(\theta_m)}{\sigma_m(0)} = 1, \qquad \lim_{m\to\infty}\frac{\sigma_n(\theta_m)}{\sigma_n(0)} = 1$$

を満たす対立仮説 H_{1m} の系列 $\{\theta_m\}$ が存在する．

3　標本標準偏差と標本平均絶対偏差の漸近相対効率

本節では, Fisher (1920), Tukey (1961) による推定の漸近相対効率の観点からの標本標準偏差と標本平均絶対偏差における優劣の比較とその問題点について整理する．また, 検定における Pitman の漸近相対効率としてそれらが解釈可能であることを述べる．

3.1　標本標準偏差の標本平均絶対偏差に対する推定の漸近相対効率

母集団分布が正規分布の場合, 汚染正規分布, 指数分布のときの結果をここでは述べる．

定理 3.1

$(X_1, X_2, \dots, X_n)$ をある分布 F からの無作為標本とする．標本平均 $\overline{X}$, 標本標準偏差 S, 標本平均絶対偏差 D を

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}, \quad D = \frac{1}{n}\sum_{1=1}^{n}\left|X_i - \overline{X}\right| \tag{3.1}$$

とする．分布 F が正規分布 $N(\mu, \sigma^2)$ のとき, 漸近相対効率 ARE は

$$\mathrm{ARE} = \lim_{n\to\infty}\frac{V(S)/\left\{E[S]\right\}^2}{V(D)/\left\{E[D]\right\}^2} \approx \frac{1}{\pi - 2} \approx 0.876$$

で与えられる．また, 分布 F が汚染正規分布 $(1-\varepsilon)N(\mu, \sigma^2) + \varepsilon N(\mu, k^2\sigma^2)$, $(0 \le \varepsilon \le 1, k > 1)$ のとき, 漸近相対効率 ARE は

$$\mathrm{ARE}_{(\varepsilon)} = \lim_{n\to\infty}\frac{V(S)/\{E[S]\}^2}{V(D)/\{E[D]\}^2} \approx \frac{\frac{1}{4}\left\{\frac{3\{1+\left(k^4-1\right)\varepsilon\}}{\{1+(k^2-1)\varepsilon\}^2} - 1\right\}}{\frac{\pi\{1+(k^2-1)\varepsilon\}}{2\{1+(k-1)\varepsilon\}^2} - 1}$$

で与えられる．また, 分布 F が指数分布 $E_X(\lambda)$, $(\lambda > 0)$ のとき, 漸近相対効率 ARE は

$$\mathrm{ARE} = \lim_{n\to\infty}\frac{V(S)/\{E[S]\}^2}{V(D)/\{E[D]\}^2} \approx \frac{8}{e^2 - 4} \approx 2.36$$

で与えられる．

上記の定理からわかるように, ここで扱っている漸近相対効率は (2.1) 式の分散比と異なり, 変動係数の 2 乗の比である．

3.2 標本標準偏差の標本平均絶対偏差に対する Pitman の漸近相対効率

定理 3.1 で述べた漸近相対効率は, 検定問題における Pitman の漸近相対効率として解釈することができる.

定理 3.2

$(X_1, X_2, \ldots, X_n)$ をある分布 F からの無作為標本とする. 帰無仮説 $H_0 : \sigma = \sigma_0$, 対立仮説 $H_1 : \sigma > \sigma_0$ としたとき, (3.1) の $\{S\}$, $\{D\}$ を検定統計量として考える. 分布 F が正規分布のとき, (2.3) 式で定義される Pitman の漸近相対効率は

$$\mathrm{ARE_P}(S|D) \approx \pi - 2$$

で与えられる. また, 分布 F が汚染正規分布のとき,

$$\mathrm{ARE_P}(S|D) \approx \frac{4\left\{1+(k^2-1)\varepsilon\right\}^2}{3\left\{1+(k^4-1)\varepsilon\right\}-\left\{1+(k^2-1)\varepsilon\right\}^2} \Bigg/ \frac{2\left\{1+(k-1)\varepsilon\right\}^2}{\pi\left\{1+(k^2-1)\varepsilon\right\}-2\left\{1+(k-1)\varepsilon\right\}^2}$$

で与えられる. また, 分布 F が指数分布のとき,

$$\mathrm{ARE_P}(S|D) \approx \frac{e^2-4}{8}$$

で与えられる.

4 まとめと今後の課題

本論文では, 標本標準偏差と標本平均絶対偏差に関して, Eddington (1914), Fisher (1920), Tukey (1961) による議論とその問題点について理論的に整理した. さらに, 仮説検定問題として再定義して Pitman の漸近相対効率を用いることで, Fisher らの結論の解釈が可能であることを述べた. また, いくつかの母集団分布による検証を行った.

これらの視点からは, 母集団が正規分布から外れると, 標本平均絶対偏差のほうが優れていることが示唆される. 他の分布を含めてより詳細な議論が今後の課題である.

参考文献

[1] Eddington, A.S. (1914). *Stellar movements and the structure of the universe*, London: Macmillan and co.

[2] Fisher, R.A. (1920). A mathematical Examination of the Methods of determining the Accuracy of Observation by the Mean Error, and by the Mean Square Error. *Monthly Notices of the Royal Astronomical Society,* Volume 80, Issue 8, pp.758-770.

[3] 前園宣彦. (2001). 統計的推測の漸近理論. 九州大学出版会.

[4] Tukey, J.W. (1961). "A Survey of Sampling from Contaminated Distributions," in I. Olkin *et al.* (eds.), *Contributions to Probability and Statistics — Essays in Honor of Harold Hotelling,* pp.448-485.

[5] 柳川堯. (1982). ノンパラメトリック法. 培風館.