フィッシャー情報量とクラメル・ラオの下限

清家大嗣

2023 年 5 月 15 日

1　前提

パラメータ θ の分布関数 f から得られた, 標本 x についての尤度関数 $L(\theta; x)$ は次式で表される (ここで, 分布関数 f から得られた標本から, パラメータ θ を推測している. 因果が逆転していることに注意したい.).

$$L(\theta; x) = f(x; \theta) \tag{1}$$

また, 確率密度関数の性質から, 尤度関数は常に正の値を取るため, 以下の対数尤度関数を定義できる.

$$l(\theta; x) = \log L(\theta; x) \tag{2}$$

対数尤度関数をパラメーター一回微分した値は, スコア関数 $U(\theta; x)$ と呼ばれる (スコア関数が 0 になるパラメータ θ を選択するケースを授業ではよく学ぶ).

$$U(\theta; x) = \frac{\partial l(\theta; x)}{\partial \theta} \tag{3}$$

スコア関数の二乗の期待値をフィッシャー情報量 $I(\theta)$ と呼ぶ.

$$I(\theta) = E[U(\theta; x)^2] = \int_{x \in \Omega} \{\frac{\partial}{\partial \theta} \log f(x; \theta)\}^2 f(x; \theta) dx \tag{4}$$

ここで, Ω は標本空間である.

2　スコア関数の性質

2.1　期待値

定義式の通りに計算すると, 0 になることが分かる (標本空間全体に対する確率密度関数の積分結果は 1, 積分と微分の可換が成立することを仮定).

$$E[U(\theta; x)] = \int_{x \in \Omega} \frac{\partial \log f(x; \theta)}{\partial \theta} f(x; \theta) dx \tag{5}$$

$$= \int_{x \in \Omega} \frac{\partial f(x; \theta)}{\partial \theta} dx \tag{6}$$

$$= \frac{\partial}{\partial \theta} \int_{x \in \Omega} f(x; \theta) dx = \frac{\partial}{\partial \theta} 1 = 0 \tag{7}$$

2.2　分散

分散の公式と, 期待値が 0 であることより, フィッシャー情報量と一致する.

$$Var(U(\theta;x)) = E[U^2(\theta;x)] = \int_{x\in\Omega} \{\frac{\partial}{\partial\theta} \log f(x;\theta)\}^2 dx \tag{8}$$

$$= I(\theta) \tag{9}$$

3　クラメル・ラオの下限 (n 個の標本データから予測する不偏推定量の分散の下限評価)

3.1　前提

パラメータ θ の分布関数 f から得られる n 個の標本 $\boldsymbol{x} = (x_1, x_2, ..., x_n)^T$ について考える. この場合の結合確率密度 $f_n(\boldsymbol{x};\theta)$ は次式で与えられる (各データを確率変数とみなし, 互いに独立していると考える).

$$f_n(\boldsymbol{x};\theta) = f(x_1;\theta)\cdot f(x_2;\theta)\cdot \ \cdots \ \cdot f(x_n;\theta) \tag{10}$$

また, 対数尤度関数 $L_n(\theta;\boldsymbol{x})$ は次式で与えられる.

$$L_n(\theta;\boldsymbol{x}) = \log f_n(\boldsymbol{x};\theta) = \sum_{i=1}^{n} \log f(x_i;\theta) \tag{11}$$

従って, n 個分の標本のスコア関数は次式のようになる.

$$U_n(\theta;\boldsymbol{x}) = \frac{\partial L_n(\theta;\boldsymbol{x})}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(x_i;\theta) \tag{12}$$

和の期待値は期待値の和になり, 互いに独立な確率変数の和の分散は分散の和になる. 従って, スコア関数 $U_n(\theta;\boldsymbol{x})$ の期待値・分散はそれぞれ以下のようになる ($I(\theta)$ は, フィッシャー情報量).

$$E[U_n(\theta;\boldsymbol{x})] = 0 \tag{13}$$

$$Var[U_n(\theta;\boldsymbol{x})] = nI(\theta) \tag{14}$$

同様に, スコア関数の二乗の期待値をフィッシャー情報量 $I_n(\theta)$ と定義する.

$$I_n(\theta) = E[U_n(\theta;\boldsymbol{x})^2] = \int_{\boldsymbol{x}\in\Omega} \{\frac{\partial}{\partial\theta} \log f_n(\boldsymbol{x};\theta)\}^2 f_n(\boldsymbol{x};\theta) d\boldsymbol{x} \tag{15}$$

最後に, n 個の標本データから得られる θ の不偏推定量を $\hat{\theta}_n(\boldsymbol{x})$ で表記する. 不偏推定量のため, 下式を満たす.

$$\theta = E[\hat{\theta}_n(\boldsymbol{x})] \tag{16}$$

3.2　主張

パラメータ θ の不偏推定量 $\hat{\theta}_n(\boldsymbol{x})$ の分散の下限がフィッシャー情報量 $I(\theta)$ で評価できる.

$$Var[\hat{\theta}_n(\boldsymbol{x})] \geq \frac{1}{I(\theta)} \tag{17}$$

3.3 証明

不偏推定量の定義より, 次式が成立する.

$$\theta = E[\hat{\theta}_n(\boldsymbol{x})] = \int_{\boldsymbol{x}\in\Omega^n} \hat{\theta}_n(\boldsymbol{x}) f_n(\boldsymbol{x};\theta) d\boldsymbol{x} \tag{18}$$

両辺を θ で微分して整理する (各標本データを互いに独立な確率変数と見做していることを 2 行目の変形で利用).

$$1 = \int_{\boldsymbol{x}\in\Omega^n} \hat{\theta}_n(\boldsymbol{x}) \left\{ \frac{\partial}{\partial\theta} f_n(\boldsymbol{x};\theta) \right\} d\boldsymbol{x} \tag{19}$$

$$= \int_{\boldsymbol{x}\in\Omega^n} \hat{\theta}_n(\boldsymbol{x}) \left\{ \frac{\partial}{\partial\theta} \log f_n(\boldsymbol{x};\theta) \right\} f_n(\boldsymbol{x};\theta) d\boldsymbol{x} \tag{20}$$

従って, スコア関数の定義 (3 式) より, 以下の式が成立する.

$$1 = E[\hat{\theta}_n(\boldsymbol{x}) U_n(\theta;\boldsymbol{x})] \tag{21}$$

$U_n(\theta;\boldsymbol{x})$ は, 各標本データ x_i が互いに独立な確率変数であるため, 期待値は 0 となる. それ故, 次式が成り立つ (θ は, 母数であるため定数).

$$1 = E[(\hat{\theta}_n(\boldsymbol{x}) - \theta) U_n(\theta;\boldsymbol{x})] \tag{22}$$

コーシー・シュワルツの不等式 (この証明は, $\sum_{i=1}^{n}(a_i x - b)^2 \geq 0$ の判別式を利用することで簡単に与えられる) を利用して, 次式が成立する.

$$E[(\hat{\theta}_n(\boldsymbol{x}) - \theta)^2] E[U_n(\theta;\boldsymbol{x})^2] \geq E[(\hat{\theta}_n(\boldsymbol{x}) - \theta) U_n(\theta;\boldsymbol{x})]^2 = 1 \tag{23}$$

これは, 次式の成立を主張していることに他ならない.

$$Var[\hat{\theta}_n(\boldsymbol{x})] = E[(\hat{\theta}_n(\boldsymbol{x}) - \theta)^2] \geq \frac{1}{E[U_n(\theta;\boldsymbol{x})^2]} = \frac{1}{I_n(\theta)} \tag{24}$$

4 フィッシャー情報量の公式

$$\frac{\partial^2}{\partial^2\theta} \log f(x;\theta) = \frac{\partial}{\partial\theta} \left\{ \frac{1}{f(x;\theta)} \frac{\partial f(x;\theta)}{\partial\theta} \right\} \tag{25}$$

$$= -\frac{1}{\{f(x;\theta)\}^2} \{ \frac{\partial f(x;\theta)}{\partial\theta} \}^2 + \frac{1}{f(x;\theta)} \frac{\partial^2 f(x;\theta)}{\partial\theta^2} \tag{26}$$

$$= -\{ \frac{\partial \log f(x;\theta)}{\partial\theta} \}^2 + \frac{1}{f(x;\theta)} \frac{\partial^2 f(x;\theta)}{\partial\theta^2} \tag{27}$$

上式の平均を取ることで, 次の公式が得られる (第二項が 0 になるのは, 積分と微分の順序が可換である場合のみを考えているため).

$$E[\frac{\partial^2}{\partial^2\theta} \log f(x;\theta)] = -I(\theta) + \int_{x\in\Omega} \frac{\partial f(x;\theta)}{\partial\theta^2} dx \tag{28}$$

$$= -I(\theta) \tag{29}$$

5 クラメル・ラオ限界の応用例

母数の推定値のバラツキは小さい程よい. これは, 自明と考えられる. 24 式の主張は, このバラツキ改善に限界があり, それはフィッシャー情報量 (スコア関数の二乗値) に依存するということである (つまり, 元々の分布関数に依存).

具体例として, 互いに独立に正規分布に従う n 個分の標本データに関する, 母平均 μ を標本平均 $\bar{x}$ で推定することを考える. つまり, 先ほどの $\theta \to \mu, \hat{\theta} \to \bar{x}$ としたケースである. n 標本データが得られた場合のフィッシャー情報量 $I_n(\theta)$ は次式で与えられる (28, 29 式を利用).

$$I_n(\theta) = nI(\theta) = -nE[\frac{\partial^2}{\partial\theta^2} \log f(x;\theta)] \tag{30}$$

$$= n \cdot \frac{1}{\sigma^2} \tag{31}$$

従って,

$$Var[\hat{\mu}] \geq \frac{1}{I_n(\theta)} = \frac{\sigma^2}{n} \tag{32}$$

つまり, 標本平均の分散が σ^2/n であるという事実から, 推定量の分散がクラメールラオの下限となっていることが分かる. この条件を満たすとき, 推定量は漸近有効であるという.

参考文献

なし