応用統計学 2017 第6回 不偏推定，クラメル・ラオの不等式

2017年11月1日(水)

清 智也 sei@mist.i.u-tokyo.ac.jp

http://ur0.pw/yTzt

- 推定量，不偏推定量，平均二乗誤差，最小分散不偏推定量，フィッシャー情報量，クラメル・ラオの不等式（情報量不等式），有効推定量[1]。

演習問題

問題 6-1. $X_1,\dots,X_n$ は平均 μ，分散 σ^2 の（正規分布とは限らない）独立同一分布に従う確率変数列とする。ただし μ,σ^2 は未知のパラメータである。

(i) 標本平均 $\bar{X}=n^{-1}\sum_{t=1}^n X_t$ は μ の不偏推定量であることを示せ。

(ii) 不偏標本分散

$$\hat{\sigma}^2=\frac{1}{n-1}\sum_{t=1}^n (X_t-\bar{X})^2$$

は σ^2 の不偏推定量であることを示せ。

なお，$\bar{X}$ の母標準偏差 $\sqrt{\sigma^2/n}$ の推定量 $\sqrt{\hat{\sigma}^2/n}$ のことを，$\bar{X}$ の標準誤差という。これは以前扱ったモンテカルロ法の標準誤差と数学的には同じ概念である[2]。ただし，$X_1,\dots,X_n$ が「観測データ」か「自分で生成できる乱数」か，という点で根本的に異なる。

問題 6-2 (重回帰モデル)**.** $\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(n)}$ はそれぞれ p 次元の列ベクトルとし，それらを並べてできる $n\times p$ 行列 $\boldsymbol{X}=(\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(n)})'$ はランク p であるとする。また，$Y_1,\dots,Y_n$ は

$$Y_t=\boldsymbol{\beta}'\boldsymbol{x}^{(t)}+\varepsilon_t,\quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0,\sigma^2)$$

により定まる確率変数列とする。ここで $\boldsymbol{\beta}\in\mathbb{R}^p$ と $\sigma^2>0$ は未知のパラメータである。最小二乗法により求められる回帰係数 $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の不偏推定量であることを示せ。また

$$\hat{\sigma}^2=\frac{1}{n-p}\sum_{t=1}^n (Y_t-\hat{\boldsymbol{\beta}}'\boldsymbol{x}_{(t)})^2$$

は σ^2 の不偏推定量であることを示せ。

なお，任意の $\boldsymbol{x}\in\mathbb{R}^p$ に対して $\hat{\boldsymbol{\beta}}'\boldsymbol{x}$ が $\boldsymbol{\beta}'\boldsymbol{x}$ の最小分散不偏推定量であることが，多次元版のクラメル・ラオの不等式から示される。実は $\hat{\sigma}^2$ も σ^2 の最小分散不偏推定量であることが知られている[3]。

[1] estimator, unbiased estimator, mean square error, minimum variance unbiased estimator, Fisher information, Cramér-Rao inequality (information inequality), efficient estimator.

[2] 以前は不偏性を気にしていなかったので $\hat{\sigma}^2$ の式の分母が n であった。この違いはあまり本質的ではない。

[3] レーマン・シェッフェの定理による。たとえば宮川「統計技法」の p.56 を参照せよ。

問題 6-3. 以下の各統計モデルのフィッシャー情報量を求めよ。

(i) 分散 1 の正規分布：$f(x;\theta) = (2\pi)^{-1/2} e^{-(x-\theta)^2/2}$.

(ii) ポアソン分布：$f(x;\theta) = (\theta^x/x!)e^{-\theta}$.

(iii) 中央値 θ のコーシー分布：$f(x;\theta) = \pi^{-1}(1+(x-\theta)^2)^{-1}$.

問題 6-4. X の確率密度関数を $f(x;\theta)$ とする。また $\partial/\partial\theta$ を ∂_θ と略記する。次の各等式を示せ：

$$\mathrm{E}\left[\partial_\theta \log f(X;\theta)\right] = 0, \qquad \mathrm{E}\left[\{\partial_\theta \log f(X;\theta)\}^2\right] = \mathrm{E}\left[-\partial_\theta^2 \log f(X;\theta)\right].$$

ただし，微積分の順序交換が可能であると仮定する[4]。

問題 6-5. 密度 $f(x;\theta)$ と $f(x;\phi)$ の間の Kullback-Leibler ダイバージェンス（分離度）は

$$\mathrm{KL}(\theta,\phi) = \int f(x;\theta) \log \frac{f(x;\theta)}{f(x;\phi)} dx$$

と定義される[5]。$\mathrm{KL}(\theta,\phi)$ は非負値であること，および

$$\mathrm{KL}(\theta,\theta+\delta) = \frac{1}{2} I(\theta)\delta^2 + \mathrm{o}(\delta^2), \quad \delta \to 0,$$

が成り立つことを示せ。フィッシャー情報量はこの意味で分布の離れ具合を表す量である。

問題 6-6. パラメータ θ は整数で，X は $\theta-1, \theta, \theta+1$ を等確率 $1/3$ で取る確率変数とする。

(i) $\hat{\theta}(X) = X$ は θ の不偏推定量であることを示せ。

(ii) θ の最小分散不偏推定量は存在しないことを示せ。

宿題 6

問題 6-7. 推定量の不偏性はパラメータの変換について不変 (invariant) でないこと，すなわち $\mathrm{E}[\hat{\theta}] = \theta$ であっても，$\mathrm{E}[h(\hat{\theta})] \neq h(\theta)$ となるような関数 h があることを，具体例によって確認せよ。

[4]数学的に気になる人は，成り立つための条件をルベーグの収束定理を援用して書き下してみるとよい。
[5]相対エントロピーとも呼ばれる。