2013年後期
応用確率統計

⑩　最尤推定量

河野 行雄
kawano@pe.titech.ac.jp

2013年12月19日

復習

■ 誤差分布未知のときの最小分散推定量

$$\hat{\theta}^* = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

■ 誤差分布既知のときの最小分散位置共変推定量

$$\hat{\theta}^* = \frac{\int \theta \prod_{i=1}^{n} f(X_i - \theta) d\theta}{\int \prod_{i=1}^{n} f(X_i - \theta) d\theta}$$

■ 不偏分散

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

推定量の分類

■ 観測データ

$$X_i = \theta + \varepsilon_i \qquad (i = 1, \cdots, n)$$

誤差の分布：$f(\varepsilon_i)$　平均：$E(\varepsilon_i) = 0$　分散：$V(\varepsilon_i) = \sigma^2$

■ 推定量の分類

最適（不偏・分散最小）な推定量

誤差の分布：
- 未知：$\hat{\theta}^* =$ 標本平均
- 既知：$\hat{\theta}^* =$ ピットマン推定量

誤差の分布既知のときの一般的な推定量

最尤推定量 ⟶ 漸近不偏、漸近有効（分散最小）

最尤推定量

■ 観測値の同時確率密度関数

$$f(X_1, X_2, \cdots X_n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

確率変数 X_i の確率密度関数 $P(X_i \mid \theta)$ を θ の関数として $f(X_i, \theta)$ と表記

■ 尤度関数

$$L(\theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

■ 最尤推定量

$$\hat{\theta}^* = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \prod_{i=1}^{n} f(X_i, \theta)$$

$$= \arg\max_{\theta} \log L(\theta) = \arg\max_{\theta} \sum_{i=1}^{n} \log f(X_i, \theta)$$

最尤推定の例

■ 問題

硬貨を100回投げて57回表が出たとき二項分布における a を求めよ

二項分布: $f(m,a) = {}_nC_m a^m (1-a)^{n-m}$

■ 最尤推定

$$L(a) = {}_{100}C_{57}\, a^{57}(1-a)^{43}$$

$$\log L(a) = \log\frac{100!}{57!43!} + 57\log a + 43\log(1-a)$$

$$\frac{\partial}{\partial a}\log L(a) = \frac{57}{a} - \frac{43}{1-a} = 0 \longrightarrow \hat{a} = \frac{57}{100}$$

フィッシャー情報量

■ フィッシャー情報量

最尤推定量 $\hat{\theta}^*$ と真の母数 θ_0 の関係を表す量

■ 最尤推定量の性質

最尤推定の性質

$$\frac{\partial}{\partial\theta}\log L(\hat{\theta}^*) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i,\hat{\theta}^*) = 0$$

$\hat{\theta}^* = \theta_0$ でテイラー展開

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i,\theta_0) + (\hat{\theta}^* - \theta_0)\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(X_i,\theta_0) + \cdots = 0$$

フィッシャー情報量

$$I = \lim\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(X_i,\theta_0)\right) = E\left(-\frac{\partial^2}{\partial\theta^2}\log f(X,\theta_0)\right)$$

フィッシャー情報量の変形

■ $\int f(x,\theta)dx = 1$ の変形

$$0 = \int\frac{\partial}{\partial\theta}f(x,\theta)dx = \int\frac{\frac{\partial}{\partial\theta}f(x,\theta)}{f(x,\theta)}f(x,\theta)dx$$

$$= \int\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)f(x,\theta)dx = E\left(\frac{\partial}{\partial\theta}\log f(X,\theta)\right)$$

$$0 = \int\left(\frac{\partial^2}{\partial\theta^2}\log f(x,\theta)\right)f(x,\theta)dx + \int\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)\frac{\partial}{\partial\theta}f(x,\theta)dx$$

$$= \int\left(\frac{\partial^2}{\partial\theta^2}\log f(x,\theta)\right)f(x,\theta)dx + \int\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2 f(x,\theta)dx$$

$$\Downarrow$$

$$I = E\left(\left(\frac{\partial}{\partial\theta}\log f(X,\theta_0)\right)^2\right) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(X,\theta_0)\right)$$

漸近有効性

■ 最尤推定量のテイラー展開

$$\sqrt{n}(\hat{\theta}^* - \theta_0)\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(X_i,\theta_0)\right) = \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i,\theta_0) + \cdots$$

$$E\left(\frac{\partial}{\partial\theta}\log f(X_i,\theta_0)\right) = 0 \quad V\left(\frac{\partial}{\partial\theta}\log f(X_i,\theta_0)\right) = E\left(\left(\frac{\partial}{\partial\theta}\log f(X_i,\theta_0)\right)^2\right) = I$$

$\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i,\theta_0) \xrightarrow{n\to\infty}$ 平均 0 分散 I の正規分布

$\sqrt{n}(\hat{\theta}^* - \theta_0) \xrightarrow{n\to\infty}$ 平均 0 分散 $\frac{I}{I^2} = \frac{1}{I}$ の正規分布

■ 漸近有効性

漸近不偏

$$E(\hat{\theta}^* \mid \theta_0) = \theta_0 + o\left(\frac{1}{\sqrt{n}}\right)$$

漸近有効

$$V(\hat{\theta}^* \mid \theta_0) = \frac{1}{nI} + o\left(\frac{1}{n}\right)$$

クラメール・ラオの不等式

■ フィッシャー情報量

$$I = E\left(\left(\frac{\partial}{\partial\theta}\log f(X,\theta_0)\right)^2\right) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(X,\theta_0)\right)$$

■ 推定量の分散の下界

$$V(\hat{\theta}\,|\,\theta_0) \ge \frac{1}{nI}$$

← 不偏性の定義式とシュワルツの不等式から証明

⇩

最尤推定量は推定量の下界を漸近的に達成（漸近有効）

正規分布の尤度関数

■ 観測データ

$X_1, \cdots X_n$ は i.i.d. で平均 m 分散 σ^2 の正規分布

■ 尤度関数

$$f(X_i, m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(X_i - m)^2}{2\sigma^2}\right)$$

$$\log f(X_i, m, \sigma^2) = -\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}(X_i - m)^2$$

$$\tilde{L}(m, \sigma^2) = \sum_{i=1}^{n}\log f(X_i, m, \sigma^2) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - m)^2$$

平均の最尤推定量

■ 母平均の推定

m : 未知母平均、 $\sigma^2 = 1$ のとき

$$\tilde{L}(m) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}(X_i - m)^2$$

■ 最尤推定

$$\frac{\partial}{\partial m}\tilde{L}(m) = \sum_{i=1}^{n}(X_i - m) = \sum_{i=1}^{n}X_i - nm = 0$$

$$\longrightarrow \quad \hat{m} = \frac{1}{n}\sum_{i=1}^{n}X_i = \bar{X}$$

← 標本平均に一致

分散の最尤推定量

■ 母分散の推定

$m = 0$ 、 σ^2 : 未知母分散のとき

$$\tilde{L}(\sigma^2) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}X_i^2$$

■ 最尤推定

$$\frac{\partial}{\partial\sigma^2}\tilde{L}(\sigma^2) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}X_i^2 = 0$$

$$\longrightarrow \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2$$

← 標本分散に一致（漸近不偏）

まとめ

- 尤度関数

$$L(\theta)=\prod_{i=1}^{n} f(X_i,\theta)$$

- 最尤推定量

$$\hat{\theta}^* = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \log L(\theta)$$

← 漸近不偏、漸近有効（クラメールラオの下界を達成）

- クラメール・ラオの不等式

$$V\left(\hat{\theta} \mid \theta_0\right) \geq \frac{1}{nI}$$

$$I = E\left(\left(\frac{\partial}{\partial\theta}\log f\left(X,\theta_0\right)\right)^2\right) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f\left(X,\theta_0\right)\right)$$