

第2章付録: 最尤推定法

最終更新：2025年11月3日

上武康亮・遠山祐太・若森直樹・渡辺安虎

「実証ビジネス・エコノミクス」

第2章「消費者の購買パターンを浮き彫りにする」

の付録

© 上武康亮・遠山祐太・若森直樹・渡辺安虎

1 最尤法の解説

本付録においては最尤法 (Maximum Likelihood Estimation, MLE) の概略について解説を行う。最尤法は第2章で扱う離散選択モデルを含め、非線形モデルのパラメタ推定において幅広く用いられている手法である。詳細な解説については以下の教科書等を参考にされたい。

- 末石直也 (2015) 『計量経済学：マイクロデータ分析へのいざない』 日本評論社。
- Hayashi, F (2000) *Econometrics*, Princeton University Press.
- Newey, W. K. and McFadden, D. (1994) “Large Sample Estimation and Hypothesis Testing,” *Handbook of Econometrics*, Vol.4, North Holland: 2111–2245.

1.1 最尤推定量の導入

ある確率変数 Y を考えよう。この確率変数 Y は密度関数 $f(y; \theta)$ を持ち、 θ は未知パラメタとする。今、データとして $\{y_i\}_{i=1}^n$ が存在し、各観測 y_i は独立かつ同一 (i.i.d.) に密度関数 $f(y; \theta_0)$ から生成されているとしよう。ここで、 θ_0 は真のパラメタであり、データからこのパラメタを推定することが目標である。

ここで、手元のデータ $\{y_i\}_{i=1}^N$ を観察する確率をパラメタ θ の関数として表現する尤度関数を考えよ

う。尤度関数 $L(\theta)$ は以下のように定義される。

$$L(\theta) \equiv \prod_{i=1}^n f(y_i; \theta).$$

また、尤度関数について自然対数を取った対数尤度を以下のように定義する

$$l(\theta) \equiv \log L(\theta) = \sum_{i=1}^n \log f(y_i; \theta).$$

最尤推定量 (Maximum Likelihood Estimator) は、対数尤度を最大化するようなパラメタ θ として定義される。

$$\hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \log f(y_i; \theta).$$

1.2 最尤推定量の例

最尤推定量に関するいくつかの例を見ていこう。

1.2.1 コイントス

確率変数 Y として、確率 p で 1, 確率 $1 - p$ で 0 をとるものを考える。すなわち、 $P(Y = 1) = p, P(Y = 0) = 1 - p$ とする。今、データとして、 $\{y_i\}_{i=1}^n$ があるとしよう。このとき、尤度関数は

$$L(\theta) = \prod_{i=1}^n P(Y = y_i) = p^{N_1} (1 - p)^{n - N_1}$$

と与えられる。ここで、 N_0, N_1 はそれぞれ、 $Y = 0, Y = 1$ となるような観測の個数である。すなわち、 $N_0 = \sum_{i=1}^n 1\{y_i = 0\}, N_1 = \sum_{i=1}^n 1\{y_i = 1\}$ である。

対数尤度関数は

$$l(\theta) = N_1 \log(p) + (n - N_1) \log(1 - p)$$

となる。この対数尤度関数をパラメタ p について微分して一階条件を解くと、

$$\hat{p} = \frac{N_1}{n}$$

が得られる。これは、 n 回の試行の中で表が出た回数が N_1 のとき、表が出る確率の推定量を、表が出

た回数の比率として与えられることを意味する。

1.2.2 正規分布の平均と分散

確率変数 $y \sim N(\mu, \sigma^2)$ を考える。未知パラメタを $\theta = (\mu, \sigma^2)$ とする。今、正規分布から i.i.d. に生成されたデータセット $\{y_i\}_{i=1}^n$ があるとき、尤度関数は以下のように与えられる。

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right).$$

また対数尤度は

$$\begin{aligned} \log L(\theta) &= \sum_{i=1}^n \left\{ -\frac{1}{2} \log 2\pi\sigma^2 - \frac{(y_i - \mu)^2}{2\sigma^2} \right\} \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2}. \end{aligned}$$

対数尤度を最大化するようなパラメタについて一回条件を用いて求めよう。

$$\begin{aligned} \frac{\partial \log L(\theta)}{\partial \sigma^2} &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 = 0, \\ \frac{\partial \log L(\theta)}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) = 0. \end{aligned}$$

これらの式を解くと、最尤推定量は

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n y_i, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2. \end{aligned}$$

となる。ここで $\hat{\mu}$ は標本平均に一致し、また $\hat{\sigma}^2$ は標本分散に一致する。

1.2.3 誤差項に正規分布を仮定した線形回帰モデル

以下の線形回帰モデルを考えよう。

$$y_i = x_i' \beta + u_i, \quad u_i \sim N(0, \sigma^2).$$

なお、 $x_i = (x_{1i}, \dots, x_{Ki})'$, $\beta = (\beta_1, \dots, \beta_K)'$ とする。ここでは誤差項に正規分布を仮定している。推定するパラメタは $\theta = (\beta, \sigma^2)$ としてまとめられる。

ここでは説明変数 x_i について条件付をした、被説明変数 y_i に関する尤度関数を考えよう。

$$L(\theta|x_i) = \prod_{i=1}^n f(y_i; \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i'\beta)^2}{2\sigma^2}\right).$$

対数尤度

$$\begin{aligned} \log L(\theta|x_i) &= \sum_{i=1}^n \left\{ -\frac{1}{2} \log 2\pi\sigma^2 - \frac{(y_i - x_i'\beta)^2}{2\sigma^2} \right\} \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \sum_{i=1}^n \frac{(y_i - x_i'\beta)^2}{2\sigma^2}. \end{aligned}$$

対数尤度を最大化するべく、一回条件を考えると

$$\begin{aligned} \frac{\partial \log L(\theta)}{\partial \sigma^2} &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - x_i'\beta)^2 = 0, \\ \frac{\partial \log L(\theta)}{\partial \beta} &= \frac{1}{\sigma^2} \sum_{i=1}^n x_i(y_i - x_i'\beta) = 0. \end{aligned}$$

ここで回帰係数 β に着目すると

$$\begin{aligned} \hat{\beta}^{MLE} &= \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \sum_{i=1}^n x_i y_i \\ &= \hat{\beta}^{OLS}. \end{aligned}$$

となり、最小二乗推定量と同じものが得られる。

1.3 最尤推定量の性質

最尤推定量は一定の仮定のもとで、一致性や漸近正規性などの漸近的性質を持つ。以下ではパラメタを K 次元のベクトル $\theta = (\theta_1, \dots, \theta_K)' \in \mathbb{R}^K$ とする。

まず一致性として、最尤推定量はサンプルサイズが大きくなるに連れて真のパラメタ θ_0 へ確率収束する。

$$\hat{\theta} \xrightarrow{p} \theta_0$$

また漸近正規性として以下の式が成立する

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \underbrace{V}_{(K \times K)}),$$

ここで、 $(K \times K)$ 行列 V は $V = H^{-1}$ であり、

$$H = -E \left[\frac{\partial^2 \log f(y_i; \theta)}{\partial \theta \partial \theta'} \bigg|_{\theta = \theta_0} \right]$$

として与えられる。行列 H は期待ヘシアンと呼ばれる。なお、ヘシアンの中身は

$$\underbrace{\frac{\partial^2 \log f(y_i; \theta)}{\partial \theta \partial \theta'}}_{(K \times K)} = \begin{pmatrix} \frac{\partial^2 \log f(y_i; \theta)}{\partial \theta_1 \partial \theta_1} & \dots & \frac{\partial^2 \log f(y_i; \theta)}{\partial \theta_1 \partial \theta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \log f(y_i; \theta)}{\partial \theta_K \partial \theta_1} & \dots & \frac{\partial^2 \log f(y_i; \theta)}{\partial \theta_K \partial \theta_K} \end{pmatrix}$$

となる。^{*1}

この結果にもとづいて、漸近分散の計算を以下のように行うことができる。まず、漸近分散の推定量として以下の \hat{A} を考える。

$$\begin{aligned} \hat{A} &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta \partial \theta} \log f(y_i, \hat{\theta}) \\ &= -\frac{\partial^2}{\partial \theta \partial \theta} \frac{1}{N} l(\hat{\theta}) \end{aligned}$$

これは目的関数 (をサンプルサイズで割った値) のヘシアンに相当する。なお、目的関数 (対数尤度) のヘシアンは、数値計算などにおいて付随的に計算されることが多い。^{*2}

よって漸近分散の推定量 $ase(\hat{\theta})$ は、

$$ase(\hat{\theta}) = \sqrt{\text{diag}\left(\frac{1}{n} \hat{A}^{-1}\right)}$$

^{*1} なお、分散共分散行列 V はしばしば、 $V = H^{-1} I H^{-1}$ とも与えられる。ここで、行列 I は $\underbrace{I}_{(K \times K)} =$

$$E \left[\frac{\partial \log f(y_i; \theta)}{\partial \theta} \bigg|_{\theta = \theta_0} \frac{\partial \log f(y_i; \theta)}{\partial \theta'} \bigg|_{\theta = \theta_0} \right] = E \left(\begin{pmatrix} \frac{\partial \log f(y_i; \theta)}{\partial \theta_1} \frac{\partial \log f(y_i; \theta)}{\partial \theta_1} & \dots & \frac{\partial \log f(y_i; \theta)}{\partial \theta_1} \frac{\partial \log f(y_i; \theta)}{\partial \theta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial \log f(y_i; \theta)}{\partial \theta_K} \frac{\partial \log f(y_i; \theta)}{\partial \theta_1} & \dots & \frac{\partial \log f(y_i; \theta)}{\partial \theta_K} \frac{\partial \log f(y_i; \theta)}{\partial \theta_K} \end{pmatrix} \right) \text{である。一定}$$

の条件下 (定式化の正しさ及びいくつかの正則条件) では、情報等式 (information equality) $I = H^{-1}$ が成立するため、 $V = H^{-1}$ となる。

^{*2} 第 2 章における離散選択モデルの推定のコードを参照されたい。

として与えられる。ここで、 $\text{diag}(X)$ は行列 X の対角要素のみを抜き出したベクトルである。