補足：最尤法入門

講師：遠山祐太

最終更新：2024-12-16

最尤法

セットアップ

- 確率変数 y を考える。その密度関数を $f(y;\theta)$ とし、 θ は未知パラメタ。
- 密度関数は正しく定式化されていると仮定する。
- データとして $\{y_i\}_{i=1}^n$ が得られ、これは独立かつ同一 (i.i.d.) に $f(y;\theta_0)$ から生成されている。
    - θ_0 は未知のパラメタ
- ゴール: 未知パラメタ θ_0 を データ $\{y_i\}_{i=1}^n$ を用いて推定する。

最尤推定量 (Maximum Likelihood Estimator, MLE)

- データ $\{y_i\}_{i=1}^n$ に関する尤度関数を以下のように定義する。

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

  - データ $\{y_i\}_{i=1}^n$ を観測する確率を パラメタ θ の関数として表したもの。

- 対数を取った**対数尤度関数**を用いることが多い。

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

- 最尤推定量を以下のように定義する。

$$\hat{\theta}_n = \mathrm{argmax}_{\theta \in \Theta} \log L(\theta).$$

例1:コイン投げ

- 以下の確率変数 Y_i を考える。

$$Y_i = \begin{cases} 1 & with\ prob\ p \\ 0 & with\ prob\ 1-p \end{cases}$$

- データとして N 回のコイン投げの結果がわかるとしよう。 $\{y_i\}_{i=1}^N$

- 尤度関数は

$$L(\theta) = \prod_{i=1}^{n} P(Y = y_i; p) = \prod_{i=1}^{n} p^{1\{y_i=1\}}(1-p)^{1\{y_i=0\}}$$

例 1 の対数尤度

- 対数尤度関数は

$$
\begin{aligned}
\log L(\theta) &= \sum_{i=1}^{n} \log \left\{ p^{1\{y_i=1\}} (1-p)^{1\{y_i=0\}} \right\} \\
&= \sum_{i=1}^{n} 1\{y_i = 1\} \log p + 1\{y_i = 0\} \log(1-p) \\
&= N_1 \log p + N_0 \log(1-p)
\end{aligned}
$$

ここで、 $N_1 = \sum_{i=1}^{N} \mathbf{1}\{y_i = 1\}, N_0 = \sum_{i=1}^{N} \mathbf{1}\{y_i = 0\}$

- 練習問題：対数尤度を最大にする $\hat{p}_{MLE}$ を求めよう。

例2: 正規分布

- $y \sim N(\mu, \sigma^2)$. 未知パラメタを $\theta = (\mu, \sigma^2)$ とする。
- 尤度関数は

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

正規確率変数における最尤推定量

- 一階条件

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

$$\frac{\partial \log L(\theta)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) = 0.$$

- 最尤推定量は

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ (sample mean)},$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \text{ (sample variance)}.$$

例3: 誤差項が正規分布に従う線形回帰

- 線形回帰モデル

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

- 説明変数 x で条件付けた尤度は

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

対数尤度

- 対数尤度関数は

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

- 対数尤度関数を最大にするパラメタを考えよう。

正規性の元ではMLE = OLS

- 一階条件は

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

$$
\frac{\partial \log L(\theta)}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(y_i - x_i'\beta) = 0.
$$

変形すると

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

最尤推定量の性質

- 最尤推定量の定義：

$$\hat{\theta} = \mathrm{argmax}_{\theta \in \Theta} \log L(\theta | x_i).$$

- 最尤推定量も一定の仮定のもとで、以下のような性質を満たす。
  - 一致性 (Consistency): As the sample size N gets larger, the estimator is getting closer to the true value θ_0
  - 漸近正規性 (Asymptotic normality): The distribution of $\hat{\theta}$ is approximated by a normal distribution as the sample size gets larger.

一致性

- サンプルサイズが大きくなるに連れて、真のパラメタに収束する。

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

漸近正規性

- パラメタベクトル $\theta = (\theta_1, \ldots, \theta_K)' \in \mathbb{R}^K$.
- 漸近正規性:

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

- 漸近分散行列は $V = A^{-1}BA^{-1}$

$$\underbrace{A}_{(K\times K)} = -E\left[\left.\frac{\partial^2 \log f(y_i;\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0}\right],$$

$$\underbrace{B}_{(K\times K)} = E\left[\left.\frac{\partial \log f(y_i;\theta)}{\partial\theta}\right|_{\theta=\theta_0} \left.\frac{\partial \log f(y_i;\theta)}{\partial\theta'}\right|_{\theta=\theta_0}\right]$$

漸近分散行列の中身の詳細

- 中身を詳しく見ると

$$
\underbrace{\left(\frac{\partial \log f(y_i;\theta)}{\partial \theta}\right)}_{(K\times 1)}\underbrace{\left(\frac{\partial \log f(y_i;\theta)}{\partial \theta'}\right)}_{(1\times K)} = \begin{pmatrix} \frac{\partial \log f(y_i;\theta)}{\partial \theta_1}\frac{\partial \log f(y_i;\theta)}{\partial \theta_1} & \cdots & \frac{\partial \log f(y_i;\theta)}{\partial \theta_1}\frac{\partial \log f(y_i;\theta)}{\partial \theta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial \log f(y_i;\theta)}{\partial \theta_K}\frac{\partial \log f(y_i;\theta)}{\partial \theta_1} & \cdots & \frac{\partial \log f(y_i;\theta)}{\partial \theta_K}\frac{\partial \log f(y_i;\theta)}{\partial \theta_K} \end{pmatrix}
$$

漸近正規性の解釈

- 単純化のためにパラメタが一次元のケースを考えよう。

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N(0, V)$$

  - 確率変数 $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$ が、サンプルサイズ n が大きくなるにつれて正規分布へ収束する。

- **近似的**に $\hat{\theta} \sim N(\theta_0, V)$ と扱う。

- 用語の整理：
  - 確率変数 $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$ の漸近分散 (Asymptotic variance) は V
  - 確率変数 $\hat{\theta}$ の漸近分散は V/n で与えられる。