最尤法について†

担当: 長倉 大輔
(ながくらだいすけ)

† この資料は私のゼミおよび講義において使用するために作成した資料です。WEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし間違いがあった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。間違いは発見次第、継続的に直していますが、まだ存在する可能性があります。

最尤法について

- 最尤法

最尤法とはモデルの未知パラメーターの推定法の一つであり、その適用範囲が広いため統計分析において非常によく用いられる方法である。

ここでは最尤法の性質についてやや理論的な測面も含めて解説する。

- 最尤法の定義

最尤法とは一言で定義するなら**尤度関数**を**最大化**する未知パラメーターの値を推定値とする推定方法である。そこでまず尤度関数について述べる。

最尤法について

分析される変数は $n \times 1$ の確率ベクトル

$$\mathbf{y}_n = [y_1, y_2, \ldots, y_n]^{\mathrm{T}}$$

であるとする。$\mathbf{y}_n$ は**結合確率密度関数**(もしくは$\mathbf{y}_n$が離散型確率変数の場合は**結合確率関数**)

$$f(\mathbf{y}_n ; \boldsymbol{\theta}) = f(y_1, y_2, \ldots, y_n; \boldsymbol{\theta})$$

を持つとする。ここで $\boldsymbol{\theta}$ はこの密度関数を特徴づける $p \times 1$ の**未知パラメーターのベクトル**である。

この時、**尤度関数**とはこの結合密度関数をこの未知パラメーターベクトル $\boldsymbol{\theta}$ の関数と見なしたものである。

すなわち、尤度関数を $L_n(\boldsymbol{\theta})$ と表すとすると

$$L_n(\boldsymbol{\theta}) = f(\mathbf{y}_n; \boldsymbol{\theta})$$

である。

最尤法について

- 最尤推定と最尤推定量

最尤推定値とは観測値 $\mathbf{y}_n$ が与えられた時に、尤度関数を最大化する未知パラメーター θ の値であり、**最尤推定量**とはそれを与える観測値 $\mathbf{y}_n$ の**関数**のことである。最尤推定量は確率変数であり、最尤推定値はその実現値のことである。

最尤法について

- 例1: 正規分布

確率変数 y_i , i =1,…,n は i.i.d. 確率変数で $y_i \sim N(\mu, \sigma^2)$ であるとし、未知パラメータ $\boldsymbol{\theta} = [\mu, \sigma^2]^T$ を推定したいとする。この時、y_i が独立であることに注意すると $\mathbf{y}_n = [y_1, \ldots, y_n]^T$ の結合密度関数は y_i の密度関数の積で

$$f(\mathbf{y}_n; \boldsymbol{\theta}) = f(y_1; \boldsymbol{\theta})\, f(y_2; \boldsymbol{\theta}) \ldots f(y_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta})$$

となる。ここで

$$f(y_i; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \mu)^2}{2\sigma^2} \right]$$

である。

最尤法について

- 例2 ベルヌーイ確率変数

確率変数 y_i , $i = 1,\ldots, n$ はベルヌーイ確率変数、すなわち、確率 p で $y_i = 1$ をとり、確率 $1-p$ で 0 をとる i.i.d. 確率変数であるとする。このとき、その結合確率関数は(ややくだけた書き方だが)

$$f(\mathbf{y}_n ; p) = \prod_{i=1}^{n} f(y_i ; p)$$

と表せる。ここで

$$f(y_i ; p) = p^{y_i}(1-p)^{1-y_i}$$

である。推定する未知パラメータは p である。

最尤法について

■ 例3 ポアソン分布

確率変数 y_{i} , i =1,…, n は i.i.d. 確率変数でポアソン分布に従っているとする。X_i がポアソン分布に従ってるとするとその確率関数は

$$\Pr(X_i = y) = \frac{\lambda^y \exp(-\lambda)}{y!}, y = 0, 1, 2, \ldots$$

で与えられる。$\mathbf{y}_n = [y_1, \ldots, y_n]^{\mathrm{T}}$ の同時確率関数は(これもやや砕けた書き方だが)

$$f(\mathbf{y}_n; \lambda) = \prod_{i=1}^n f(y_i; \lambda)$$

であり、ここで $f(y_i; \lambda) = \dfrac{\lambda^{y_i} \exp(-\lambda)}{y_i!}$ である。

最尤法について

- 例4　指数分布

確率変数 y_i は指数分布に従う i.i.d. 確率変数であるとする。このとき、y_i の密度関数は

$$f(y_i; \lambda) = \lambda \exp(-\lambda y_i), \quad y_i \geq 0$$

であり、その結合密度関数は

$$f(\mathbf{y}_n; \lambda) = \prod_{i=1}^{n} f(y_i; \lambda)$$

である。

最尤法について

- 対数尤度関数

実際に最尤推定値を計算する際には、尤度関数を直接最大化するよりも、尤度関数の自然対数をとった**対数尤度関数**、すなわち

$$\ell_n(\boldsymbol{\theta}) = \log L_n(\boldsymbol{\theta})$$

を最大化する方が計算が簡単である。対数は単調増加関数であるから、このように計算した最尤推定値は尤度関数を最大化して計算した最尤推定値と同じになる。

最尤法について

対数尤度関数はそれぞれの観測値に関する密度関数の対数の和になることに注意。すなわち

$$
\begin{aligned}
\ell_n(\boldsymbol{\theta}) &= \log L_n(\boldsymbol{\theta}) \\
&= \log\left[\prod_{i=1}^{n} f(y_i;\boldsymbol{\theta})\right] \\
&= \sum_{i=1}^{n} \log f(y_i;\boldsymbol{\theta})
\end{aligned}
$$

である。

最尤法について

- 最尤推定法の例 $1-4$

先ほどの正規分布の例において実際に最尤推定量を導出してみよう。

ホワイトボードで導出

最尤法について

- 最尤法の性質

最尤法の性質として主に以下のものがあげられる

(1) 一致性

(2) 漸近正規性

(3) 漸近有効性

ここでは(1) について見てみる。

最尤法について

- 最尤法の一致性

最尤推定量を $\hat{\theta}_n$ と表すとする。**一致性**とは $\hat{\theta}_n$ が観測数 n が増えるにつれて真の値 θ_0 に確率収束する、すなわち、どのような $\varepsilon > 0$ に対しても $n \to \infty$ の時

$$\Pr(|\hat{\theta}_n - \theta_0| > \varepsilon) \to 0$$

が成り立つことである。以下では最尤推定量の一致性の証明のスケッチを与える。

　ある数列 x_n が x_0 に確率収束するとき

$$x_n \to_p x_0 \quad や \quad \mathrm{plim}_{n\to\infty} x_n = x_0$$

というような書き方をする。

最尤法について

- ヤンセンの不等式

最尤推定量の一致性の証明には以下の**ヤンセンの不等式**と呼ばれる不等式を用いるのが一般的である。

(ヤンセンの不等式) 確率変数 X と凹関数 $h(.)$ に対して

$$E(h(X)) \leq h(E(X))$$

が成り立つ。

凹関数とは $0 \leq \alpha \leq 1$ に対して、

$$h((1-\alpha)x + \alpha y) \geq (1-\alpha)h(x) + \alpha f(y)$$

が成り立つ関数のこと。

最尤法について

- 一致性

ここでは、最尤推定量の一致性の証明を、対数尤度関数の期待値を最大化する θ の値は真の値 θ_0 と等しく、さらに最尤推定量 $\hat{\theta}_n$ はその値に確率収束することを示すという手順で行う。以下の期待値は全て真の値 θ_0 における期待値とする。

まず対数関数は凹関数なのでヤンセンの不等式より

$$E\left\{\log\left[\frac{L_n(\theta)}{L_n(\theta_0)}\right]\right\} \le \log\left\{E\left[\frac{L_n(\theta)}{L_n(\theta_0)}\right]\right\}$$

が任意の θ で成り立つ。等号は$L_n(\theta)/L_n(\theta_0)$が確率変数でないとき。

最尤法について

$L_n(\theta_0)$は $\mathbf{y}_n$ の(真の値の下での)結合密度関数であるから

$$E\left[\frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)}\right]=\int_{Y^n}\frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)}L(\boldsymbol{\theta}_0)d\mathbf{y}_n=\int_{Y^n}L_n(\boldsymbol{\theta})d\mathbf{y}_n=1$$

を得る。ここで $\int_{Y^n}\ldots d\mathbf{y}_n$ は n 次元ベクトルの(その取りうる値の範囲である) Y^n 上での積分を表す。最後の等号は $L_n(\theta)$ も任意の θ で結合確率密度関数であることによる。

最尤法について

よって 任意の$\theta^* \neq \theta_0$ において

$$E\left\{\log\left[\frac{L_n(\boldsymbol{\theta}^*)}{L_n(\boldsymbol{\theta}_0)}\right]\right\} = E[\log L_n(\boldsymbol{\theta}^*)] - E[\log L_n(\boldsymbol{\theta}_0)]$$

$$= E[\ell_n(\boldsymbol{\theta}^*)] - E[\ell_n(\boldsymbol{\theta}_0)] < 0$$

$$\Leftrightarrow E[\ell_n(\boldsymbol{\theta}^*)] < E[\ell_n(\boldsymbol{\theta}_0)]$$

が成り立っている。これは任意の n で成り立っていることに注意。これより対数尤度関数の期待値は真の値 $\boldsymbol{\theta}_0$ で最大化されていることがわかる。

最尤法について

ここである条件の下で

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\theta)=\operatorname*{plim}_{n\to\infty}\frac{1}{n}E[\ell_n(\theta)]$$

が成り立つとすると(例えば y_i がi.i.d. の時など)、先ほどの不等式の極限をとると

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}E[\ell_n(\boldsymbol{\theta}^*)]\le\operatorname*{plim}_{n\to\infty}\frac{1}{n}E[\ell_n(\boldsymbol{\theta}_0)]$$

$$\Leftrightarrow\quad\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\boldsymbol{\theta}^*)\le\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\boldsymbol{\theta}_0)$$

が成り立つことがわかる(ここで極限を取っているので不等号は等式を含むようになることに注意)

最尤法について

一方、最尤推定量 $\hat{\boldsymbol{\theta}}_n$は定義により常に

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\hat{\boldsymbol{\theta}}_n)\geq\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\boldsymbol{\theta}_0)$$

を満たす。この2つの不等式がともに成立するのは

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\hat{\boldsymbol{\theta}}_n)=\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\boldsymbol{\theta}_0)$$

の時だけであるので最尤推定量はこの等式を満たすことがわかる。

最尤法について

ただしこの式自体は $\hat{\theta}_n$ の一致性を保証しない。なぜなら

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\boldsymbol{\theta}_0)=\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\boldsymbol{\theta})$$

となる $\boldsymbol{\theta}$ の値が $\boldsymbol{\theta}_0$ 以外にもある可能性があるからである。これを排除するためには $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$ の時

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\boldsymbol{\theta}_0)\neq\operatorname*{plim}_{n\to\infty}\frac{1}{n}\ell_n(\boldsymbol{\theta})$$

というような仮定を置く。この仮定は**漸近識別条件**と呼ばれる。