

最尤推定

確率・統計 - 第9講

村田 昇

前回のおさらい

- 平均の点推定
 - 点推定の考え方
 - 良い不偏推定量
 - 正規分布モデルの標本平均の性質
- 平均の区間推定
 - 区間推定の考え方
 - 正規分布モデルにおける区間推定

統計的推測の考え方

- 観測データは確率変数の集合
- 確率変数列 X_1, X_2, \dots, X_n に対する考察が重要
- 現象の理解のためには観測された実現値より確率分布にこそ興味がある
- 一般に分析対象のデータには独立性と同分布性が同時に仮定される
- 観測データの背後の確率分布を推定
 - 分布のもつ特性量 (平均や分散など) を評価する
 - 分布そのもの (確率質量関数や確率密度関数) を決定する

点推定

- 定義
- 点推定とは母数 (パラメタ) θ を X_1, \dots, X_n の関数

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$$

で推定することで, $\hat{\theta}$ を θ の推定量と呼ぶ.

Cramer-Rao の不等式

- 定理

1次元母数 θ を含む連続分布を考え, その確率密度関数 $f_\theta(x)$ は θ に関して偏微分可能であるとする. このとき, 緩やかな仮定の下で, θ の任意の不偏推定量 $\hat{\theta}$ に対して以下の不等式が成り立つ.

$$\text{Var}(\hat{\theta}) \geq \frac{1}{nI(\theta)},$$

$$I(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \theta} \log f_\theta(x) \right)^2 f_\theta(x) dx.$$

一様最小分散不偏推定量

- 定理 (Cramer-Rao の不等式の系)

θ の不偏推定量 $\hat{\theta}$ で分散が Cramer-Rao 下界に一致するものが存在すれば、一様最小分散不偏推定量となる。

推定誤差

- 推定量 $\hat{\theta}$ には推定誤差が必ず存在
- 推定結果の定量評価には推定誤差の評価が重要
 - “誤差 $\hat{\theta} - \theta$ が区間 $[l, u]$ の内側にある確率が $1-\alpha$ 以上”
 (“外側にある確率が α 以下”)

$$P(l \leq \hat{\theta} - \theta \leq u) \geq 1-\alpha$$

- この確率の厳密あるいは近似的な評価を利用

区間推定

- 定義

区間推定とは未知母数 θ とある値 $\alpha \in (0, 1)$ に対して以下を満たす確率変数 L, U を観測データから求めることをいう。

$$P(L \leq \theta \leq U) \geq 1-\alpha$$

- 区間 $[L, U]$: $1-\alpha$ 信頼区間 ($100(1-\alpha)$ % とも書く)
- L : $1-\alpha$ 下側信頼限界
- U : $1-\alpha$ 上側信頼限界
- $1-\alpha$: 信頼係数 ($\alpha = 0.01, 0.05, 0.1$ など)

加法的正規雑音モデル

- 加法的雑音モデル

$$X_i = \theta + \varepsilon_i, \quad i = 1, \dots, n$$

(確率変数) (未知母数) (誤差)

- 以下の仮定を加える
 - $\varepsilon_1, \dots, \varepsilon_n$ は平均 0, 分散 σ^2 の正規分布に従う。
- 観測値の分布
 - このとき X は平均 θ , 分散 σ^2 の正規分布に従う。

標本平均

- 平均 θ , 分散 σ^2 の正規分布
 - 平均母数 θ に関する Fisher 情報量 :

$$I(\theta) = \frac{1}{\sigma^2}$$

- 標本平均 \bar{X} の分散: $\sigma^2/n = \text{Cramer-Rao 下界}$
- \bar{X} は平均の一様最小分散不偏推定量

平均の区間推定

- 分散既知の場合の信頼区間

$z_{1-\alpha/2}$ を標準正規分布の $1-\alpha/2$ 分位点とすれば

$$P\left(-z_{1-\alpha/2} \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq z_{1-\alpha/2}\right) = 1-\alpha$$

となるので、 σ が既知の場合の平均 μ の $1-\alpha$ 信頼区間は以下で構成される.

$$\left[\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

最尤法

離散分布の場合

- 観測値 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ の同時確率
 - 確率質量関数: $f_{\theta}(x)$
 - 確率質量関数の母数: $\theta = (\theta_1, \dots, \theta_p)$
 - 独立な確率変数の同時確率:

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= \prod_{i=1}^n P(X_i = x_i) \\ &= \prod_{i=1}^n f_{\theta}(x_i) = f_{\theta}(x_1) \cdot f_{\theta}(x_2) \cdots f_{\theta}(x_n) \end{aligned}$$

尤度関数

- 定義

母数 θ に対して観測データ X_1, X_2, \dots, X_n が得られる理論上の確率

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

を θ の**尤度** と言い、 θ の関数 L を **尤度関数** と呼ぶ.

- 観測データ X_1, X_2, \dots, X_n が現れるのに母数 θ の値がどの程度 **尤もらしいか** を測る尺度

連続分布の場合

- 確率密度関数 $f_{\theta}(x)$ を用いて尤度を定義
- 尤度関数:

$$L(\theta) = \prod_{i=1}^n f_{\theta}(x_i) = f_{\theta}(x_1) \cdot f_{\theta}(x_2) \cdots f_{\theta}(x_n)$$

- 確率データ X_i が微小な区間 $[x_i - \delta, x_i + \delta]$ に含まれる確率に比例する量

最尤法

- 最尤法

観測データに対して「最も尤もらしい」母数の値を θ の推定量として採用する方法を **最尤法** という。

- 最尤推定量

Θ を尤度関数の定義域として、尤度関数を最大とする $\hat{\theta}$

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta).$$

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta).$$

を θ の **最尤推定量** という。

最尤推定量の計算

- 対数尤度関数：

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f_{\theta}(X_i).$$

- 対数関数は狭義増加
- $\ell(\theta)$ の最大化と $L(\theta)$ の最大化は同義
- 扱い易い和の形なのでこちらを用いることが多い
- 大数の法則を用いて対数尤度関数の収束が議論できる

演習

練習問題

- 観測データ X_1, X_2, \dots, X_n が以下の分布に従うとき母数 λ の最尤推定量を求めよ。
 - $\lambda > 0$ の Poisson 分布 (確率質量関数)

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (x \text{ は } 0 \text{ 以上の整数})$$

- $\lambda > 0$ の指数分布 (確率密度関数)

$$f(x) = \lambda e^{-\lambda x} \quad (x > 0)$$

解答例：Poisson 分布の最尤推定

- 対数尤度関数 (未知母数： λ)

$$\ell(\lambda) = \sum_{i=1}^n \log \frac{\lambda^{X_i}}{X_i!} e^{-\lambda} = \sum_{i=1}^n (X_i \log \lambda - \log X_i!) - n\lambda$$

- 少なくとも 1 つの i について $X_i > 0$ を仮定する

- $\ell(\lambda)$ の微分：

$$\ell'(\lambda) = \frac{1}{\lambda} \sum_{i=1}^n X_i - n, \quad \ell''(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n X_i < 0$$

- 上に凸なので方程式 $\ell'(\lambda) = 0$ の解が $\ell(\lambda)$ を最大化
- λ の最尤推定量:

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i$$

解答例：指数分布の最尤推定

- 対数尤度関数 (未知母数: λ)

$$\ell(\lambda) = \sum_{i=1}^n \log \lambda e^{-\lambda X_i} = n \log \lambda - \lambda \sum_{i=1}^n X_i$$

- $\ell(\lambda)$ の微分:

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n X_i, \quad \ell''(\lambda) = -\frac{n}{\lambda^2} < 0$$

- 方程式 $\ell'(\lambda) = 0$ の解が $\ell(\lambda)$ を最大化
- λ の最尤推定量:

$$\hat{\lambda} = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}$$

参考：ガンマ分布の最尤推定

- 母数 $\nu, \alpha > 0$ のガンマ分布 (調べてみよ)
- 対数尤度関数 (未知母数: ν, α)

$$\begin{aligned} \ell(\nu, \alpha) &= \sum_{i=1}^n \log \frac{\alpha^\nu}{\Gamma(\nu)} X_i^{\nu-1} e^{-\alpha X_i} \\ &= n\nu \log \alpha - n \log \Gamma(\nu) + \sum_{i=1}^n \{(\nu-1) \log X_i - \alpha X_i\} \end{aligned}$$

– $\ell(\nu, \alpha)$ を最大化する ν, α は解析的に求まらないので実際の計算では数値的に求める

最尤推定量の性質

一貫性

- 定理

全ての x に対して $f(x) > 0$ で f が連続ならば、最尤推定量 $\hat{\theta}^*$ は一致推定量になる。すなわち、真の母数の値が θ_0 のとき、任意の $\varepsilon > 0$ に対して

$$P(|\hat{\theta}^* - \theta_0| < \varepsilon) \rightarrow 1 \quad (n \rightarrow \infty)$$

が成り立つ。

- 観測データが十分多ければ、最尤推定量は真の値に一致する

証明の概略

- 対数尤度の大数の法則

$$\frac{1}{n} \log L(\theta) \xrightarrow{n \rightarrow \infty} \mathbb{E}_{\theta_0}[\log f(X, \theta)]$$

- \mathbb{E}_{θ_0} は真の母数 (X を生成している分布) での平均を表す

- 対数密度・質量の性質

$$\mathbb{E}_{\theta_0}[\log f(X, \theta)] \leq \mathbb{E}_{\theta_0}[\log f(X, \theta_0)]$$

- 対数関数の凸性を用いて証明 (資料参照)
- 情報理論のエントロピー (平均符号長) とも関係

$$\mathbb{E}_{\theta_0}[-\log f(X, \theta)] \geq \mathbb{E}_{\theta_0}[-\log f(X, \theta_0)]$$

- 対数尤度の性質

n が十分大きければ

$$\frac{1}{n} \log L(\theta) \leq \frac{1}{n} \log L(\theta_0)$$

がほぼ確実に成り立つ

- n が大きければ、真の値で対数尤度は最大になる

漸近正規性

- 定理

$f(x) > 0$ が連続で 2 階微分可能ならば $\sqrt{n}(\hat{\theta}^* - \theta_0)$ は $n \rightarrow \infty$ で正規分布 $\mathcal{N}(0, I(\theta_0)^{-1})$ に近づく。

- $I(\theta_0)$ は Fisher 情報量
- 観測データが十分多ければ、最尤推定量の誤差の分散は Cramer-Rao 下界に一致する

証明の概略

- 真の母数の性質

$$\mathbb{E}_{\theta_0} \left[\frac{\partial}{\partial \theta} \log f(X, \theta_0) \right] = 0$$

- Fisher 情報量 (以下が等しいことは講義資料を参照)

$$\begin{aligned}
I(\theta_0) &= \mathbb{E}_{\theta_0} \left[-\frac{\partial^2}{\partial \theta^2} \log f(X, \theta_0) \right] \\
&= \mathbb{E}_{\theta_0} \left[\left(\frac{\partial}{\partial \theta} \log f(X, \theta_0) \right)^2 \right]
\end{aligned}$$

- 最大値の性質

$$L(\hat{\theta}^*) = \max_{\theta \in \Theta} L(\theta)$$

$$\ell(\hat{\theta}^*) = \max_{\theta \in \Theta} \ell(\theta)$$

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

- Taylor 展開による近似

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

– $\tilde{\theta}$ は $\{X_i\}$ に依存して決まる θ_0 と $\hat{\theta}^*$ の間の値

- 誤差の近似

$$\begin{aligned}
\sqrt{n}(\hat{\theta}^* - \theta_0) &\left\{ -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f(X_i, \tilde{\theta}) \right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(X_i, \theta_0)
\end{aligned}$$

- 左辺

n が大きくなると $\tilde{\theta} \rightarrow \theta_0$ となり、大数の法則により以下が成り立つ。

$$\begin{aligned}
&-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f(X_i, \tilde{\theta}) \\
&\rightarrow \mathbb{E}_{\theta_0} \left[-\frac{\partial^2}{\partial \theta^2} \log f(X_i, \theta_0) \right] = I(\theta_0)
\end{aligned}$$

- 右辺

$n \rightarrow \infty$ のとき中心極限定理により右辺

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(X_i, \theta_0) \\
&= \sqrt{n} \times \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(X_i, \theta_0) \\
&= \sqrt{n} \times \left(\frac{\partial}{\partial \theta} \log f(X, \theta_0) \text{ の標本平均} \right)
\end{aligned}$$

は平均 0 分散 $I(\theta_0)$ の正規分布 $\mathcal{N}(0, I(\theta_0))$ に近づく。

- 両辺を整理

$n \rightarrow \infty$ のとき

$$\sqrt{n}I(\theta_0)(\hat{\theta}^* - \theta_0) \sim \mathcal{N}(0, I(\theta_0)) \quad (n \rightarrow \infty)$$

であるので

$$\begin{aligned} \sqrt{n}(\hat{\theta}^* - \theta_0) &\sim \mathcal{N}\left(0, I(\theta_0)^{-1}I(\theta_0)I(\theta_0)^{-1}\right) \\ &= \mathcal{N}\left(0, I(\theta_0)^{-1}\right) \end{aligned}$$

となる.

漸近正規性にもとづく区間推定

推定量の漸近正規性

- 漸近正規性

多くの推定量 $\hat{\theta}$ の分布は正規分布で近似できる

- モーメントに基づく記述統計量は漸近正規性をもつ
- 最尤推定量は広い範囲の確率分布に対して漸近正規性をもつ
- いずれも中心極限定理にもとづく

- 正規分布を用いて近似的に信頼区間を構成可能

漸近正規性にもとづく区間推定

- 推定量の分布

観測データ数 n が十分大きいとき, 母数 θ の推定量 $\hat{\theta}$ が

$$\mathbb{E}[\hat{\theta}] = \theta_0, \quad \text{Var}(\hat{\theta}) = s^2$$

の正規分布で近似できるとする.

- 信頼区間の構成

母数 θ の $1-\alpha$ 信頼区間は以下で構成される.

$$[\hat{\theta} - z_{1-\alpha/2} \cdot s, \hat{\theta} + z_{1-\alpha/2} \cdot s]$$

標本平均の区間推定

- 定理 (標本平均の漸近正規性)

確率分布が2次のモーメントを持てば, 分布の平均 μ の推定量である標本平均

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

は漸近正規性をもつ. 確率変数 X の標準偏差の一致推定量を $\hat{\sigma}$ とすれば, ϕ を標準正規分布の確率密度関数として, 任意の $a \leq b$ に対して以下が成立する.

$$P\left(a \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}} \leq b\right) \rightarrow \int_a^b \phi(x) dx \quad (n \rightarrow \infty)$$

- σ の一致推定量

例えば不偏分散

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

の平方根を用いれば良い.

- 推定量の分散

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \simeq \frac{\hat{\sigma}^2}{n}$$

- 信頼区間の構成

平均 μ の $1-\alpha$ 信頼区間は以下で構成される.

$$\left[\bar{X} - z_{1-\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}} \right]$$

- サンプル数 n が十分大きい場合に近似的に正しい

最尤推定量の区間推定

- 定理 (最尤推定量の漸近正規性)

観測データ数 n が十分大きいとき, 1 次元母数 θ を含む連続分布の最尤推定量 $\hat{\theta}$ は

$$\mathbb{E}[\hat{\theta}] = \theta_0, \quad \text{Var}(\hat{\theta}) = \frac{1}{nI(\hat{\theta})}$$

の正規分布で近似できる.

- 信頼区間の構成

母数 θ の $1-\alpha$ 信頼区間は以下で構成される.

$$\left[\hat{\theta} - z_{1-\alpha/2} \cdot \frac{1}{\sqrt{nI(\hat{\theta})}}, \hat{\theta} + z_{1-\alpha/2} \cdot \frac{1}{\sqrt{nI(\hat{\theta})}} \right]$$

- サンプル数 n が十分大きい場合に近似的に正しい

演習

練習問題

- 選挙の投票率を調べるために, 400 人にアンケートを実施した結果, 320 人が投票に行き, 残りが行かなかったと答えた. 以下の問に答えよ.
 - 投票率を最尤推定しなさい.
 - 投票率の最尤推定量の分散を求めよ.
 - 投票率の 0.9 信頼区間を求めよ. ただし $z_{0.95} = 1.64$ として計算せよ.
 - ヒント: 以下のような確率変数を考えるとよい.

$$X = \begin{cases} 1, & \text{投票に行った} \\ 0, & \text{行かなかった} \end{cases}$$

解答例

- 確率質量関数

投票率を θ とし、確率変数 X の確率質量関数を以下で定義する.

$$\begin{aligned}f(1) &= \theta \\f(0) &= 1 - \theta\end{aligned}$$

以下のようにまとめられる. (他にも書き方はある)

$$f(x) = \theta^x (1 - \theta)^{1-x}$$

- 対数尤度

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{400} \log f(X_i) \\&= \sum_{i=1}^{400} \{X_i \log \theta + (1 - X_i) \log(1 - \theta)\} \\&= n\bar{X} \log \theta + n(1 - \bar{X}) \log(1 - \theta)\end{aligned}$$

ただし \bar{X} は標本平均.

- 最尤推定量

$$\begin{aligned}\ell'(\hat{\theta}) &= \frac{n\bar{X}}{\hat{\theta}} - \frac{n(1 - \bar{X})}{1 - \hat{\theta}} = 0 \\ \hat{\theta} &= \bar{X}\end{aligned}$$

- 推定値

$$\hat{\theta} = \bar{X} = \frac{320}{400} = 0.8$$

- Fisher 情報量を用いた場合

$$\begin{aligned}I(\theta) &= \mathbb{E} \left[-\frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right] \\&= \mathbb{E} \left[\frac{X}{\theta^2} + \frac{(1 - X)}{(1 - \theta)^2} \right] \\&= \frac{1}{\theta} + \frac{1}{(1 - \theta)} = \frac{1}{\theta(1 - \theta)}\end{aligned}$$

したがって真の母数が θ_0 のときの推定量の分散は以下で与えられる.

$$\text{Var}(\hat{\theta}) = \frac{\theta_0(1 - \theta_0)}{n}$$

- 標本平均の性質を用いた場合

確率変数の平均と分散は

$$\begin{aligned}\mathbb{E}[X] &= 1 \times \theta + 0 \times (1 - \theta) = \theta \\ \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \theta - \theta^2 = \theta(1 - \theta)\end{aligned}$$

であるので、真の母数が θ_0 のときの推定量の分散は以下で与えられる。

$$\text{Var}(\hat{\theta}) = \text{Var}(\bar{X}) = \frac{\theta_0(1 - \theta_0)}{n}$$

- 推定量の分散の一致推定量

真の値 θ_0 の代理として最尤推定量 $\hat{\theta}$ を用いる

$$\begin{aligned}s^2 &= \frac{\hat{\theta}(1 - \hat{\theta})}{n} = \frac{0.8 \times 0.2}{400} = 4 \times 10^{-4} \\ s &= 0.02\end{aligned}$$

- 0.9 信頼区間 (90%信頼区間)

$$[\hat{\theta} - z_{0.95} \cdot s, \hat{\theta} + z_{0.95} \cdot s] = [0.7672, 0.8328]$$

今回のまとめ

- 最尤法の考え方
- 最尤推定量の性質
 - 一致性
 - 漸近正規性
- 最尤推定量による区間推定
 - 標本平均の場合
 - 母数推定の場合