政治学方法論 I

最尤法

矢内 勇生

神戸大学　法学部/法学研究科

2014 年 12 月 10 日

今日の内容

尤度関数 (likelihood function)

特定の母数 θ が与えられたとき、データ D を得る確率（下の式の右辺）を、θ の関数として表現したもの

$$L(\theta|D) = \Pr(D|\theta)$$

$\rightarrow$ 与えられた D に対する、θ の尤もらしさを表す！

- D：データ
- θ：母数（のベクトル）

定数 k を用いて同値類をまとめて扱う場合もある：

$$l(\theta|D) = k\Pr(D|\theta) \propto \Pr(D|\theta)$$

尤度 (likelihood)

尤度関数 $L(\theta|D)$ を、特定の母数 $\theta = \theta_i$ で評価したものを、D に対する θ_i の尤度と呼ぶ

- $L(\theta_1|D)$：D が観測されたとき、母数が θ_1 であるのはどの程度尤もらしいか
- $L(\theta_2|D)$：D が観測されたとき、母数が θ_2 であるのはどの程度尤もらしいか

注意：尤度は絶対的な基準ではない！ $\rightarrow$ モデル内で比較したとき、尤度が大きい方がより尤もらしい

ベイズの公式と尤度

ベイズの公式：

$$\begin{aligned}\Pr(\theta|D) &= \frac{\Pr(D|\theta)\Pr(\theta)}{\Pr(D)} \\ &\propto \Pr(D|\theta)\Pr(\theta) \\ &\propto L(\theta|D)\Pr(\theta)\end{aligned}$$

- $\Pr(\theta)$：θ の事前確率（D を得る前の θ に対する信念）
- $\Pr(\theta|D)$：θ の事後確率（D で得た情報を考慮したことにより更新された θ に対する信念）

事前/事後確率を受け入れられない $\rightarrow$ 不確実性の一つの源泉として、尤度を尤度 ($\neq$ 確率）として扱う

問題の設定

例題：確率 θ で表、$1-\theta$ で裏が出るコイン

あるコインを 10 回投げたところ、「表」が 8 回出た。このコインを 1 回投げたときに表が出る確率 θ はいくつか？

- データ D： D の内容は、
  - コインを投げた回数： $n = 10$
  - 表が出た回数：$x = 8$
- 推定したい母数：θ
- 検討する尤度：$L(\theta) = \Pr(D|\theta)$

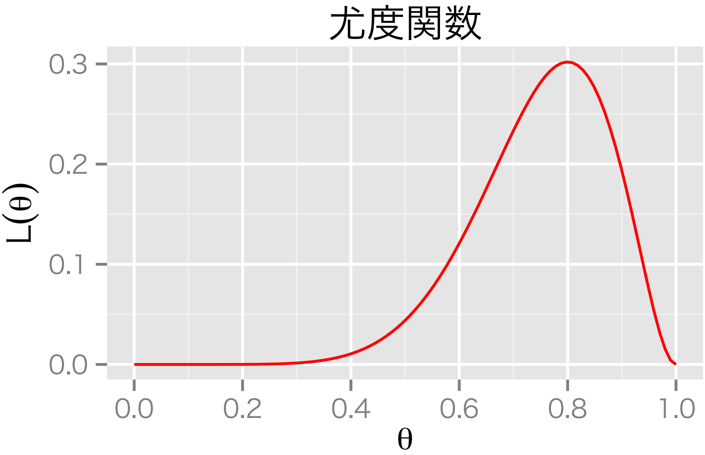
尤度関数の特定

$$\begin{aligned} L(\theta|D) &= \Pr(D|\theta) = \binom{10}{8}\theta^8(1-\theta)^{10-8} \\ &= 45\theta^8(1-\theta)^2 \end{aligned}$$

この式が最大になる θ を見つける：観察されたデータ D を生み出すのに最も尤もらしい θ を見つける

- $\theta = 0 \to L(\theta) = 0$：尤もらしくない
- $\theta = 0.2 \to L(\theta) = 0.000073$：尤もらしい？
- $\theta = 0.6 \to L(\theta) = 0.12$：尤もらしい？
- $\theta = 0.8 \to L(\theta) = 0.30$：尤もらしい？
- $\theta = 0.9 \to L(\theta) = 0.19$：尤もらしい？
- $\theta = 1 \to L(\theta) = 0$：尤もらしくない

尤度関数 $L(\theta|D)$



尤度関数の最大化

この例のように単純な尤度関数の最大化は簡単

$$L(\theta|D) = 45\theta^8(1-\theta)^2 = 45(\theta^{10} - 2\theta^9 + \theta^8)$$

最大化条件は、

$$\begin{aligned}
\frac{dL(\theta|D)}{d\theta} = 90(5\theta^9 - 9\theta^8 + 4\theta^7) &= 0 \\
\Leftrightarrow 5\theta^9 - 9\theta^8 + 4\theta^7 &= 0 \\
\Leftrightarrow \theta^7(\theta-1)(5\theta-4) &= 0 \\
\Leftrightarrow \theta &= \frac{4}{5} \qquad (\because \theta \neq 0, 1)
\end{aligned}$$

問題の設定

例題：確率 θ で表 (H)、$1-\theta$ で裏 (T) が出るコイン

あるコインを 10 回投げたところ、{H, H, T, H, H, H, H, H, H, T} という結果になった。このコインを 1 回投げたときに表が出る確率 θ はいくつか？

- データ D：

$$
\begin{aligned}
D &= \{H,H,T,H,H,H,H,H,H,T\} \\
&= \{1,1,0,1,1,1,1,1,1,0\}
\end{aligned}
$$

- 推定したい母数：θ
- 検討する尤度：$L(\theta) = \Pr(D|\theta)$

尤度関数の特定 (1)

各ベルヌーイ試行が独立だとすると、

$$L(\theta|D) = \Pr(D|\theta) = \prod_{i=1}^{10} \Pr(D_i|\theta) = \prod_{i=1}^{10} L_i(\theta|D_i)$$

ただし、$D = \{D_1, D_2, \dots, D_{10}\}$.

各試行 i について

$$L_i(\theta|D_i) = \Pr(D_i|\theta) = \theta^{D_i}(1-\theta)^{1-D_i}$$

だから、

$$L(\theta|D) = \prod_{i=1}^{10} [\theta^{D_i}(1-\theta)^{1-D_i}]$$

尤度関数の特定 (2)

D_i の値は 0 か 1 のいずれかであり、

$$
\begin{aligned}
L_i(\theta|D_i = 1) &= \theta^1(1-\theta)^0 = \theta, \\
L_i(\theta|D_i = 0) &= \theta^0(1-\theta)^1 = 1-\theta.
\end{aligned}
$$

となるから、結局、

$$
L(\theta|D) = \prod_{i=1}^{10} L_i(\theta|D_i) = \theta^8(1-\theta)^2
$$

となる。最大化条件は、

$$
\begin{aligned}
\frac{dL(\theta|D)}{d\theta} = 2\theta^7(\theta-1)(5\theta-4) &= 0 \\
\therefore \theta &= 4/5 \qquad (\because \theta \neq 0, 1)
\end{aligned}
$$

対数尤度 (log likelihood)

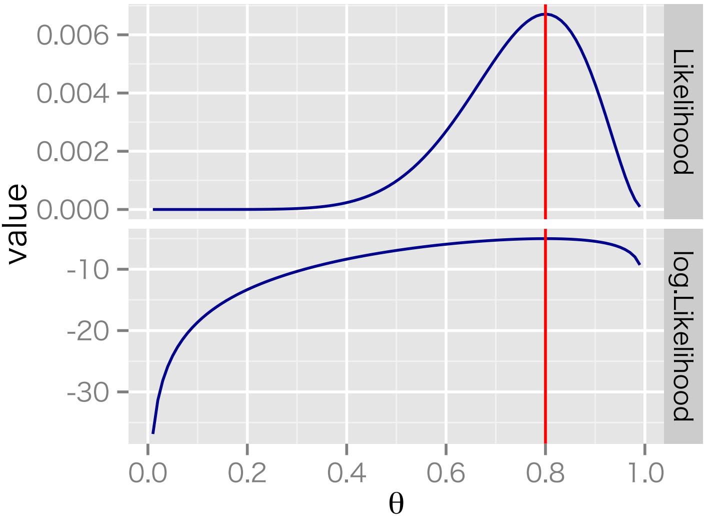
対数をとっても数値の大小関係は変わらない、つまり、$x_1 < x_2$ ならば $\log(x_1) < \log(x_2)$ なので、最大値を求めるときは尤度の自然対数をとってもかまわない。自然対数をとると、

$$\begin{aligned}\log[L(\theta|D)] &= \log\left(\prod_{i=1}^{10}[\theta^{D_i}(1-\theta)^{1-D_i}]\right) \\ &= \sum_{i=1}^{10}\log[\theta^{D_i}(1-\theta)^{1-D_i}] = 8\log\theta + 2\log(1-\theta)\end{aligned}$$

最大化条件は、

$$\begin{aligned}\frac{d}{d\theta}\log[L(\theta|D)] = \frac{8}{\theta} - \frac{2}{1-\theta} &= 0 \\ \Leftrightarrow \theta &= \frac{4}{5}\end{aligned}$$

尤度と対数尤度


0.006
0.004
0.002
0.000
-10
-20
-30
value
Likelihood
log.Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
θ

連続分布の場合の問題

$\Pr(X = x|\theta) = 0 \rightarrow$ これを使うと尤度が常に 0 になってしまう

- 観測値には誤差 ϵ（精度限界）がある
- 観測値 x の意味：$x \in (x - \epsilon/2, x + \epsilon/2)$
- 連続分布の確率密度関数を $p(x|\theta)$ とすると、ϵ が十分小さければ、

$$
\begin{aligned}
L(\theta|X) &= \Pr[X \in (x - \epsilon/2, x + \epsilon/2)] \\
&= \int_{x-\epsilon/2}^{x+\epsilon/2} p(x|\theta)dx \approx \epsilon p(x|\theta)
\end{aligned}
$$

- モデル内で θ の比較をするとき、尤度は定数倍してもその有用性が変わらない（同値類は一緒くたに扱える） $\rightarrow$ 右辺の ϵ は無視してかまわない

$\rightarrow$ 連続分布の尤度関数は、確率密度関数を使って表す

正規分布の例

問題の設定

$x_i \sim N(\theta, \sigma^2), i = 1, 2, \ldots, n$ で、σ^2 を知っているとする。このとき、観察された x_i に対する θ の尤度関数を求める

- $N(\theta, \sigma^2)$ の確率密度関数：

$$p(x|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2}\right]$$

尤度関数の特定

- 観測値 x_i に対する θ の尤度関数は、

$$L_i(\theta|x_i,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\theta)^2}{2\sigma^2}\right]$$

- 全体の対数尤度は、

$$\begin{aligned} \log L(\theta) &= \log\left[\prod_{i=1}^{n} L_i(\theta|x_i,\sigma^2)\right] \\ &= \sum_{i=1}^{n} \log L_i(\theta|x_i,\sigma^2) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2 \end{aligned}$$

尤度比 (likelihood ratio)

2 つの尤度, $L(\theta_1|D)$ と $L(\theta_2|D)$ をどうやって比較する？

- データ x に対して、x に 1 対 1 対応する y を考えると、

$$\frac{L(\theta_2|y)}{L(\theta_1|y)} = \frac{L(\theta_2|x)}{L(\theta_1|x)}$$

- $L(\theta_1|D)$ と $L(\theta_2|D)$ の「比」が重要である（「差」の大きさではなく）
- $\rightarrow$ 尤度の絶対的な大きさに意味はない：尤度関数全体を k 倍して考えてかまわない (k は正の実数)
- より一般的に、尤度関数 L の代わりに $f'(.) > 0$ となる $f(L)$ で置き換えられる：対数尤度を利用する
- $\rightarrow$ 尤度関数のうち、母数が含まれていない項は無視できる

最尤推定値：Maximum Likelihood Estimate (MLE)

MLE: 尤度関数の最大値 → 最尤法による推論における点推定値

- MLE は尤度関数の最も簡単な要約
- MLE は最尤法による推論過程の「一部」を表現しているに過ぎない
- MLE だけでは尤度関数の特徴を表せない → 尤度関数全体を使って推論を行うべき
- 対数尤度関数が二次関数で近似できるとき（“regular” likelihood）、尤度関数の特徴を表すには少なくとも 2 つの数値が必要
  1. 最大値の位置 (location)：MLE
  2. 最大値での曲率 (curvature)

スコア関数とフィッシャー情報量

- スコア関数：対数尤度の一次導関数

$$S(\theta) \equiv \frac{\partial}{\partial \theta} \log L(\theta)$$

- MLE $\hat{\theta}$ は $S(\theta) = 0$ の解である
- $\hat{\theta}$ における曲率を $I(\hat{\theta})$ とする。ただし、

$$I(\theta) \equiv -\frac{\partial^2}{\partial \theta^2} \log L(\theta)$$

(最大値では二次導関数の値が負なので、マイナスをかけておく)

- $I(\hat{\theta})$：フィッシャー情報量 (observed Fisher information)：この値が大きいほど、θ に関する不確実性が低い

スコア関数とフィッシャー情報量の例（1-1）

正規分布

$x_i \sim N(\theta, \sigma^2), i = 1, 2, \ldots, n$ で、σ^2 を知っているとする。このとき、観察された x_i に対する θ の MLE とフィッシャー情報量を求める。

- θ が含まれていない項は無視できるので、

$$\log L(\theta|x, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2$$

- スコア関数は、

$$S(\theta) = \frac{\partial}{\partial\theta}\log L(\theta|x, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \theta)$$

スコア関数とフィッシャー情報量の例（1-2）

- 対数尤度を 2 回微分すると、フィッシャー情報量は、

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

  - $\mathrm{Var}(\hat{\theta}) = \sigma^2/n = I^{-1}(\hat{\theta})$：情報量が多いほど、推定値の分散が小さい
  - $\mathrm{se}(\hat{\theta}) = \sigma/\sqrt{n} = I^{-1/2}(\hat{\theta})$

スコア関数とフィッシャー情報量の例（2-1）

二項分布

確率 θ で成功するベルヌーイ試行を n 回実施した結果、成功が x 回、失敗が $n-x$ 回だった。このとき、観察された x に対する θ の MLE とフィッシャー情報量を求める。

- 定数項を無視すると、対数尤度は、

$$\log L(\theta) = x \log \theta + (n-x) \log(1-\theta)$$

- スコア関数は、

$$S(\theta) = \frac{\partial}{\partial \theta} \log L(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

- $S(\theta) = 0$ を解くと、

$$\hat{\theta} = \frac{x}{n}$$

スコア関数とフィッシャー情報量の例（2-2）

- 対数尤度を 2 回微分すると、

$$I(\theta) \equiv -\frac{\partial^2}{\partial \theta^2} \log L(\theta) = \frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2}$$

- したがって、フィッシャー情報量は、

$$I(\hat{\theta}) = \frac{n}{\hat{\theta}(1-\hat{\theta})} = \frac{n^3}{x(n-x)}$$

尤度区間 (likelihood intervals)

MLE だけでは推定の不確実性が不明 $\rightarrow$ 「区間推定」が必要

- 尤度区間を、以下の条件を満たす θ の集合として定める

$$\left\{ \theta : \frac{L(\theta)}{L(\hat{\theta})} > c \right\}$$

  - $c \in (0, 1)$：任意の閾値
  - $L(\theta)/L(\hat{\theta})$：~~標準化~~正規化された尤度関数

尤度区間の例

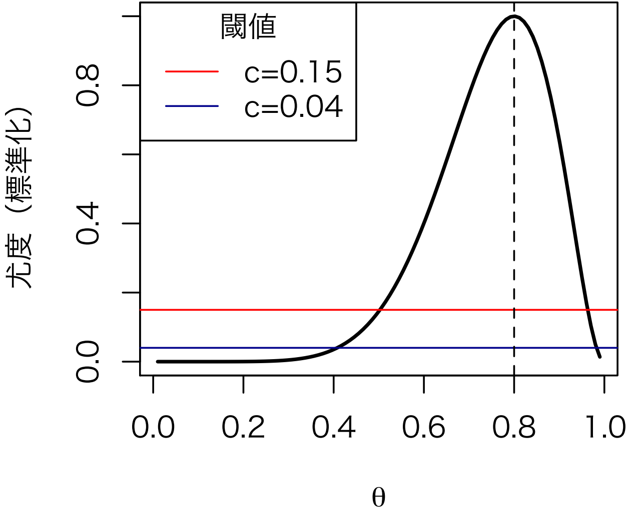
確率 θ で表が出るコインを $n = 10$ 回投げて $x = 8$ 回表がでたときの、θ の尤度区間

- $c = 0.15$：尤度区間は $(0.50, 0.96)$
- $c = 0.04$：尤度区間は $(0.41, 0.98)$

尤度区間の問題

- c を選ぶ基準がわからない
- 特定の尤度区間の意味（解釈）がわからない

尤度区間の例：$n = 10,\ x = 8$ の二項分布



確率に依拠した区間推定：正規分布の場合 (1)

- 正規分布の場合、先に求めた対数尤度より、

$$\log \frac{L(\theta)}{L(\hat{\theta})} = -\frac{n}{2\sigma^2}(\bar{x} - \theta)^2$$

- $\bar{x} \sim N(\theta, \sigma^2/n)$ だから、

$$\frac{n}{\sigma^2}(\bar{x} - \theta)^2 \sim \chi^2_1$$

- つまり、

$$W \equiv 2 \log \frac{L(\hat{\theta})}{L(\theta)} \sim \chi^2_1$$

- W：ウィルクス (Wilks) の尤度比統計量

（正規分布以外でも、n が十分大きければ、χ^2_1 で近似できる）

確率に依拠した区間推定：正規分布の場合 (2)

- 特定の尤度区間をとる確率を考えると、

$$\begin{aligned}\Pr\left(\frac{L(\theta)}{L(\hat{\theta})} > c\right) &= \Pr\left(2\log\frac{L(\hat{\theta})}{L(\theta)} < -2\log c\right) \\ &= \Pr(\chi_1^2 < -2\log c)\end{aligned}$$

- ここで、$0 < \alpha < 1$ によって c を決める

$$c = \exp\left(-\frac{1}{2}\chi^2_{1,(1-\alpha)}\right)$$

ただし、$\chi^2_{1,(1-\alpha)}$ は χ_1^2 の $100(1-\alpha)$ パーセンタイル

確率に依拠した区間推定：正規分布の場合 (3)

- このとき、

$$\Pr\left(\frac{L(\theta)}{L(\hat{\theta})} > c\right) = \Pr(\chi_1^2 < \chi^2_{1,(1-\alpha)}) = 1 - \alpha$$

- これで、$100(1-\alpha)$ 信頼区間と同等の区間が得られる
- 特に、$\alpha = 0.05$ ならば $c = 0.15$, $\alpha = 0.01$ ならば $c = 0.04$ である

$c = 0.15$ $(c = 0.04)$ の尤度区間は 95% (99%) 信頼区間と同等とみなすことができる

尤度比検定

- 帰無仮説 $\mathrm{H}_0 : \theta = \theta_0$ を考える
- 以下の尤度比が「小さ過ぎる」とき、帰無仮説を棄却する

$$\frac{L(\theta_0)}{L(\hat{\theta})}$$

- 「小さ過ぎる」の基準は？ $\rightarrow$ 確率的に考える必要
- ウィルクスの尤度比統計量を使うと、帰無仮説の尤度比が c のとき、p 値は、

$$p = \Pr(\chi_1^2 > -2\log c)$$

- ただし、この式はいつも成り立つわけではない

標準誤差

- 対数尤度が二次関数で近似できるとき、

$$\log \frac{L(\theta)}{L(\hat{\theta})} \approx -\frac{1}{2} I(\hat{\theta})(\theta - \hat{\theta})^2$$

- よって、$\{\theta : L(\theta)/L(\hat{\theta}) > c\}$ となる区間は近似的に

$$\theta \pm \sqrt{-2 \log c} \cdot I(\hat{\theta})^{-1/2}$$

- 一般的に、MLE $\hat{\theta}$ の標準誤差は、

$$\mathrm{se}(\hat{\theta}) = I(\hat{\theta})^{-\frac{1}{2}}$$

ワルド統計量 (Wald statistic)

- MLE の標準誤差を使うと、Wald 統計量 z は、

$$z = \frac{\hat{\theta} - \theta_0}{\mathrm{se}(\hat{\theta})}$$

- $|z|$ が大きいほど、帰無仮説 $\theta = \theta_0$ の尤度は小さくなり、p 値も小さくなる
- 95%ワルド信頼区間は、

$$\hat{\theta} \pm 1.96\mathrm{se}(\hat{\theta})$$

- ワルド信頼区間の利点：$\hat{\theta}$ に関して左右対称
- ワルド信頼区間の弱点：対数尤度が二次関数で近似できないと、近似が成り立たない

$\rightarrow$ （確率に基づいた）尤度区間のほうが望ましい場合が多い

来週の内容

最尤法（続き）

- 最尤法によるロジスティック回帰
- プロビット回帰