

推定

確率分布を推定する

村田 昇

講義の内容

- 点推定
 - 不偏推定量
 - Cramér-Rao の不等式
- 最尤推定量
- 区間推定
 - 信頼区間
 - 正規母集団の区間推定
 - 漸近正規性にもとづく区間推定

推定とは

統計解析の目的

- 観測データを確率変数の実現値と考えてモデル化
- 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 (確率関数や確率密度) を決定する
- 統計学で広く利用されている推定方法を説明
 - **点推定**
 - **区間推定**

推定の標準的な枠組

- 観測データは独立同分布な確率変数列 X_1, X_2, \dots, X_n
- X_i の従う共通の法則 \mathcal{L} を想定
 - \mathcal{L} として全ての分布を考察対象とすることは困難
 - * 対象とする範囲が広くなりすぎる
 - * データ数 n が大きくなると意味のある結論を導き出せない
 - 確率分布 \mathcal{L} を特徴づけるパラメタ θ を考察対象
 - * \mathcal{L} の平均・分散・歪度・尖度など
 - * \mathcal{L} の確率関数・確率密度関数のパラメタ

点推定

点推定

- 定義

\mathcal{L} に含まれるパラメタ 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.

- 記述統計量は分布のパラメタの 1 つ
- 推定量の例:

\mathcal{L} の平均 μ を標本平均 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ によって推定することが点推定であり, \bar{X} は μ の推定量となる.

良い推定量

- 一般に 1 つのパラメタの推定量は無数に存在
- 推定量の良さの代表的な基準: **不偏性・一致性**
 - $\hat{\theta}$ が θ の不偏推定量:

$$E[\hat{\theta}] = \theta$$

- $\hat{\theta}$ が θ の (強) 一致推定量:

$$\hat{\theta} \text{ が } \theta \text{ に収束する確率が } 1 \quad (n \rightarrow \infty)$$

- 良い推定量の例:

標本平均, 不偏分散はそれぞれ \mathcal{L} の平均, 分散の不偏かつ一致性をもつ推定量

良い不偏推定量

- 一般に不偏推定量も複数存在
 - 例: \mathcal{L} の平均 μ の不偏推定量:
 - 標本平均 \bar{X}
 - X_1
 - X_1, \dots, X_n のメディアン (\mathcal{L} が $x = \mu$ に関して対称な場合)
- 不偏推定量の良さを評価する基準が必要
- **一様最小分散不偏推定量:**
 - θ の任意の不偏推定量 $\hat{\theta}'$ に対して推定値のばらつき (分散) が最も小さいもの

$$\text{Var}(\hat{\theta}) \leq \text{Var}(\hat{\theta}')$$

Cramér-Rao の不等式

- 定理

\mathcal{L} は 1 次元パラメタ θ を含む連続分布とし、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:

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

ただし

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

一様最小分散不偏推定量

- 用語の定義

- 下界 $1/(nI(\theta))$: **Cramér-Rao 下界**
- $I(\theta)$: **Fisher 情報量**

- 定理 (Cramér-Rao の不等式の系)

θ の不偏推定量 $\hat{\theta}$ で分散が Cramér-Rao 下界 $1/(nI(\theta))$ に一致するものが存在すれば、それは一様最小分散不偏推定量となる。

例: 正規分布モデルの標本平均

- \mathcal{L} は平均 μ , 分散 σ^2 の正規分布
- 平均パラメタ μ に関する Fisher 情報量:

$$I(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma^4} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma^2}$$

- Cramér-Rao 下界: σ^2/n
- 標本平均 \bar{X} の分散: σ^2/n (=Cramér-Rao 下界)
- \bar{X} は μ の一様最小分散不偏推定量

演習

練習問題

- X を一様乱数に従う確率変数とし、平均値の推定量として以下を考える。それぞれの推定量の分散を比較しなさい。
 - 標本平均 (mean)
 - 中央値 (median)
 - 最大値と最小値の平均 $((\max + \min)/2)$
- ヒント: 以下のような関数を作り、Monte-Carlo 実験を行えばよい

```
myMeanEst <- function(n, min, max){ # 観測データ数
  x <- runif(n, min=min, max=max) # 一様乱数を生成, 範囲は引数から
  return(c(xbar=mean(x), med=median(x), mid=(max(x)+min(x))/2))
} # 3つまとめて計算する関数
```

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: $f_{\theta}(x)$
 - 確率関数のパラメタ: $\theta := (\theta_1, \dots, \theta_p)$
 - 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
パラメタ 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 が現れるのにパラメタ θ の値がどの程度尤もらしいかを測る尺度

最尤法

- 最尤法:
観測データに対して「最も尤もらしい」パラメタ値を 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 $\hat{\theta}$

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

を θ の **最尤推定量** という. 以下のように書くこともある.

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

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
- 最尤推定量の性質
広い範囲の確率分布に対して最尤推定量は **一貫性** を持つ

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:

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

- 対数尤度関数:

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

例: Poisson 分布の最尤推定

- \mathcal{L} をパラメタ $\lambda > 0$ の Poisson 分布でモデル化
 - 対数尤度関数 (未知パラメタ: 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
- (Poisson 分布のつづき)
 - $\ell(\lambda)$ の微分:

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

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

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

例: 指数分布の最尤推定

- \mathcal{L} をパラメタ $\lambda > 0$ の指数分布でモデル化
 - 対数尤度関数 (未知パラメタ: λ)

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

- (指数分布のつづき)
 - $\ell(\lambda)$ 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:

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

例: ガンマ分布の最尤推定

- \mathcal{L} をパラメタ $\nu, \alpha > 0$ のガンマ分布でモデル化
 - 対数尤度関数 (未知パラメタ: 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
- R での計算例 (ガンマ分布の最尤推定量の例)

```
library(stats4) # 関数 mle を利用するため
## 数値最適化のためには尤度関数を最初に評価する初期値が必要
mle.gamma <- function(x, # 観測データ
  nu0=1, alpha0=1){ # nu, alpha の初期値
  ## 負の対数尤度関数を定義 (最小化を考えるため)
  ll <- function(nu, alpha) # nu と alpha の関数として定義
    suppressWarnings(-sum(dgamma(x, nu, alpha, log=TRUE)))
  ## suppressWarnings は定義域外で評価された際の警告を表示させない
  ## 最尤推定 (負の尤度の最小化)
  est <- mle(minuslogl=ll, # 負の対数尤度関数
    start=list(nu=nu0, alpha=alpha0), # 初期値
    method="BFGS", # 最適化方法 (選択可能)
    nobs=length(x)) # 観測データ数
  return(coef(est)) # 推定値のみ返す
}
```

演習

練習問題

- 東京都の気候データ (tokyo_weather.csv) の風速 (wind) の項目について以下の問に答えよ.
 - 全データを用いてヒストグラム (密度) を作成しなさい.
 - ガンマ分布でモデル化して最尤推定を行いなさい.
 - 推定した結果をヒストグラムに描き加えて比較しなさい.
- 自身で収集したデータを用いて, モデル化と最尤推定を試みよ.

区間推定

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 ”

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

- “外側にある確率が α 以下” と言い換えてもよい
- パラメタの範囲の推定に書き換え
 - “ θ が含まれる確率が $1-\alpha$ 以上となる区間 $[\hat{\theta} - u, \hat{\theta} - l]$ ”

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

区間推定

- 定義
区間推定とは未知パラメタ 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** ($100(1-\alpha)$ % と書くことも多い)
- L : $1-\alpha$ **下側信頼限界**
- U : $1-\alpha$ **上側信頼限界**
- $1-\alpha$: **信頼係数** ($\alpha = 0.01, 0.05, 0.1$ とすることが多い)

信頼区間の性質

- 信頼区間は幅が狭いほど推定精度が良い
 - 真のパラメタが取りうる値の範囲を限定することになるため
- 最も推定精度の良い $1-\alpha$ 信頼区間 $[L, U]$

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

- 信頼区間の幅が狭いほど $P(L \leq \theta \leq U)$ は小さくなるため
- 実行可能である限り $1-\alpha$ 信頼区間 $[L, U]$ は上式を満たすように L, U を決定する

正規母集団の区間推定

平均の区間推定 (分散既知)

- 正規分布に従う独立な確率変数の重み付き和は正規分布に従う
- 一般の場合

Z_1, Z_2, \dots, Z_k を独立な確率変数列とし、各 $i = 1, 2, \dots, k$ に対して Z_i は平均 μ_i 、分散 σ_i^2 の正規分布に従うとする。このとき a_0, a_1, \dots, a_k を $(k+1)$ 個の 0 でない実数とすると、 $a_0 + \sum_{i=1}^k a_i Z_i$ は平均 $a_0 + \sum_{i=1}^k a_i \mu_i$ 、分散 $\sum_{i=1}^k a_i^2 \sigma_i^2$ の正規分布に従う。

- 同分布の場合

$$k = n, \mu_i = \mu, \sigma_i^2 = \sigma^2, a_0 = 0, a_i = 1/n \ (i = 1, \dots, n)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (\text{標本平均})$$

は平均 μ , 分散 σ^2/n の正規分布に従う.

- 同分布を標準化した場合

$$k = 1, \mu_1 = \mu, \sigma_1^2 = \sigma^2/n, a_0 = -\sqrt{n}\mu/\sigma, a_1 = \sqrt{n}/\sigma$$

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

は標準正規分布に従う.

- 標準化した確率変数の確率

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

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

- 信頼区間の構成

μ について解くと

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

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.

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

平均の区間推定 (分散未知)

- χ^2 分布の特徴付け

– 標準正規分布に従う k 個の独立な確率変数の二乗和は自由度 k の χ^2 分布に従う

- t 分布の特徴付け

– Z を標準正規分布に従う確率変数, Y を自由度 k の χ^2 分布に従う確率変数とし, Z, Y は独立であるとする. このとき確率変数

$$\frac{Z}{\sqrt{Y/k}}$$

は自由度 k の t 分布に従う

- 標本平均と不偏分散の性質

X_1, X_2, \dots, X_n は独立同分布な確率変数列で, 平均 μ , 分散 σ^2 の正規分布に従うとする. 不偏分散を

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

とすると, \bar{X} と s^2 は独立であり, 確率変数 $(n-1)s^2/\sigma^2$ は自由度 $n-1$ の χ^2 分布に従う.

- 標準化した確率変数の性質

前の命題と $\sqrt{n}(\bar{X} - \mu)/\sigma$ が標準正規分布に従うことから、確率変数

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{s} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{(n-1)s^2/\sigma^2/(n-1)}}$$

は自由度 $n-1$ の t 分布に従う.

- 信頼区間の構成

$t_{1-\alpha/2}(n-1)$ を自由度 $n-1$ の t 分布の $1-\alpha/2$ 分位点とすれば

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

となるので、分散が未知の場合の平均 μ の $1-\alpha$ **信頼区間** は以下で構成される.

$$\left[\bar{X} - t_{1-\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}} \right]$$

分散の区間推定

- 不偏分散の性質

$(n-1)s^2/\sigma^2$ は自由度 $n-1$ の χ^2 分布に従う

- 不偏分散の確率

$\chi^2_{\alpha/2}(n-1)$, $\chi^2_{1-\alpha/2}(n-1)$ をそれぞれ自由度 $n-1$ の χ^2 分布の $\alpha/2, 1-\alpha/2$ 分位点とすれば

$$P\left(\chi^2_{\alpha/2}(n-1) \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}(n-1)\right) = 1-\alpha$$

- 信頼区間の構成

σ^2 について解くと

$$P\left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}\right) = 1-\alpha$$

となるので、 σ^2 の $1-\alpha$ **信頼区間** は以下で構成される.

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} \right]$$

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することができる

標本平均の漸近正規性

- 定理

確率分布 \mathcal{L} が2次のモーメントを持てば, \mathcal{L} の平均 μ の推定量である標本平均は漸近正規性をもつ.

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

\mathcal{L} の標準偏差を σ とすれば, 任意の $a \leq b$ に対して以下が成立する. (ϕ は標準正規分布の確率密度関数)

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

平均の区間推定 (分散既知)

- 標本平均の確率

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

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

となるので, μ について解くと以下が成り立つ.

$$P\left(\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \rightarrow 1-\alpha \quad (n \rightarrow \infty)$$

- 信頼区間の構成

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.

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

(サンプル数 n が十分大きい場合に近似的に正しい)

平均の区間推定 (分散未知)

- σ をその一致推定量 $\hat{\sigma}$ で置き換えてもそのまま成立する
– $\hat{\sigma}$ としては例えば不偏分散の平方根を用いる

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

- 実問題で平均はわからないが, 分散はわかるという場合はあまりない
- t -分布は自由度 $n \rightarrow \infty$ で標準正規分布になる

- 信頼区間の構成

σ が未知の場合の平均 μ の $1-\alpha$ **信頼区間**は以下で構成される.

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

- 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)

\mathcal{L} が 1 次元パラメタ θ を含む連続分布とすると、最尤推定量 $\hat{\theta}$ は平均 θ (真の値)、分散 $1/(nI(\hat{\theta}))$ の正規分布で近似できる。

- 信頼区間の構成

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

- 東京都の気候データ (`tokyo_weather.csv`) の日射量 (`solar`) の項目について以下の間に答えよ。
 - 全データによる平均値を計算しなさい。
 - ランダムに抽出した 50 点を用いて、平均値の 0.9(90%) 信頼区間を求めなさい。
 - 上記の推定を 100 回繰り返した際、真の値 (全データによる平均値) が信頼区間に何回含まれるか確認しなさい。
- 自身で収集したデータで区間推定を試みよ。