

平均の推定

確率・統計 - 講義 9

村田 昇

前回のおさらい

- 確率モデルと推定量
 - 観測データの生成モデル
 - 独立同分布性
 - 推定量と推定値
- 推定量の評価
 - 推定量の平均値の不偏性
 - 推定量の分散の大きさ

復習

確率と統計

- 確率

ある決まった確率法則のもとで、確率変数がどのような性質を持つか論じる
- 統計

ある確率法則に従うと考えられる確率変数の実現値を観測して、それを生成する確率法則について何らかの推測を行う

 - 工学的な問題としては後者の統計的な問題設定に接する機会が多い

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 (確率関数や確率密度) を決定する

大数の法則

- 定理

X_1, X_2, \dots を独立同分布な確率変数列とし、その平均を μ とする。このとき、 X_1, \dots, X_n の標本平均

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

が $n \rightarrow \infty$ のとき μ に収束する確率は 1 である。

中心極限定理

- 定理

X_1, X_2, \dots を独立同分布な確率変数列とし、その平均を μ 、標準偏差を σ とする。このとき、すべての実数 $a < b$ に対して

$$P\left(a \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \quad (n \rightarrow \infty)$$

が成り立つ。

推定量と推定値

- 推定量

未知の母数を推定する方法。

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$$

- 推定値

$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ という観測値が得られたとき、推定量に代入して得られる値。

$$\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$$

良い推定量

- 不偏性

推定量の平均が真の母数 θ に一致すること。

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

– 観測値に基づく推定値 (見本点) が真の値のまわりにばらつくと考えればよい

良い不偏推定量

- 不偏推定量は複数存在

例: 平均の不偏推定量:

- 標本平均 \bar{X}
- X_n
- X_1, \dots, X_n の中央値 (分布が対称な場合)

- 一般に推定量の分散を用いて良さを評価

良い不偏推定量 θ とは任意の不偏推定量 $\hat{\theta}'$ に対して推定値のばらつき (分散) が最も小さいもの

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

対称な不偏推定量

- 定理

X_1, X_2, \dots, X_n が互いに独立に同じ分布に従い, $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ を母数 θ の一つの不偏推定量とすると, X_1, X_2, \dots, X_n の対称な関数として表される不偏推定量 $\hat{\theta}^*$ で, つねに

$$\text{Var}(\hat{\theta}^*) \leq \text{Var}(\hat{\theta})$$

となるものが存在する.

— 観測データ X_1, X_2, \dots, X_n の順序に関して対称な推定量だけ考えればよい

点推定

点推定

- 定義

母数 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.

- 平均や分散は分布の母数の1つ

- 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 $\hat{\mu}$ となる.
- 分散 σ^2 を不偏分散 $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ によって推定することが点推定であり, s^2 は σ^2 の推定量 $\hat{\sigma}^2$ となる.

一様最小分散不偏推定量

- 定義

θ の任意の不偏推定量 $\hat{\theta}'$ に対して推定量の分散が最も小さいもの

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

が存在すれば, それを一様最小分散不偏推定量という.

— 常に存在するとは限らない

Cramér-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.

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

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

— 証明は講義ノート参照

一様最小分散不偏推定量の存在

- 用語
 - 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。

演習

練習問題

- 平均 θ , 分散 σ^2 の正規分布に従う独立な確率変数 X_1, X_2, \dots, X_n から求められる標本平均の分散を求めよ。
- 上記の正規分布の平均 θ に関する Fisher 情報量を求めよ。

$$f_{\theta}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

解答例

- 独立性に注意して計算する

$$\begin{aligned} \mathbb{E}[\bar{X}] &= \mathbb{E}\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \theta \\ \text{Var}(\bar{X}) &= \text{Var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n} \sigma^2 \end{aligned}$$

- 定義通りに計算する

$$\begin{aligned} I(\theta) &= \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \theta} \log f_{\theta}(x) \right)^2 f_{\theta}(x) dx \\ &= \int_{-\infty}^{\infty} \left(-\frac{x - \theta}{\sigma^2} \right)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma^4} \int_{-\infty}^{\infty} (x - \theta)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma^4} \text{Var}(X) = \frac{1}{\sigma^2} \end{aligned}$$

- 正規分布の分散について
 - 3行目の積分の中身は分散の計算であることに注意
 - 特性関数を用いると簡単に計算できる

$$\begin{aligned}\phi(s) &= \mathbb{E}[e^{isX}] \\ \phi'(s) &= \frac{d}{ds} \mathbb{E}[e^{isX}] = \mathbb{E}[iX e^{isX}] \\ \phi''(s) &= \mathbb{E}[-X^2 e^{isX}] \\ \phi''(0) &= \mathbb{E}[-X^2]\end{aligned}$$

- 正規分布の特性関数を用いて確認せよ

正規分布モデルにおける点推定

観測値の生成モデル

- 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)

- 誤差の仮定
 - $\varepsilon_1, \dots, \varepsilon_n$ は互いに独立に分布する.
 - $\varepsilon_1, \dots, \varepsilon_n$ は同じ分布に従う.
 - $\mathbb{E}[\varepsilon_i] = 0, i = 1, \dots, n$
 - $\mathbb{E}[\varepsilon_i^2] < \infty, i = 1, \dots, n$

誤差の正規性

- 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.
- 観測値の分布が議論できる

このとき X は平均 θ , 分散 σ^2 の正規分布に従う.

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

正規分布モデルの標本平均

- 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:

$$I(\theta) = \frac{1}{\sigma^2}$$
 - Cramér-Rao 下界: $1/nI(\theta) = \sigma^2/n$
 - 標本平均 \bar{X} の分散: σ^2/n (= Cramér-Rao 下界)
- \bar{X} は平均の一樣最小分散不偏推定量

一般の分布の場合の注意

- 標本平均の平均, 分散は同様に計算される
- Fisher 情報量は誤差の分布によって異なる
- 一般に標本平均の分散は Cramér-Rao 下界と異なる
- したがって標本平均が必ずしも良い不偏推定量になるとは限らない

演習

練習問題

- 加法的雑音モデルにおいて誤差 $\varepsilon_1, \dots, \varepsilon_n$ が区間 $[-a, a]$ の一様分布に従うとする。独立な観測データ X_1, X_2, \dots, X_n から求められる標本平均の分散を求めよ。
- 上記のモデルにおいて θ に関する Fisher 情報量を求めよ。

解答例

- ε の分散

$$\begin{aligned}\text{Var}(\varepsilon) &= \mathbb{E}[\varepsilon^2] \\ &= \int_{-a}^a \varepsilon^2 \frac{1}{2a} d\varepsilon = \frac{a^2}{3} \\ \text{Var}(\bar{X}) &= \frac{a^2}{3n}\end{aligned}$$

- 密度関数が滑らかでない (微分できない) ので Fisher 情報量は定義されない。
- この分布における良い推定量 (Pitman 推定量) は

$$\hat{\theta}(X_1, X_2, \dots, X_n) = \frac{\min_i X_i + \max_i X_i}{2}$$

となることが知られている (講義ノート参照)。

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”
 (“外側にある確率が α 以下” と言い換えてもよい)

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

- 母数の範囲の推定に書き換え
 - “ θ が含まれる確率が $1-\alpha$ 以上となるような区間 $[\hat{\theta} - u, \hat{\theta} - l]$ を推定”

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

区間推定

- 定義

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。

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

- 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
 - 真の母数を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

演習

練習問題

- 確率変数 X, Y は独立で、それぞれ平均 μ_1, μ_2 , 分散 σ_1^2, σ_2^2 の正規分布に従うとする。このとき a, b を実数として新しい確率変数 Z を

$$Z = aX + bY$$

で定義するとき、 Z の平均と分散を求めよ。

- 確率変数 Z がどのような分布になるか答えよ。

解答例

- 独立性から計算できる

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}[aX + bY] \\ &= a\mu_1 + b\mu_2 \\ \text{Var}(Z) &= \text{Var}(aX + bY) \\ &= \text{Var}(aX) + \text{Var}(bY) \\ &= a^2\text{Var}(X) + b^2\text{Var}(Y) \\ &= a^2\sigma_1^2 + b^2\sigma_2^2\end{aligned}$$

- 特性関数を考えるのがわかりやすい

$$\begin{aligned}\phi_Z(s) &= \mathbb{E}[e^{isZ}] = \mathbb{E}[e^{is(aX+bY)}] \\ &= \mathbb{E}[e^{is(aX)}]\mathbb{E}[e^{is(bY)}] \\ &= \phi_X(as)\phi_Y(bs) \\ &= e^{\mu_1 i(as) - \frac{\sigma_1^2(as)^2}{2}} e^{\mu_2 i(bs) - \frac{\sigma_2^2(bs)^2}{2}} \\ &= e^{(a\mu_1 + b\mu_2)is - \frac{(a^2\sigma_1^2 + b^2\sigma_2^2)s^2}{2}}\end{aligned}$$

これは平均 $a\mu_1 + b\mu_2$, 分散 $a^2\sigma_1^2 + b^2\sigma_2^2$ の正規分布の特性関数に他ならない。

正規分布モデルにおける区間推定

標本平均の性質

- 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

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

– 確率密度 f の α 分位点 z_α ($0 < \alpha < 1$)

$$\int_{-\infty}^{z_\alpha} f(x)dx = \alpha$$

– 標準正規分布の分位点の性質 (確率密度の対称性より)

$$z_\alpha = -z_{1-\alpha}$$

平均の区間推定 (分散既知)

- 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.

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

標本平均と不偏分散の性質

- 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関係

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.

平均の区間推定 (分散未知)

- 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.

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

今回のまとめ

- 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