

母数推定

確率・統計 第8回

村田 昇

2020年8月23日

復習

演習問題

- 偏りのないコインを投げたとき表なら1, 裏なら-1となる確率変数 X を考える. n 回投げたときに得られた結果を X_1, X_2, \dots, X_n と書くことにする.

$$Y_n = \frac{X_1 + X_2 + \dots + X_n}{\sqrt{n}}$$

とするとき, 以下の間に答えよ.

- Y_n の平均と分散を求めよ.
- Y_n の特性関数 $\phi_n(s)$ を求めよ.
- 特性関数 $\phi_n(s)$ の極限 $\lim_{n \rightarrow \infty} \phi_n(s)$ を求めよ.

解答例

- 題意より

$$P(X=1) = \frac{1}{2}, \quad P(X=-1) = \frac{1}{2}$$

X_1, X_2, \dots, X_n は独立として計算

- X の平均と分散が

$$\mathbb{E}[X] = 0, \quad \text{Var}(X) = \mathbb{E}[X^2] = 1$$

であることより

$$\mathbb{E}[Y_n] = 0, \quad \text{Var}(Y_n) = \frac{1}{n} \cdot n \text{Var}(X) = 1$$

- 特性関数は

$$\begin{aligned} \phi_n(s) &= \mathbb{E}[e^{isY_n}] \\ &= (\mathbb{E}[e^{isX/\sqrt{n}}])^n = \left(\frac{e^{is/\sqrt{n}} + e^{-is/\sqrt{n}}}{2} \right)^n \\ &= \cos^n \left(\frac{s}{\sqrt{n}} \right) \end{aligned}$$

- 余弦関数の Taylor 展開は以下となる

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{24} + \dots$$

* 定義通り計算してもよいし、指数関数の展開と Euler の式を用いてもよい

$$\begin{aligned}\lim_{n \rightarrow \infty} \cos^n \left(\frac{s}{\sqrt{n}} \right) &= \lim_{n \rightarrow \infty} \left(1 - \frac{s^2}{2n} + \cdots \right)^n \\ &= e^{-\frac{s^2}{2}}\end{aligned}$$

となり標準正規分布の特性関数となる。

統計的推測

統計的推測の考え方

- 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

確率変数と確率分布

- 確率変数 X :
 - 値がランダムに決定される変数
 - 任意の区間 $[a, b]$ に含まれる確率が定められている
- 確率分布 P (または単に分布):
 - 区間 $[a, b]$ ($a \leq b$) に X が含まれる確率
$$P(a \leq X \leq b)$$
 - 「確率変数 X は確率分布 P に従う」という
 - 現象の理解のためには実現値より確率分布にこそ興味がある

同時分布

- 観測データは確率変数の集合
- 確率変数の列 X_1, X_2, \dots, X_n に対する考察が重要
- 定義
“ X_1 が x_1 という値をとり、 X_2 が x_2 という値をとり、 \dots 、 X_n が x_n という値をとる”
という事象が起きる確率を**同時分布**という

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

独立性

- 無関係にサンプリングされた観測データの性質
- 定義
 n 個の確率変数 X_1, X_2, \dots, X_n が**独立**であるとは、 $a_i \leq b_i$ ($i = 1, \dots, n$) なる任意の実数 $a_1, b_1, \dots, a_n, b_n$ に対して

$$\begin{aligned}P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \dots, a_n \leq X_n \leq b_n) \\ = P(a_1 \leq X_1 \leq b_1)P(a_2 \leq X_2 \leq b_2) \cdots P(a_n \leq X_n \leq b_n)\end{aligned}$$

が成り立つことをいう。

同分布性

- 同一の法則に従って生成された観測データの性質
- 定義

n 個の確率変数 X_1, X_2, \dots, X_n が **同分布** であるとは, $a \leq b$ なる任意の実数 a, b に対して

$$P(a \leq X_1 \leq b) = P(a \leq X_2 \leq b) = \dots = P(a \leq X_n \leq b)$$

が成り立つことをいう.

独立同分布性

- 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
- 定義
独立かつ同分布な確率変数列を**独立同分布**もしくは **i.i.d.** であるという
(i.i.d. は independent and identically distributed の略)

極限定理

極限定理

- 大数の法則, 中心極限定理など
- 緩やかな条件のもとで確率変数の列について成立する法則
 - 大数の法則は平均が存在すれば成立する
 - 2 次のモーメントをもたない (計算できない) 場合には中心極限定理は成立しない

無限列の独立性と同分布性

- 無限列に対しては任意の部分列について考える
- 独立性
 X_1, X_2, \dots が **独立** であるとは, 任意の正整数 n に対して X_1, X_2, \dots, X_n が独立であることをいう
- 同分布性
 X_1, X_2, \dots が **同分布** であるとは, 任意の正整数 n に対して X_1, X_2, \dots, X_n が同分布であることをいう
- 独立同分布性
 X_1, X_2, \dots が **独立同分布**もしくは **i.i.d.** であるとは, X_1, X_2, \dots が独立かつ同分布であることをいう

大数の法則の概要

- 同一の法則に従って生成された集団から**ランダムな** 観測を多数繰り返すと, 観測値の標本平均は真の平均値に近づく
- 具体的な例:
 - 歪みの無いコインの表が出た回数の割合
 - 視聴率の調査
- この法則を数学的に定式化した定理が **大数の法則**

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。

中心極限定理の概要

- 大数の法則の主張:
 - 推定誤差 $\bar{X}_n - \mu$ は n を大きくすると 0 に近づく
 - どの程度の大きさになるのか定量的な評価は与えていない
- 誤差の評価の定量化:
 - 推定誤差がある区間 $[\alpha, \beta]$ に入る確率で定量的に評価可能

$$P(\alpha \leq \bar{X}_n - \mu \leq \beta)$$

- 上式の正確な計算は一般には困難
- サンプル数が大きい場合の定量的な評価の近似方法を述べたのが **中心極限定理**

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

推定の標準的な枠組

- 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{P} を想定
 - \mathcal{P} 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{P} を特徴づける母数 θ を考察対象
 - * \mathcal{P} の平均・分散・歪度・尖度など
 - * \mathcal{P} の確率関数・確率密度関数の母数

推定量

- 定義

統計では未知の母数を推定する方式を**推定量** (estimator) と呼び、通常未知母数に $\hat{\cdot}$ を付けて表す.

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

は n 個の確率変数 X_1, X_2, \dots, X_n を観測し、それらを用いて未知母数を推定する計算方法を表わしている.

推定量は確率変数 X_1, X_2, \dots, X_n の関数であるので、それ自体確率変数である.

推定値

- 定義

実際に実験 (試行) を行い $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ という観測値が得られたとき、これらの確率変数の実現値を上の推定量に代入して得られる

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

を**推定値** (estimate) と呼び、推定量と区別する.

推定値は確率変数の実現値であり、確率変数ではなく固定の値である.

確率モデルと推定量の例

観測値の生成モデル

- 加法的雑音モデル

同じ条件の元で繰り返し測定した実験データを考える. データには測定毎に誤差が生じるが、この偶然に変動するばらつきを確率的なものとして捉えることによって

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

(確率変数) (未知母数) (誤差)

という確率モデルを設定することができる.

- 母数推定

実験を何回か行なって得られる観測値 X_1, X_2, \dots, X_n を用いて未知母数 θ を推定する方法を考える.

誤差の性質の仮定

- 仮定 0

$\varepsilon_1, \dots, \varepsilon_n$ はある同時確率分布に従う確率変数である.

(誤差自体が確率変数であるという確率モデルの前提条件の言い換え)

- 仮定 1 (独立性)

$\varepsilon_1, \dots, \varepsilon_n$ は互いに独立に分布する.

(実験が十分注意深く行われていることを前提とした仮定)

- 仮定 2 (同分布性)

$\varepsilon_1, \dots, \varepsilon_n$ は同じ分布に従う.

(各回の実験が完全に同じ条件で行われることを仮定)

– 仮定 1, 2 をあわせて **独立同分布 (i.i.d.)** であると言う

- 仮定 3 (誤差の不偏性)

$$E[\varepsilon_i] = 0, i = 1, \dots, n$$

観測に偏りが無いことを仮定 (偏りがあってもデータからは知ることができない)

- 仮定 4 (極限定理のための条件)

$$E[\varepsilon_i^2] < \infty, i = 1, \dots, n$$

(大数の法則や中心極限定理などを使ってデータの性質を調べる際に必要な仮定. 実験データでは測定値が発散するような状況はほとんどないので特に不合理な制約ではない)

推定量の例

- 推定量: 例 1

$$\hat{\theta} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

確率変数の算術平均で**標本平均** (sample mean) と呼ばれる. 誤差に偏りが無いのであれば直感的にも θ に非常に近い値になるであろうことが予想される.

- 推定量: 例 2

$$\hat{\theta} = (X_1, X_2, \dots, X_n \text{ の中央の値})$$

中央値 (median) と呼ばれるもので, 誤差が正の側と負の側に均等に出るのであれば θ をうまく近似することが予想される.

- 推定量: 例 3 (トリム平均)

$$\hat{\theta} = \frac{1}{n - 2m} \sum_{i=m+1}^{n-m} X_{(i)}$$

($\{X_{(i)}\}$ は $\{X_i\}$ を小さい順に並べ換えたもの)

トリム平均 (trimmed mean) と呼ばれるもので, X_1, X_2, \dots, X_n を小さい順に並べ, 小さい方の m 個と大きい方の m 個を捨てた算術平均である. 極端に大きな, あるいは小さな値として現れる観測値を捨てることによって, 特異的な値による誤差の影響を取り除いて平均を推定しようというものである.

- 推定量: 例 4

$$\hat{\theta} = c_1 X_1 + c_2 X_2 + \dots + c_n X_n, \quad \sum_{i=1}^n c_i = 1$$

各観測の重みを変えて平均を取った加重平均

- 推定量: 例 5

$$\hat{\theta} = X_n$$

最後のデータのみ使った推定量

推定量の良さ

同じ問題においてもいろいろな推定量を構成することができる. 多くの候補の中で望ましい推定量とはどういった性質をもつものであるかを考える.

不偏性

- 定義

推定量 $\hat{\theta}$ が **不偏** (unbiased) であるとは、真の母数が θ であるときに、そこから得られた観測値に基づく推定値の平均値が真の母数 θ に一致すること

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

である。

演習

- 標本平均、荷重平均、最後のデータのみ使った推定量が平均の不偏推定量になることを確かめよ。
- 標本分散

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

が分散の不偏推定量とならないことを確かめよ。

解答例

- 標本平均

標本平均は

$$\begin{aligned}\mathbb{E}[\hat{\theta}] &= \mathbb{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\ &= \frac{\mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]}{n} \\ &= \frac{\theta + \theta + \cdots + \theta}{n} \\ &= \theta\end{aligned}$$

となり、平均値の不偏な推定量になっていることがわかる。他も同様に計算できる。

- 標本分散

標本分散の分母である観測値と標本平均の差の2乗和の平均を計算すると以下のようになる。

$$\begin{aligned}&\mathbb{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n \{X_i - \theta - (\bar{X} - \theta)\}^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n (X_i - \theta)^2 - 2\sum_{i=1}^n (X_i - \theta)(\bar{X} - \theta) + \sum_{i=1}^n (\bar{X} - \theta)^2\right]\end{aligned}$$

- 標本分散 (つづき)

ここで $\sum_{i=1}^n (X_i - \theta) = n(\bar{X} - \theta)$ を用いると

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

- 標本分散 (つづき)

したがって

$$\mathbb{E} \left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \right] = \frac{n-1}{n} \sigma^2$$

となり, 分散 σ^2 の不偏な推定量とはなっていない.

- 不偏分散

分散の不偏推定量は

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

で与えられる.

不偏推定量の良さ

推定量の分散

- 一般に不偏推定量も複数存在

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)

- 不偏推定量の良さを評価する基準が必要

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) が最も小さいものが良い

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

不偏推定量の平均

- 定理

$\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ がいずれも不偏推定量であり, その分散が等しく σ^2 であるとする. このときこれらの不偏推定量の単純な平均を

$$\hat{\theta}^* = \frac{1}{k} \sum_{i=1}^k \hat{\theta}_i$$

とおけば

$$\mathbb{E}[\hat{\theta}^*] = \theta \quad (\text{不偏性が保存される})$$

$$\text{Var}(\hat{\theta}^*) \leq \sigma^2 \quad (\text{分散が小さくなる可能性がある})$$

が成り立つ.

証明

- 証明

平均値の線形性から不偏性は明らか。ここで

$$\begin{aligned}
 & \sum_i (\hat{\theta}_i - \theta)^2 \\
 &= \sum_i (\hat{\theta}_i - \hat{\theta}^* + \hat{\theta}^* - \theta)^2 \\
 &= \sum_i (\hat{\theta}_i - \hat{\theta}^*)^2 + 2 \sum_i (\hat{\theta}_i - \hat{\theta}^*)(\hat{\theta}^* - \theta) + \sum_i (\hat{\theta}^* - \theta)^2 \\
 &= \sum_i (\hat{\theta}_i - \hat{\theta}^*)^2 + k(\hat{\theta}^* - \theta)^2
 \end{aligned}$$

途中 $\hat{\theta}^* = \frac{1}{k} \sum_{i=1}^k \hat{\theta}_i$ より $\sum_i (\hat{\theta}_i - \hat{\theta}^*) = 0$ を用いた。

- (証明のつづき)

両辺の平均を取ると

$$\begin{aligned}
 \sum_i \mathbb{E}[(\hat{\theta}_i - \theta)^2] &= \sum_i \mathbb{E}[(\hat{\theta}_i - \hat{\theta}^*)^2] + k\mathbb{E}[(\hat{\theta}^* - \theta)^2] \\
 \sum_i \text{Var}(\hat{\theta}_i) &= \sum_i \mathbb{E}[(\hat{\theta}_i - \hat{\theta}^*)^2] + k\text{Var}(\hat{\theta}^*) \\
 k\sigma^2 &= \sum_i \mathbb{E}[(\hat{\theta}_i - \hat{\theta}^*)^2] + k\text{Var}(\hat{\theta}^*)
 \end{aligned}$$

となる。右辺の第1項は0または正なので

$$\sigma^2 \geq \text{Var}(\hat{\theta}^*)$$

となり、分散が小さくなることがわかる。

対称な推定量

- 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。

- 意味

推定量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ が X_1, X_2, \dots, X_n について対称でない場合には、対称化することによって性質を良くすることができる。

証明

- 証明

$\hat{\theta}$ を X_1, X_2, \dots, X_n について対称化、すなわち

$$\hat{\theta}^* = \frac{1}{n!} \sum \hat{\theta}(\{X_1, X_2, \dots, X_n \text{ のあらゆる並べ替え} \})$$

を考え、前定理を使えばよい。

まとめ

- 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

注意

- 確認テスト (講義 8 の中)
 - 期間: 6/23 講義終了-6/30 13:00
 - 時間制限: 1 時間
 - 5 問 (確率変数と特性関数)
 - 試験の意味を考えて誠実な対応をすること
 - ネットワーク環境等には自身で配慮すること