

推測統計（母集団・標本・標本抽出・推定）・点推定

樋口さぶろお <https://hig3.net>

龍谷大学 先端理工学部 数理・情報科学課程

確率統計Ⅰ L10(2026-06-22 Mon)

最終更新: Time-stamp: "2026-06-23 Tue 09:04 JST hig"

今日の目標

- 母集団, 標本抽出, 推定を説明できる 久保川 統計学入門 §7.1
- 母平均値, 母比率, 母分散を点推定できる
- 久保川 統計学入門 §7.2
- 不偏推定量, 一致推定量の意味を説明できる



L09-Q1

TA Prob and Sol: 二項分布の再生性

確率変数 X_1, X_2 が独立なとき，次のうち，正しいものを選ぼう． $\Rightarrow$ は「ならば」を意味する．

- $X_1 \sim \text{Bin}(10, \frac{2}{6}), X_2 \sim \text{Bin}(20, \frac{1}{6}) \Rightarrow X_1 + X_2 \sim \text{Bin}(30, \frac{3}{6})$
- $X_1 \sim \text{Bin}(10, \frac{2}{6}), X_2 \sim \text{Bin}(10, \frac{2}{6}) \Rightarrow X_1 + X_2 \sim \text{Bin}(10, \frac{2}{6})$
- $X_1 \sim \text{Bin}(10, \frac{2}{6}), X_2 \sim \text{Bin}(20, \frac{2}{6}) \Rightarrow X_1 + X_2 \sim \text{Bin}(30, \frac{2}{6})$
- $X_1 \sim \text{Bin}(10, \frac{2}{6}), X_2 \sim \text{Bin}(10, \frac{1}{6}) \Rightarrow X_1 + X_2 \sim \text{Bin}(10, \frac{3}{6})$

略解

$$X_1 \sim \text{Bin}(10, \frac{2}{6}), X_2 \sim \text{Bin}(20, \frac{2}{6}) \Rightarrow X_1 + X_2 \sim \text{Bin}(30, \frac{2}{6})$$

一般に， $X_1 \sim \text{Bin}(n_1, p_1), X_2 \sim \text{Bin}(n_2, p_2), p = p_1 = p_2 \Rightarrow X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$.

L09-Q2

TA Prob and Sol: 独立同一分布と中心極限定理

確率変数 X_i ($i = 1, \dots, 400$) は独立同一分布にしたがい, $E[X_i] = \frac{1}{10}$, $\text{Var}(X_i) = \frac{9}{100}$ である.

$S = X_1 + \dots + X_{400}$ とする.

S の分布を正規分布で近似し, $P(S > 31)$ の確率を, 標準正規分布の累積分布関数 $\Phi(z)$ を用いて表し, さらに `scipy.stats.norm.cdf()` (または正規分布表) を用いて小数値として近似的に求めよう.

略解

$n = 400$ が大きいと考えると, 中心極限定理より, S は近似的に正規分布 $N(n\mu, n\sigma^2)$ すなわち $N(40, 6^2)$ にしたがう. $Z = \frac{S-40}{6}$ は近似的に標準正規分布 $N(0, 1^2)$ にしたがう. よって, 求める確率は, $P(S > 31) = P(Z > -\frac{9}{6}) = P(-\frac{9}{6} < Z < +\infty) = \Phi(\infty) - \Phi(-\frac{9}{6}) = 1 - \Phi(-\frac{9}{6}) = 0.9332$.

L09-Q3

TA Prob and Sol: 独立同一分布と中心極限定理

確率変数 X_i ($i = 1, \dots, 400$) は独立同一分布にしたがい, $X_i \sim U(3, 5)$ である. $S = X_1 + \dots + X_{400}$ とする.

S の分布を正規分布で近似し, $P(S \leq 1580)$ の確率を, 標準正規分布の累積分布関数 $\Phi(z)$ を用いて表し, さらに `scipy.stats.norm.cdf()` (または正規分布表) を用いて小数値として近似的に求めよう.

略解

$$\mu = E[X_i] = \frac{3+5}{2}, \sigma^2 = \text{Var}(X_i) = \frac{(5-3)^2}{12}.$$

$n = 400$ が大きいと考えると, 中心極限定理より, S は近似的に正規分布 $N(n\mu, n\sigma^2)$ すなわち $N(1600, \frac{400}{3})$ にしたがう. よって, $Z = \frac{S-1600}{20/\sqrt{3}}$ は近似的に標準正規分布 $N(0, 1^2)$ にしたがう. よって, 求める確率は,
 $P(-\infty < Z \leq -\sqrt{3}) = \Phi(-\sqrt{3}) - \Phi(-\infty) = \Phi(-\sqrt{3}) - 0 = 0.0416.$

L09-Q4

TA Prob and Sol: ベルヌーイ分布の独立同一分布の和と中心極限定理

表が $\frac{4}{5}$, 裏が $\frac{1}{5}$ の確率で出る超いびつなコインを, 100 回投げる. 表が 73 回より多く 79 回以下で出る確率を求めたい.

標準正規分布の累積分布関数 $\Phi(z)$ を用いて表し, さらに `scipy.stats.norm.cdf()` (または正規分布表) を用いて小数値として近似的に求めよう.

略解

100 回中表の出る回数は, $Y = X_1 + \cdots + X_{100}$, $X_i \sim \text{Bin}(1, \frac{4}{5})$, 独立同一分布, $E[X_i] = \frac{4}{5}$, $\text{Var}(X_i) = \frac{4}{5} \frac{1}{5}$. よって, $E[Y] = 80$, $\text{Var}(Y) = 4^2$ である (これは, $Y \sim \text{Bin}(100, \frac{4}{5})$ から求められる)

$n = 100$ が大きいと考えると, 中心極限定理より, Y は近似的に正規分布 $N(80, 4^2)$ にしたがる.

標準化された $Z = \frac{Y-80}{4}$ は近似的に標準正規分布 $N(0, 1^2)$ にしたがる. よって, 求める確率は,

$$P(73 < X \leq 79) = P(-\frac{7}{4} < Z \leq -\frac{1}{4}) = \Phi(-\frac{1}{4}) - \Phi(-\frac{7}{4}) = 0.4599 - 0.0987 = 0.3612.$$

L10-Q1

Quiz(中心極限定理)

X_i ($i = 1, \dots, n$) は独立で、同一の「ある分布」にしたがう。

$U = \frac{1}{n}(X_1 + \dots + X_n)$ について、 n が大きいとき、中心極限定理の結果として言えることをひとつ選ぼう。

- ① 「ある分布」が任意の分布のとき、 U は、近似的に、連続一様分布に従う
- ② 「ある分布」が一般の正規分布のとき、 U は、近似的に、標準正規分布に従う
- ③ 「ある分布」が一般の連続一様分布のとき、 U は、近似的に、標準連続一様分布に従う
- ④ 「ある分布」が標準連続一様分布のとき、 U は、近似的に、正規分布に従う

ここまで来たよ

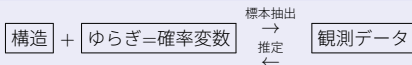
- 9 中心極限定理・独立同一分布の和の正規近似
- 10 推測統計（母集団・標本・標本抽出・推定）・点推定
 - 母集団と標本
 - 母平均値・母比率の(点)推定
 - 母分散の(点)推定

統計モデル

久保川 統計学入門 §7.1

統計モデル

久保川 統計学入門 §7.1



記述統計の母集団と標本, 有限母集団

久保川 統計学入門 §7.1

某アイドルグループ全員 (→ 有限母集団, $N = 46$)

身長 x_i の平均値 $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ を求めたい!

某アイドルグループの滋賀県出身率 $p_1 = m/N$ を求めたい!

握手券を 5 枚買って質問する. 標本 $X_1, \dots, X_5, Y_1, \dots, Y_5$

標本平均値 $\frac{1}{5}(X_1 + \dots + X_5)$, 標本比率 $\frac{1}{5}(Y_1 + \dots + Y_5)$.

データ分析

- 母集団サイズ = 46 の母集団から,
- 標本サイズ = 5 の標本 (sample) を
- 標本の個数 = 1 個を標本抽出する

推定

- 身長之母平均値 $\mu = E[X]$ を求めたい.
- 全員の身長を知れば定義の式使うだけ.
- しかし, いま, 全員の身長 (母集団) を知ることはできない.
- 握手券 46 枚買わず (実際は重複するかも), 5 枚で何とかすませたい.

推測統計の母集団と標本, 無限母集団

久保川 統計学入門 §7.1

某アイドルグループの各メンバーの身長は, 何かの分布にしたがう

滋賀県出身の正否は, ベルヌーイ分布 $Ber(p_1)$ にしたがう.

確率変数と思えば, 何回でも引くことができる (46 を超えて). (→ 無限母集団)

「確率分布 $Ber(p_1)$ にしたがう試行を 5 回する」

＝「5 個の確率変数 $Y_1, \dots, Y_n, \text{i.i.d.} \sim Ber(p_1)$ の実現値を得る」 (p_1 は不明)

「平均値」 $\bar{Y} = \frac{1}{5}(Y_1 + \dots + Y_5)$ も確率変数.

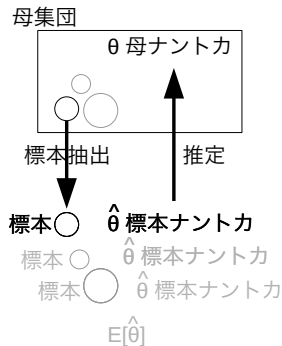
- 母集団サイズ = $+\infty$ の母集団から,
- 標本サイズ = 5 の標本 (sample) を
- 標本の個数 = 1 個を標本抽出する

推定

- 分布の母平均値を知りたい.
- 分布を知れば定義の式を使うだけ
- 分布を知ることはできない.
- 有限個の実現値から何とか「推測」

推測統計：母集団・標本・標本抽出・推定

- **母集団** population = 考えたい集団. どんな分布, 母平均値, 母分散, などわかっていないことがあるが, 全体を調べるわけにはいかない集団. パラメタ θ が備わる.
- **標本**=sample (名詞) = 母集団から '無作為に' とってきた一部分
- **標本抽出**する sample(動詞)=母集団から '無作為に' とってくる $\leadsto$ sampling (動名詞)
- **推定**する estimate(動詞) = 標本を調べて母集団について, 特に θ について, 正しい事実を見つける $\leadsto$ estimation (名詞)
- **確率変数** X , $\bar{X}$ 分布をもつ変数
- **実現値, 観測値** x , $\bar{x}$ 標本を1つにとって確定した値
- **推定量** $\hat{\theta}(x)$ 母集団の量 θ を推定する式
- **点推定** 1個の値を提出する方式の推定



推定には**誤差**ある. 標本の選び方ごとに答は違

科目参加者全体から抽出した標本: 身長, 滋賀県内高校

3 変量データ

- 身長 X = 身長 (参加者)(cm) 量的データ 連続
- $Y = \begin{cases} 1 & (\text{滋賀県高校 Yes}) \\ 0 & (\text{No}) \end{cases}$. 質的データ 名義尺度 久保川 統計学入門 p.34
- $W = \text{A,B,C,D NotebookLM 使う?}$ 質的データ 順位尺度 久保川 統計学入門 p.34

→ 導かれるデータ「身長高い」

$$W = \begin{cases} 1 & (X \geq 170.7) \\ 0 & (\text{No}) \end{cases}$$

統計モデル

$X_i \sim$ 何かの独立同一分布.

$Y_i \sim \text{Ber}(p_1)$, 独立同一分布

NotebookLM $\sim$ 多項分布, 独立同一分布

$W_i \sim \text{Ber}(p_2)$, 独立同一分布. X_i の言い換え.

こういう文脈での, ベルヌーイ分布のパラメータ, 母平均値 p を, **母比率** (ratio) という.

- $p_1 = E[Y] = P(Y = 1)$.
- $p_2 = E[W] = P(X \geq 170.7)$.

母集団=回答者全体

- ① 母集団サイズ 101 (クラス全体が回答すれば 118 だった)

サブチームに割り当てられた標本

- ① 1 サブチームに割り当てる標本の個数 1 個
- ② 標本サイズ 様々

久保川 統計学入門 p.34

- よい標本抽出には実はいろいろなハイテクがある $\leadsto$ 社会調査法, 社会統計学
- 本当は, 母集団が無限でないとき, 標本を作るのに重複選出を禁止するとき, 数学的にいろいろな複雑なことが起きるのだが, この授業では知らん顔する. $\leadsto$ 確率・統計

ここまで来たよ

9 中心極限定理・独立同一分布の和の正規近似

10 推測統計（母集団・標本・標本抽出・推定）・点推定

- 母集団と標本
- 母平均値・母比率の（点）推定
- 母分散の（点）推定

母平均値の (点) 推定 高校 数学 C 久保川 統計学入門 §7.2

組 $(X_1, X_2, \dots, X_n)$ はサイズ n の標本. 各 X_i は母平均値 $\mu = E[X_i]$, 母分散 $\sigma^2 = \text{Var}(X_i)$ の独立同一分布にしたがう確率変数.

定義 (標本平均値 久保川 統計学入門 公式 7.2)

$$\text{標本平均値 } \hat{\theta}(X_1, \dots, X_n) = \bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n) = \text{先週の } U_n$$

母平均値 $\mu = E[X_i]$ の 'よい' 推定量になっている. 'よい' $\simeq$ 不偏推定量
Pandas/Python では `.mean()`, Excel, Google スプレッドシートでは関数 `average()`

定義 (標本期待値)

$$g(X) \text{ の標本期待値 } \overline{g(X)} = \frac{1}{n}(g(X_1) + \dots + g(X_n))$$

$E[g(X_i)]$ の 'よい' 推定量になっている.

復習：中心極限定理

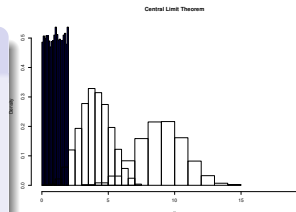
久保川 統計学入門 §7.4

定理 (中心極限定理 (いいかげんバージョン))

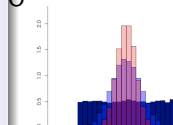
$X_1, \dots, X_n$ が母平均値 μ , 母分散 σ^2 の独立同分布に従うとき, $n \rightarrow +\infty$ で

- $S_n = X_1 + \dots + X_n$ の確率分布は,
正規分布 $N(n\mu, n\sigma^2)$ に似る
- $U_n = \frac{1}{n}(X_1 + \dots + X_n)$ の確率分布は,
正規分布 $N(\mu, \sigma^2/n)$ に似る
- 標準化した $Z_n = \frac{U_n - \mu}{\sigma/\sqrt{n}}$ の確率分布は,
標準正規分布 $N(0, 1^2)$ に似る

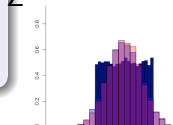
S



U



Z



‘よい’推定量の資格 1

パラメタ θ を持つ分布の

母平均値 $E[X]$, 母期待値 $E[g(X)]$ はひとつに定まっているが,

標本平均値 $\bar{X}$, $g(\bar{X})$ は確率変数で, 試行=標本抽出のたびに変わる ($\bar{X}$ は確率分布をもつ)

一般に推定量 $\hat{\theta}(X_1, \dots, X_n)$ は確率分布を持つ.

定義 (推定量の不偏性 久保川 統計学入門 定義 7.1)

$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ が, パラメタ θ の不偏推定量であるとは,

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

が成り立つこと.

標本平均値 $\hat{\theta} = \bar{X}_n$ は, 母平均値 $\theta = \mu$ の不偏推定量である: $E[\bar{X}_n] = \mu$. 中心極限定理
また, $\text{Var}(\bar{X}_n) = \sigma^2/n$. 中心極限定理

‘よい’ 推定量の資格 2

定義 (推定量の一致性 久保川 統計学入門 定義 7.5)

$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ が, パラメタ θ の一致推定量であるとは,

$$\forall \epsilon > 0 \quad \lim_{n \rightarrow +\infty} P(|\hat{\theta}(X_1, \dots, X_n) - \theta| > \epsilon) = 0.$$

推定量と母ナントカに一定の差がある確率が, 標本サイズ n を大きくすると zero になること.

大数の法則 確率統計 I(2026)L09 より, 標本平均値 $\hat{\theta} = \bar{X}_n$ は, 母平均値 $\theta = \mu$ の一致推定量である.

大数の (弱) 法則

確率統計 I(2026)L9

$X_1, \dots, X_n$ が独立同一分布にしたがい (n 回のくじの賞金), $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$ とする.

確率変数 $\bar{X} = U_n = \frac{1}{n} \sum_{i=1}^n X_i$ を考える (n 回の賞金の相加平均) と

$$E[U_n] = \mu, \text{Var}(U_n) = \sigma^2/n.$$

定理 (大数の (弱) 法則アバウト版) 久保川 統計学入門 公式 7.8

n が十分大きいとき確率変数 U_n は「 μ に近い値ばかり」(U_n が μ から外れる確率はゼロに近づく)

弱法則の正確な表現

定理 (大数の (弱) 法則) 久保川 統計学入門 公式 7.8

$$\forall \epsilon > 0 \quad \lim_{n \rightarrow +\infty} P(|U_n - \mu| \geq \epsilon) = 0.$$

「 U_n は μ に確率収束する」

L10-Q2

Quiz(母平均値, 母分散, 母比率の点推定)

フライドチキン屋さんのフライドチキンの大量の在庫 (=母集団) から, 無作為に 6 本のチキンを取り出したところ, 重さは次のようだった.

118g, 109g, 109g, 119g, 101g, 113g.

- ① 重さの母平均値を点推定しよう.
- ② 重さの二乗の母期待値を点推定しよう.
- ③ 重さの母分散を点推定しよう.
- ④ 105g 以上のものの母比率を点推定しよう.

記述上の注意 厳しく弾圧します. × き

- 母平均値 $= \mu = E[X] \neq$ 標本平均値 $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$.
- 母分散 $= \sigma^2 = \text{Var}(X) \neq$ 不偏標本分散 $V^2 = \frac{1}{n-1}(\dots)$
- 母比率 $= p \neq$ 標本比率 $\hat{p} = \frac{k}{n}$.
- 「母ナントカを…と × 求めた○推定する」

比率=ratio

久保川 統計学入門 p.12

確率変数 $Y \sim \text{Ber}(p)$ ベルヌーイ分布, を考える.

こういう Y は, いろんな母集団を, 「 X は…である」という条件の成立不成立で 2 つに類別して作れる. **名義データ カテゴリ変数**

- $X \sim$ ある分布, $Y = \begin{cases} 1 & (X \text{ の条件成立}) \\ 0 & (X \text{ の不成立}) \end{cases}$. 例 $X > 10$ なら $Y = 1$.
- 母集団=日本国民, その国民血液型が A であるなら $Y = 1$.

定義 (母比率)

$\text{Ber}(p)$ の p を母比率という. 母集団で X の条件「…」から $Y \sim \text{Ber}(p)$ を作ったとき, '**母集団の「…である」ものの母比率 p** ', ともいう.

有限母集団なら,

母集団の「…である」母比率 $p = \frac{\text{「…である」メンバー } x \text{ の個数}}{\text{すべてのメンバーの個数}} = E[Y]$

母比率の (点) 推定

定義 (標本比率)

ベルヌーイ分布にしたがう $Y_1, \dots, Y_n \sim \text{Ber}(p)$ に対しては, (推定量である) 標本平均値を, 特に標本比率 $\hat{p}$ という. $Y_i = 1$ のものが K 個とすると,

$$\hat{p}(Y_1, \dots, Y_n) = \frac{1}{n}(Y_1 + \dots + Y_n) = \frac{K}{n}$$

標本比率は, 母比率 p の不偏推定量, 一致推定量

$$\frac{1}{n} \left[\underbrace{1 + \dots + 1}_k + \underbrace{0 + \dots + 0}_{n-k} \right] = \frac{k}{n}.$$

ここまで来たよ

9 中心極限定理・独立同一分布の和の正規近似

10 推測統計（母集団・標本・標本抽出・推定）・点推定

- 母集団と標本
- 母平均値・母比率の（点）推定
- 母分散の（点）推定

母分散の (点) 推定

母分散 $\text{Var}(X) = E[(X - \mu)^2]$ もしよせん母期待値じゃん．データ分析で使った

$$\text{標本分散 } S^2 = \frac{1}{n} [(X_1 - \bar{X}_n)^2 + \cdots + (X_n - \bar{X}_n)^2]$$

がよい推定値じゃないの？…ちょっと待った！

$\bar{X}_n$ の中には $X_1, \dots, X_n$ がぜんぶ入ってるので，定数 μ であるかのような計算はできない．

分布の重心がちょっとずれるので修正が必要．

今後，その分散や分布を考えようと思っても，

$(X_1 - \bar{X}_n)^2, \dots, (X_n - \bar{X}_n)^2$ は互いに独立ではないから簡単ではない．

母分散の (点) 推定

定義 (不偏標本分散 久保川 統計学入門 公式 7.2)

$$\begin{aligned}\text{不偏標本分散 } \hat{\theta}(X_1, \dots, X_n) = V^2 &= \frac{1}{n-1} [(X_1 - \bar{X}_n)^2 + \dots + (X_n - \bar{X}_n)^2] \\ &= \frac{n}{n-1} \left[\frac{1}{n} \sum_i X_i^2 - (\bar{X}_n)^2 \right]\end{aligned}$$

不偏標本分散は母分散 σ^2 の 'よい' 推定値 不偏性 $E[V^2] = \sigma^2$.

不偏標本分散は 久保川 統計学入門 p.155 の不偏分散のこと. 久保川 統計学入門 p.154 の標本分散は $n-1$ でなく n で割ったもの.

ここで, $\bar{X}_n$ は母平均値でなく, 上の標本平均値.

$n-1$ の理由 こうしないと不偏にならない

直観的理由 $\bar{X}$ は X_i の重心だから, μ より近くにある. $(X_i - \bar{X})^2$ は $(X_i - \mu)^2$ より小さくなりがち ($\frac{n-1}{n}$ 倍) なので修正.

不偏標本分散の不偏性の確認

$$n = 2. \text{Var}(X_i) = \sigma^2. \bar{X} = \frac{1}{2}(X_1 + X_2).$$

$$\begin{aligned} E[V^2] &= E\left[\frac{1}{2-1}((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2)\right] \\ &= E\left[(X_1 - \frac{1}{2}(X_1 + X_2))^2 + (X_2 - \frac{1}{2}(X_1 + X_2))^2\right] \\ &= 2 \cdot \frac{1}{4} E[(X_1 - X_2)^2] \\ &= 2 \cdot \frac{1}{4} E[X_1^2 - 2X_1X_2 + X_2^2] \\ &= 2 \cdot \frac{1}{4} ((\sigma^2 + \mu^2) - 2\mu\mu + (\sigma^2 + \mu^2)) \\ &= 2 \cdot \frac{1}{4} (2\sigma^2) = \sigma^2. \end{aligned}$$

不偏標本分散は、Pandas/Python では `.var(ddof=1)`, Excel では関数 `var.s()`, Google スプレッドシートでは, `var()`, `var.s()`s for sample(標本)

標本分散は、Pandas/Python では `.var()`, Excel では関数 `var.p()`, Google スプレッドシートでは, `varp()`, `var.p()`p for population(母集団)