母比率の推定再訪・離散-連続型確率変数

樋口さぶろお https://hig3.net

龍谷大学大学院 先端理工学研究科 数理・情報科学コース

計算科学特論 L01(2026-04-13 Mon)

最終更新: Time-stamp: "2026-04-19 Sun 07:31 JST hig"

今日の目標

- この授業ののりを説明できる
- （従来方式の）母比率の信頼区間とその意味を説明できる
- 離散-連続型確率分布を説明できる

ここまで来たよ

学習目標など

到達目標 $\rightarrow$ シラバス

- I. ベイズ統計
  - ベイズ統計学に基づいて統計ソフトウェアを用いて標本からベイズ的な推定を行える
  - フィッシャー流頻度主義統計学 確率統計 I(2023)L?? 実現値，確率変数，母数（パラメタ）$x_1, \dots,\ X \sim \mathrm{N}(\mu, \sigma^2)$
  - $\leftrightarrow$ ベイズ統計学 母数も確率変数
- II. 多変量解析，時系列解析

I ベイズ推定，と思っていいです．

資料など

`https://hig3.net`

- → 計算科学特論 (配布資料).
- → LearnMoodle `https://learn.hig3.net` (Google でログイン) → 計算科学特論
  - → Teams

`https://hig3.net`

参考書

- I. [岩薩林] 岩佐, 薩摩, 林, 理工系の数理 確率統計, 2018
- I. [原] 原，原理と意味から理解するベイズ統計入門講義, 2025
- I. [赤石] 赤石, Python でスラスラわかるベイズ推論「超」入門，講談社，2023 `https://github.com/makaishi2/python_bayes_intro`

授業のやり方とピーナッツカウント

授業のり対面，Teams 画面共有下で，PDF 資料に GoodNotes で書き込み，録画公開するかも．
資料はたぶん直前まで完成しません．
Python on Google Colab (来週以降)
グループワークあるかも．

ピーナッツ計算ののり
平常点 100 ピーナッツ．（手計算撮影 and/or 数式入力）Quiz 12.5x8(授業内または次の授業までに Moodle で)

欠席届
毎回出席を前提に進めます．やむを得ず欠席する場合や，課題を提出できない場合で，ピーナッツ的に考慮されたい場合は事後 2 週間以内に LearnMoodle にある専用のリンクから届けてください．

担当者ののり

- なまえ: 樋口さぶろお a00010 at mail, chat はいつでも．
- へや: 1-513
- Web ページ．`https://hig3.net`
- オフィスアワー 後期木昼, 1-539 or Teams chat a00010

ここまで来たよ

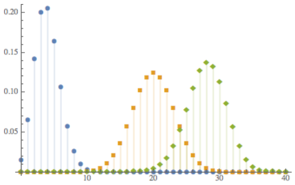
二項分布 高校 数学 B 岩薩林 §3.4

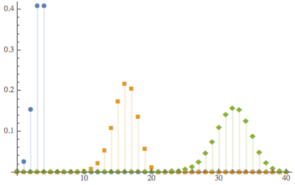
定義 (二項分布 岩薩林 (3.24)p.66)

離散型確率変数 X が次の確率関数を持つとき, X はパラメタ n, p の二項分布 $\mathrm{B}(n,p)$ にしたがうという．$\mathrm{B}(1,p)$ ベルヌーイ分布．

$$f(x) = \begin{cases} \frac{n!}{x!(n-x)!}\, p^x (1-p)^{n-x} & (x = 0, 1, 2, 3, \ldots, n) \\ 0 & (\text{他}) \end{cases}$$

意味: 確率 p で表の出るコインを n 回投げたとき, x 回表が出る確率.





二項係数 高校 数学 A

$${}_n\mathrm{C}_x = \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$\mathrm{B}(40, 0.1), \mathrm{B}(40, 0.5), \mathrm{B}(40, 0.7)$, $\mathrm{B}(4, 0.8), \mathrm{B}(20, 0.8), \mathrm{B}(40, 0.8)$

身分制：母数・確率変数・実現値

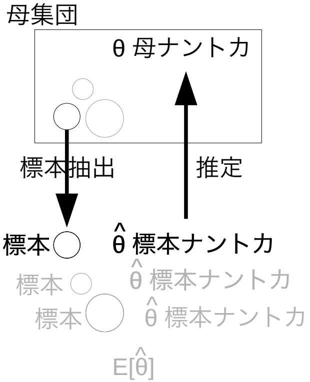
（暗に測度空間が仮定されている）

- X 確率変数
- $f(x)$ 確率関数（が確率分布を決める）
- n, p パラメタ，母数（確率変数ではない）
- $X_1, \ldots, X_N$ 独立同分布にしたがう確率変数（N 回の試行）
- $x, x_1, \ldots, x_N$ 確率変数の実現値（確率変数ではない）
- $(x_1, \ldots, x_N)$ サイズ N の標本
- $\mathrm{E}[X]$ 母平均値
    - 二項分布 $\mathrm{B}(n, p)$ の場合 np.
- $\overline{X} = \frac{1}{N}(X_1 + \cdots + X_N)$ 標本平均値（母平均値の推定量）
    - $\mathrm{B}(1, p)$ のとき，$= \frac{K}{N}$. そのまま母比率 p の推定量になっている
- $\overline{x} = \frac{1}{N}(x_1 + \cdots + x_N)$ 標本平均値の実現値

母ナントカ 対 標本ナントカ

予告：ベイズ統計学においては，X, p の身分が民主化される

岩薩林 図 p.109,115,137,167



連続型確率変数 岩薩林 §4.1

X: 実数に値をとる確率変数 (確率空間から実数への関数)
確率 $P(X$ の条件$)$ が考えられる (大部分の条件に対して)

定義 (確率密度関数 (pdf=probability density function))

確率密度関数 $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ があって, 次のように書ける.

$$P(a < X \leq b) = \int_a^b f(x)\,\mathrm{d}x.$$

定義 (母期待値)

X: 連続型確率変数, $g(x)$: 関数に対して, $g(X)$ の母期待値

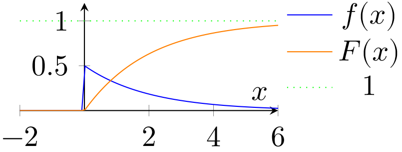
$$\mathrm{E}[g(X)] := \int_{-\infty}^{+\infty} g(x) f(x) \mathrm{d}x.$$

累積分布関数

定義 (累積分布関数 (cdf=cumulative density function))

$$F(x) := \int_{-\infty}^{x} f(x')\,\mathrm{d}x'.$$

指数分布の確率密度関数と累積分布関数



命題 (累積分布関数の性質)

- $F(x)$ は広義単調増加.
- $P(X \leq b) = F(b)$.
- $P(a < X \leq b) = F(b) - F(a)$
- $F(-\infty) = 0, \quad F(+\infty) = 1.$

ここまで来たよ

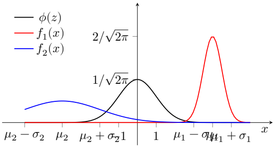
一般の正規分布 $\mathrm{N}(\mu, \sigma^2)$

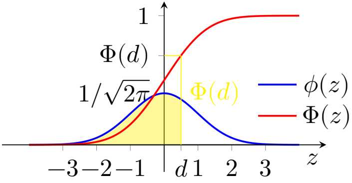
定義 (一般の正規分布 $\mathrm{N}(\mu, \sigma^2)$)

次の確率密度関数を持つ確率変数 X を，母平均値 μ, 母分散 σ^2 の正規分布 $\mathrm{N}(\mu, \sigma^2)$ (normal distribution) にしたがうという．

$$f_{\mathrm{X}}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$X \sim \mathrm{N}(\mu, \sigma^2)$ を標準化 $Z = \frac{X-\mu}{\sigma}$ すると標準正規分布 $Z \sim \mathrm{N}(0, 1^2)$.
標準正規分布の確率密度関数 $\phi(z)$, 累積分布関数 $\Phi(z)$.





信頼区間 (confidence interval)

定義 (頻度論的な信頼区間)

母数 θ の区間で，上下限が確率変数であるようなもの $[\ell(X_1,\dots,X_N), u(X_1,\dots,X_N)]$ が，パラメタ θ に対する，信頼係数 $1-\alpha$ の信頼区間であるとは，

$$P(\ell(X_1,\dots) < \theta < u(X_1,\dots)) = 1-\alpha$$

であることをいう．

これに，実現値 $X_i = x_i$ を代入したものも信頼区間とよばれる．
しばしば，$P(\theta < \ell(X_1,\dots)) = P(\theta > u(X_1,\dots)) = \frac{\alpha}{2}$ であるように（上下対称であるように）とる．

例

$\mathrm{B}(1,\theta)$ のサイズ N の標本．$\theta = p$

θ の 点推定 標本平均値 $\overline{X} = \frac{1}{N}\sum_i x_i$. $\mathrm{E}[\overline{X}] = \theta$, $\mathrm{V}[\overline{X}] = \theta(1-\theta)/N$.

θ の区間推定

中心極限定理による近似

$$P(\theta + \Phi^{-1}(\frac{\alpha}{2})\sqrt{\frac{\theta(1-\theta)}{N}} < \overline{X} < \theta + \Phi^{-1}(1-\frac{\alpha}{2})\sqrt{\frac{\theta(1-\theta)}{N}}) = 1-\alpha$$

移項

$$P(\overline{X} - \Phi^{-1}(1-\frac{\alpha}{2})\sqrt{\frac{\theta(1-\theta)}{N}} < \theta < \overline{X} - \Phi^{-1}(\frac{\alpha}{2})\sqrt{\frac{\theta(1-\theta)}{N}}) = 1-\alpha$$

近似 $\theta = \overline{X}$

$$P(\overline{X} - \Phi^{-1}(1-\frac{\alpha}{2})\sqrt{\frac{\overline{X}(1-\overline{X})}{N}} < \theta < \overline{X} - \Phi^{-1}(\frac{\alpha}{2})\sqrt{\frac{\overline{X}(1-\overline{X})}{N}}) = 1-\alpha$$

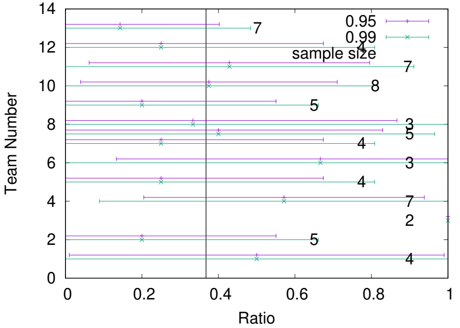
L01-Q1

Quiz(母比率の区間推定)

選挙で出口調査をしたところ, 50 人中 35 人が A 候補に投票したと答えた. 母集団を投票した人 2500 人とする. そのうち A 候補に投票した人の母比率 (得票率) を考える.

1. A 候補の得票率を, (点) 推定しよう
2. A 候補の得票率を, 信頼係数 $1-\alpha=0.95$ で区間推定しよう.
3. A 候補の得票率を, 信頼係数 $1-\alpha=0.99$ で区間推定しよう.

標本抽出を繰り返したとき，（未知だが定数の）θ が（毎回得られる）区間に含まれる確率が $1-\alpha$.



ここまで来たよ

2 つの離散型確率変数の同時確率分布 高校 数学 B 岩薩林 p.56

例 6 枚のカードから無作為に 1 枚引く. ♡7 ♡8 ♡9 ⋄8 ♠9 ♣9

2 つの離散型確率変数の同時分布

$X=$ 数, $Y=0$(赤札), 1(黒札) とすると (x,y) を得る確率は 2 変数の確率関数で書ける. 同時分布, 結合分布, joint distribution という.

コンマ前が X, 後が Y の値. 明示のため $p_{\mathrm{XY}}(x,y)$ と書くことも. 岩薩林 では, p_{xy} とも. $P(X=x, Y=y)$ とも. P(条件) 型の表記.

表で書いた方がまし. ここでは,「他」は省略.

表の縦軸横軸に意味なし. ラベル X, Y をよく見て.

$$p(x,y)=\begin{cases}\frac{1}{3} & ((x,y)=(8,0))\\ \frac{1}{6} & ((x,y)=(9,0))\\ \frac{1}{3} & ((x,y)=(9,1))\\ \frac{1}{6} & ((x,y)=(7,0))\\ 0 & (\text{他})\end{cases}$$

$y \backslash x$	7	8	9	計
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	
1	0	0	$\frac{1}{3}$	
計				

周辺分布 岩薩林 §3.3

定義 (確率変数の周辺分布)

同時分布 $p(x, y)$ に対して,
X の周辺分布 $p_{\mathrm{X}}(x)$, Y の周辺分布 $p_{\mathrm{Y}}(y)$ は,

$$p_{\mathrm{X}}(x) = p(x, \bullet) = \sum_y p(x, y), \quad p_{\mathrm{Y}}(y) = p(\bullet, y) = \sum_x p(x, y)$$

要するに一方を無視した分布．小計．

岩薩林 例題 3.4(p.57)

$y \backslash x$	7	8	9	計
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	
1	0	0	$\frac{1}{3}$	
計				

条件付き確率 岩薩林 §2.3

定義 (条件付き確率 岩薩林 (2.12))

条件 (事象) A のもとでの, 事象 B の条件付き確率

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

縦棒「|=given」の前後は不平等: P(確率を考える事象|条件の事象)
$P(B|A) \neq P(A|B), \neq P(A \cup B)$.

意味 $\boxed{A\text{ が起きる場合に限定した, }B\text{ が起きる確率}}$

$P(B|A)$ を P (of) B given A と発音する人がいる. given 条件=条件が与えられたとき, という言い方から.

縦棒の前だけ見る (縦棒の後を固定する) と, ただの確率. 全事象は 1.

離散型確率変数に対する条件付き確率

同時確率分布

$p(x,y) = P(X = x$かつ$Y = y) =$

$y\backslash x$	150	160	165
45	3/8	0	1/12
50	1/8	1/3	1/12

離散型確率変数に対する条件付き確率 岩薩林 (3.15)

条件 (事象)$Y = y$ のもとでの，事象 $X = x$ の条件付き確率の記号
$P(X = x|Y = y) = p_{\mathrm{X}|\mathrm{Y}}(x$ の値$|y$ の値$)$

例 $p_{\mathrm{X}|\mathrm{Y}}(150|45), p_{\mathrm{Y}|\mathrm{X}}(45|150)$．縦棒の前だけ見る (縦棒の後ろを固定する) と，前の変数についての，ただの確率分布．意味 自分の言葉でどうぞ

命題 (条件付き確率を同時確率分布と周辺分布で表す式 岩薩林 (2.12))

$$P(X = x|Y = y) = p_{\mathrm{X}|\mathrm{Y}}(x|y) = \frac{p(x,y)}{p_{\mathrm{Y}}(y)},$$

$$P(Y = y|X = x) = p_{\mathrm{Y}|\mathrm{X}}(y|x) = \frac{p(x,y)}{p_{\mathrm{X}}(x)}.$$

命題 (乗法公式 岩薩林 (2.13)：実現値 (x, y) ペアを得るのに便利な式)

$$\begin{aligned} p(x, y) &= p_{\mathrm{X|Y}}(x|y)p_{\mathrm{Y}}(y) \\ &= p_{\mathrm{Y|X}}(y|x)p_{\mathrm{X}}(x) \end{aligned}$$

X の実現値が 2 種類，Yの実現値が 2 種類であるとき，(X, Y) の 4 面サイコロを使う必要はなく，
(1 行目)Yの $\mathrm{B}(1, p)$ コインをまず投げ，次に，Yの実現値 y に応じて, X の $\mathrm{B}(1, p'(y))$ コインを投げればよい．
(2 行目)X の $\mathrm{B}(1, p)$ コインをまず投げ，次に，X の実現値 x に応じて, X の $\mathrm{B}(1, p'(x))$ コインを投げればよい．

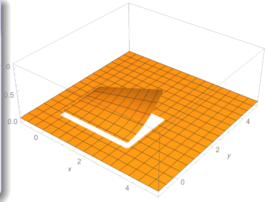
2 次元の連続型確率変数 岩薩林 §4.6

X, Y: 連続型確率変数, $D \subset \mathbb{R}^2$ の, ある性質を満たす部分集合.

定義 (同時確率密度関数)

次の性質を持つ関数 $f : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ を同時確率密度関数という.

$$P((X, Y) \in D) = \iint_D f(x, y) \mathrm{d}x \mathrm{d}y.$$



定義 (母期待値)

X, Y: 連続型確率変数, $g(x, y)$: 関数に対して, $g(X, Y)$ の母期待値

$$\mathrm{E}[g(X, Y)] := \iint_{\mathbb{R}^2} g(x, y) f(x, y) \mathrm{d}x \, \mathrm{d}y.$$

ここまで来たよ

はじめに
- この授業どんなのり?

1. 母比率の推定再訪・離散-連続型確率変数
   - 1 変数の確率変数
   - 母比率の区間推定
   - 2 変数の確率分布
   - 離散-連続型確率分布

離散-連続型確率変数

確率変数の組 (X, Y) を考える．
X: 連続型確率変数，
Y: 離散型確率変数, $Y = y_1, \dots, y_n$.
次を満たす確率密度関数 $f(x, y)$ で指定される．

定義 (離散-連続混合型分布)

$$P(a < X \le b, Y = y_i) = \int_a^b f(x', y_i)\, \mathrm{d}x'.$$

$$P(a < X \le b, y_i \le Y \le y_j) = \int_a^b \sum_{y_i \le y' \le y_j} f(x', y')\, \mathrm{d}x'$$

気分を出すために, 表や場合分けで $f(x, y)$. ただし, $Y = 0, 1$. x についての確率密度関数 $h_0(x), h_1(x)$.

$$f(x, y) = \begin{cases} h_0(x) & (y = 0) \\ h_1(x) & (y = 1) \\ 0 & (y \neq 0, 1) \end{cases}$$

$y \backslash x$	$-\infty < x < +\infty$	
0	$h_0(x)$	
1	$h_1(x)$	

全 $= \int_{-\infty}^{+\infty} h_0(x)\,\mathrm{d}x + \int_{-\infty}^{+\infty} h_1(x)\,\mathrm{d}x = 1.$

周辺分布は？ 条件付き分布は？

独立である例

同時分布

$$f(x, y) = u(x; c, d) \times b(y; p)$$

略記 $u(x; c, d)$ は，連続型一様分布 $\mathrm{U}(c, d)$ の確率密度関数

$$u(x; c, d) = \begin{cases} \frac{1}{d-c} & (c < x \le d) \\ 0 & (\text{他}) \end{cases}$$

略記 $b(y; p)$ は，ベルヌーイ分布 $\mathrm{B}(1, p)$ の確率関数

$$b(y; p) = \begin{cases} 1 - p & (y = 0) \\ p & (y = 1) \end{cases}$$

$\mathrm{U}(0, 1)$ のルーレット，$\mathrm{B}(1, p)$(p 任意) のコインが使えるなら，X にルーレット，Y にコインを別々に使えばよい．

L01-Q2

Quiz(離散-連続混合型分布の周辺分布)

連続型確率変数 X, 離散型確率変数 Y の同時分布が次のように与えられる．ただし，$u(x; c, d)$ は連続型一様分布 $\mathrm{U}(c, d)$ の確率密度関数．

$$f(x, y) = \begin{cases} (1-x)\cdot u(x; 0, 1) & (y = 0) \\ x \cdot u(x; 0, 1) & (y = 1) \\ 0 & (\text{他}) \end{cases}$$

1. X の周辺分布の確率密度関数を描こう．
2. Y の周辺分布を求めよう．
3. $Y = 1$ という条件のもとでの，X の分布 $p_{\mathrm{X}|\mathrm{Y}}(x|1)$ を求めよう．
4. $X = \frac{7}{10}$ という条件のもとでの，Y の分布 $p_{\mathrm{Y}|\mathrm{X}}(y|\frac{7}{10})$ を求めよう．
5. $\mathrm{B}(1, p)$(p 任意) のコインと，$\mathrm{U}(0, 1)$ のルーレットが使えるとき，X, Y の実現値は，X, Yの順に求めるほうが簡単であることを納得しよう．

L01-Q3

Quiz(離散-連続混合型分布の周辺分布)

連続型確率変数 X, 離散型確率変数 Y の同時分布が次のように与えられる．ただし，$u(x;c,d)$ は連続型一様分布 $\mathrm{U}(c,d)$ の確率密度関数．

$$f(x,y) = \begin{cases} \frac{3}{10}u(x;1,5) & (y=0) \\ \frac{7}{10}u(x;4,6) & (y=1) \\ 0 & (\text{他}) \end{cases}$$

1. X の周辺分布の確率密度関数を描こう．
2. Y の周辺分布を求めよう．
3. $Y=1$ という条件付きの，X の分布 $p_{\mathrm{X}|\mathrm{Y}}(x|1)$ を求めよう．
4. $X=4.5$ という条件付きの，Y の分布を求めよう．
5. $\mathrm{B}(1,p)$(p 任意) のコインと，$\mathrm{U}(0,1)$ のダーツが使えるとき，X,Yの実現値は Y,X の順に求めるほうが簡単であることを納得しよう．