

# 母比率の推定再訪+PyMC

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(2025-09-24 Wed)

最終更新: Time-stamp: "2025-10-15 Wed 17:26 JST hig"

## 今日の目標

- この授業ののりを説明できる
- Colab, PyMC で確率分布を可視化できる
- (従来方式の) 母比率の信頼区間とその意味を説明できる



# 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+PyMC

- 1 変数の離散型確率変数
- 1 変数の連続型確率変数

# 学習目標など

## 到達目標 → 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(2021)L?? 実現値, 確率変数, 母数 (パラメタ)  $x_1, \dots, X \sim N(\mu, \sigma^2)$
- ▶  $\leftrightarrow$  ベイズ統計学 母数も確率変数

### ● II. マルコフ連鎖モンテカルロ法=MCMC.

- ▶ MCMC の中身までわかって, MCMC でベイズ事後分布を求められる

I ベイズ推定, と思っていいです.

II は I(や他の目的) のための計算手段の話. ただし, マルコフ連鎖は, モンテカルロ法やベイズ統計に限らず, 必然的に確率論・確率過程的に興味深い対象です.

## 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,II. 岩薩林 岩佐, 薩摩, 林, 理工系の数理 確率統計, 2018
- I,II. 赤石 ベイズ推論 赤石雅典, 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](https://github.com/makaishi2/python_bayes_intro)
- II. 鎌谷 MC 鎌谷研吾 モンテカルロ統計計算 講談社 2020  
<https://kkamatani.github.io/MonteCarloText/>

# 授業のやり方とピーナッツカウント

授業のり オンラインライブ, Teams 画面共有下で, PDF 資料に GoodNotes で書き込み, 録画公開. GoodNotes の共有 URL は Web から閲覧, アプリから編集可能.

資料はたぶん直前まで完成しません. 印刷しません.

Python on Google Colab (来週以降)

参加者発声あり. グループワークあり.

ピーナッツ計算のり

平常点 100 ピーナッツ. グループワーク参加 2.5x8(直後にグループ同内容で Moodle に提出), (手計算撮影 and/or 数式入力) Quiz 10x8(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.

担当者のり

- なまえ: 樋口さぶろお a00010 at mail, chat はいつでも.
- へや: 1-507
- 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

### ● はじめに

- この授業どんなのり?

### ① 母比率の推定再訪+PyMC

- 1 変数の離散型確率変数
- 1 変数の連続型確率変数

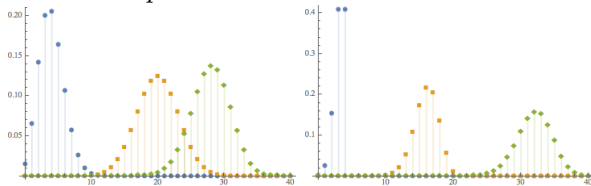
## 二項分布 高校 数学 B 岩薩林 §3.4

### 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  $B(n, p)$  にしたがうという.  $B(1, p)$  ベルヌーイ分布.

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

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

$${}_n C_x = \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$B(40, 0.1), B(40, 0.5), B(40, 0.7), B(4, 0.8), B(20, 0.8), B(40, 0.8)$

## 身分制：母数・確率変数・実現値

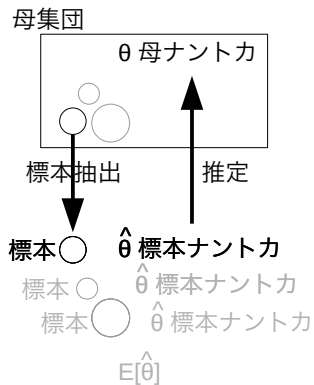
(暗に測度空間が仮定されている)

- $X$  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)
- $X_1, \dots, X_N$  独立同分布にしたがう確率変数 ( $N$  個のコイン)
- $x, x_1, \dots, x_N$  確率変数の実現値 (確率変数ではない)
- $(x_1, \dots, x_N)$  サイズ  $N$  の標本
- $E[X]$  母平均値
  - ▶ 二項分布  $B(n, p)$  の場合  $np$ .
- $\bar{X} = \frac{1}{N}(X_1 + \dots + X_N)$  標本平均値 (母平均値の推定量)
  - ▶  $B(1, p)$  のとき,  $= \frac{K}{N}$ . そのまま母比率  $p$  の推定量になっている
- $\bar{x} = \frac{1}{N}(x_1 + \dots + x_N)$  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



## 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+PyMC

- 1 変数の離散型確率変数
- 1 変数の連続型確率変数

## 確率密度関数と累積分布関数 岩薩林 §4.1 Replace

$X$ : 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  $P(X \text{ の条件})$  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  $f: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  があって、次のように書ける.

$$P(a < X \leq b) = \int_a^b f(x) dx.$$

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

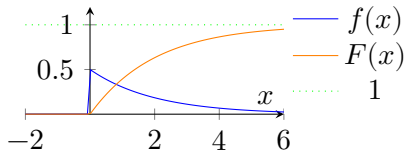
$$E[g(X)] := \int_{-\infty}^{+\infty} g(x)f(x)dx.$$

## 累積分布関数

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

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

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.
- $P(X \leq b) = F(b)$ .
- $P(a < X \leq b) = F(b) - F(a)$
- $F(-\infty) = 0, \quad F(+\infty) = 1$ .

## 経験累積分布関数

定義 (経験累積分布関数 (empirical cumulative density function))

$X$  のサイズ  $N$  の標本の実現値  $\{x_i\}_{i=1,\dots,N}$  に対して,

$$\hat{F}_N(x) := \frac{|\{i|x_i \leq x\}|}{N}$$

確率密度関数/ヒストグラムのほうが直観的だが (?), 累積分布関数のほうが数学的に扱いやすい

| 母ナントカ                                          | 標本ナントカ              |
|------------------------------------------------|---------------------|
| 確率密度関数                                         | ヒストグラム              |
| 累積分布関数                                         | ‘標本累積分布関数’=経験累積分布関数 |
| ヒストグラムは階級によるし, $N \rightarrow +\infty$ で母に似ない. |                     |
| 経験累積は $N \rightarrow +\infty$ で, ある意味, 母に収束する. |                     |

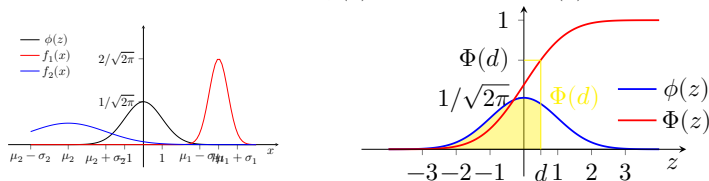
# 一般の正規分布 $N(\mu, \sigma^2)$

## 定義 (一般の正規分布 $N(\mu, \sigma^2)$ )

次の確率密度関数を持つ確率変数  $X$  を、母平均値  $\mu$ 、母分散  $\sigma^2$  の正規分布  $N(\mu, \sigma^2)$  (normal distribution) にしたがるという。

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

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



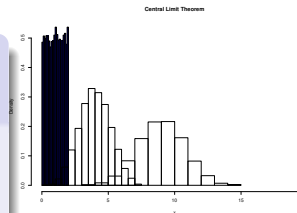
# 中心極限定理 岩薩林 §4.4

定理 (中心極限定理 (いいかげんバージョン)) 岩薩林 定理 4.2(p.87)

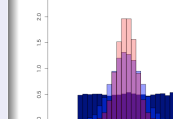
$X_1, \dots, X_n$ : 独立同一分布にしたがう.  $E[X_i] = \mu$ ,  $V[X_i] = \sigma^2$ .  $n \rightarrow +\infty$  で,

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

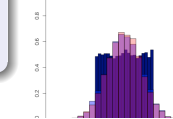
S



U



Z



## 信頼区間 (confidence interval)

### 定義 (頻度論的な信頼区間 Replace)

母数  $\theta$  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  $\theta$  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

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

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

$\theta$  の区間推定

中心極限定理による近似

$$P\left[\theta + \Phi^{-1}\left(\frac{\alpha}{2}\right)\sqrt{\frac{\theta(1-\theta)}{N}} < \bar{X} < \theta + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{\theta(1-\theta)}{N}}\right] = 1 - \alpha$$

移項

$$P\left[\bar{X} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{\theta(1-\theta)}{N}} < \theta < \bar{X} - \Phi^{-1}\left(\frac{\alpha}{2}\right)\sqrt{\frac{\theta(1-\theta)}{N}}\right] = 1 - \alpha$$

近似  $\theta = \bar{X}$

$$P\left[\bar{X} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{\bar{X}(1-\bar{X})}{N}} < \theta < \bar{X} - \Phi^{-1}\left(\frac{\alpha}{2}\right)\sqrt{\frac{\bar{X}(1-\bar{X})}{N}}\right] = 1 - \alpha$$

## 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。

- ➊ A 候補の得票率を, (点) 推定しよう
- ➋ A 候補の得票率を, 信頼係数  $1 - \alpha = 0.95$  で区間推定しよう。
- ➌ 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)  $\theta$  が (毎回得られる) 区間に含まれる確率が  $1 - \alpha$ .

